

**UNIVERSIDAD DE LAS CIENCIAS INFORMÁTICAS**

**Centro de Consultoría y Desarrollo de Arquitecturas Empresariales**



**Modelo para la réplica de datos en sistemas de  
bases de datos distribuidas**

---

Trabajo final presentado en opción al título de  
Máster en Informática Avanzada

**Autora: Ing. Gloria Raquel Leyva Jerez**  
**Tutor: Dr.C. Arturo César Arias Orizondo**

**La Habana, Noviembre de 2018**

## **DECLARACIÓN JURADA DE AUTORÍA**

Declaro por este medio que yo Gloria Raquel Leyva Jerez, con carné de identidad 90051145295, soy la autora principal del trabajo final de maestría Modelo para la réplica de datos en sistemas de bases de datos distribuidas, desarrollada como parte de la Maestría en Informática Avanzada y que autorizo a la Universidad de las Ciencias Informáticas a hacer uso de la misma en su beneficio, así como los derechos patrimoniales con carácter exclusivo.

Y para que así conste, firmo la presente declaración jurada de autoría en La Habana a los \_\_\_\_ días del mes de \_\_\_\_\_ del año \_\_\_\_\_.

---

Ing. Gloria Raquel Leyva Jerez  
Autora

---

Dr.C. Arturo César Arias Orizondo  
Tutor

## RESUMEN

La creciente informatización de los procesos en los diversos sectores de la sociedad ha propiciado el aumento y constante manejo de la información. Esta situación ha favorecido el desarrollo de sistemas de información, sistemas de bases de datos distribuidos y soluciones de réplica de datos para mantener actualizadas los objetos de bases de datos. Sin embargo, la replicación es más que una simple copia de datos entre diferentes almacenamientos; comprende el análisis, diseño, implementación, administración y monitorización de un servicio que debe garantizar la consistencia de los datos replicados. Precisamente, la presente investigación propone un modelo de réplica de datos que contribuye a mantener la consistencia de los datos replicados en ambientes distribuidos, proceso sustancial en sistemas hospitalarios, gubernamentales, bancarios, entre otros como empresariales, con alto flujo de datos. La propuesta comprende principios, características, componentes que describen procesos vinculados con la replicación de datos, modelados con la técnica estratégica de Gestión de Procesos de Negocio (BPM) y que pueden implementarse con tecnologías modernas y libres; y por indicaciones metodológicas para su aplicación. La propuesta fue validada con métodos cuantitativos y cualitativos, entre ellos: Likert, ladov, Grupos focales, Triangulación metodológica y método de prueba. Las validaciones realizadas evidencian que existe una correspondencia satisfactoria entre el objetivo y los resultados obtenidos. La aplicación de la propuesta en el desarrollo del Replicador de datos REKO corroboró la contribución del modelo para mantener la consistencia de los datos, y constituye un aporte a la soberanía tecnológica y a la sustitución de importaciones.

**Palabras claves:** bases de datos distribuidas, consistencia, gestión de procesos de negocio, réplica de datos, sistemas de información.

## **ABSTRACT**

The growing computerization of processes in various sectors of society has led to an increase and constant handling of information. This situation has favored the development of information systems, distributed database systems and data replication solutions to keep database objects updated. However, replication is more than a simple copy of data between different storages. It includes the analysis, design, implementation, administration and monitoring of a service that must guarantee the consistency of the replicated data. Precisely, this research proposes a data replication model that contributes to maintaining the consistency of replicated data in distributed environments, important process in hospital, government, banking systems, among others, as process in enterprises, with a high data flow. The proposal includes principles, characteristics, components that describe processes related to data replication, modeled with the strategic technique of Business Process Management (BPM) and that can be implemented with modern and free technologies; and by methodological indications for its application. The proposal was validated with quantitative and qualitative methods, among them: Likert, ladov, Focus groups, Methodological triangulation and test method; evidencing that there is a satisfactory correspondence between the objective and the results obtained. The validations carried out show that there is a satisfactory correspondence between the objective and the results obtained. The application of the proposal in the development of the REKO Data Replicator corroborated the contribution of the model to maintain data consistency, and it is a contribution to technological sovereignty and import substitution.

**Key words:** distributed databases, consistency, business process management, data replication, information systems.

# ÍNDICE

<b>RESUMEN .....</b>	<b>III</b>
<b>ÍNDICE.....</b>	<b>V</b>
<b>INTRODUCCIÓN .....</b>	<b>1</b>
<b>CAPÍTULO 1. FUNDAMENTACIÓN TEÓRICA .....</b>	<b>6</b>
1.1 Elementos asociados al dominio del problema .....	6
1.1.1 Base de datos.....	6
1.1.2 Homogeneidad .....	6
1.1.3 Consistencia .....	7
1.1.4 Coherencia de los datos .....	7
1.1.5 Integridad .....	7
1.1.6 Sincronización .....	8
1.2 Réplica de datos.....	8
1.2.1 Administración de la réplica de datos .....	9
1.2.2 Monitorización de los datos replicados .....	10
1.2.3 Conflictos en la réplica.....	10
1.3 Modelos y soluciones de réplica de datos.....	12
<b>CAPÍTULO 2. MODELO PARA LA RÉPLICA DE DATOS .....</b>	<b>20</b>
2.1 Concepción del modelo .....	20
2.2 Principios, características y premisas del modelo propuesto .....	20
2.3 Modelo propuesto para la réplica de datos .....	21
2.3.1 Componente Réplica de datos.....	22
2.3.2 Componente Administración de configuraciones .....	26
2.3.3 Componente Métodos de replicación.....	30
2.3.4 Componente Gestión tecnológica .....	40
2.3.5 Indicaciones metodológicas.....	41
<b>CAPÍTULO 3. VALIDACIÓN DE LOS RESULTADOS.....</b>	<b>43</b>
3.1 Instrumentación del modelo en el Replicador de datos REKO .....	43
3.2 Descripción de la validación de la hipótesis.....	43
3.3 Valoración de los expertos.....	45
3.3.1 Proceso de selección de los expertos.....	45
3.3.2 Proceso de aplicación de Likert .....	46
3.4 Validación de satisfacción del usuario .....	47
3.5 Grupo focal.....	49
3.6 Validación de la consistencia.....	52
<b>CONCLUSIONES GENERALES .....</b>	<b>56</b>
<b>RECOMENDACIONES .....</b>	<b>57</b>
<b>REFERENCIAS BIBLIOGRÁFICAS .....</b>	<b>58</b>
<b>ANEXOS.....</b>	<b>63</b>

## ÍNDICE DE FIGURAS

Figura 1 Representación porcentual de cubrimiento detallado del criterio 1 .....	16
Figura 2 Concepción metodológica para elaborar el modelo .....	20
Figura 3 Representación gráfica del modelo .....	22
Figura 4 Propuesta general para la replicación de datos .....	22
Figura 5 Esquema lógico del MRD propuesto .....	24
Figura 6 Esquema físico del MRD propuesto .....	25
Figura 7 Proceso “Administrar configuraciones generales” .....	28
Figura 8 Proceso “Administrar configuraciones de nodos” .....	29
Figura 9 Proceso “Administrar configuraciones de réplica” .....	30
Figura 10 Proceso “Sincronizar” .....	32
Figura 11 Proceso “Exportar nodo” .....	33
Figura 12 Proceso “Importar nodo” .....	33
Figura 13 Proceso “Exportar configuración de réplica” .....	34
Figura 14 Proceso “Importar configuración de réplica” .....	34
Figura 15 Proceso “Exportar cambios” .....	35
Figura 16 Proceso “Importar cambios” .....	36
Figura 17 Proceso “Resolver conflictos” .....	37
Figura 18 Proceso “Realizar transformaciones” .....	38
Figura 19 Componentes de una infraestructura tecnológica para réplica de datos .....	41
Figura 20 Fases de implementación del modelo propuesto .....	42
Figura 21 Técnicas utilizadas para evaluar el modelo .....	44
Figura 22 Resultados de la aplicación de la escala Likert .....	46
Figura 23 Resultados de la aplicación de la técnica ladov .....	48
Figura 24 Resultados obtenidos en el escenario 1 .....	52
Figura 25 Resultados obtenidos en el escenario 2 .....	53
Figura 26 Resultados obtenidos en el escenario 3 .....	54
Figura 27 Resultados obtenidos en el escenario 4 .....	54
Figura 28 Resultados obtenidos en el escenario 5 .....	55

## ÍNDICE DE TABLAS

Tabla 1 Resultado de la evaluación de los modelos y soluciones de replicación .....	15
Tabla 2 Definiciones operacionales de la variable independiente .....	43
Tabla 3 Definiciones operacionales de la variable dependiente .....	44
Tabla 4 Composición de los expertos involucrados en el diagnóstico .....	63
Tabla 5 Fuentes de argumentación del conocimiento de los expertos .....	66
Tabla 6 Resultados de la encuesta para determinar nivel de competencia .....	67
Tabla 7 Respuesta emitida por los expertos para cada indicador .....	69
Tabla 8 Varianza de los resultados de la encuesta por preguntas .....	70
Tabla 9 Cuadro Lógico de ladov modificado por la autora .....	72
Tabla 10 Distribución de los usuarios por escalas de años de experiencia .....	72
Tabla 11 Resultados de la satisfacción individual .....	72
Tabla 12 Composición de los participantes en el grupo focal .....	73

## INTRODUCCIÓN

La creciente informatización de los procesos en los diversos sectores de la sociedad ha propiciado el aumento y constante manejo de la información, promoviendo la utilización de sistemas de información. El sistema de información es el conjunto formal de procesos que, operando sobre una colección de datos estructurada, de acuerdo con las necesidades de una empresa, recopila, elabora y distribuye la información necesaria para la operación de la empresa y para las actividades de dirección y control correspondientes, apoyando, al menos en parte, los procesos de toma de decisiones necesarios para desempeñar las funciones de negocio de la empresa de acuerdo con su estrategia (Lapiedra et al., 2011).

La necesidad de las empresas u organizaciones de compartir datos ubicados en ordenadores descentralizados geográficamente permitió que los sistemas de información evolucionaran desde pequeños sistemas de Bases de Datos (BD) hasta complejos Sistemas de Bases de Datos Distribuidas (SBDD), estructurados bajo una lógica territorial o geográfica, y favorecidos por la utilización de redes de comunicación. Esta tendencia al desarrollo de sistemas distribuidos ha impulsado vertiginosamente el desarrollo de soluciones de réplica para mantener actualizados los objetos de las BDs, facilitar la recuperación y la tolerancia a fallos en la distribución de los datos. Constituyen una forma de aumentar el rendimiento y disponibilidad de los datos.

La replicación en el contexto de BD se entiende como el proceso de copiar y mantener objetos de una BD, como tablas, secuencias y/o triggers, en múltiples BDs, de tal forma que los cambios realizados localmente sean enviados a las BDs remotas y luego sean aplicados (Urbano, 2013). Este proceso abarca el análisis, diseño, implementación, administración y monitorización de un servicio que debe garantizar la consistencia de los datos replicados (Len, 2001; Pérez, 2006; Fuentes, 2015).

La consistencia es el grado en que los datos poseen atributos que están libres de contradicciones y son coherentes con otros datos en un contexto de uso específico (ISO/IEC, 2008). Tiene un carácter decisivo en cualquier organización; ello se puede ilustrar con sencillos ejemplos que representan situaciones comunes (PowerData, 2013):

- Administración de medicamentos en un hospital: cualquier modificación en los registros electrónicos del tratamiento de un paciente podría tener gravísimas consecuencias para su salud si, por ejemplo, se multiplica o reduce la dosis, al cambiar alguna de las cifras de la dosis diaria que este necesita de un determinado medicamento.
- Gestión de incentivos desde un departamento de recursos humanos: modificar o eliminar datos dentro de la información de nóminas puede conllevar a un impago o a practicar un cómputo inexacto que no coincida con las condiciones negociadas con el trabajador.
- Elaboración de un reporte gerencial: si se carece de la integridad necesaria entre sistemas, esta falta de integridad afectará a sus relaciones, en primer lugar, lo que provocará que el

reporte no pueda beneficiarse de la inteligencia de negocio y no sea, en ningún caso, fidedigno ni válido.

- En un sistema bancario en el que se replica en varios sitios la información de las cuentas, es necesario asegurarse de que el saldo de cada cuenta concuerde en todos los sitios (Silberschatz et al., 2002).

La consistencia es un aspecto clave en la replicación (Vila, 2017; Escobar, 2017). Si los sistemas de información no emplean una solución de réplica que garantice la consistencia de los datos, se desencadenan problemas de integridad de la información. Estos problemas afectan la toma de decisiones y la calidad de los procesos gestionados en los sistemas de información por las diferentes organizaciones.

Para la construcción de soluciones de réplica coexisten diversas herramientas, escenarios, técnicas y modelos. Esta variedad introduce una complejidad técnica, pues se requieren amplios conocimientos y habilidades para ofrecer una solución que experimente un alto grado de consistencia. No tomar las decisiones correctas puede conducir a problemas en el rendimiento, el funcionamiento, la calidad de los servicios ofrecidos y la elevación de los costos de desarrollo y mantenimiento.

El estudio realizado mediante un análisis documental y comparativo entre modelos y soluciones de replicación de datos, evidenció un conjunto de insuficiencias que atentan contra la capacidad de los modelos para contribuir a mantener la consistencia en la replicación de datos, debido a que:

- La documentación de calidad que se encuentra publicada en la mayoría de los casos pertenece a compañías líderes que se refieren específicamente a las características de sus herramientas.
- La variedad de marcos de trabajo y tecnologías para el desarrollo de soluciones de réplica dificulta la toma de decisiones en el proceso de selección y para sugerir un método que gestione la consistencia de los datos.
- Algunas soluciones guían a los equipos en “qué” hacer sobre determinados procesos, pero no especifican “cómo” hacerlo. Esto provoca falta de entendimiento en los analistas y desarrolladores, que limita las capacidades que contribuyen directamente a lograr los resultados esperados.
- Ninguna de las soluciones y modelos describe en su totalidad los procesos lo cual limita las capacidades de los sistemas de replicación para responder de manera eficaz ante cada proceso.
- Todos los casos no evidencian explícitamente cómo gestionan la consistencia en la distribución de los datos, en consecuencia, las capacidades de las soluciones para ofrecer un servicio de calidad con alto grado de consistencia se ven limitadas.



Por otra parte, debido a la avalancha de información de la que han sido testigos las organizaciones, la gestión documental<sup>1</sup> juega un papel fundamental en el adecuado manejo de los sistemas de información (Chávez et al., 2013). Una estrategia utilizada por los Sistemas de Gestión Documental para la generación de reportes, consiste en la creación y manejo de nuevas tablas en la BD relacional, para agilizar los procesos de indexado y consulta de los datos de los reportes (Abadal et al., 2010). Esto trae como consecuencia que la estructura de la BD cambie constantemente de forma dinámica y se generan diferencias de estructuras con respecto al resto de las bases de datos distribuidas que deben ser replicadas. Por lo tanto, a las limitaciones anteriormente mencionadas se suman las siguientes insuficiencias:

- El proceso de actualización de los cambios DDL (Lenguaje de Definición de Datos, del inglés Data Definition Language) es descrito a manera de síntesis y con pocas precisiones, lo cual limita las capacidades de las soluciones de réplica para mantener la consistencia de los datos bajo este escenario.
- Si los sistemas de información que generan reportes no emplean una solución de réplica capaz de actualizar una copia que contiene cambios DDL, se generan inconsistencias en los datos que pueden derivar en problemas de integridad referencial de la información, tablas y datos incompletos e incoherencias entre las BDs que se encuentran distribuidas.

Dada la situación antes expuesta, se define el siguiente **problema de la investigación** a resolver: ¿Cómo contribuir a mantener la consistencia de los datos replicados en ambientes distribuidos?

A partir del problema enunciado, se define como **objeto de estudio**: la replicación de datos en bases de datos distribuidas, y dentro de esta se enmarca como **campo de acción**: consistencia en los procesos relacionados con la distribución de los datos.

Para dar solución al problema planteado como **objetivo general** de la investigación se plantea: Desarrollar un modelo para la réplica de datos en sistemas de bases de datos distribuidas que contribuya a mantener la consistencia de los datos replicados.

El cual ha sido desglosado en los siguientes **objetivos específicos**:

- Elaborar el marco teórico de la investigación relacionado con las soluciones de réplica y buenas prácticas que siguen los modelos de replicación.
- Desarrollar un modelo para la construcción de soluciones de réplica de datos en ambientes distribuidos que experimente la consistencia en la distribución de datos.
- Diseñar e implementar procesos relacionados con la réplica de datos que soporten el modelo y faciliten su aplicación.

---

<sup>1</sup> Conjunto de actividades que permiten coordinar y controlar los aspectos relacionados con creación, recepción, organización, preservación almacenamiento, acceso y difusión de documentos (Russo, 2011).

- Validar la propuesta a través de los métodos definidos en la investigación.

Para solucionar el problema y como resultado de la construcción del marco teórico referencial de la presente investigación se formula la siguiente **hipótesis**: si se desarrolla un modelo para la réplica de datos en sistemas de bases de datos distribuidas que describa los procesos de sincronización, resolución de conflictos, monitorización y administración de configuraciones, se contribuirá a mantener la consistencia de los datos replicados.

Para dar cumplimiento a los objetivos trazados en el presente trabajo se utilizaron los siguientes métodos científicos de investigación:

### **Métodos teóricos**

- Hipotético-deductivo: para formular la hipótesis de la investigación y proponer nuevas líneas de trabajo a partir de los resultados.
- Analítico-sintético: permitió descomponer el problema de la investigación en partes, profundizar por separado en su estudio y luego sintetizarlos en la solución propuesta.
- Histórico-lógico: permitió realizar un estudio crítico sobre la evolución y tendencias de las herramientas de replicación, gestores que realizan trabajos similares y modelos de replicación, para emplearlo como punto de referencia y comparar los resultados alcanzados.
- Modelación: se aplicó en la elaboración del modelo de replicación propuesto.
- Sistémico: para integrar lógicamente los componentes del modelo.

### **Métodos empíricos**

- Análisis documental: permitió la elaboración del marco teórico de la investigación, en sus aspectos conceptuales y metodológicos, mediante la revisión de la literatura especializada, tanto académica como empresarial.
- Entrevista a profundidad: en la obtención de la información a partir de la experiencia de especialistas acerca de aspectos relevantes del desarrollo de soluciones de réplica.
- Encuesta: para conocer los criterios de especialistas y expertos en torno a la propuesta de solución.
- Grupos focales: para refinar el modelo y evaluar su utilidad, y en la validación de la propuesta como un método cualitativo para tener en cuenta en la triangulación metodológica.
- Likert e Iadov: para evaluar y corroborar por expertos y potenciales usuarios, la factibilidad y pertinencia del modelo, así como los aportes fundamentales de la investigación.
- Triangulación metodológica: hizo posible analizar las coincidencias y divergencias en los criterios obtenidos entre los métodos aplicados.
- Observación participante: la autora de la investigación fue participante activa en una gran parte de los procesos desarrollados.

**Nivel estadístico:** mediante la estadística descriptiva, con el empleo de gráficos, tablas, porcentos, hizo posible la organización y regularización de la información obtenida.

La novedad de la investigación se expresa mediante los siguientes aportes:

**Aportes prácticos:**

- Concepción y fundamentación de un modelo para la replicación de datos, que contribuye a mantener la consistencia de los datos replicados en ambientes distribuidos.
- Una herramienta de replicación de datos basada en el modelo desarrollado.
- Propuesta tecnológica basada en software libre para el desarrollo de soluciones de replicación de datos.

**Aporte social y económico:**

La investigación está acorde a la Política Económica y Social del Partido Comunista de Cuba en sus lineamientos 131 y 132 de la política de Ciencia, Tecnología, Innovación y Medio Ambiente. Permite cubrir las necesidades de réplica de datos y diseminación de la información entre múltiples localizaciones, proceso sustancial en los sistemas hospitalarios, gubernamentales, bancarios, entre otros como: empresariales. El modelo propuesto puede implementarse con tecnologías modernas y libres, por lo que representa un paso importante en la estrategia de migración hacia este tipo de tecnologías que sigue el país. Además, realiza un aporte a la soberanía tecnológica y a la sustitución de importaciones.

El presente documento consta de tres capítulos, conclusiones, recomendaciones, referencias bibliográficas y anexos.

El **Capítulo 1** presenta la fundamentación teórica de la investigación, donde se realiza un análisis del estado del arte sobre herramientas y modelos de réplica de datos. Se muestran las bases teóricas sobre las tendencias y tecnologías actuales que permiten comprender la necesidad de desarrollar la propuesta.

El **Capítulo 2** aborda la concepción y desarrollo del modelo para la réplica de datos. Se diseña el modelo y definen sus principios, componentes, estructura y funcionamiento. Finalmente, se plantean las indicaciones metodológicas que guían su aplicación.

El **Capítulo 3** muestra la validación de los resultados de la investigación, los que permitieron confirmar la pertinencia, comprensión, coherencia, adaptabilidad, satisfacción del usuario y contribución del modelo propuesto para mantener la consistencia de los datos en ambientes distribuidos.

## CAPÍTULO 1. FUNDAMENTACIÓN TEÓRICA

Una vez establecido el marco metodológico que sustenta la investigación y de analizar los antecedentes, actualidad, importancia, trascendencia del problema planteado, necesidad social del tema y describir la situación problemática, se procede a describir el marco teórico de la investigación. El marco teórico además de orientar conceptualmente la investigación, permite consultar diferentes fuentes bibliográficas para finalmente analizar y exponer aquellas teorías, investigaciones y antecedentes que se consideran válidos en el contexto del problema planteado.

### 1.1 Elementos asociados al dominio del problema

En esta sección se ofrecen las definiciones y acercamientos teóricos utilizados en la investigación para el mejor entendimiento y comprensión de la investigación.

#### 1.1.1 Base de datos

En la investigación se asumen los siguientes conceptos definidos en el libro “Fundamentos de bases de datos” por Abraham Silberschatz y otros autores, porque los considera válidos y acordes a los objetivos de la investigación. Las ideas expuestas sobre estos conceptos en (Silberschatz et al., 2002) convergen con las emitidas por otros autores en (Date, 2001; Carranza, 2006; Marqués, 2011).

*“Una base de datos es un almacén o sistema formado por un conjunto de datos pertenecientes a un mismo contexto, almacenados de forma organizada y relacionada entre sí, y que permite almacenar información para un posterior uso”.*

*“Una base de datos relacional consiste en un conjunto de tablas, a cada una de las cuales se le asigna un nombre exclusivo. Cada fila de la tabla representa una relación entre un conjunto de valores”.*

*“Una base de datos distribuida es un conjunto de bases de datos parcialmente independientes que (idealmente) comparten un esquema común y coordinan el procesamiento de transacciones que acceden a datos remotos. Los procesadores se comunican entre sí a través de una red de comunicación que gestiona el encaminamiento y las estrategias de conexión”.*

#### 1.1.2 Homogeneidad

En el Diccionario de la lengua española se define el término como cualidad de homogéneo que se refiere a (DLE, 2017b):

1. *adj. Perteneciente o relativo a un mismo género, poseedor de iguales caracteres.*
2. *adj. Dicho de una sustancia o de una mezcla de varias: De composición y estructura uniformes.*
3. *adj. Dicho de un conjunto: Formado por elementos iguales.”*

*“La homogeneidad de las transacciones en la replicación obliga a que los datos sean idénticos en todos los sitios que participan en la replicación, como si la transacción fuera a ejecutarse sobre todos los sitios”* (Gabillaud, 2015).

A partir de las definiciones antes expuestas, en el contexto de replicación, el SBDD mantiene la homogeneidad si sus estructuras son iguales respondiendo a la configuración de réplica realizada sobre las tablas.

### 1.1.3 Consistencia

Ángel Cobo (2007) enuncia que: *“La consistencia consiste en impedir que exista información inconsistente o contradictoria en la base de datos. La inconsistencia surge cuando existen varias copias del mismo dato y tras la modificación de una de ellas, las demás no son actualizadas, o sí lo son, pero de manera incorrecta”*.

Francisco López (2015) expresa que uno de los principales retos en los sistemas distribuidos es hacer que las réplicas se mantengan consistentes, lo que implica garantizar que todas las copias del sistema sean actualizadas.

En la norma internacional 25012 publicada por la ISO/IEC<sup>2</sup> se define como: *“El grado en que los datos poseen atributos que están libres de contradicciones y son coherentes con otros datos en un contexto de uso específico. Puede ser uno o ambos entre los datos con respecto a una entidad y entre datos similares para entidades comparables”* (ISO/IEC, 2008).

A los efectos de la presente investigación se entiende que la consistencia se refiere a la ejecución correcta de los cambios durante la actualización de réplicas, de tal forma que las BDs que intervienen en este proceso y se encuentran distribuidas mantienen la homogeneidad en un contexto de uso específico.

### 1.1.4 Coherencia de los datos

José Antonio Ocampo (2004) expresa que: *“... es la idoneidad de los datos para ser combinados en forma fiable de diferentes maneras y distintos usos, procedan de una fuente única o investigaciones estadísticas de diversas naturalezas”*.

En el contexto de réplica, se entiende por coherencia a la igualdad entre los datos que se envían desde un nodo local y se aplican en un nodo remoto para mantener la integridad de los datos.

### 1.1.5 Integridad

Para Ángel Cobo (2007) proteger la integridad de una BD consiste en: *“... asegurar que la información almacenada es correcta, es decir, refleja fielmente la realidad”*. Plantea que no existe

---

<sup>2</sup> Organización Internacional para la Normalización (ISO, por sus siglas en inglés) y Comisión Electrotécnica Internacional (IEC, por sus siglas en inglés)

integridad de datos cuando existe inconsistencia o cuando existe información imposible o falsa que no se ajusta a la realidad.

Especialistas de gestión de datos plantean que: *“La integridad de un dato alude a ese atributo o cualidad que es inherente a la información cuando se considera exacta, completa, homogénea, sólida y coherente con la intención de los creadores de los datos que la conforman”* (PowerData, 2013).

Mientras, la integridad referencial protege las relaciones definidas entre las tablas cuando se crean o se eliminan filas. La integridad referencial garantiza que los valores de clave sean coherentes en las distintas tablas (Mato, 2005).

### 1.1.6 Sincronización

El colectivo de Microsoft expone que: *“Sincronizar es actualizar dos miembros de un conjunto de réplicas mediante el intercambio de todos los registros y objetos actualizados en cada miembro. Dos miembros de un conjunto de réplicas se encuentran sincronizados cuando los cambios de cada uno han sido aplicados al otro”* (Microsoft Corporation, 2014).

Por su parte el equipo de Microsoft SQL Server plantea que: *“... se refiere al proceso de propagación de los cambios en los datos y el esquema entre el publicador y los suscriptores después de haber aplicado la instantánea inicial en el suscriptor”* (SQL Server 2014, 2014).

Los autores del Diccionario de Informática y Tecnología se refieren a la sincronización como: *“... el proceso del establecimiento de consistencia entre datos en fuentes remotas y de la continua armonización de los datos en el tiempo”* (Diccionario de Informática y Tecnología, 2017).

En correspondencia con los autores citados anteriormente, se puede sintetizar que sincronización es el proceso de establecimiento de consistencia entre los datos en diferentes fuentes remotas mediante la actualización de todos los registros de un conjunto de réplicas. Dos instancias se encuentran sincronizadas cuando los registros de una han sido aplicados en la otra.

## 1.2 Réplica de datos

La réplica de datos permite mantener actualizados los objetos de las BDs y facilita la recuperación y la tolerancia a fallos en la distribución de los datos.

En la literatura se pueden encontrar varias definiciones que se consideran válidas, tales como:

- Para el autor Marius Cristian Mazilu (2010) representa: *“... el proceso de intercambio de información para garantizar la coherencia entre los recursos redundantes, tales como: componentes de software o hardware. Permite mejorar la confiabilidad, tolerancia a fallos y accesibilidad”*.

- Para un grupo de desarrolladores de Microsoft SQL Server la replicación es un conjunto de tecnologías destinadas a la copia y distribución de datos y objetos de BDs desde una BD a otra, para luego sincronizar ambas BDs y mantener su coherencia (SQL Server 2012, 2012).
- Para Randy Urbano (2013) el proceso de copiar y mantener objetos de una BD, como tablas, secuencias y/o triggers, en múltiples BDs, de tal forma que los cambios realizados localmente sean enviados a las BDs remotas y luego sean aplicados.

Las soluciones de réplica adquieren características para lograr la actualización de los cambios de manera correcta y es precisamente mediante los procesos de administración, sincronización, resolución de conflictos, monitorización de las acciones replicables y con el empleo de mecanismos de protección y recuperación de los datos, que a continuación se enuncian:

- Mediante la administración se planifica la distribución de los datos según el entorno de réplica definido, para que la actualización de los datos se realice de forma correcta (SL, 2014).
- Con la sincronización es posible establecer la consistencia entre datos en fuentes remotas (Diccionario de Informática y Tecnología, 2017).
- Cualquier irregularidad originada, en el proceso de aplicación de la información que impide que la transacción pueda ser aplicada, es solucionada mediante el consenso o variantes definidas dentro de la resolución de conflictos (Pérez, 2008).
- Con las tareas de monitorización se controla el estado de la réplica de datos, analizando si el envío y captura de los datos se realiza de forma correcta o incorrecta, y gestionando la detección de conflictos en la sincronización o propagación de los cambios con conexión o fuera de línea (Microsoft, 2014).
- Los mecanismos de protección y recuperación, ante eventualidades como caída de una instancia o fallos en la red, constituyen un componente fundamental para evitar pérdida de los datos replicados y por ende inconsistencias entre las BDs que se encuentran distribuidas.

A continuación, se abordan soluciones adquiridas por replicadores de datos que favorecen la correcta distribución de los datos en ambientes distribuidos.

### 1.2.1 Administración de la réplica de datos

En las soluciones de réplica se realizan tareas de administración donde se consideran aspectos como (SL, 2014):

- Forma de transmisión de la información.
- Cantidad de tablas a replicar.
- Estado de los datos que contienen las tablas replicadas.
- Acciones a realizar sobre los datos.
- Verificación del registro de réplica de datos.

Estas tareas no se realizan de la misma forma porque dependen del tipo de réplica de datos que se desea implementar, de la herramienta de réplica de datos empleada, del entorno de réplica de datos definido y de las necesidades del negocio implicado. Pero independientemente de la forma, están orientadas a (SL, 2014):

- Validar datos periódicamente.
- Visualizar los hilos de empuje y tirado de los datos en el envío y captura de los datos.
- Verificar los lotes de datos que entran y salen de un nodo.
- Verificar las propiedades del sistema donde está montada la réplica de datos.
- Gestionar nuevos nodos en la réplica de datos.
- Definir transformaciones e instrucciones basadas en SQL (Lenguaje de Consulta Estructurada, del inglés Structured Query Language) que se puedan ejecutar sobre las tablas utilizando condiciones de filtros además de agregaciones (Silberschatz et al., 2002).

Por tanto, en la investigación se considera que la administración es el proceso de planificar la distribución de los datos mediante configuraciones, para que la actualización de los datos se realice de forma correcta según los requerimientos del cliente.

### 1.2.2 Monitorización de los datos replicados

El diccionario de Oxford define la monitorización como: *“Observar, supervisar o mantener en examen, a medida o prueba a intervalos, especialmente con el propósito de la regulación o control. Cualquier instrumento o dispositivo para el seguimiento de algún proceso o cantidad”*.

Precisamente en la replicación, las tareas de monitorización permiten controlar el estado de la réplica de datos. Este proceso analiza si el envío y la captura de los datos se realiza de forma correcta o incorrecta. Además, gestiona la detección de conflictos en la sincronización, fragmentación o enrutamiento de los datos (Microsoft, 2014).

### 1.2.3 Conflictos en la réplica

Para Fabiola Ocampo (2001): *“... dos operaciones están en conflicto si operan sobre el mismo dato, una de ellas es una operación de escritura y son emitidas por diferentes transacciones”*.

Para Jofman Pérez Tarancón (2008) se entiende como conflicto en la réplica a: *“... cualquier irregularidad originada en el proceso de aplicación de la información que impide que la transacción pueda ser aplicada. Los conflictos normalmente son asociados a inconsistencia en la información existente en un ambiente con relación al otro”*.

En el contexto de la investigación se consideran válidos ambos conceptos y se asume como conflicto en la replicación de datos al error generado en la BD que impide que la transacción pueda ser aplicada. En consecuencia, se introducen inconsistencias entre las BDs que se encuentran distribuidas. La transacción puede contener acciones DML (Lenguaje de Manipulación de Datos, del inglés Data Manipulation Language) y/o DDL.



Los conflictos pueden ser reparados manualmente por un administrador, simplemente por usuarios, o mediante transacciones de compensación. Estas últimas son generadas por el sistema porque necesitan reconciliación, para mantener la consistencia de los datos sobre todas las réplicas (Kempe et al., 2000).

Emplear transacciones compensatorias significa que a veces las actualizaciones de una transacción no son durables<sup>3</sup>. La pérdida de durabilidad crea un efecto en el cual un usuario puede tomar una decisión sobre información que es subsecuentemente la perdedora en la resolución de un conflicto, es decir, las transacciones que son posteriormente deshechas, son vistas por otras transacciones antes de que el proceso ocurra. Dependiendo del grado de consistencia deseado, puede suceder que las transacciones secundarias también requieran acciones compensatorias para asegurar la consistencia de la base (Len, 2001).

Javier Len (2001) presenta un mecanismo para gestionar los conflictos de actualización de datos que se basa en la detección de conflictos y un mecanismo de resolución, para así mantener la consistencia de los datos.

En el esquema Síncrono<sup>4</sup> es posible prevenir mediante espera o bloqueos que dos actualizaciones generadas en diferentes sitios choquen al actualizar la misma réplica, principalmente con dos tipos de protocolos dependiendo si usan bloqueos distribuidos o broadcast (difusión) atómicos para ordenar las operaciones en conflictos (Len, 2001). Estos protocolos para el control de concurrencia son ampliamente descritos en (Jiménez et al., 2015; Miaton, 2001; García, 2017; Ocampo, 2001; Martínez, 2007; Ilarri, 2012).

Para los esquemas Asíncronos<sup>5</sup> se propone utilizar la técnica de timestamps (marcas de tiempo) para garantizar la convergencia entre las réplicas, donde cada actualización propagada lleva un timestamp de cada réplica original. Si el timestamp de la réplica local viola el orden, existe un conflicto y se necesita alguna forma de reconciliación. La reconciliación consiste en lograr convergencia entre los datos, y enviar la información de los conflictos ocurridos a un nuevo proceso denominado Recovery (recuperación), con el fin de ser utilizada en los índices de evaluación. Este proceso tiene como función recibir mensajes provenientes de aquellas actualizaciones que tienen conflictos y habría que recuperar, grabando luego los datos de dichas actualizaciones en almacenamiento permanente. El timestamp está compuesto por el "IdHost" que es el identificador del host y por un "Lc" que sería el contador local para ordenar las propagaciones.

En (Arora, 2005) se plantea que en la replicación asincrónica las inconsistencias potenciales son una preocupación común. Pero los usuarios pueden configurar la replicación para eliminar

---

<sup>3</sup> Significa que los efectos de una transacción cometida sobreviven a las fallas.

<sup>4</sup> En la replicación Síncrona los cambios ejecutados en un nodo maestro son aplicados instantáneamente en el nodo origen y los nodos destino dentro de una misma transacción.

<sup>5</sup> En la replicación Asíncrona los cambios ejecutados en una tabla son almacenados y enviados posteriormente al resto de los nodos del entorno cada cierto intervalo de tiempo.

virtualmente la posibilidad de divergencia. A través de las técnicas anteriormente referenciadas se evidencia que es posible evitar o resolver los conflictos ante actualizaciones concurrentes de réplicas, incluso aplicar modelos que contribuyen a evitar conflictos, y mantener con alguna de estas variantes la consistencia de los datos. Sin embargo, se debe considerar que la BD siempre va a estar expuesta a cambios que puedan ser introducidos por administradores. Los administradores o usuarios de réplica manualmente pueden ejecutar operaciones DML o DDL, generando diferencias en el SBDD sobre el mismo dato en diferentes sitios, que al ser propagadas crean inconsistencias. Por esta razón, las soluciones de réplica han diseñado otros tipos de conflictos y formas de resolución para cada una de las situaciones que se puedan generar.

Precisamente, en (Urbano, 2013) además de la técnica descrita anteriormente, se plantean los siguientes tipos de conflictos:

- Conflicto de unicidad: se produce cuando la replicación de una fila intenta violar la integridad de la entidad, como una restricción Primary key (llave primaria) o Unique key (llave única). Por ejemplo, cuando dos transacciones se originan en dos sitios diferentes, cada uno insertando una fila en una réplica de tabla respectiva con el mismo valor de clave primaria. En este caso, la replicación de las transacciones provoca un conflicto de exclusividad.
- Conflicto de borrado o conflicto de eliminación: ocurre cuando dos transacciones se originan en sitios diferentes, con una transacción eliminando una fila y otra transacción actualizando o eliminando la misma fila, porque en este caso la fila no existe para ser actualizada o borrada.

La clasificación anterior converge con la expuesta en (Pérez, 2006), donde adicionalmente definen el conflicto de actualización, que ocurre cuando una fila está siendo actualizada por dos transacciones diferentes simultáneamente. Además, emplean los siguientes métodos de resolución propuestos por Oracle:

- Ignorar: ignoran el error, aunque no se haya actualizado.
- Aplicar: aplica los cambios basándose en la llave de la tabla involucrada.
- Suma y promedio: utiliza una fórmula para obtener el valor a actualizar. Solamente es aplicable para un grupo de columnas formadas por campos numéricos.
- Mayor: inserta al final el valor que sea mayor.
- Menor: inserta al final el valor que sea menor.
- Prioridad: en el momento del conflicto, inserta al final los valores de las columnas de mayor prioridad.
- Función: define la lógica de resolución de conflictos con un programa, tal vez complejo.

### 1.3 Modelos y soluciones de réplica de datos

Un gran número de empresas comercializan soluciones de réplica en el mercado del software, algunas incluidas dentro del software, en el propio gestor de BD y otras como opciones que se pueden licenciar de forma independiente. Estas soluciones implementan modelos de replicación

que abarcan diferentes formas de replicación. Se realizó un diagnóstico inicial del conocimiento explícito abordado en modelos y soluciones de réplica de datos para analizar los procesos de replicación, tales como: distribución de datos, sincronización, resolución de conflictos, mecanismos de tolerancia a fallos y recuperación, manejo de la réplica desatendida, así como la descripción de técnicas y métodos diseñados para cumplir con la consistencia de los datos en cada uno de estos procesos.

Un diagnóstico es el proceso mediante el cual se lleva a cabo un análisis integral que evalúa cada una de las áreas para recopilar información que conlleven al logro de una misión, objetivos estratégicos en un contexto real (Domínguez et al., 2014). Mediante un diagnóstico se trata de focalizar y evaluar un conjunto de variables que juegan un importante papel en la comprensión, predicción y control del comportamiento de un fenómeno determinado (Shull et al., 2008).

Como parte del diagnóstico realizado en la presente investigación, se realizó un análisis de la siguiente literatura especializada, tanto académica con los modelos examinados, como empresarial mediante las soluciones comerciales:

- M1: Modelo para la experimentación en técnicas de actualización de réplicas (Len, 2001).
- M2: Modelo de replicación de datos en sistemas distribuidos de bases de datos relacionales (Pérez, 2006).
- M3: Modelo de servicios web para replicación de instantáneas sobre motores múltiples de bases de datos (Erazo et al., 2014).
- M4: Modelo de replicación de Microsoft SQL Server (Microsoft Corporation, 2016).
- R1: Oracle Streams (Arora, 2005; Oracle Corporation, 2007; Oracle Help Center, 2018).
- R2: pglogical (Peter et al., 2016; PostgreSQL, 2017; 2ndQuadrant PostgreSQL, 2018).
- R3: SymmetricsDS (Figueroa, 2011; Long, 2012; Long et al., 2013; Community SymmetricDS, 2018).
- R4: Tungsten Replicator (Continuent Ltd, 2018).

La selección de las soluciones estuvo basada en un grupo de características que estas presentan, identificadas en la literatura y ratificada por especialistas que han trabajado directamente con herramientas de replicación.

En (Ramírez, 2012) se plantea que los métodos son los medios que capacitan al ser humano para el manejo de pensamientos y objetos de una manera adecuada. Aplicar métodos científicos permite abordar la investigación con carácter objetivo y sistemático, y constituye un ejercicio verificable, reproducible y fundamentado en un razonamiento lógico (Barboza, 2015). Con tal propósito en el diagnóstico se aplicaron los siguientes métodos y técnicas de investigación:

- Análisis documental: para revisar la literatura relacionada con la réplica de datos y enfoques para gestionar la consistencia en la distribución de los datos.
- Análisis comparativo: para detectar similitudes, diferencias e insuficiencias en los modelos y

herramientas representativas de réplica de datos.

- Nivel estadístico: mediante la estadística descriptiva, con el empleo de gráficos y porcentos se cuantifica la información obtenida.
- Entrevista a profundidad: participan 12 especialistas con experiencia en el desarrollo o empleo de soluciones de replicación de datos, arquitectos, analistas, desarrolladores de software y administrativos, relacionados con la actividad del objeto de estudio. Para la selección de los especialistas se tomaron en cuenta varios criterios basados en la experiencia, roles que desempeñan, y la relación entre las temáticas laborales en las que trabajan y el objetivo principal de esta investigación. La composición de los expertos involucrados se encuentra detallada en el Anexo 1.
- Encuesta: se aplicó una encuesta a los 12 especialistas seleccionados con un cuestionario previamente elaborado para conocer la valoración de expertos con respecto a temas relacionados con el objeto de estudio y determinar los criterios para comparar los modelos y soluciones de replicación de datos. El cuestionario se muestra en el Anexo 2.
- Observación participante: la autora de la investigación fue participante activa en los procesos desarrollados.

Una investigación puede ser conveniente por diversos motivos: tal vez ayude a resolver un problema social o a construir una nueva teoría. Lo que algunos consideran relevante y debe ser investigado, puede no serlo para otros. Respecto a ello, suele diferir la opinión de las personas. Sin embargo, se puede establecer una serie de criterios para evaluar la utilidad de un estudio propuesto, criterios que evidentemente sean flexibles (Sampieri et al., 2010). Por las razones antes mencionadas, para la comparación fueron definidos un conjunto de criterios, identificados en la literatura y ratificados por expertos. Los criterios identificados fueron:

- **C1. Procesos de réplica:** se detallan los procesos asociados a la réplica de datos en ambientes distribuidos.
- **C2. Aplicabilidad:** se describen técnicamente los procesos hacia soluciones prácticas del desarrollo para valorar la aplicabilidad de la propuesta y la factibilidad de su uso.
- **C3. Software libre:** está orientado hacia una infraestructura tecnológica soportada por soluciones libres que faciliten su aplicación.
- **C4. Consistencia:** se reconoce la consistencia como un aspecto fundamental para la correcta distribución de los datos.
- **C5. Evaluación de la consistencia:** se fundamentan y validan las capacidades del sistema para contribuir a mantener la consistencia en los procesos asociados a la réplica.

Los primeros tres atributos responden a buenas prácticas en función de las características de los modelos y herramientas de replicación, y a criterios para evaluar la utilidad y facilidades de uso de las soluciones. Mientras que el cuarto y quinto se refieren a las capacidades que deben ser adquiridas para evitar inconsistencias en ambientes distribuidos.

Para evaluar cualitativamente cada variable existen cuatro niveles de medición ampliamente fundamentados en (Sampieri et al., 2010). Para evaluar los criterios anteriores se establece un nivel de medición nominal. En este nivel se tienen dos o más categorías del ítem o variable. En la presente investigación la variable estará comprendida por las siguientes categorías:

- Sí: cuando se verificó que el atributo evaluado es considerado explícitamente por el modelo.
- Parcialmente (P): cuando no existe evidencia explícita, pero se aprecia que el atributo es considerado por el modelo.
- Pobremente (P-): cuando no se evidencia que el atributo evaluado es incorporado intencionalmente en el modelo, pero no lo desconoce del todo.
- No: cuando definitivamente el modelo no lo considera.

Como resultado del análisis comparativo realizado se detectaron insuficiencias que atentan contra la capacidad de los modelos para contribuir a mantener la consistencia en la replicación de datos. No se identificaron modelos ni soluciones que consideraran todos los atributos definidos. A continuación, se describen las insuficiencias detectadas:

Tabla 1 Resultado de la evaluación de los modelos y soluciones de replicación

-Modelo y replicadores-	-C1-	-C2-	-C3-	-C4-	-C5-
<b>M1</b>	Sí	P-	No	Sí	No
<b>M2</b>	Sí	P-	No	Sí	Sí
<b>M3</b>	Sí	P-	No	P	No
<b>M4</b>	Sí	P-	No	Sí	No
<b>R1</b>	Sí	P	P	P-	No
<b>R2</b>	Sí	P	P	P-	No
<b>R3</b>	Sí	P	P	P-	No
<b>R4</b>	Sí	P	P	P-	No

- **C1. Procesos de réplica:** se detallan los procesos asociados a la réplica de datos en ambientes distribuidos.

Todas las soluciones describen procesos vinculados a la distribución de los cambios en ambientes distribuidos, aunque no todos los procesos identificados en la literatura son abordados por los modelos y replicadores. Para una mejor valoración de este criterio, como se ilustra en la Figura 1, se representó porcentualmente el cubrimiento de los procesos identificados en la literatura y ratificados por especialistas en el tema.

La Administración de configuraciones ha sido comprendida por el 50 % de las soluciones correspondientes a M1, M3, M4 y R4. El modelo M2 no hace alusión a este proceso, pglogical ajusta las configuraciones mediante filtros establecidos manualmente en el gestor de BD, Oracle Streams posee un conjunto de reglas predefinidas que determinan las acciones a replicar, y en SymmetricsDS se realizan a través de comandos por consola. Los procesos de construcción de

metadatos y dialectos no son considerados por ninguna de las soluciones. Solamente R1 presenta un acercamiento con la descripción de columnas LCR (Unidad básica de Captura de Cambios, del inglés Logical Change Records). La sincronización de los cambios es realizada por M4, R1, R3, M1 y M3, el resto de las soluciones no describen esta funcionalidad. A pesar de que Oracle Streams, SymmetricsDS y Microsoft SQL Server soportan la replicación entre diferentes estructuras, en las fuentes consultadas no se identifica un proceso para describir las transformaciones realizadas. Solamente M1, M4 y R4 presentan y describen un monitor de réplica para visualizar su comportamiento. A pesar de que las soluciones presentan características para recuperarse ante fallos y operar sin conexión, estas no son fundamentadas. Y M3 es la única propuesta que no brinda soporte para la resolución de conflictos.

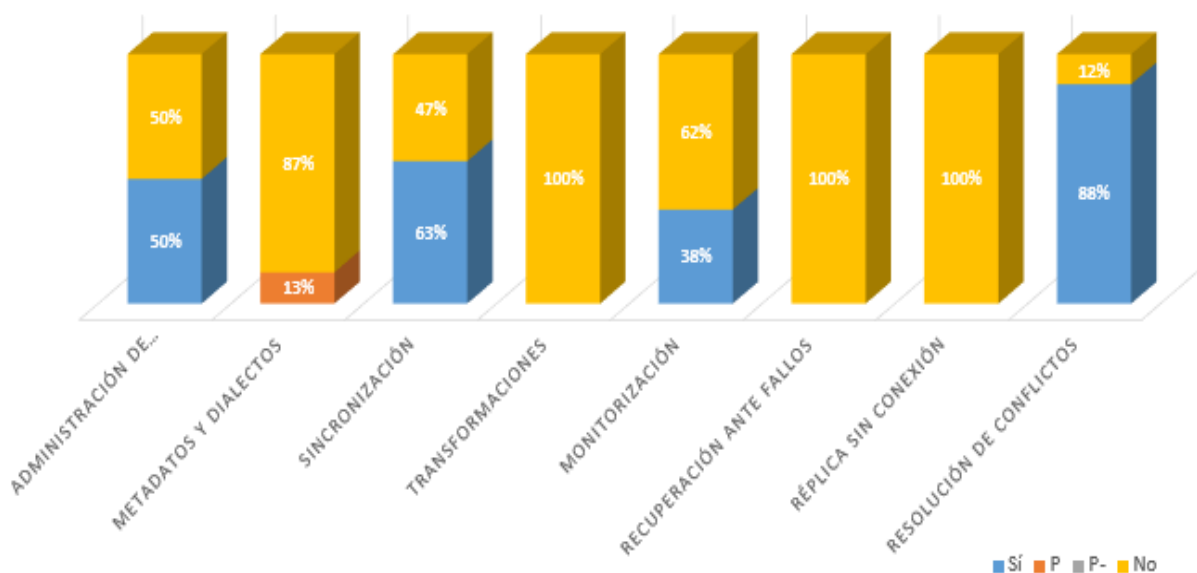


Figura 1 Representación porcentual de cubrimiento detallado del criterio 1 (procesos de réplica)

Ningunas de las soluciones y modelos describen en su totalidad los procesos comprendidos, lo cual evidencia la ausencia de un modelo que generalice todos los procesos implicados, e introduce una complejidad técnica que puede convertirse en algo abrumador, pues es necesario tener amplios conocimientos y habilidades para ofrecer una solución de réplica acorde a las necesidades planteadas. No tomar las decisiones correctas puede conducir a problemas en el rendimiento, el funcionamiento, la calidad de los servicios ofrecidos y la elevación de los costos de desarrollo y mantenimiento.

- **C2. Aplicabilidad:** se describen técnicamente los procesos hacia soluciones prácticas del desarrollo para valorar la aplicabilidad de la propuesta y la factibilidad de su uso.

La documentación de calidad que se encuentra publicada en la mayoría de los casos pertenece a compañías líderes que se refieren específicamente a las características de sus herramientas, como es el caso de Microsoft SQL Server y Oracle. Los modelos si bien se limitan a aspectos teóricos y en cuanto a lo que los equipos deben hacer, las herramientas de replicación que sí son soluciones técnicas, abordan sus procesos a modo de síntesis y con pocas precisiones. Esto provoca falta de

entendimiento en los analistas y desarrolladores, que limita las capacidades para contribuir directamente al logro de los resultados esperados.

En ninguna de las soluciones abordadas describen las tecnologías empleadas para valorar su aplicación en el desarrollo de soluciones de replicación, y favorecer con ello la reutilización y mejores tiempos de desarrollo.

- **C3. Software libre:** está orientado hacia una infraestructura tecnológica soportada por soluciones libres que faciliten su aplicación.

Aunque en la literatura se identifican *pglogical*, *Tungsten Replicator* y *SymmetricsDS* como herramientas Open Source (código abierto), no cumplen con todos los requerimientos básicos planteados.

En *pglogical* las tablas del proveedor y el suscriptor deben tener los mismos nombres y estar en el mismo esquema, por lo que no permite transformaciones o mapeos entre nodos. Por lo tanto, no brinda soporte para replicación entre BDs con diferentes estructura. No permite replicación de esquemas, es decir, operaciones DDL pues solamente admite replicar los cambios realizados sobre tablas. Tampoco ofrece replicación Multi-Master<sup>6</sup> (Jelínek, 2016; Peter et al., 2016).

*Tungsten Replicator* ignoran deliberadamente cambios en el lenguaje de definición de datos (DDL) y otras declaraciones SQL para que la replicación no se detenga. Por tanto, en el SBDD se pueden presentar problemas de integridad referencial de la información, tablas y datos incompletos e incoherencias (Continuent Ltd, 2016).

*SymmetricsDS* a pesar de ser una herramienta de código abierto la solución *SymmetricDS Pro*, que es la alternativa para facilitar al usuario la configuración y gestión de la solución de réplica de datos, posee una licencia de software privativo. En consecuencia, su utilización implicaría costos económicos según el número de instancias involucradas (Alfonso, 2017). La Herramienta administrativa para mecanismo de replicación *SymmetricDS* presentada en (Figuroa, 2011), Módulo para la administración y monitorización de réplica propuesto en (Alfonso, 2017), y la aplicación Interfaz web para la administración y monitorización para la herramienta de réplica de datos *SymmetricDS* presentada en (Carcasés et al., 2014), son soluciones libres. Pero, la fiabilidad<sup>7</sup> sobre el uso de estas herramientas depende de su compatibilidad con las arquitecturas de los sistemas que necesitan replicar datos (Alfonso, 2017). Las soluciones consultadas, para superar la limitante de *SymmetricDS* referente al uso de comandos para realizar las configuraciones y monitorización, terminan desarrollando aplicaciones web propias a su arquitectura.

Las soluciones más robustas en estas temáticas son las herramientas propuestas por Microsoft

---

<sup>6</sup> Un esquema Maestro-Maestro también denominado como Múltiples-Maestros permite configurar los nodos como maestro, permitiendo la réplica bidireccional; y enviar consultas de lectura/escritura a múltiples servidores replicados.

<sup>7</sup> "Probabilidad de buen funcionamiento de algo" (DLE, 2017a).

SQL Server y Oracle pero constituyen herramientas propietarias cuya utilización representa un alto costo por concepto de licencias y soporte (Oracle Help Center, 2018; Oracle Corporation, 2007).

- **C4. Consistencia:** se reconoce la consistencia como un aspecto fundamental para la correcta distribución de los datos.

M1, M2 y M4 consideran explícitamente que la consistencia es importante. M1 tiene como objetivo medir características de rendimiento, sin embargo, en él se considera que la replicación es un proceso que debe garantizar la consistencia de los datos replicados.

M2 se enfoca en el rendimiento, disponibilidad, escalabilidad y consistencia en sistemas tolerantes a fallos, donde se plantea que las organizaciones pueden presentar problemas si no cuentan con técnicas y métodos que garanticen una aceptable consistencia y oportunidad en los datos replicados, trabajando en un entorno tolerante a fallos, donde el coste económico de la red de comunicación es relativamente bajo.

M4 considera que la consistencia es uno de los requisitos más importantes que se debe considerar al implementar la replicación en un sistema distribuido. Sin embargo, solamente M2 realizó una mejor valoración del tema al comprender la consistencia desde el objetivo hasta la validación.

- **C5. Evaluación de la consistencia:** se fundamentan y validan las capacidades del sistema para contribuir a mantener la consistencia en los procesos asociados a la réplica.

Todos los casos no evidencian explícitamente cómo gestionan la consistencia en la distribución de los datos, lo cual limita las capacidades de los modelos para ofrecer un servicio de calidad con alto grado de consistencia.

En M1 no se simula un escenario donde ocurran fallas. Por tanto, no existe garantía de que el sistema se recupere automáticamente ante fallos. Se define el esquema de replicación de manera estática, por lo que no ofrece replicación para operaciones DDL y por ende se generarán inconsistencias en estos entornos.

M2 es el que mejor describe en su modelo el grado de la consistencia. Pero, el agente para la resolución de conflictos modelado no garantiza en su totalidad la resolución de conflictos, puesto que los resultados de las pruebas arrojaron ciertas inconsistencias a mayor número de transacciones. Simplemente se realizó una consulta sobre 50 vehículos, lo cual no responde a las características en cuanto al volumen de datos gestionados en un sistema replicado. Tampoco ofrece información sobre las acciones replicadas y aun así y experimenta un 2 % de error.

Mientras que M3 y M4, aunque reconocen la consistencia como un aspecto clave, no fundamentan ni evalúan el comportamiento de los procesos para valorar la consistencia de los datos.



### Conclusiones parciales del capítulo

Después de elaborar el marco teórico de la investigación y como resultado del análisis del estado del arte sobre los referentes teóricos de replicación de datos, se concluye que:

- La conceptualización de los elementos asociados al dominio permitió contextualizar los principales términos abordados en el capítulo y la investigación en general.
- La consistencia de los datos es un aspecto clave en la replicación para mantener actualizadas las BDs que se encuentran distribuidas.
- Por razones de rendimiento, la mayoría de los productos comerciales e investigaciones han direccionado hacia la utilización de esquemas Asíncronos. Sin embargo, esta mejora de rendimiento se ve empañada por la posibilidad de trabajar con datos inconsistentes. Por esta razón, es imprescindible proveer un mecanismo muy riguroso de resolución de conflictos.
- El análisis de modelos y soluciones de replicación de datos permitió sentar las bases e identificar buenas prácticas para la propuesta de solución.
- El diagnóstico realizado del conocimiento explícito abordado en modelos y herramientas de replicación de datos sobre los procesos implicados en la replicación, evidenció un conjunto de insuficiencias que atentan contra la capacidad de los modelos para contribuir a mantener la consistencia en la replicación de datos.

## CAPÍTULO 2. MODELO PARA LA RÉPLICA DE DATOS

Finalizada la fundamentación teórica de la investigación se procede a realizar el marco metodológico para guiar la concepción y elaboración del modelo propuesto. En este capítulo se desarrolla la propuesta que comprende: principios, características, premisas, estructura y componentes del modelo, así como indicaciones metodológicas para guiar su aplicación.

### 2.1 Concepción del modelo

El significado de modelo ha sido discutido, entre otros, por científicos, filósofos de la ciencia, psicólogos, lingüistas y educadores. El punto de vista más aceptado es que un modelo es una representación de una idea, objeto, acontecimiento, proceso o sistema, creado con un objetivo específico. Se pueden utilizar para: simplificar fenómenos complejos, ayudar en la visualización de entidades abstractas, servir de apoyo en la interpretación de resultados experimentales, servir de ayuda en la elaboración de explicaciones y en la propuesta de previsiones (Justi, 2006).

No existen reglas generales para la construcción de modelos, la capacidad de construir modelos es una destreza tácita, que debe de ser aprendida y no enseñada (Justi, 2006). Pero analizando, por una parte, lo que los filósofos de la ciencia y educadores dicen sobre cómo el conocimiento científico se desarrolla y por otra, cómo el trabajo de los científicos ha contribuido significativamente a ese proceso de desarrollo, es posible extraer algunos elementos que, organizados, producen un modelo cognitivo de la ciencia en forma de esquema, que permite discutir sobre el proceso que siguen los científicos en la construcción de modelos en general. Para la concepción del modelo se realizó un diseño metodológico que se basa en una combinación teórico práctica a partir de la propuesta para elaborar modelos representada en (Justi, 2006). La vida del modelo está definida por la dinámica de su funcionamiento y su aplicabilidad, por lo que se fue perfeccionando su construcción y desarrollo a medida que se elaboraba.



Figura 2 Concepción metodológica para elaborar el modelo

### 2.2 Principios, características y premisas del modelo propuesto

Para superar las insuficiencias anteriormente detectadas y con el propósito de desarrollar un modelo de replicación de datos con capacidad para contribuir a mantener la consistencia en la distribución de los datos en ambientes distribuidos, fueron definidos los siguientes principios:

- a) Orientado a contribuir a mantener la consistencia de los datos replicados en ambientes distribuidos, a partir de las prácticas positivas, métodos representados, y experiencias de expertos, en materia de modelos y herramientas de replicación de datos.
- b) La pertinencia como garantía de la adecuación del modelo en el contexto de la replicación de datos en ambientes distribuidos.
- c) La coherencia con las dimensiones y capacidades arquitectónicas usualmente empleadas para describir las soluciones de replicación.
- d) La adaptabilidad que se logra utilizando componentes con funcionalidades genéricas y sobre una estructura que facilita su aplicación a particularidades de los sistemas existentes.
- e) La actualización mediante la retroalimentación de la información que nutre al modelo que se logra analizando los elementos que faltan o que requieren ser adaptados cuando se identifican nuevos factores.

Las características del modelo propuesto son:

- a) Integralidad que comprende fundamentos teóricos, buenas prácticas, marcos de trabajo y experiencia de expertos; sobre modelos, herramientas de replicación y elementos necesarios para mantener la consistencia en ambientes distribuidos.
- b) Enfoque sistémico que se expresa mediante la gestión de los procesos asociados a la replicación de datos y la consistencia, a partir de la interacción de sus componentes e interdependencias de forma coherente y en conjunto.
- c) Mejora continua con el perfeccionamiento constante del modelo a partir de los resultados que se van obteniendo tras su aplicación en entornos de replicación y según las variaciones del contexto de réplica de datos.
- d) Amplitud que brinda la capacidad de describir múltiples procesos asociados a la replicación de datos y la posibilidad de emplearse en variados escenarios de replicación distribuidos.

Para la aplicación del modelo propuesto se establecieron las siguientes premisas:

- a) Clasificar el entorno de replicación de datos para identificar los métodos de replicación de datos propuestos en el modelo que serán necesarios aplicar en el desarrollo de la solución.
- b) Se deberá establecer un modelo de datos adecuado para la replicación de datos en ambientes distribuidos.

### **2.3 Modelo propuesto para la réplica de datos**

Sobre los principios anteriormente establecidos se elaboró la propuesta del modelo de réplica de datos (MRD) que como se muestra en la Figura 3 está compuesto por cuatro componentes relacionados entre sí. Estos integran fundamentos teóricos, buenas prácticas, marcos de trabajo y experiencia de expertos, sobre modelos y herramientas de replicación. El componente Réplica de datos describe el funcionamiento general del proceso de replicación de datos mediante la especificación del esquema lógico y distribución física que sustenta la propuesta. El componente Métodos de replicación, está diseñado para describir los procesos asociados a la réplica. Ambos

componentes requieren de la Administración de configuraciones, y para la construcción de todos los componentes se emplean herramientas, normas y servicios propuestos en el componente Gestión tecnológica. Finalmente, se plantean las indicaciones metodológicas para la implementación del modelo.

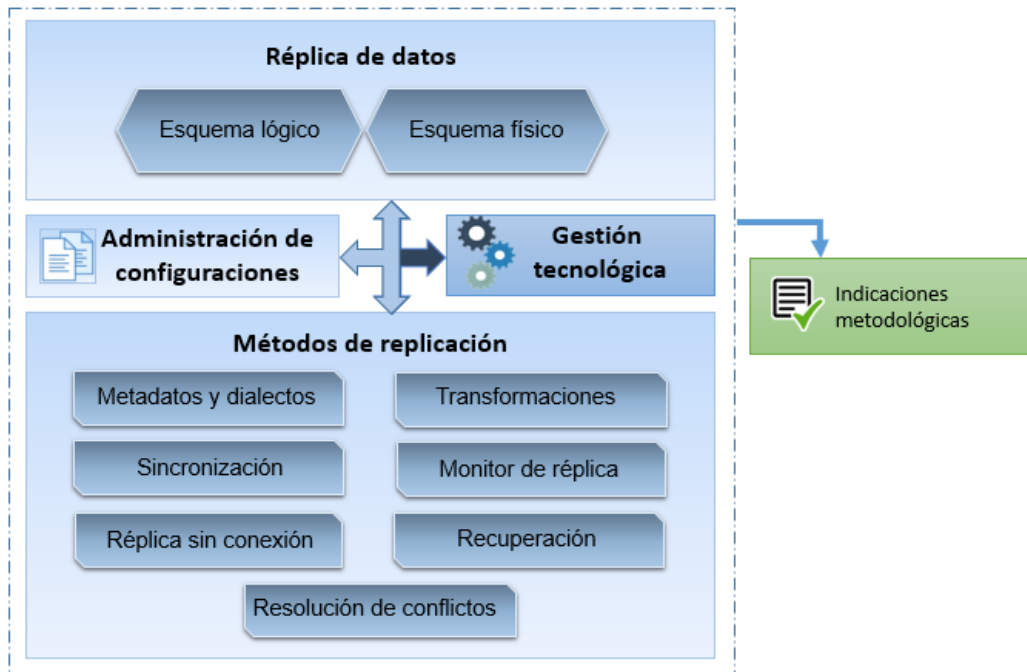


Figura 3 Representación gráfica del modelo

### 2.3.1 Componente Réplica de datos

El modelo de replicación de datos propuesto está compuesto por cuatro componentes: Consola de administración responsable de realizar las configuraciones principales del software como el registro de nodos y configuración de los datos a replicar, Capturador encargado de capturar los cambios que se realizan sobre la BD, Distribuidor responsable de realizar la propagación de los cambios empleando para ello el Servidor de mensajería ActiveMQ, y el Aplicador concepto designado para aplicar los cambios en la BD.

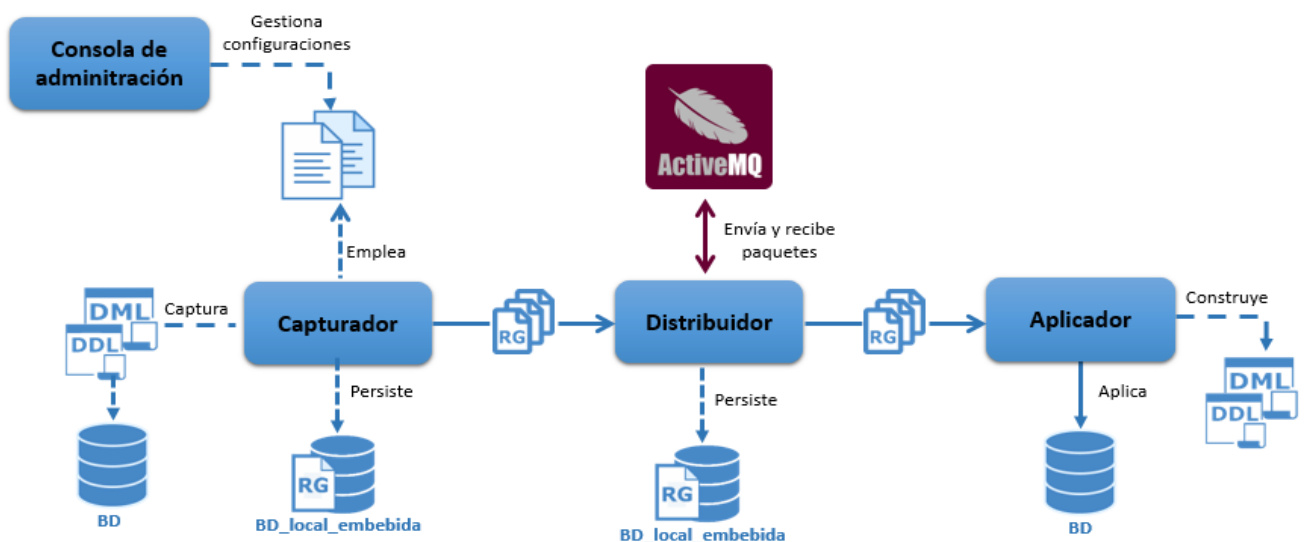


Figura 4 Propuesta general para la replicación de datos

En el modelo de replicación de datos propuesto intervienen varios conceptos que se detallan a continuación para su mejor comprensión:

- Base de datos: es la BD que contiene los datos a replicar. El software de réplica envía los cambios que se realizan sobre ella y aplica los cambios que provienen de otros nodos de réplica.
- BD local embebida: está diseñada para guardar las configuraciones propias de la réplica, las acciones sobre la BD que han dado conflicto al aplicarse y las acciones o transacciones que no han podido llegar a su destino.
- Consola de administración: para realizar las configuraciones principales del software como el registro de nodos, configuración de las tablas a replicar, datos a replicar y la monitorización del funcionamiento del software.
- Servidor JMS ActiveMQ: Servidor de Mensajería bajo la especificación Java Message Service, utilizado como punto intermedio en la distribución de la información enviada bajo JMS.
- Nodo: cada instancia del mecanismo se identifica como un Nodo con un identificador único. Los nodos pueden estar agrupados en etiquetas. El Nodo origen es aquella instancia que captura y envía el cambio, mientras que el Nodo destino, es la instancia que recibe y aplica el cambio.
- Replicable Action (Acción replicable, RA por sus siglas en inglés): estructura que almacena las modificaciones realizadas sobre las tuplas o filas de una tabla producto a la ejecución de sentencias DML. Está compuesta por la tabla y el tipo de acción que puede ser insert (insertar), update (actualizar) o delete (eliminar).
- Replicable Group (Grupo replicable, RG por sus siglas en inglés): almacena un grupo de acciones replicables con los cambios a realizar. Está compuesto por el listado de acciones y el o los nodos destinos.
- Replicable Structure Action (Acción de estructura replicable, RSA por sus siglas en inglés): estructura que almacena las modificaciones realizadas sobre las tablas, columnas, y otras definiciones como los esquemas producto a la ejecución de sentencias DDL. Está compuesta por el esquema, la tabla y el tipo de acción que puede ser create (crear), alter (modificar) o drop (borrar).
- Replicable Structure Group (Grupo de estructura replicable, RSG por sus siglas en inglés): almacena un grupo de acciones de estructura replicables con los cambios a realizar sobre las definiciones. Está compuesto por el listado de acciones y el o los nodos destinos.

### 2.3.1.1 Esquema lógico

El esquema lógico describe la esencia del sistema. Representa la estructura del modelo propuesto y el funcionamiento interno de sus componentes. Como se muestra en la Figura 5 la dinámica del proceso de replicación consiste en que cada ciertos intervalos de tiempo, definidos por el administrador de réplica, el Capturador de cambios captura y registra los cambios realizados sobre

la BD. Para realizar el proceso de captura se emplean los triggers DML y DDL introducidos en la BD previamente cuando se guardó la configuración de réplica. En las instancias donde los gestores configurados no soporten el manejo con triggers, deberán implementarse funciones que detecten dichos cambios que no puedan ser capturados con triggers.

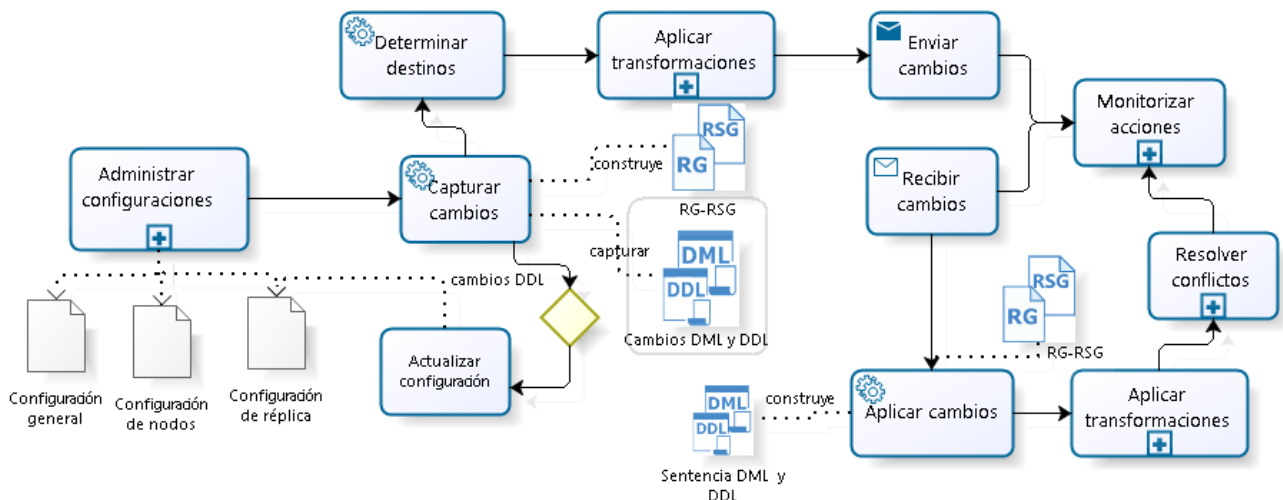


Figura 5 Esquema lógico del MRD propuesto

Una vez que se capturan los cambios, debe almacenarse en las estructuras de control toda la información correspondiente al cambio, de tal forma que cada tupla de la tabla de control se refiera a un cambio, y a partir de cada tupla se construirá una RA. Los cambios a replicar se construyen a partir de los cambios registrados en la BD teniendo en cuenta las operaciones DML o DDL ejecutadas en la BD. Si el cambio se produce por la ejecución de sentencias insert, update o delete se construye una RA para almacenar los cambios realizados sobre los datos contenidos en las tablas de la BD. Mientras que si el cambio se produce por la ejecución de sentencias create, alter o drop se construye una RSA para almacenar los cambios realizados sobre las definiciones de tablas, columnas y esquemas de la BD. Es necesario señalar que si en la captura se detectan cambios sobre las definiciones de las BDs debe actualizarse la configuración de réplica, los triggers y las estructuras de control de la BD, para garantizar que se realice la captura de los cambios de manera correcta, no se pierdan los datos, y por ende evitar inconsistencias entre las BDs que se encuentran distribuidas.

A partir de estas entidades, las configuraciones registradas y los filtros asociados a estas, se determinan los destinos hacia donde se enviará el grupo replicable. De la propagación de los grupos replicables se encarga el Distribuidor de datos empleando el servidor de JMS. Si el entorno de replicación es heterogéneo, antes de realizar el envío de los grupos deben aplicarse las transformaciones, con el objetivo de mapear las bases de datos según la configuración de réplica previamente almacenada. Al enviar y recibir las transacciones se controla el estado de la réplica de datos mediante el proceso de monitorización. En cada nodo destino, se comprueba que la transacción es la que corresponde aplicar debido a que, para evitar conflictos e incoherencias en las BDs, las acciones se deben aplicar en el mismo orden en que se ejecutaron en la BD y se

capturaron en el sistema. Para conocer si la acción es la que corresponde, cada instancia local lleva un control sobre los grupos replicados hacia las instancias remotas, a través de un contador de acciones enviadas entre todas las instancias involucradas. Si es el grupo replicable esperado, se pasa al Aplicador de cambios para ejecutar en la BD del destino sus acciones. Para ello, se construye un SQL por cada acción replicable almacenada en el grupo replicable. Si al ejecutar el SQL se produce un error en la BD se procede a detectar y resolver los conflictos. Si el contador es menor que el esperado se descarta la agrupación y si es mayor se envía una notificación a la instancia remota de contador incorrecto, para solicitar el grupo con el contador esperado. De esta manera, se establece el orden de ejecución de las acciones y se garantiza que los grupos no se pierdan ante la ocurrencia de fallos en la red u otra eventualidad que impida que el grupo arribe a su destino.

Se debe comprobar además el envío, confirmación de recepción y registro en BDs locales embebidas, y de los grupos replicables; para garantizar la integridad y persistencia de los datos que se replican ante posibles eventualidades. La confirmación de envío y recepción se propone realizar mediante el envío de notificaciones de acuse de recibo que a su vez se visualizarían en el Monitor de réplica, de tal forma que el administrador pueda observar que el grupo fue enviado, recibido y aplicados sus cambios. Y se propone hacer uso de bases de datos embebidas en el propio sistema para persistir las acciones construidas, enviadas y recibidas, así como cualquier operación de actualización sobre las acciones y agrupaciones, para evitar pérdida de los datos que conlleven a inconsistencias entre las BDs.

### 2.3.1.2 Esquema físico

El esquema físico muestra cómo puede ser desplegado el sistema, incluyendo el hardware, software y los archivos, involucrados en el sistema.

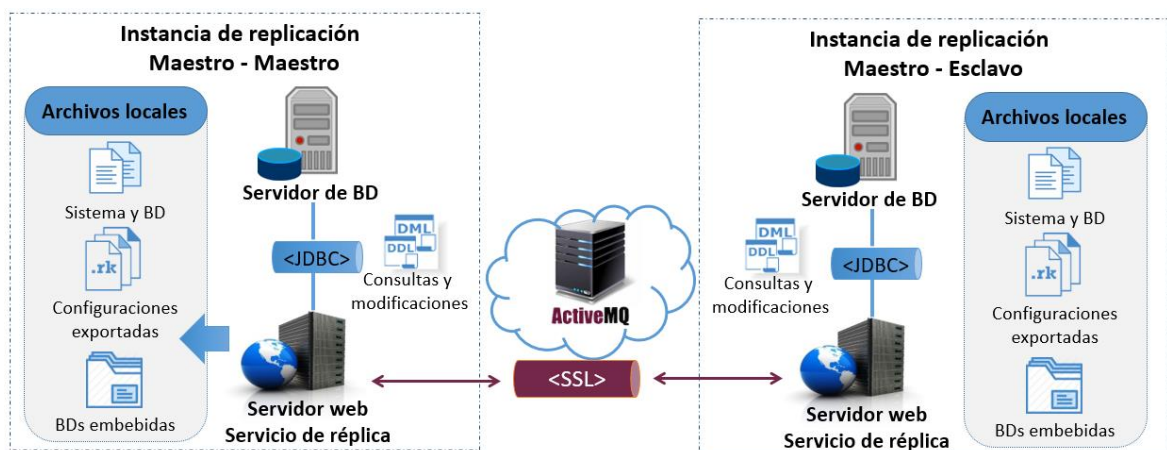


Figura 6 Esquema físico del MRD propuesto

Como se ilustra en la Figura 6 el modelo que se propone comprende un conjunto de computadoras independientes, llamadas nodos o sitios, conectadas vía una red de comunicaciones. Inicialmente los administradores de réplica deben configurar el entorno de réplica deseado. Cada instancia de replicación estará compuesta por un servidor de aplicaciones con el servicio de réplica desplegado

y el servidor de BD sobre el cual se realizan las consultas y modificaciones de los datos. Los datos son almacenados en todos o algunos de los sitios, dependiendo de la configuración de réplica realizada. Los sitios se comunican entre ellos a través de mensajes, empleando el servidor de mensajería ActiveMQ. Los mensajes son intercambiados entre nodos del sistema de replicación a través de una red de comunicación fiable y asíncrona que verifica los conflictos, y la pérdida de transacciones.

### 2.3.1.3 Almacenamiento de datos permanentes

En los componentes propuestos se realiza el almacenamiento de datos permanentes, tales como: logs (registros) de la BD y del software de replicación, ficheros con extensión “*rk*” (formato de los archivos del sistema REKO) exportados, ficheros con extensión “*xml*” para almacenar el historial de exportaciones y ficheros del directorio “*db*” referentes a la persistencia en BDs embebidas.

Se propone almacenar los paquetes y transacciones en BDs embebidas en el propio sistema, con el objetivo de evitar la pérdida de los paquetes ante eventualidades como fallos en la red o caída de una instancia de réplica. Se debe realizar la persistencia en BDs embebidas en los siguientes momentos:

- Captura: al finalizar la captura, cuando se construyen los grupos replicables que contienen los cambios DML y/o DDL a replicar. De tal forma que, si se interrumpe en ese momento el servicio de réplica, el sistema al restablecerse puede reiniciar el envío de las acciones pendientes, sin necesidad de reconstruirlas o comparar las estructuras de las BDs para verificar la consistencia y determinar las diferencias, debido a que en el nodo local se encuentra persistido el grupo con las acciones replicables pendientes por ser enviadas.
- Envío: cuando se conforman las acciones por destinos y se construyen los grupos para cada destino, antes de ejecutar el envío hacia el servidor de mensajería. De tal forma que, si no hay conexión con la instancia remota, al iniciar el sistema es posible realizar el envío de los grupos pendientes por ser enviados, sin necesidad de realizar nuevamente el proceso anteriormente descrito.
- Recepción: inmediatamente cuando el grupo que contiene los cambios es recibido. De tal forma que, si se interrumpe el sistema en ese momento, es posible iniciar la aplicación de los cambios cuando este reinicie. Con ello se evita que el cambio no se aplique en la BD y por ende que se generen inconsistencias, y también se evita esperar porque el nodo local nuevamente realice el envío del grupo replicable.

### 2.3.2 Componente Administración de configuraciones

Para elaborar el componente Réplica de datos y poder iniciar el proceso de captura de los cambios es necesario establecer el entorno de replicación. En el entorno de replicación se deben:

- Identificar los nodos que forman parte del sistema distribuido y las conexiones a los servidores para poder establecer las comunicaciones entre las instancias, que se proponen establecer mediante configuraciones generales del sistema.



- Registrar los nodos y establecer la dirección o sentido de transmisión de los datos, que se proponen realizar mediante la configuración de nodos.
- Especificar los cambios a replicar entre instancias que se propone condicionar mediante la configuración de réplica.

Se puede ofrecer una solución que, de acuerdo con la dirección en que fluyen los datos, brinde réplica en un solo sentido (unidireccional) o réplica en ambos sentidos (bidireccional). En ambos casos es posible determinar el sentido mediante configuraciones de réplica, donde en cada nodo se definirían las tablas a replicar y hacia qué nodo se enviarán los cambios. Además, mediante las configuraciones enunciadas anteriormente es posible realizar el proceso de réplica de forma desatendida, donde no se requiera la presencia de un administrador de réplica, porque mediante las configuraciones el sistema puede responder a las necesidades del cliente de manera automática.

### 2.3.2.1 Configuraciones generales

La configuración general incluye una serie de parámetros iniciales, necesarios para definir el entorno de replicación, tales como:

- Datos de identificación del nodo (identificador, nombre, descripción): dado que pueden existir múltiples clientes, cada cliente ha de poseer un identificador único.
- Base de datos (nombre del driver, url, usuario, contraseña): porque pueden converger varias BDs en el mismo servidor de BD, por esta razón debe especificarse el nombre de la BD sobre la cual se realizan las consultas y modificaciones, así como el driver (controlador de dispositivo) debido a que se describe una solución para gestores diferentes (PostgreSQL, MySQL, Microsoft SQL Server y Oracle).
- Resolución de conflictos: para especificar la forma de resolver los conflictos detectados (Nodo ganador o Nodo perdedor) y variante de solución.
- Asignación de permisos a objetos de réplica: dado que, para evitar ciclos en la réplica bidireccional, se determinan los usuarios con permisos a las tablas.
- Programar captura y envío: debido a que el modelo general de réplica presentado se basa en replicación asíncrona podría proveer las siguientes alternativas para elegir el momento de replicar:

Automáticamente: cuando se capturen los cambios, se envían automáticamente.

Programado: los cambios se envían según una frecuencia específica que se puede determinar en las configuraciones generales.

Manual: el administrador manualmente iniciaría el envío de los cambios.

Para realizar la configuración general se deben establecer, guardar y actualizar los parámetros anteriormente mencionados en una configuración general, y además actualizar el estado de las conexiones a los servidores y estado de réplica en un panel de notificaciones. El panel de notificaciones le permitiría al usuario visualizar desde cualquier interfaz el estado de las conexiones.

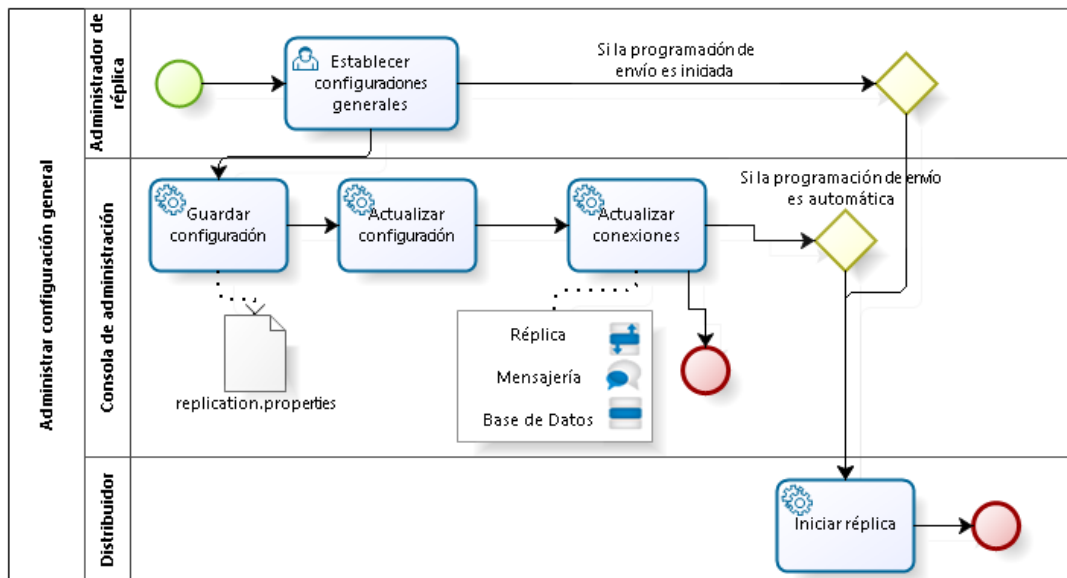


Figura 7 Proceso “Administrar configuraciones generales”

Si en la programación y captura de los cambios se determina el modo Automático, inmediatamente se inicia el envío de los grupos replicables que contiene las modificaciones realizadas en la BD hasta el momento. Esta información se consume de la BD embebida que almacena todos los cambios realizados en la BD. Al determinar el modo Automático se crea una tarea programada en el sistema con el objetivo realizar el envío de transacciones cada cierto tiempo definido en dicha tarea.

### 2.3.2.2 Configuraciones de nodos

Las configuraciones de nodos permiten especificar las instancias involucradas en el entorno de replicación. Para realizar la solicitud de registro se especifica el identificador del nodo con el que desea efectuar el registro (nodo remoto). La petición de registro de nodos transita por los siguientes estados: “Esperando respuesta remota”, “Registrado”, “Rechazado por el nodo local”, “Rechazado por el nodo remoto”, “Esperando respuesta local”, “Rechazado permanentemente por el nodo local” y “Rechazado permanentemente por el nodo remoto”.

De cada nodo se almacena el identificador, descripción y un contador que almacena la cantidad de envíos que el nodo origen realiza hacia el nodo remoto. La configuración de nodo debe almacenarse en una BD embebida para prevenir que, ante cualquier interrupción de una instancia, los usuarios de réplica se vean obligados a realizar nuevamente el registro de los nodos. A continuación, se representa el proceso para realizar la configuración de nodos.

En un escenario de replicación, varias instancias pueden contener la misma configuración de réplica debido a que existe la posibilidad de que una instancia replique el mismo cambio hacia un grupo de nodos, procediendo de igual forma. Por esta razón y con el objetivo de brindar mayor capacidad a las soluciones en cuanto a usabilidad, se propone emplear etiquetas. La etiqueta es el concepto que maneja una configuración de réplica y que, al aplicarse a varios nodos, se evita crear la misma configuración para cada nodo.

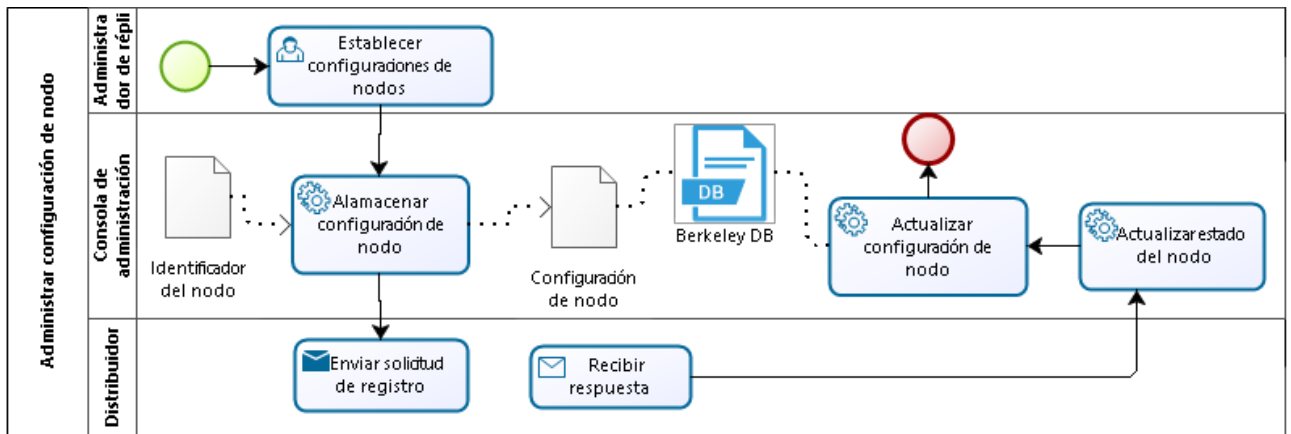


Figura 8 Proceso “Administrar configuraciones de nodos”

El proceso anteriormente descrito se aplica en aquellos entornos que trabajan con conexión. Sin embargo, se propone adicionar la posibilidad de que el registro de los nodos pueda realizarse de manera offline (sin conexión) mediante tareas de exportación de configuración mediante un fichero en el nodo local e importación del fichero en el nodo remoto. Ambos métodos se detallan más adelante en el acápite referente a la Réplica offline.

### 2.3.2.3 Configuraciones de réplica

A partir de un nodo origen o fuente pueden realizarse distintas configuraciones según los nodos y/o etiquetas destino que se definan, realizando de esta manera configuraciones de réplica para nodos y/o etiquetas. En cada configuración se condicionan los cambios según el entorno de réplica deseado. Se determinan las acciones DML y DDL a replicar mediante:

- Configuración sobre tablas (DML): se realizan configuraciones independientes de las tablas que se desean replicar. Cada tabla puede ser configurada para replicar según la acción que se efectúe sobre ella: inserción, actualización y/o eliminación. Además, pueden definirse filtros SQL para determinar por cada acción si se replicará o no el cambio, y realizar transformaciones a partir de una configuración de entrada, para posteriormente replicar datos hacia otro tipo de gestor de BD.
- Configuración sobre definiciones (DDL): cada instancia puede ser configurada para replicar, según las acciones: creación, modificación y eliminación, de las definiciones de la BD.

En cualquiera de los casos al establecer una configuración de réplica, deben crearse los triggers correspondientes a la acción especificada y por ende iniciar la detección de los cambios. Se propone crear una estructura de control en la BD para almacenar toda la información referente al cambio. Esta tabla almacenaría: nombre de la tabla, usuario con permisos a la tabla, tipo de acción y parámetros referentes a la sentencia.

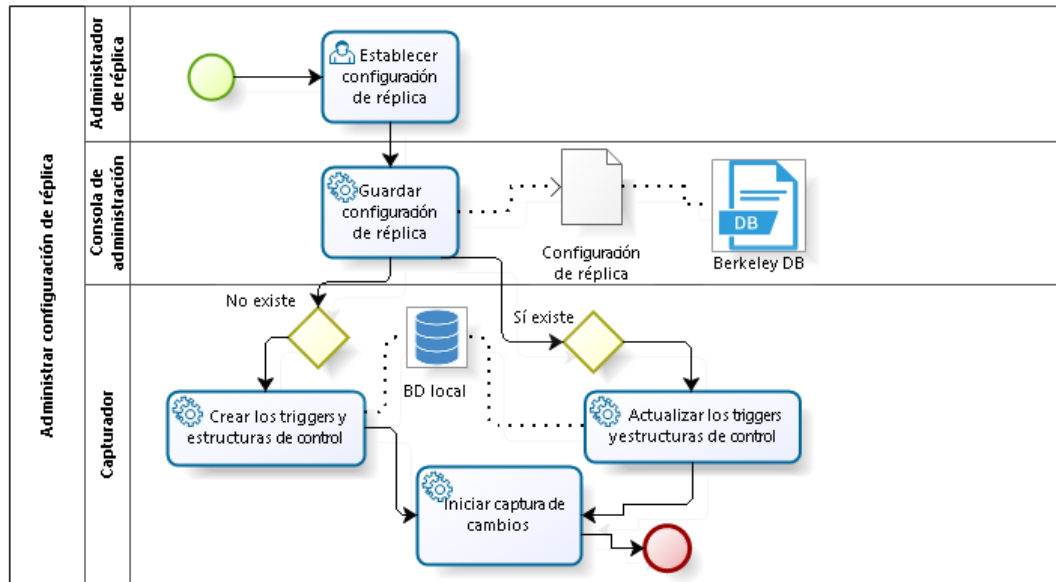


Figura 9 Proceso “Administrar configuraciones de réplica”

Si se modifica una configuración de réplica previamente almacenada, es importante señalar, que el sistema debe, además de actualizar los datos de la configuración de réplica persistida en la BD local embebida, actualizar los triggers y estructuras de control para garantizar que se realice la captura de los cambios de manera correcta, no se pierdan los datos y que por ende se eviten inconsistencias entre las BDs que se encuentran distribuidas.

### 2.3.3 Componente Métodos de replicación

En (Pérez, 2006) se enuncian los servicios que debe proveer la replicación de datos y en la revisión bibliográfica sobre las tendencias actuales de replicación se caracterizan las soluciones. A partir de la fundamentación teórica, para aumentar las capacidades de las soluciones de replicación de ofrecer un servicio que mantenga la consistencia de los datos replicados, se definieron los métodos referentes a los procesos de replicación que a continuación se describen.

#### 2.3.3.1 Metadatos y dialectos para el manejo de SQL

El metadato es un dato estructurado sobre la información, o sea, información sobre información, o de forma más simple, datos sobre datos (Lapuente, 2013). En el contexto de BD, son información sobre los datos y definiciones de la BD, que se pueden guardar, intercambiar y procesar en el sistema de replicación, y que están estructurados de tal forma que permiten ayudar a la identificación, descripción, clasificación y localización el contenido de los cambios realizados en la BD para su replicación.

Los metadatos son indispensables para que las soluciones de réplica almacenen las definiciones y datos de la BD cuando el sistema inicia. Para construir las sintaxis de las operaciones DML y de las DDL se requiere que el sistema almacene previamente las estructuras y datos de la BD. Para construir las acciones DML basta con conocer la información sobre las tablas y columnas, mientras que para construir las acciones DDL se recomienda construir todas las definiciones posibles de la BD tales como: esquemas, tablas, columnas, restricciones, funciones, secuencias, y otros como los

índices; para ofrecer una solución más completa y evitar incoherencias entre las BDs al no gestionar los cambios sobre una definición específica.

Para obtener la información de la BD se puede utilizar la API<sup>8</sup> [DatabaseMetaData](#) mediante las funcionalidades que aprovisiona, detalladas en (IBM Knowledge Center, 2015), y el resto de la información podrá capturarse mediante las consultas definidas en los dialectos desarrollados por la solución.

El Diccionario Oxford define un dialecto como una forma particular de un lenguaje que es propio de una determinada región o grupo social. Una versión de un idioma que se habla en una zona geográfica en particular o de un grupo particular de personas (Oxford University Press, 2013). A partir de este concepto se concluye que los dialectos informáticos, para los gestores desarrollados por una solución de replicación, constituyen la definición de la sintaxis de los SQL correspondientes para cada gestor, que se diseñan e implementan en el sistema para ejecutar cualquier operación DML o DDL sobre los gestores de BD, respectivamente.

Por tanto, la solución de replicación deberá diseñar un dialecto para cada gestor de BD con el cual necesita operar. A través de cada dialecto se definirán las sintaxis SQL correspondientes al gestor.

### 2.3.3.2 Sincronización

El proceso de sincronización se realiza entre el nodo local y un nodo remoto. El administrador determina el nodo remoto que desea sincronizar. Al realizar la sincronización se debe comprobar que exista una configuración de réplica del nodo local hacia el nodo remoto. Por lo tanto, los nodos con los cuales se permitirá sincronizar serán aquellos que posean una configuración de réplica. El proceso de sincronización radica en homogeneizar los datos y estructuras del nodo remoto con el nodo local según la configuración de réplica registrada.

Cuando el administrador del nodo local determina que desea sincronizar con el nodo remoto, se envía una petición de sincronización hacia el nodo remoto para que autorice la modificación de su BD. La petición transita por los siguientes estados:

- Aceptar: el nodo remoto permite la sincronización.
- Rechazar: el nodo remoto se opone a la sincronización y rechaza permanentemente cualquier solicitud de ese nodo.
- Cancelar: el nodo remoto en ese momento anula la petición.

Cuando el nodo local recibe la confirmación, se construyen los grupos replicables con los cambios y se envían a través del distribuidor de cambios, quien se responsabiliza de su propagación con el empleo del servidor de mensajería Apache ActiveMQ. Recibidos los grupos por el nodo remoto se procede a aplicar los cambios arribados, teniendo en cuenta los tipos de acciones siguientes:

---

<sup>8</sup> Interfaz de Programación de Aplicaciones, del inglés [Application Programming Interface](#).

- Para las acciones replicables, correspondientes a las operaciones DML, por cada acción se construye una acción de inserción sobre la tabla configurada. Si al ejecutar la acción se produce un error en la BD se generará un conflicto de unicidad automáticamente.
- Para las acciones de estructuras replicables, correspondientes a las operaciones DDL, se deben extraer las definiciones contenidas en cada RSA que arribaron del nodo, y se procede a compararlas con las estructuras locales de los esquemas de la BD. Al finalizar el proceso de comparación se entregan las diferencias encontradas al Aplicador de cambios, quien se encargará de ejecutarlas en la BD.

Si se produce un error en la BD durante la aplicación de los cambios se procede a resolver el conflicto generado según la configuración establecida, para lo cual se debe brindar la posibilidad de configurar la resolución de conflictos para cada nodo.

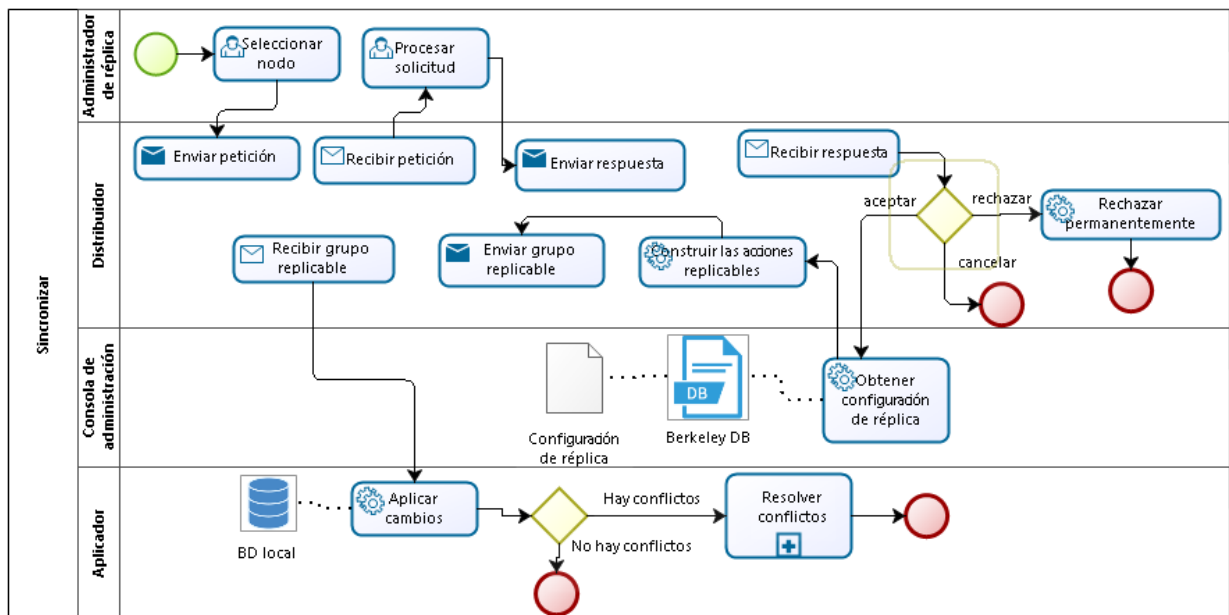


Figura 10 Proceso "Sincronizar"

Para automatizar el proceso de sincronización e impulsar la réplica desatendida se propone adicionar a la configuración general o al mecanismo de sincronización, la programación para la sincronización mediante una configuración, donde se defina en el nodo local y remoto su comportamiento y de esta forma responder a la solicitud de sincronización de manera automática.

### 2.3.3.3 Réplica offline

En ambientes distribuidos eventualmente pueden ocurrir fallas como interrupciones en los servicios de réplica desplegados. Estas fallas generan períodos de tiempo en los cuales las bases de datos pueden almacenar datos inconsistentes. Para mantener la consistencia de los datos bajo este escenario, se propone una solución que consiste en realizar la réplica sin conexión de forma manual. Para describir la solución se ha diseñado el siguiente algoritmo que cuenta con cuatro pasos detallados a continuación:

**Paso 1. Registrar nodos de manera offline:** para efectuar la réplica sin conexión es necesario registrar primeramente los nodos en estas condiciones. Para ello el administrador del sistema debe exportar la configuración del nodo local, y mediante correo u otra alternativa, debe enviar esta configuración al administrador de la instancia remota, quien a su vez importa esta configuración, logrando realizar el registro de nodos sin conexión.

Para exportar la configuración del nodo, como se muestra en la Figura 11, el sistema debe permitir descargar los datos del nodo local para un archivo. El archivo debe almacenar la configuración del nodo local que se encuentra almacenada en la BD embebida. El archivo puede ser almacenado en un directorio específico de la computadora teniendo por nombre el identificador del nodo en cuestión.

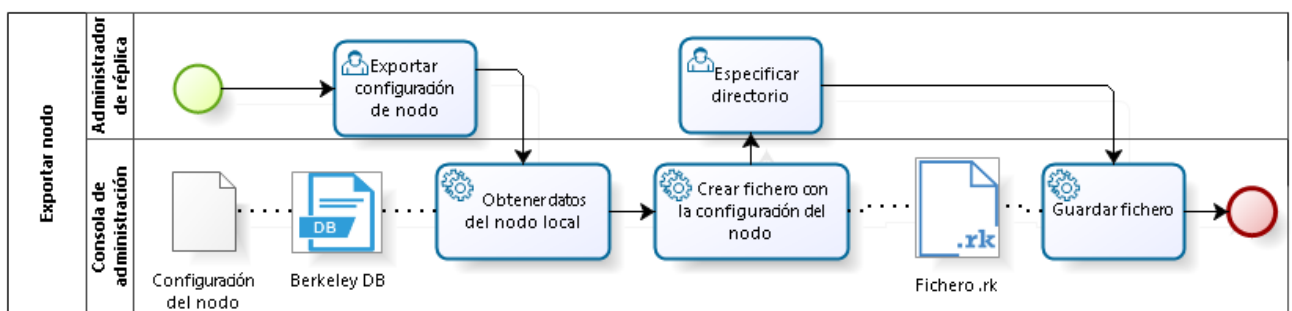


Figura 11 Proceso "Exportar nodo"

Para importar la configuración del nodo, como se muestra en la Figura 12, el sistema debe permitir que el administrador pueda cargar el archivo que contiene la configuración del nodo con el que se desea efectuar el registro. El sistema debe permitir que el archivo pueda ser cargado desde un directorio específico de la computadora.

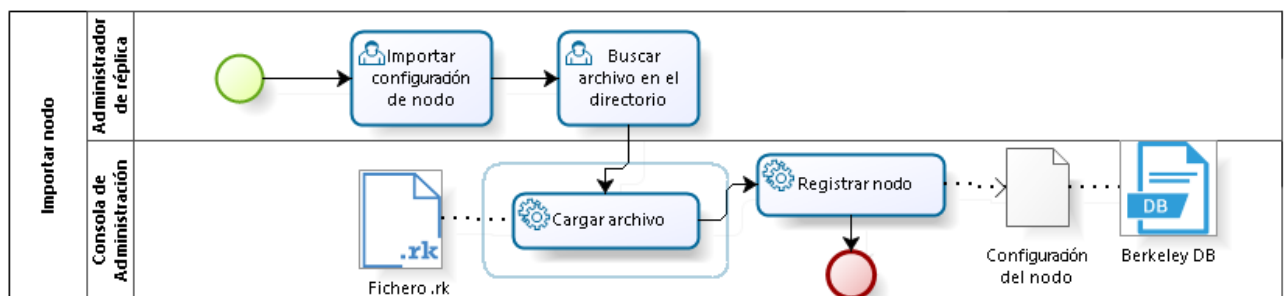


Figura 12 Proceso "Importar nodo"

**Paso 2. Registrar configuración de réplica:** se deben establecer las configuraciones de réplica según el entorno de replicación a desplegar. Si se desea registrar una configuración, a partir de otra previamente establecida en un nodo remoto, se debe exportar la configuración de réplica en cuestión e importarla en el nodo donde se realiza el nuevo registro. Para ello el administrador del sistema debe exportar la configuración de réplica, y mediante el correo u otra alternativa, debe enviar esta configuración al administrador de la instancia. El administrador realiza el registro mediante la importación del fichero que almacena la configuración.

Para exportar la configuración de réplica como se muestra en la Figura 13, se debe permitir al usuario exportar la configuración a partir del listado que contiene todas las configuraciones registradas en el nodo local. Cuando el usuario selecciona la configuración deseada, el sistema obtiene los datos almacenados en la BD embebida de la configuración indicada. Posteriormente el sistema crea un fichero con la configuración que se almacena en el directorio especificado por el usuario.

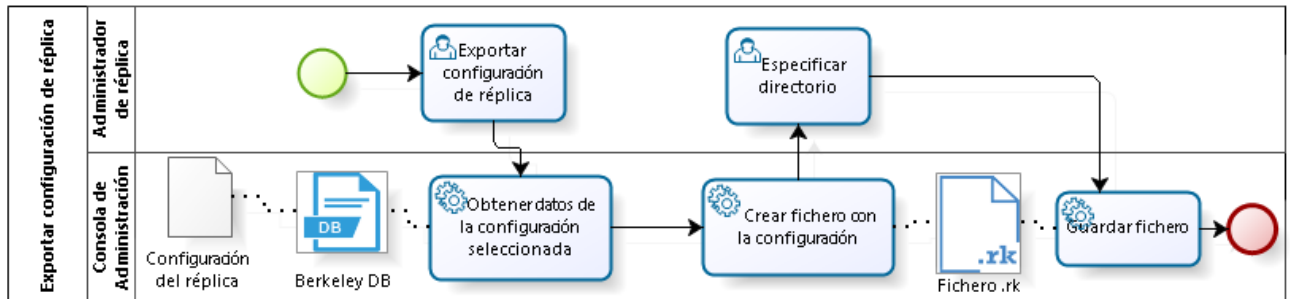


Figura 13 Proceso “Exportar configuración de réplica”

Para importar la configuración de réplica, como se muestra en la Figura 14, el sistema debe permitir al usuario o administrador cargar una configuración de réplica y asignarla a un Nodo o Etiqueta previamente registrados. Primeramente, el sistema debe mostrar una vista que le permita al usuario buscar la ubicación del fichero que contiene la configuración. Una vez que el usuario selecciona el fichero, el sistema debe cargar automáticamente la configuración de réplica almacenada en el fichero y permitirle al usuario que la modifique si así lo desea. Cuando el usuario seleccione el Nodo o Etiqueta deseado, el sistema debe almacenar la configuración correspondiente al Nodo o Etiqueta en la BD embebida.

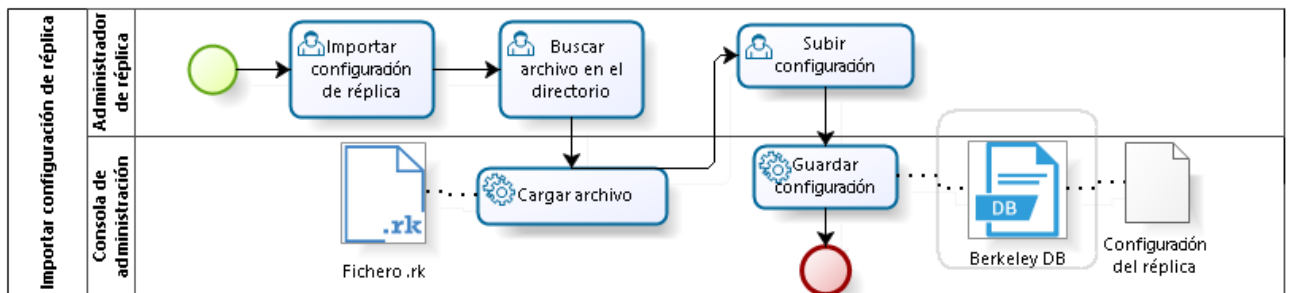


Figura 14 Proceso “Importar configuración de réplica”

**Paso 3. Exportar cambios de un nodo local o nodo origen de datos:** cuando el nodo local tiene almacenados cambios que necesitan ser replicados, pero no hay conexión entre las instancias con el servidor de mensajería, las bases de datos pueden almacenar inconsistencias temporales. Para solucionar este problema se propone realizar la replicación de datos fuera de línea, es decir, de forma offline mediante la exportación de un archivo que almacenará los cambios a replicar del nodo local y que será importado en el nodo remoto.

Para realizar este proceso, como se muestra en la Figura 15, deben estar registrados los nodos involucrados y el nodo remoto debe contar con la configuración de réplica en el nodo local. Se deben construir los cambios a replicar (acciones replicables) a partir del nodo seleccionado, su



configuración registrada, y también la fecha especificada. La fecha determina que el grupo a replicar almacenará las acciones correspondientes a los cambios realizados hasta dicha fecha. El fichero se crea, posteriormente se guarda en la ruta especificada por el usuario cuando lo descarga y se monitoriza mediante un histórico de exportaciones a través de un fichero “xml”.

Se pueden realizar en cada nodo todas las exportaciones que se deseen realizar con diferentes fechas, por lo que el fichero transitará por los siguientes estados:

- **Aceptar:** se acepta en el nodo local cuando se ha confirmado que el nodo remoto ha importado los cambios, se elimina del histórico el fichero y se eliminan del sistema las acciones que se encuentran almacenadas en el fichero y que han sido aceptadas teniendo la certeza que han sido aplicadas exitosamente en el nodo remoto.
- **Cancelar:** se elimina del histórico el fichero cancelado, pero las acciones contenidas se mantienen en el sistema para que posteriormente puedan ser replicadas en otro momento ya sea con conexión o sin conexión.

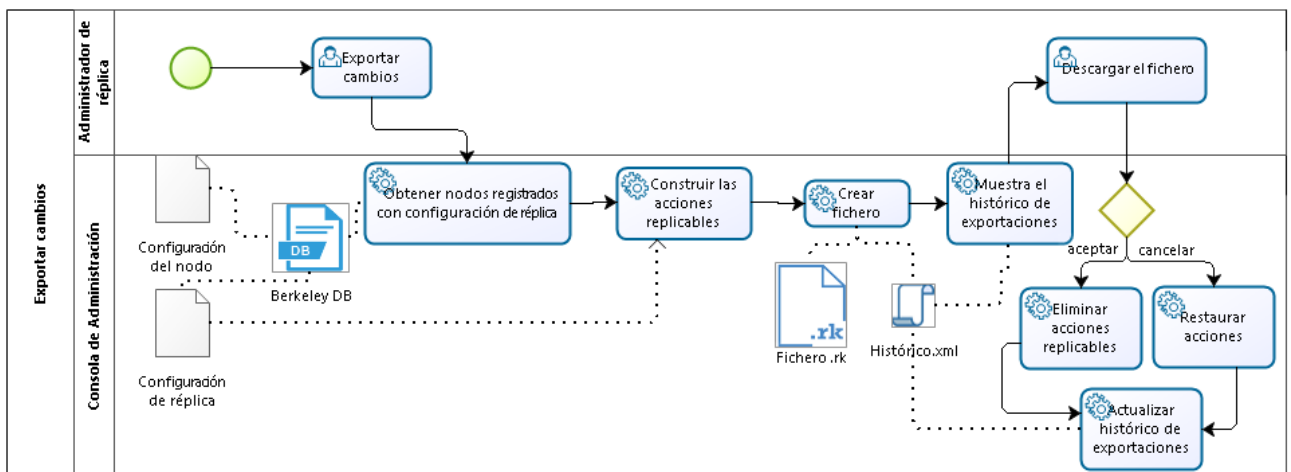


Figura 15 Proceso “Exportar cambios”

**Paso 4. Importar cambios de un nodo remoto o nodo destino:** para importar los cambios almacenados en el fichero exportado, como se muestra en la Figura 16, el sistema debe permitir al usuario o administrador importar el archivo que contiene las acciones replicables, que deben aplicarse en la BD para mantener la consistencia de los datos. El sistema debe mostrar una vista que le permita al usuario buscar la ubicación del fichero. Al importar el fichero se deben construir las acciones replicables correspondiente a la información de los cambios almacenada en el fichero. Finalmente, las acciones construidas deben aplicarse en la BD para mantener la coherencia con respecto a la BD del nodo local.

De producirse un error al aplicar los cambios en la BD, se resuelve el conflicto de la forma y variante previamente configurada. Se deben monitorizar las acciones contenidas en el fichero mediante un histórico de importaciones que almacena por cada tipo de acción la cantidad de acciones recibidas y aplicadas en la BD.

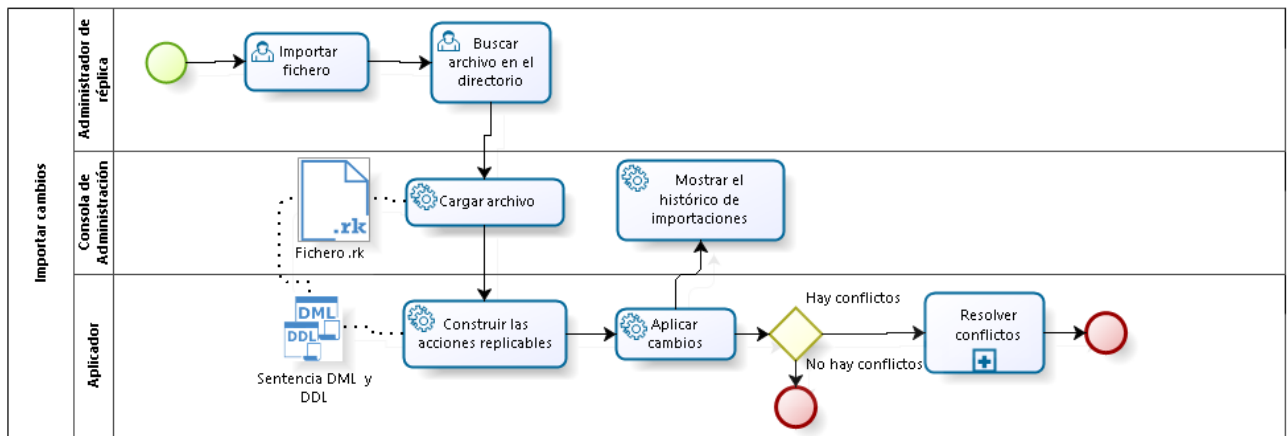


Figura 16 Proceso "Importar cambios"

### 2.3.3.4 Resolución de conflictos

Al aplicar los cambios en la BD, como parte de los procesos de réplica con conexión, sin conexión y sincronización, es posible que la ejecución de la sentencia DML o DDL no resulte exitosa y con ello se generen datos y/o definiciones incompletas que provocan inconsistencias entre las BDs. Por esta razón, se hace necesario diseñar un mecanismo de resolución de conflictos que lleve a la BD a un estado consistente.

La dinámica del proceso consiste en capturar el conflicto cuando se produce el error al aplicar la sentencia DML o DDL en la BD, luego se clasifica y finalmente se resuelve según el tipo de conflicto y variante configurada para la resolución en las configuraciones generales. El conflicto se clasifica mediante el SQL State<sup>9</sup> que genera la BD cuando se produce el error al aplicar el SQL y las variantes determinan la manera de resolver los conflictos.

Para definir las variantes de solución al conflicto capturado, se consideraron las siguientes condiciones:

- Si se define el nodo remoto como Ganador entonces la información recibida que entra en conflicto tiene prioridad y prevalece sobre la existente en el nodo local.
- Si se define el nodo remoto como No Ganador entonces la información recibida que entra en conflicto no tiene prioridad sobre la existente en el nodo local.
- En cualquier variante de solución que sea aplicada, debe mantenerse la consistencia entre las BDs.

En el escenario de réplica cada nodo local puede comportarse de varias formas con respecto a las restantes instancias involucradas en el proceso de replicación; para algunos nodos como nodo Ganador y para otros nodos como nodo No Ganador. Por esta razón en cada instancia debe especificarse, en la configuración general, las variantes de solución y además el tipo de nodo. Las variantes de solución propuestas se encuentran detalladas en (Martínez et al., 2015).

<sup>9</sup> Error que se muestra cuando se ejecuta una acción incorrecta en el gestor de BD, en el panel de salida en la pestaña mensaje.

Cada vez que sea capturado un conflicto debe ser resuelto. Si el nodo local no está en conflicto y no tiene grupos pendientes por ser aplicados se procederá a resolverlo, de lo contrario será persistido en una BD embebida para esperar a que pueda ser aplicado. Si la resolución del conflicto es automática se resolverá automáticamente según la variante que se haya definido en las configuraciones generales. Si la configuración definida es manual, en el momento que se genere el conflicto el usuario podrá decidir qué variante aplicar para dar solución al mismo, teniendo en cuenta las variantes definidas para el tipo de conflicto generado. Si existe una configuración manual para un conflicto de eliminación y el nodo que recibe la información es ganador el sistema automáticamente deberá ignorar el cambio.

Cuando el nodo local resuelva el conflicto serán monitorizadas las acciones de las instancias involucradas con el conflicto. Para monitorizar el proceso de resolución de conflictos generados durante la réplica se mostrarán todos los eventos referentes a las acciones que se encuentren en conflicto, visualizando en cada caso el SQL y el tipo de conflicto, entre otros parámetros propios del Monitor. Se mantendrá informado al usuario local y remoto del proceso realizado para resolver el conflicto.

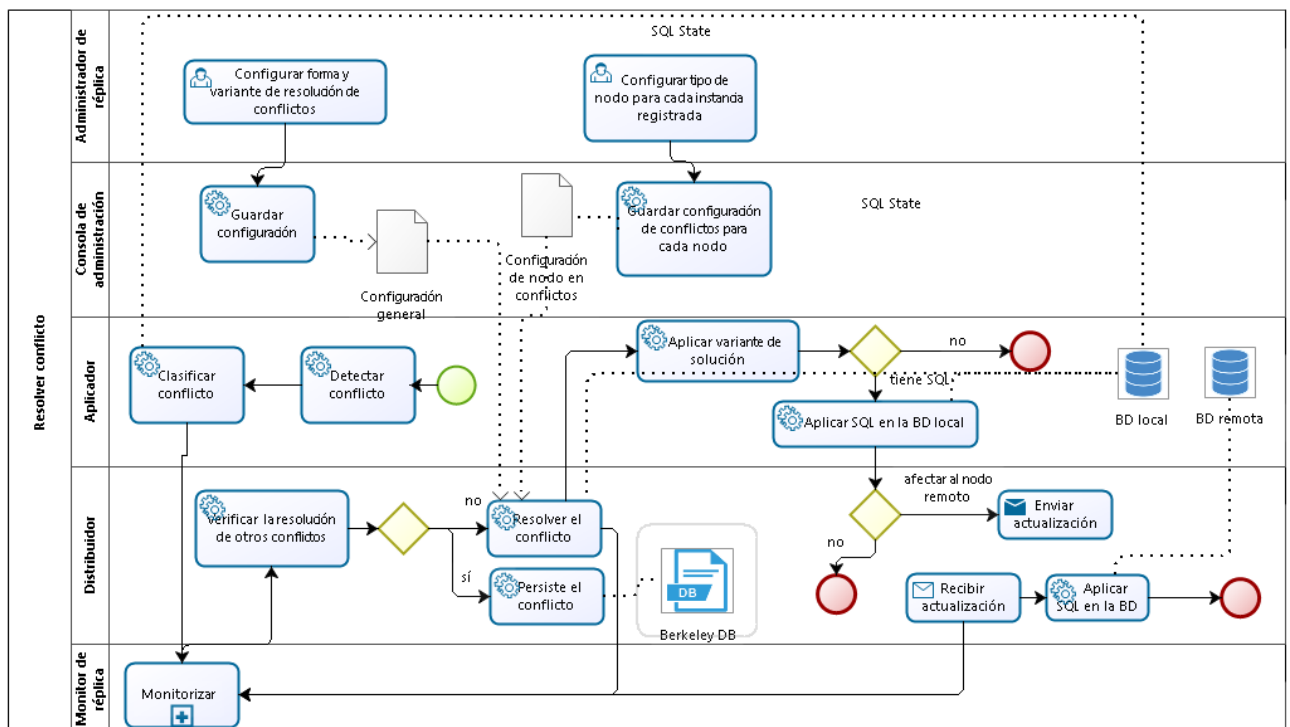


Figura 17 Proceso "Resolver conflictos"

### 2.3.3.5 Transformaciones

En determinados escenarios de replicación cuando los gestores de BDs o sus estructuras son diferentes se hace necesario transformar los datos a replicar, o los metadatos asociados a estos. A este tipo de replicación se le denomina réplica con transformación.

Como se muestra en la Figura 18 el proceso de transformación consiste en mapear las estructuras a partir de una configuración de entrada durante el proceso de registrar la configuración de réplica sobre cada tabla. Una tabla puede tener diferentes configuraciones de transformación por cada tipo de acción para diferentes destinos.

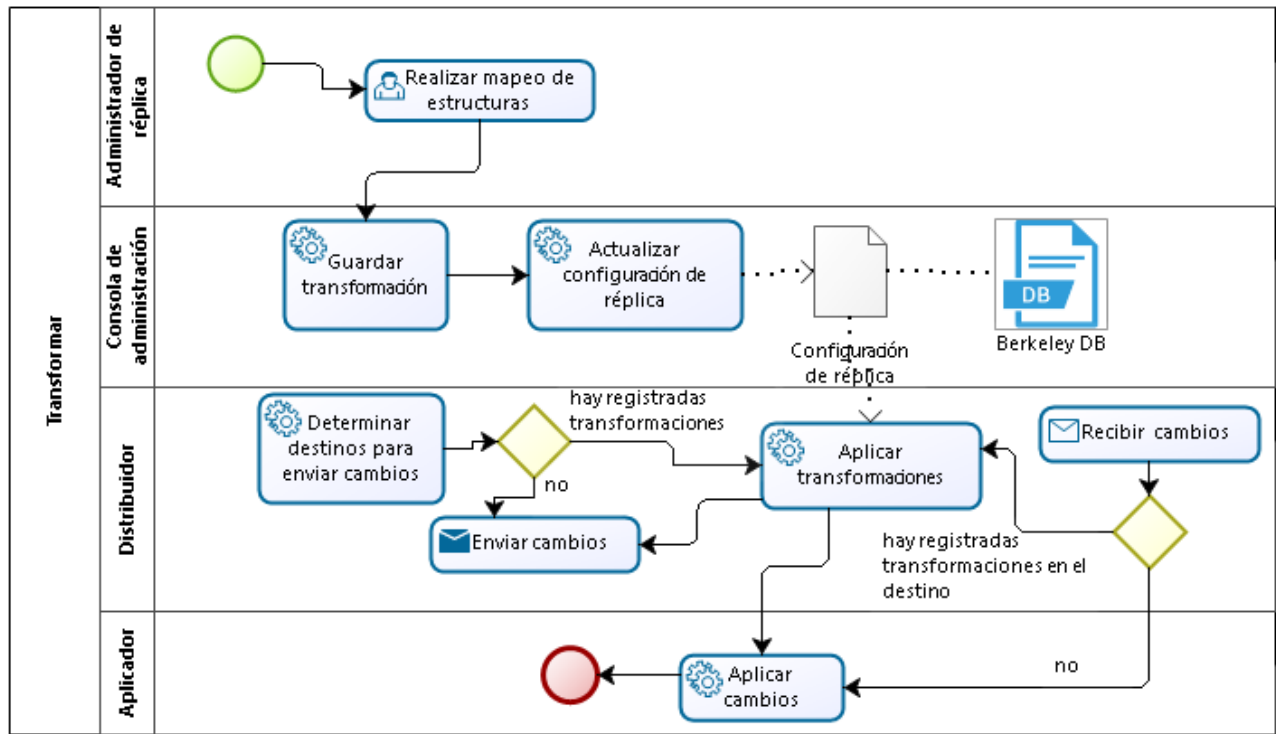


Figura 18 Proceso “Realizar transformaciones”

Para responder a múltiples entornos donde se requiera aplicar mapeo entre estructuras para poder realizar la aplicación de cambios de manera correcta, se definen un conjunto de transformaciones caracterizadas en (Pérez et al., 2010).

El proceso de transformación tiene dos cuestiones fundamentales: la validación de las transformaciones configuradas y la aplicación de transformaciones sobre las acciones replicables. Al guardar una configuración de réplica para un destino ya sea nodo o etiqueta se verifican si existen transformaciones asociadas, en caso positivo se verifica que:

- Sean válidas las precondiciones de cada una de las transformaciones: compatibilidad de tipos, manejo de campos nulos, entre otras precondiciones específicas de cada transformación.
- A cada campo en el destino esté asociada una sola transformación.
- Existan transformaciones hacia todos los campos requeridos de las tablas del destino.

Después de ser validadas las transformaciones, se determinan los campos requeridos en el origen a partir de los campos requeridos en el destino. Las transformaciones se ejecutan sobre cada una de las acciones replicables que integran un grupo replicable una vez determinados sus destinos. Por cada acción replicable original:

- Se revisan las transformaciones configuradas para cada destino.

- Se generan nuevas acciones por cada tabla del destino a partir de las transformaciones configuradas y la acción replicable original.
- En caso de no existir ningún destino sin transformaciones para la acción replicable original, esta es descartada.

Las acciones replicables resultantes de este proceso son enviadas al Distribuidor de datos para ser propagadas hacia sus destinos. Pueden definirse dos alternativas para responder a las necesidades del contexto de replicación con respecto a cuándo aplicar las transformaciones previamente definidas: antes de ser enviados los grupos replicables, una vez determinados sus destinos; o en cada uno de los nodos destinos, antes de ser aplicadas las acciones correspondientes al cambio. Aplicar la transformación implica actualizar el metadato almacenado en la acción replicable que contiene el cambio que se actualizará en el destino, y construir el SQL que se aplica en el nodo destino a partir de las estructuras mapeadas y definidas en la configuración de réplica para cada acción replicable.

### 2.3.3.6 Monitor de réplica

La monitorización es el proceso que debe controlar el estado de la réplica de datos, analizando si el envío y captura de los datos se realiza de forma correcta o incorrecta y gestionando la detección de conflictos en la sincronización o propagación de los cambios con conexión o fuera de línea (Microsoft, 2014).

El Monitor de réplica debe mostrar el comportamiento del proceso de réplica de las acciones cuando se envían y reciben, ya sean las acciones replicadas con conexión, sin conexión, o las acciones en conflictos. Para cada caso se define una acción monitorizada con los siguientes datos:

- Dirección: indica la trayectoria del cambio, si el cambio entra o sale.
- Fecha: muestra la fecha y hora en que la acción de enviar, recibir o aplicar es realizada.
- Remitente: contiene el identificador del nodo que originó el cambio.
- Destinos: contiene el identificador del nodo que procesa el cambio, además del estado en que se encuentra (enviando, enviada, recibiendo o recibida, aplicada).
- SQL: sentencia SQL de la acción correspondiente al cambio a ejecutar en la BD.
- Esquema: el esquema que se afecta con el cambio.
- Tabla: la tabla que se afecta con el cambio.
- Tipo de acción: el tipo de acción a ejecutar en la BD.
- Observación: información adicional que se desee mostrar.

### 2.3.3.7 Mecanismos de recuperación

Cada instancia almacena en las BDs embebidas los grupos replicados, los cuales antes de ser enviados almacenan el identificador del destino y un contador. El contador almacena la cantidad de envíos realizados del nodo local al nodo destino hacia el cual se replica el grupo. Además, cada instancia a su vez, almacena en una BD embebida un contador de agrupaciones recibidas de cada

destino. Cuando el grupo arriba a su destino, antes de ser aplicado, se verifica si es el grupo que corresponde ser aplicado, para que, si un grupo no fue recibido debido a una interrupción o fallos en la red, pueda solicitarse al nodo origen el grupo esperado. Aunque, para evitar pérdida de tiempo a la hora de esperar a que arribe una transacción para detectar que una agrupación no ha sido aplicada, se propone emplear:

- Tareas que se ejecuten cada vez que inicie el sistema de replicación, para enviar los grupos pendientes por ser enviados, cuando no hay conexión con el servidor de mensajería.
- Tareas programadas para notificar entre las instancias el estado de los contadores. Cuando no se correspondan los contadores se solicitan los grupos que no han sido aplicados.

Mediante este mecanismo con el empleo BDs embebidas y contadores, cuando el sistema se recupera, la réplica continúa y los grupos que contienen los cambios no se pierden, con lo cual se evita pérdida de información e inconsistencias entre las BDs que se encuentran distribuidas.

### 2.3.4 Componente Gestión tecnológica

La infraestructura tecnológica está constituida por una serie de tecnologías y herramientas que se consideran necesarias para el funcionamiento de una organización o desarrollo de una actividad (Febles, 2013).

Esta investigación contiene una propuesta que articula el desarrollo tecnológico con el uso de herramientas libres basada en la línea de investigación del centro de Consultoría y Desarrollo de Arquitecturas Empresariales (CDAE) para el desarrollo de réplica de datos. Como se ilustra en la Figura 19, el desarrollo de la solución propuesta está soportado fundamentalmente por las siguientes tecnologías:

- Como Entorno de Desarrollo Integrado (IDE) Eclipse STS (Del inglés Spring Tool Suite), herramienta libre para construir aplicaciones empresariales enriquecidas por los proyectos de Spring Source (Panamsource, Inc., 2015).
- Lenguaje de programación JAVA versión 1.8, lenguaje orientado a objeto, de código abierto e independiente de la plataforma (Weitzenfeld, 2005).
- JSF (Del inglés Java Server Faces) como marco de interfaz de usuario orientado a componentes estándares para la plataforma Java EE (Holmes et al., 2007).
- Primefaces como framework (marco de referencia) para el desarrollo de las interfaces de usuario (Çalışkan et al., 2015).
- Spring en su versión 2.5 como contenedor y framework de peso ligero, basado en inyección de dependencias y orientación a aspecto (Wheeler et al., 2013).
- Apache Tomcat como servidor web (Agrawal et al., 2014) y Apache ActiveMQ como servidor de mensajería (Snyder et al., 2011).
- Servicio de Mensajería de Java (del inglés Java Message Service), que permite a los componentes de aplicaciones basados en la plataforma JEE crear, enviar, recibir y leer

mensajes. Hace posible la comunicación confiable de manera síncrona y asíncrona (Juneau, 2013).



Figura 19 Componentes de una infraestructura tecnológica para réplica de datos

En el modelo propuesto se propone tener en cuenta para el desarrollo de las soluciones de replicación aspectos críticos como:

- Asegurar una infraestructura robusta y eficiente: un ambiente de red con un apropiado ancho de banda y con servicios de alta disponibilidad.
- El número de recursos manejado debe quedar en un número razonable dentro de los límites técnicos para el cual el sistema ofrezca un nivel aceptado de rendimiento.

Para asegurar una infraestructura que ofrezca servicios tolerantes a fallos, para el despliegue de las soluciones que apliquen el modelo propuesto deben valorarse las siguientes premisas: asignar suficiente memoria para la ejecución de los servicios, de lo contrario se detendrá la máquina virtual de Java y por ende se detendrá el servicio de réplica; monitorizar el crecimiento de los directorios donde están instaladas la solución de replicación y el gestor de BD, debido a que el modelo propuesto se basa en el uso de BDs embebidas y ficheros que con el tiempo aumentarán su tamaño; verificar que los puertos designados para la conexión a los servidores no estén bloqueados porque en ocasiones las políticas de seguridad de las instituciones bloquean los puertos por protección; y realizar salvallas periódicas del directorio del sistema correspondiente a las bases de datos embebidas para ante eventualidades regresar fácilmente el sistema al estado anterior al fallo.

### 2.3.5 Indicaciones metodológicas

Para orientar la aplicación del modelo propuesto se plantean las siguientes indicaciones metodológicas que guían su implementación. Las indicaciones metodológicas para la

implementación del modelo se elaboraron a partir del análisis de los modelos de implementación propuestos en la literatura y de la experiencia en el desarrollo de soluciones informáticas en la Universidad de las Ciencias Informáticas (UCI). Están compuestas por las fases que se describen a continuación:



Figura 20 Fases de implementación del modelo propuesto

**Inicio:** se identifican las necesidades de replicación para determinar qué procesos propuestos del modelo se aplicarán en el desarrollo.

**Elaboración:** de acuerdo con el escenario de replicación requerido se seleccionan los procesos a implementar, se seleccionan las herramientas adecuadas, se diseñan las interfaces de usuario y desarrolla la lógica interna de los componentes, aplicando el modelo propuesto y una metodología de desarrollo de acuerdo con las características del proyecto.

**Despliegue:** se instalan y configuran los servidores de BDs y mensajería, así como la herramienta de replicación. Inicialmente se realizan las configuraciones generales para conectar el sistema a los servidores u otras configuraciones para establecer el escenario de réplica requerido.

**Actualización:** se retroalimenta el modelo a partir de la experiencia tras su aplicación en entornos reales.

### Conclusiones parciales del capítulo

Después de elaborar el modelo para la réplica de datos en sistemas de bases de datos distribuidas se concluye que:

- Mediante la integración de fundamentos teóricos, buenas prácticas, marcos de trabajo y experiencias positivas en materia de modelos y herramientas de replicación fueron definidos los constructos y concebida la propuesta, que sustenta un modelo para la réplica de datos el cual supera las principales insuficiencias detectadas.
- El modelo elaborado contiene los componentes que conforman el proceso de desarrollo de una solución de replicación y establece un conjunto de recomendaciones e indicaciones metodológicas para su aplicación.
- La diagramación, mediante el empleo de la Notación de Modelado de Procesos de Negocio (BPMN, del inglés Business Process Management Notation) y la herramienta Bizagi Modeler (Modelador de procesos), favorece la comprensión y satisfacción del usuario para desarrollar una solución de replicación.



## CAPÍTULO 3. VALIDACIÓN DE LOS RESULTADOS

En el presente capítulo se exponen los principales elementos relacionados con la validación del modelo para la réplica de datos en sistemas de bases de datos distribuidas, mediante la aplicación de métodos científicos teniendo en cuenta las variables definidas en la hipótesis. Se aplicaron métodos cuantitativos y cualitativos. Entre los encuestados estuvieron especialistas de la UCI, pertenecientes a los centros: CDAE, Tecnologías de Gestión de Datos (DATEC), Dirección de Informatización (DIN) y Tecnologías para la Formación (FORTES). Además, formaron parte del grupo focal, especialistas del Centro Nacional de Calidad de Software (CALISOF) y miembros del Subcomité 7 Ingeniería de Software y Sistemas en Cuba.

### 3.1 Instrumentación del modelo en el Replicador de datos REKO

El software Replicador de datos REKO, es la aplicación informática desarrollada como aporte práctico de esta investigación. El mismo cubre las necesidades fundamentales de replicación, tales como: sincronización, transferencia de datos entre diversas localizaciones y la centralización de la información en una única localización, protección y recuperación, así como satisfacer otras necesidades relacionadas con la distribución de datos entre los gestores más populares. Para guiar su desarrollo se empleó la metodología AUP-vUCI (Proceso Unificado Ágil versión Universidad de las Ciencias Informáticas), por ser la metodología definida por la universidad para proyectos productivos; y para el desarrollo se utilizaron las herramientas propuestas en la infraestructura tecnológica. El software constituye una solución de alto impacto en la sustitución de importaciones como una alternativa construida sobre una arquitectura en tecnologías modernas y libres.

### 3.2 Descripción de la validación de la hipótesis

En la investigación fue planteada la siguiente hipótesis: el desarrollo de un modelo para la réplica de datos en sistemas de bases de datos distribuidas que describa los procesos de sincronización, resolución de conflictos, monitorización y administración de configuraciones contribuirá a mantener la consistencia de los datos replicados en los SBDD.

Al formular la hipótesis, es indispensable definir los términos o variables que están siendo incluidos en ella (Sampieri et al., 2010). A continuación, se muestran las definiciones operacionales de las variables dependientes e independientes definidas en la investigación:

Tabla 2 Definiciones operacionales de la variable independiente

-Variable independiente-	-Dimensión-	-Indicadores-	-Valor que asume-
Modelo de replicación de datos	Nivel de calidad del modelo de replicación de datos	Pertinencia	Valor real
		Comprensión	
		Coherencia	
		Adaptabilidad	
		Satisfacción del usuario	Valor real

Tabla 3 Definiciones operacionales de la variable dependiente

-Variable dependiente-	-Dimensión-	-Indicadores-	-Valor que asume-
Consistencia	Replicación de cambios	Integridad referencial de la información ----- Tablas o datos completos ----- Coherencia en las BDs	Racional
	Sincronización entre bases de datos	Homogeneidad de las bases de datos	
	Resolución de conflictos	Capacidad de resolución de conflictos	
	Restablecimiento ante eventualidades	Persistencia en BDs embebidas	
	Replicación de forma <u>offline</u> (fuera de línea)	Capacidad de operar sin conexión	

La propuesta de solución comprende un modelo y una herramienta para la replicación de datos. Para analizar si la solución es eficaz para mantener la consistencia de los datos se evalúa la solución. Para la evaluación se valora el criterio de expertos y usuarios potenciales que son también expertos en el tema, empleando los métodos Likert, ladov y Grupo focal. Mientras que para la herramienta se aplican pruebas funcionales en un entorno real. Para una mejor comprensión sobre los constructos a valorar en cada método se realiza la siguiente representación.

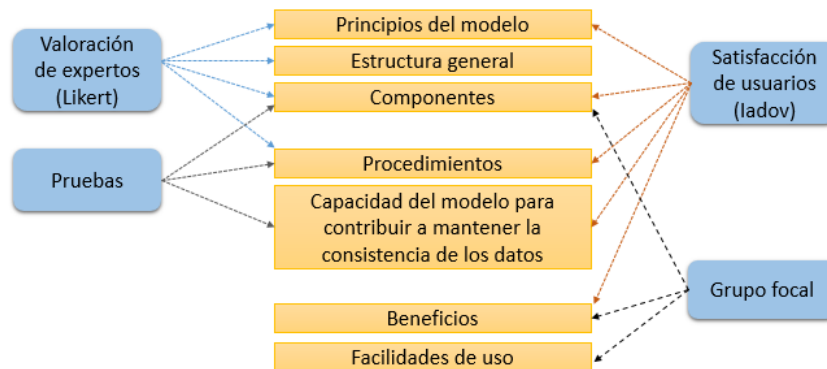


Figura 21 Técnicas utilizadas para evaluar el modelo

- El escalamiento de Likert para validar los componentes del modelo.
- La técnica ladov se aplica para medir el índice de satisfacción de los usuarios con respecto al modelo propuesto.
- Grupo focal para definir las medidas para los indicadores definidos en las dimensiones de la variable dependiente, con el propósito de evaluar la consistencia de los datos; y conocer además la opinión de los expertos sobre los beneficios y facilidades de uso del modelo propuesto.
- Se realizaron pruebas funcionales para validar la consistencia de los datos.
- La triangulación metodológica para disminuir el sesgo que se produce al comparar los resultados obtenidos en la cuantificación de variables.

### 3.3 Valoración de los expertos

Para conocer la opinión de los expertos sobre la contribución del modelo elaborado para mantener la consistencia de los datos replicados en ambientes distribuidos, fue aplicado el escalamiento de Likert. Este método fue desarrollado por Rensis Likert y consiste en un conjunto de ítems presentados en forma de afirmaciones o juicios ante los cuales se pide la reacción de los sujetos. Es decir, se presenta cada afirmación y se pide al sujeto que externé su reacción eligiendo uno de los cinco puntos de la escala. A cada punto se le asigna un valor numérico. Así, el sujeto obtiene una puntuación respecto a la afirmación y al final se obtiene su puntuación total sumando las puntuaciones obtenidas con relación a todas las afirmaciones (Sampieri et al., 2010).

La escala psicométrica Likert fue aplicada en la presente investigación a través de un cuestionario con el objetivo de conocer el nivel de acuerdo o desacuerdo con los principios, características y componentes del modelo. Para lo cual se definieron una serie de aspectos para valorar:

1. Principios del modelo, estructura general, y factibilidad de uso.
2. Pertinencia: la estructura del modelo, sus componentes y procesos son congruentes con los objetivos del mismo teniendo en cuenta las exigencias del problema de investigación planteado.
3. Coherencia: existe correlación entre los elementos del modelo para dar solución al problema de investigación planteado.
4. Comprensión: son entendidos los resultados que se obtienen a raíz de aplicar el modelo, tanto para los analistas de sistema como para los programadores de un proyecto.
5. Adaptabilidad: los componentes son descritos con funcionalidades genéricas y es soportado sobre una estructura que facilita su ajuste a las particularidades de las aplicaciones existentes para el proceso de replicación.

#### 3.3.1 Proceso de selección de los expertos

Inicialmente se realizó una valoración inicial de los expertos para validar el modelo. En el proceso de selección se consultaron especialistas con experiencia en tecnologías de gestión de datos, desarrollo y/o despliegue de soluciones de replicación de datos, pertenecientes a los centros CDAE, DATEC, DIN y FORTES. Se evaluaron sus conocimientos sobre su experiencia y conocimiento en el tema de replicación de datos, investigaciones y/o publicaciones, así como actualización en cursos de posgrado, diplomados, maestrías, entre otros como la vinculación como tutores en trabajos de diplomas. De una cantidad inicial de 20 expertos dispuestos a colaborar, se desestimaron 4 debido a que sus índices de competencia fueron inferiores o iguales a 0,5. De los expertos, 12 obtuvieron un coeficiente alto y 4 medio. De los expertos seleccionados, 12 han mantenido una vinculación activa al tema de replicación de datos. En el Anexo 3 se describe el procedimiento empleado para determinar el coeficiente de competencia de los candidatos a expertos y los resultados obtenidos.

**3.3.2 Proceso de aplicación de Likert**

Las afirmaciones del cuestionario están enfocadas a obtener las valoraciones de los expertos en función de los indicadores definidos y los problemas identificados en la investigación, y representan 8 aspectos relevantes del modelo propuesto, donde el experto expresa su valoración de cada indicador mediante la siguiente escala: 5- Muy de acuerdo (MA), 4- De acuerdo (DA), 3- Ni de acuerdo ni en desacuerdo (Sí-No), 2- En desacuerdo (ED) y 1- Completamente en desacuerdo (CD). Tanto la encuesta como la escala definida y el procesamiento de los resultados, se encuentran en el Anexo 4.

Para el procesamiento de los resultados se empleó un método que consiste en identificar la frecuencia en cada categoría de la escala de Likert definida en la encuesta, y se calculan los porcentajes de concordancia de cada categoría de acuerdo con las características propuestas por la autora. Luego se calcula el Índice Porcentual (IP) de cada afirmación, que integra en un solo valor la aceptación del grupo de evaluadores sobre las características del modelo, como lo indica la siguiente fórmula:

$$IP = \frac{5(\% \text{ de MA}) + 4(\% \text{ de DA}) + 3(\% \text{ de Sí-No}) + 2(\% \text{ de ED}) + 1(\% \text{ de CD})}{5} \quad (3.1)$$

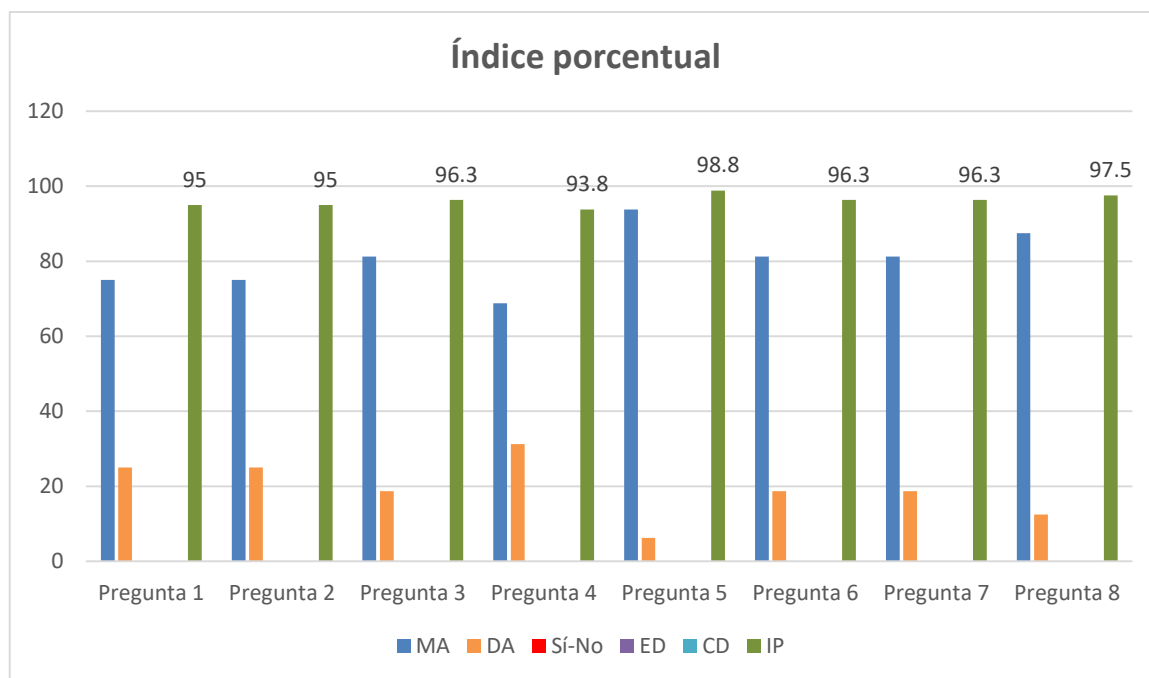


Figura 22 Resultados de la aplicación de la escala Likert

El IP de los expertos en cada una de las preguntas sobrepasa el valor de 80 %. Por tanto, se puede concluir, de manera general, que la valoración de la variable calidad del modelo obtuvo buenos resultados, por los altos índices de aceptación que tuvo por los expertos encuestados. El procesamiento realizado a través del escalamiento de Likert evidenció que tanto los elementos teóricos como los principios, características y componentes del modelo, tienen una alta valoración

por parte de los expertos. Se pudo comprobar su gran satisfacción respecto a la comprensión, pertinencia, coherencia y adaptabilidad del modelo propuesto; lo que influye positivamente en la variable calidad del modelo propuesto.

Para demostrar la confiabilidad de las respuestas dadas por los expertos en los cuestionarios se realizó el cálculo del coeficiente de confiabilidad Alfa-Cronbach sobre la base de la varianza de los ítems, procedimiento descrito en (Sampieri et al., 2010). En (Quero, 2010; González et al., 2015) se considera que valores de alfas superiores a 0,8 son suficientes para garantizar la fiabilidad de la escala. Por tanto, se evidencia un buen grado de confiabilidad, al obtener un coeficiente de 0,84. El cálculo del coeficiente de confiabilidad del cuestionario se encuentra en el Anexo 5.

### **3.4 Validación de satisfacción del usuario**

El conocimiento del estado de satisfacción del usuario respecto al modelo de replicación propuesto, es de gran utilidad para la toma de decisiones y en la validación de la propuesta. La técnica ladov constituye una vía para el estudio del grado de satisfacción de los implicados en el proceso objeto de análisis. Esta técnica fue creada para establecer el nivel de satisfacción por la profesión de carreras pedagógicas. Luego algunos autores la han modificado en parte y aplicado para valorar la satisfacción en múltiples campos y como parte de diagnósticos y validaciones en diferentes investigaciones (Febles, 2013). La técnica de ladov se basa en el análisis de un cuestionario que tiene una estructura interna determinada que sigue una relación entre tres cuestionamientos cerrados y el análisis posterior de dos cuestiones abiertas. La relación entre las preguntas cerradas se establece a través del denominado "Cuadro Lógico de ladov" (Orellana, 2016).

En el Anexo 6 se muestra en detalle cómo se aplicó la técnica empleada específicamente para medir la satisfacción de los usuarios con relación al modelo para la replicación de datos. Para su desarrollo se aplicó una encuesta que permitió conocer el grado de satisfacción sobre el modelo propuesto, en cuanto a: necesidad de mantener la consistencia de los datos replicados en ambientes distribuidos, calidad y utilidad del modelo propuesto para mantener la consistencia en la replicación de datos, y los aspectos que se consideran positivos y negativos del modelo propuesto, así como sugerencias para mejorar su aplicación en el desarrollo de soluciones de réplica.

Para medir este grado de satisfacción se tomó una muestra de 11 usuarios pertenecientes a los centros CDAE, DATEC, DIN y FORTES, teniendo en cuenta los años de experiencia en el desarrollo y/o despliegue de soluciones de replicación y/o gestión de datos, el rol que desempeñan, entre otros aspectos. El resultado de la satisfacción individual se muestra también en el Anexo 6.

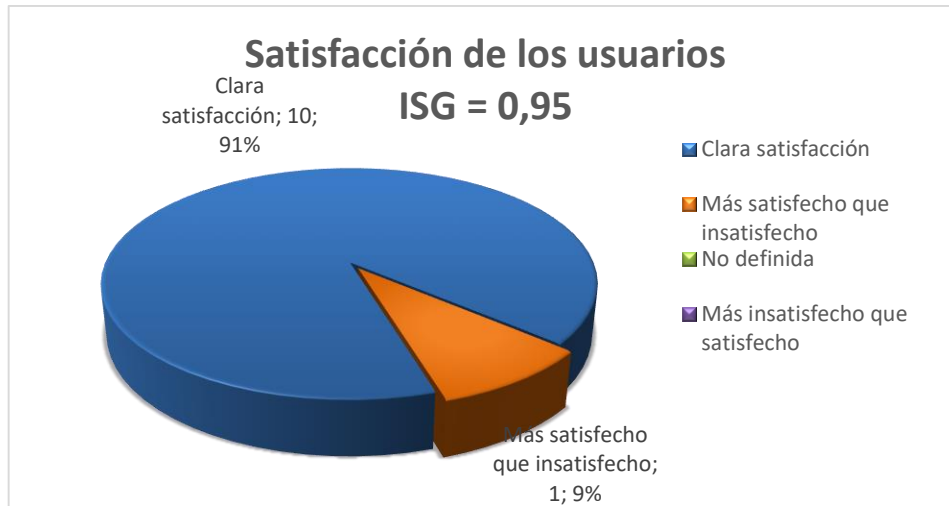


Figura 23 Resultados de la aplicación de la técnica ladov

Esta técnica permite obtener el Índice de Satisfacción Grupal (ISG), para lo cual se trabaja con los diferentes niveles de satisfacción que se expresan en la escala numérica que oscila entre (+1) y (-1). La satisfacción grupal se calcula mediante la siguiente fórmula:

$$ISG = \frac{A(+1)+B(+0,5)+C(0)+D(-0,5)+ E(-1)}{N}, \quad (3.2)$$

donde A, B, C, D, E representan el número de sujetos con índice individual 1; 2; 3 o 6; 4; 5 y N representa el número total de sujetos del grupo; que permite reconocer las categorías grupales (Cañizares, 2012):

- Insatisfacción: desde (-1) hasta (-0,5).
- Contradictorio: desde (-0,49) hasta (+0,49).
- Satisfacción: desde (+0,5) hasta (1).

Para este caso que se analiza, el índice de satisfacción grupal se obtuvo al calcular:

$$ISG = \frac{10(+1)+1(+0,5)+0(0)+0(-0,5)+ 0(-1)}{11} = 0,95 \quad \text{Evaluar en (3.2)}$$

Como resultado de la aplicación de la técnica se evidencia que el índice de satisfacción grupal que se obtuvo de 0,95; lo que significa una clara satisfacción con la propuesta y reconocimiento de su utilidad para contribuir a mantener la consistencia de los datos en ambientes distribuidos.

La técnica ladov contempla además dos preguntas complementarias de carácter abierto, las cuales permiten profundizar en las causas que originan los diferentes niveles de satisfacción (Febles, 2013). Además, porque sugieren aspectos de utilidad para investigaciones presentes y futuras. Entre las principales recomendaciones asociadas a las respuestas de los expertos a las preguntas anteriores estuvieron:

- La posibilidad de definir elementos críticos relacionados con el despliegue del sistema y documentarlos de manera explícita en el componente Gestión tecnológica.
- Flexibilizar la resolución de conflictos de unicidad para permitir que, ante la generación automática de identificadores en tablas de la BD, se brinde la posibilidad de insertar los datos en nodos ganadores.
- Para enriquecer la propuesta se recomienda definir una curva de aprendizaje sobre el desarrollo de soluciones de replicación, donde se determine el tiempo que los desarrolladores necesitan para la aplicación del modelo propuesto. Además de la incorporación de una estrategia de capacitación para el uso efectivo del modelo en sistemas de información.
- Almacenar en una base de casos las decisiones tomadas para la resolución de conflictos para recomendar configuraciones o forma de resolución cuando se encuentre en modo automático o en situaciones similares.
- Aunque se reconoció por todos los encuestados como elementos positivos la estructura y componentes del modelo, se recomienda elaborar un componente asociado a una propuesta de arquitectura para el desarrollo de soluciones de replicación, que además de contribuir a disminuir el tiempo de desarrollo oriente aún más al desarrollador y con ello la propuesta del modelo profundice en el “cómo” hacer cada uno de los componentes propuestos.

La primera recomendación fue concebida en el modelo, refinándolo mediante las especificaciones realizadas de manera explícitas en el componente señalado. La segunda recomendación se aborda como premisa del modelo, debido a que se establece como condición que los sistemas de bases de datos deben estar correctamente diseñados, es decir, deben poseer un correcto modelo de datos. Las restantes recomendaciones fueron propuestas en las recomendaciones generales de la investigación.

### 3.5 Grupo focal

El grupo focal es una técnica de investigación cualitativa, que se define en (Balcázar et al., 2013) como: *“un tipo especial de entrevista grupal que se estructura para recolectar opiniones detalladas y conocimientos acerca de un tema particular, vertidos por los participantes seleccionados”*. Esta técnica permite organizar grupos de discusión alrededor de una temática, la cual es elegida por el investigador. La información es útil para dar profundidad, detalle y explicación de los datos cuantitativos. Además, se obtienen respuestas a fondo sobre lo que piensan y sienten las personas de un grupo de forma libre y espontánea, guiados por un facilitador o moderador (Balcázar et al., 2013).

En la actualidad muchas organizaciones conducen grupos focales para mejorar servicios, evaluar cambios, además de que se han utilizado para la elaboración de diversos instrumentos de evaluación, dando pauta a la comprensión de fenómeno o constructo con el que se trabajará

(Balcázar et al., 2013). Precisamente, para la aplicación de esta técnica se definió un grupo focal con el objetivo de definir las medidas adecuadas para evaluar cada uno de los indicadores definidos en la operacionalización de la variable dependiente.

El investigador está presente en la sesión de trabajo del grupo, pero debe actuar como anotador de las ideas expresadas por los participantes de esta manera no influyen sus criterios en las opiniones de los integrantes del grupo (Cañizares, 2012).

Para el grupo focal se seleccionaron 8 especialistas, de los cuales 5 están vinculados directamente con el desarrollo de soluciones de replicación pertenecientes a los centros CDAE y DATEC, quienes han estado inmersos en temas relacionados con la replicación de datos en sus áreas de trabajo. Además, 3 de los 8 especialistas trabajan en el Centro Nacional de Calidad de Software (CALISOF), de los cuales 2 son miembros del Subcomité 7 Ingeniería de Software y Sistemas en Cuba<sup>10</sup>, con más de 8 años de experiencia y partícipes en la definición de normas cubanas. La composición de los participantes en el grupo focal puede ser consultada en el Anexo 7. Para conducir la actividad se diseñó un guion de desarrollo que se muestra en el Anexo 8. Para su confección se tuvo en cuenta que la actividad debe comportarse como una entrevista abierta pero estructurada donde se propicie el debate en función de las experiencias personales y al conocimiento que poseen sobre la temática cada uno de los especialistas. Se seleccionaron además dos personas que no tienen relación directa con la investigación para que participaran como moderador y otro para que llevara la relatoría de una forma imparcial.

Mediante una lluvia de ideas según la guía de preguntas y el análisis realizado se fueron proponiendo y definiendo las medidas para evaluar cada indicador en las dimensiones establecidas. Para establecer las medidas se analizaron las normas internacionales ISO/IEC 25012 Modelo de la calidad de los datos, donde se ejemplifica la medición de la calidad de los datos referenciada en (ISO/IEC, 2008); y la ISO/IEC 25024 Medición de la calidad de los datos, donde se definen específicamente las métricas para realizar la medición de la calidad de datos. En el proceso de definición de las medidas, para valorar la consistencia de los datos específicamente en la replicación, fue necesario ajustar la métrica emitida por la ISO/IEC 25012 al contexto de réplica de datos. Como resultado se definieron las medidas e interpretaciones siguientes:

Medida para evaluar la integridad referencial, completitud y coherencia de los datos en la **replicación de cambios**:

$I_1 = A/R$ , donde A representa la cantidad de acciones aplicadas y R la cantidad de acciones replicadas.

---

<sup>10</sup> Subcomité 7 Ingeniería de Software y Sistemas, SC7 por sus siglas en español, pertenece al Comité Técnico de Normalización No 18 Tecnologías de la Información, CTN18 por sus siglas en español. Tanto el CTN18 como el SC7 son homólogos por la parte cubana de organismos internacionales de normalización.



Medida para evaluar la homogeneidad de las bases de datos en la **sincronización de cambios**:

$I_2 = A/C$ , donde C representa la cantidad de acciones construidas para homogeneizar las BDs según la configuración de réplica registrada.

Medida para evaluar la capacidad de resolución de conflictos en la **resolución de conflictos**:

$I_3 = R/G$ , donde R representa la cantidad de conflictos resueltos y G la cantidad de conflictos generados.

Medida para evaluar la persistencia en el **restablecimiento** al estado deseado en caso de eventualidades como: de fallos en la red o interrupciones eléctricas:

$I_4 = P/E$ , donde P representa la cantidad de grupos persistidos en el origen pendientes por ser enviados y E la cantidad de grupos enviados que se encontraban pendientes.

Medida para evaluar la capacidad de operar sin conexión en la **replicación fuera de línea**:

$I_5 = M/X$ , donde M representa la cantidad de acciones importadas que se aplicaron en el destino y X la cantidad de acciones exportadas en los archivos en el origen.

Para analizar los resultados se define D como el conjunto formado por los indicadores correspondientes a las dimensiones que pueden establecerse para evaluar la consistencia de los datos ( $I_i, i = 1, 2, \dots, n: n \in \mathbb{N}$ ). Entonces, para que la consistencia de los datos sea correcta según los indicadores  $I_i \in D$ , tal que  $i = 1, 2, \dots, 5$  debe cumplirse que:

$$I_i = 1: \forall i = 1, 2, \dots, 5 \quad (3.3)$$

La propiedad anterior expresa que los indicadores en todas las dimensiones establecidas deben arrojar un valor igual a 1. De lo contrario, si existe al menos  $I_i$  diferente de 1 el resultado es incorrecto, y por ende se dice que el sistema no mantiene la consistencia.

Durante el proceso de desarrollo del grupo focal se arrojaron los siguientes resultados:

- Una total coincidencia con respecto a la ausencia de medidas para evaluar específicamente la consistencia en la replicación de los datos.
- Todos los indicadores expuestos en la operacionalización de la variable independiente fueron nombrados por el grupo focal como indicadores importantes para evaluar la consistencia de los datos en los sistemas de replicación.
- Hubo un 100 % de concordancia en establecer que para evaluar la consistencia deben considerarse todos los indicadores propuestos, y que el sistema mantendrá la consistencia si todos los indicadores son correctos.

Se realizaron algunas recomendaciones que se tuvieron en cuenta en el perfeccionamiento del modelo propuesto en la investigación. Las más significativas fueron las siguientes:

- Retroalimentar el modelo a partir de la experiencia tras su aplicación en entornos reales, para actualizar de ser necesario su estructura y componentes.
- Considerar el asesoramiento necesario para la aplicación del modelo y la documentación complementaria que describa el “cómo” desarrollar los componentes propuestos.

### 3.6 Validación de la consistencia

El proceso de instrumentación se realizó aplicando la propuesta en un escenario de replicación, del cual se extrajeron los elementos necesarios para validar la consistencia de los datos en las dimensiones definidas, mediante pruebas funcionales. El escenario de replicación se corresponde con un entorno de replicación que se encuentra desplegado por el centro Nacional de Electromedicina y las instancias provinciales: el centro de datos nacional, Holguín, Camagüey, La Habana y la Isla de la Juventud. Se valoraron las acciones realizadas en el período del 18 de marzo al 3 de abril del presente año donde se registra actividad en el sistema. Como estrategia de pruebas se efectuaron pruebas al sistema. Específicamente se realizaron pruebas funcionales aplicando la técnica de Caja Negra con el método de Partición equivalente. Para lo cual se emplearon las BDs embebidas, los datos mostrados por el Monitor de réplica, entre otras opciones del Replicador de datos REKO tales como: detalles de exportación e importación.

Escenario 1. Evaluar la integridad referencial, completitud y coherencia de los datos en la **replicación de cambios**: en la siguiente figura se muestran las acciones recibidas y aplicadas por parte del centro de datos, y replicadas por los orígenes correspondientes a las provincias Camagüey e Isla de la Juventud. Adicionalmente se inspeccionaron las configuraciones registradas para corroborar que las instancias seleccionadas cubren los posibles escenarios respecto a los tipos de acciones DML y DDL sobre múltiples estructuras de la BD.

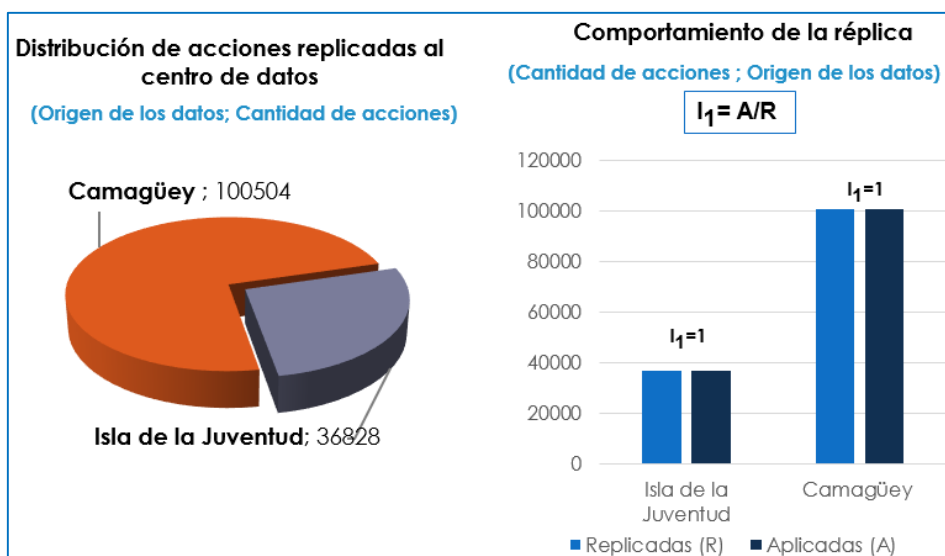


Figura 24 Resultados obtenidos en el escenario 1

Al calcular  $I=A/R$  se obtiene que el  $I=1$ , lo cual evidencia la replicación correcta de los cambios, pues las acciones replicadas de las provincias fueron aplicadas satisfactoriamente en el centro de datos.

Escenario 2. Evaluar la homogeneidad de las BDs en la **sincronización de cambios**: se realiza con dos instancias recientemente configuradas y que previamente no hayan realizado el proceso de replicación, para garantizar que el escenario construye varias acciones que satisfacen los diferentes tipos de acción y estructuras definidas.

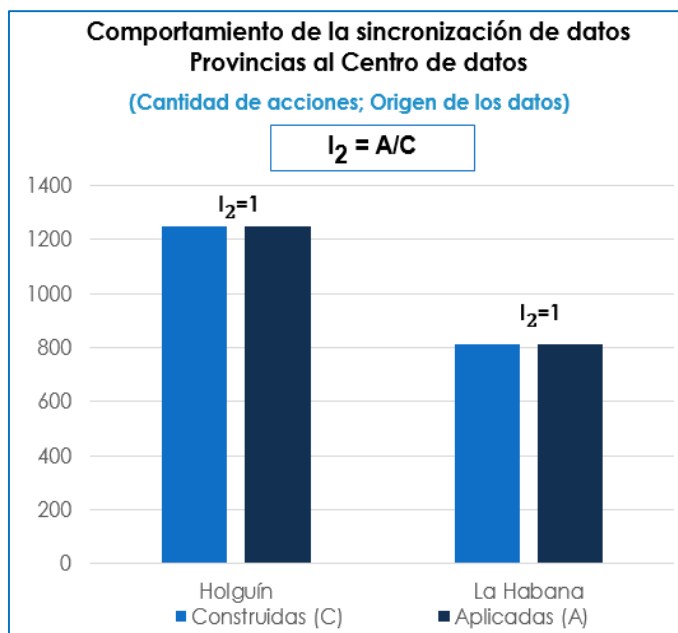


Figura 25 Resultados obtenidos en el escenario 2

Al calcular  $I=A/C$  se obtiene que el  $I=1$ , lo cual evidencia que la solución mantiene la homogeneidad en el proceso de sincronización de cambios.

Escenario 3. Evaluar la capacidad de resolución de conflictos en la **resolución de conflictos**: se selecciona el centro de datos, pues constituye el escenario más representativo en cuanto a cantidad de acciones y con ello variedad de conflictos.

Se identifica que en su mayoría los conflictos se producen en el centro de datos porque el administrador diariamente realizaba la sincronización. Se producían conflictos de unicidad debido a que los datos ya se encontraban registrados por la sincronización del día anterior. Lo correcto sería sincronizar la primera vez que se registran las instancias, y posteriormente realizar réplica de datos online (con conexión) u offline (sin conexión).

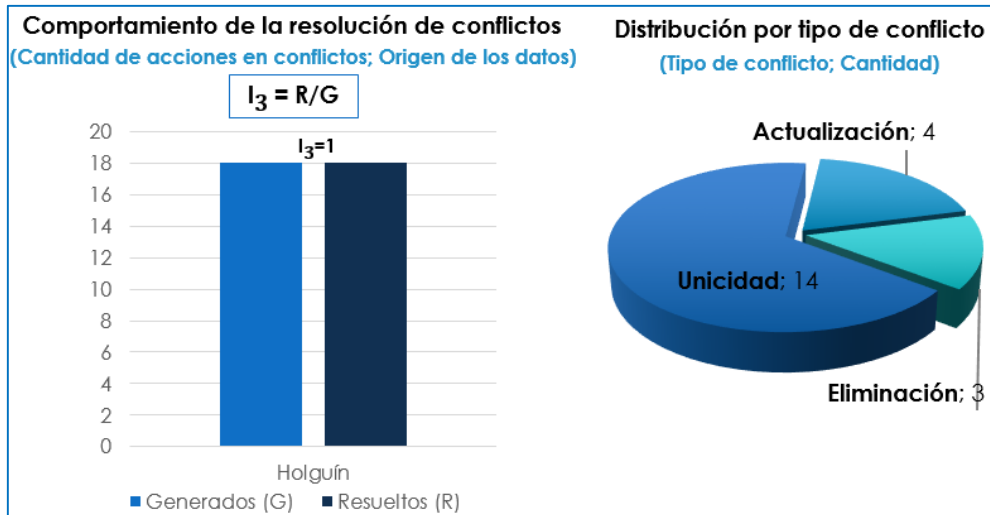


Figura 26 Resultados obtenidos en el escenario 3

Al calcular  $I=R/G$  se obtiene que el  $I=1$ , lo que evidencia que la solución es capaz de resolver conflictos de réplica.

Escenario 4. Medida para evaluar la persistencia en el **restablecimiento** al estado deseado en caso de eventualidades como fallos en la red o interrupciones eléctricas: en este escenario se desconecta manualmente al Servidor de Mensajería Active MQ del centro de datos, se realizan cambios en la BD, se procede a identificar los grupos construidos (pendientes por ser enviados y persistidos), y los grupos enviados cuando se restablece la conexión.

Al calcular  $I=E/P$  se obtiene que el  $I=1$ , lo cual evidencia la capacidad de la solución para restablecer el mecanismo de réplica ante interrupciones de conectividad.

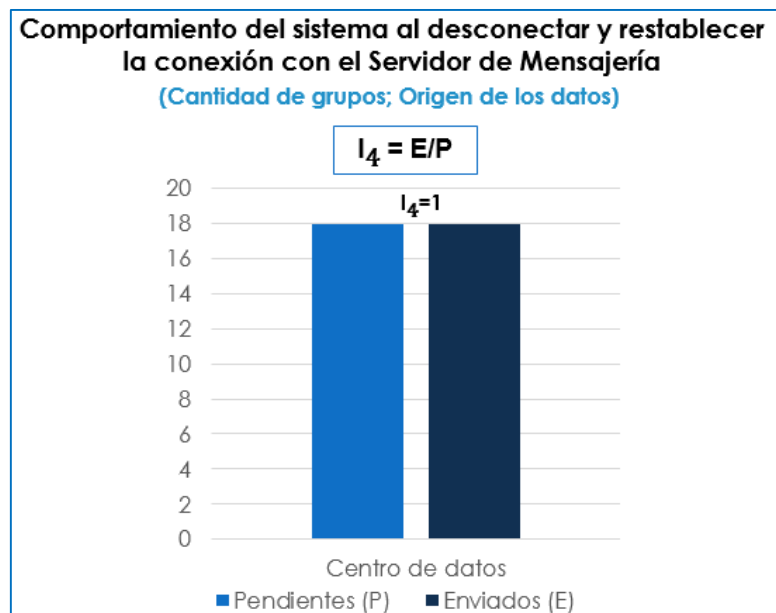


Figura 27 Resultados obtenidos en el escenario 4

Escenario 5. Medida para evaluar la capacidad de operar sin conexión en la **replicación fuera de línea**: las acciones son exportadas en el origen e importadas en el destino. Al calcular  $I=M/X$  se obtiene que el  $I=1$ , lo que evidencia que el mecanismo sin conexión funciona de manera eficaz.

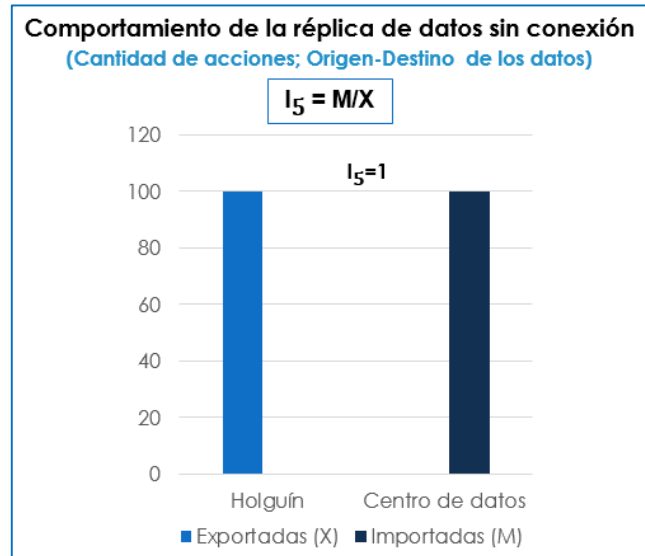


Figura 28 Resultados obtenidos en el escenario 5

Los resultados mostrados en los escenarios fueron satisfactorios, arrojando que el indicador para medir la consistencia ( $I$ ) fue igual a 1 en todos los casos. Por tanto, se corrobora que el Replicador de datos REKO, mantiene la consistencia de los datos replicados. Por ende, se confirma la contribución del modelo para mantener la consistencia de los datos en ambientes distribuidos.

### Conclusiones parciales del capítulo

Después de validar los resultados de la investigación, se concluye que:

- Fueron comprobados el rigor y la relevancia del modelo elaborado mediante la aplicación de métodos cuantitativos y cualitativos.
- El modelo elaborado fue satisfactoriamente valorado por potenciales usuarios y aceptado por expertos.
- El modelo fue refinado a partir de los criterios obtenidos tras la aplicación de los métodos Likert, ladov y Grupos focales, lo que favoreció la mejora de su diseño.
- Los resultados obtenidos en los escenarios del Centro Nacional de Electromedicina, corroboraron la aplicabilidad y utilidad del modelo propuesto y de las indicaciones metodológicas elaboradas.
- La triangulación metodológica permitió ratificar los resultados que por separado se obtuvieron a través de la técnica de ladov, Likert y los Grupos focales.

## CONCLUSIONES GENERALES

1. El análisis de los referentes teóricos relacionados con los modelos, procesos y soluciones de replicación de datos, permitió analizar las tendencias actuales en el desarrollo de soluciones de réplica e identificar un grupo de insuficiencias, que atentan contra la capacidad de los modelos para contribuir a mantener la consistencia en la replicación de datos.
2. El desarrollo de la propuesta conformada por cuatro componentes y la descripción de los procesos asociados, acordes a buenas prácticas, métodos representados y experiencias de expertos, en materia de modelos y herramientas de replicación de datos, permitió estructurar un modelo para la réplica de datos que contribuye a mantener la consistencia de los datos replicados.
3. La validación del modelo mediante la aplicación de métodos científicos y pruebas funcionales en un entorno real, corroboró que los componentes del mismo están alineados a las tecnologías más actuales; y existen opiniones favorables y una alta satisfacción de usuarios potenciales con respecto a la aplicabilidad, actualidad y factibilidad de uso, del modelo.
4. La instrumentación del modelo mediante la herramienta desarrollada y las pruebas realizadas demostraron su contribución para mantener la consistencia de los datos replicados en ambientes distribuidos.

## **RECOMENDACIONES**

1. Incorporar una base de conocimiento para resolver conflictos desconocidos de forma automática y sugerir variante de solución, mediante el estudio del comportamiento en cuanto a la variante aplicada para cada conflicto generado, que posibilite tomar decisiones según los casos almacenados.
2. Incorporar una base de conocimiento para sugerir mejores horarios de replicación a través de la programación y captura de envío de los cambios.

## REFERENCIAS BIBLIOGRÁFICAS

- 2ndQuadrant PostgreSQL. pglogical. *2ndQuadrant PostgreSQL. Expertos mundiales en soporte, capacitación, desarrollo, migración y consultoría para PostgreSQL* 2018. [27 febrero 2018]. <https://www.2ndquadrant.com/es/resources/pglogical/>.
- Abadal, E y Codina, LI. Sistemas de gestión documental y bases de datos documentales. España, Universidad Abierta de Cataluña, 2010. [http://openaccess.uoc.edu/webapps/o2/bitstream/10609/6788/8/Bases%20de%20datos\\_M%C3%B3dulo3\\_Sistemas%20de%20gesti%C3%B3n%20documental%20y%20bases%20de%20datos%20documentales.pdf](http://openaccess.uoc.edu/webapps/o2/bitstream/10609/6788/8/Bases%20de%20datos_M%C3%B3dulo3_Sistemas%20de%20gesti%C3%B3n%20documental%20y%20bases%20de%20datos%20documentales.pdf).
- Agrawal, S. y Dev, R. Development and comparison of open source based web gis frameworks on Wamp and Apache Tomcat web servers. *ISPRS Technical Commission IV Symposium*. India 4(1), 2014.
- Alfonso, R. Módulo para la configuración y monitorización de réplicas con la herramienta SymmetricDS para la arquitectura Xalix. 2017.
- Arora, S. Oracle® Streams for Near Real Time Asynchronous Replication. En *Distributed and Parallel Databases*, Vol. 14 of *Information Society and Media*, p. 17-23, EU project GORDA, 2005.
- Balcázar, P.; González, N.; Peña, Gurrola; Gurrola, G. M. y Moysén, A. Investigación cualitativa. México, Universidad Autónoma del Estado de México, 2013. <http://repositorio.minedu.gob.pe/handle/123456789/4641>.
- Barboza, M. Aplicación del método científico en la realización de peritajes médico legales. *Medicina Legal de Costa Rica* 32(1), 2015, p. 96-101.
- Çalışkan, M. y Varaksin, O. PrimeFaces Cookbook - Second Edition. 2da ed. Birmingham, Packt Publishing Ltd, 2015.
- Cañizares, R. Repositorio de recursos educativos para las instituciones de educación superior. Tesis de doctorado, Universidad de las Ciencias Informáticas, Cuba, 2012, p. 84-86.
- Carcasés, A. y Sabina, A. Interfaz web para la administración y monitorización de la herramienta de réplica de datos SymmetricDS. Tesis de grado, Universidad de las Ciencias Informáticas, Cuba, 2014, 25-46. [https://repositorio\\_institucional.uci.cu/jspui/handle/ident/9109](https://repositorio_institucional.uci.cu/jspui/handle/ident/9109).
- Carranza, F. Bases de datos distribuidas. Escuela Académico Profesional de Informática, 2006 <https://es.scribd.com/document/237548585/bases-de-datos-distribuidas-pdf>.
- Chávez, Y. y Pérez, H. Gestión documental, Gestión de información y Gestión del conocimiento: nociones e interrelaciones. *Bibliotecas Anuales de Investigación. Reflexiones* 8(9), 2013, p. 222-227.
- Cobo, A. Diseño y programación de bases de datos. España, Editorial Visión Libros, 2007. [https://books.google.com.cu/books?hl=es&lr=&id=anCDr9N-kGsC&oi=fnd&pg=PA7&dq=consistencia+de+los+datos+en+un+sistema+distribuido&ots=UYB-lp5opuW&sig=l0vwLFU9pDBIEZn7p9gDQ7p99dl&redir\\_esc=y#v=onepage&q=consistencia&f=false](https://books.google.com.cu/books?hl=es&lr=&id=anCDr9N-kGsC&oi=fnd&pg=PA7&dq=consistencia+de+los+datos+en+un+sistema+distribuido&ots=UYB-lp5opuW&sig=l0vwLFU9pDBIEZn7p9gDQ7p99dl&redir_esc=y#v=onepage&q=consistencia&f=false).
- Community SymmetricDS. SymmetricDS. 2018.<http://www.symmetricds.org/about/overview>.
- Continuent Ltd. Products and Solutions. *Continuent* 2018. <https://www.continuent.com/solutions/>.
- Continuent Ltd. Tungsten Replicator 5.0 Manual. 2016. <https://docs.continuent.com/tungsten-replicator-5.0.pdf>.



- Cowman, S., Triangulation: a means of reconciliation in nursing research, *Journal of Advanced Nursing*. 1993.
- Date, C. J. *Introducción a los Sistemas de Bases de Datos*. 7ma ed. México, Pearson Educación, 2001.
- Diccionario de Informática y Tecnología. Sincronización de Datos. *Diccionario de Informática y Tecnología* 2017. <http://www.alegsa.com.ar/Dic/sincronizar%20datos.php>.
- DLE. Fiabilidad. *Diccionario de la lengua española* 2017a. <http://dle.rae.es/?id=Hpsj999>.
- DLE. Homogeneidad. *Diccionario de la lengua española* 2017b. <http://dle.rae.es/srv/search?m=30&w=homogeneidad>.
- Domínguez, D.; Mora, C. E.; Arellano, G. G. y Torres, K. M. La alternativa del diagnóstico empresarial para la gestión directiva en las pequeñas empresas comerciales en Sinaloa. *Telos* 16(2), 2014, p. 279-281.
- Erazo, Roberto; Cobos, Carlos y Mendoza, Martha. Modelo de servicios web para replicación de instantáneas sobre motores múltiples de bases de datos. *Revista Facultad de Ingeniería* 2014, p. 144–157.
- Escobar, A. Replicación multimaestro en bases de datos PostgreSQL. Tesis de grado, Universidad de Alcalá. Escuela Politécnico Superior, España, 2017, p. 9-10.
- Escobar, D.; Aguilera, A. L. y Parra, Y. Propuesta metodológica para la orientación de aplicaciones informáticas hacia BPM y SOA. *Revista Cubana de Ciencias Informáticas* 10(1), marzo 2016, p. 39-54.
- Febles, O. MIDAC: Modelo para el desarrollo de aplicaciones compuestas basadas en arquitecturas orientadas a servicios. 2013. <http://repositorio.uci.cu/jspui/handle/ident/7910>.
- Figuroa, R. Herramienta administrativa para mecanismo de replicación SymmetricDS. 2011.
- Fuentes, Francisco de Asis López. *Sistemas distribuidos*. Universidad Autónoma Metropolitana. México, Colección Una Década, 2015. [http://dccd.cua.uam.mx/libros/archivos/03IXStream\\_sistemas\\_distribuidos.pdf](http://dccd.cua.uam.mx/libros/archivos/03IXStream_sistemas_distribuidos.pdf).
- Gabillaud, J. *SQL Server 2014. Administración de una base de datos transaccional*. Ediciones ENI, 2015. <https://www.ediciones-eni.com/libro/sql-server-2014-administracion-de-una-base-de-datos-transaccional-con-sql-server-management-studio-9782746096684>.
- García, F. *Modelos y algoritmos distribuidos*. Grupo de Arquitectura de Computadores, 2017. <https://www.arcos.inf.uc3m.es/infods/wp-content/uploads/sites/38/2017/02/2.pdf>.
- González, J. A. y Pazmiño, M. Cálculo e interpretación del Alfa de Cronbach para el caso de validación de la consistencia interna de un cuestionario, con dos posibles escalas tipo Likert. *Revista* 2(1), 2015, p. 62-77.
- Holmes, J. y Schalk, Ch. *JavaServer Faces: The Complete Reference*. 1ra ed. New York, NY, USA, McGraw-Hill, Inc., 2007.
- IBM Knowledge Center. *DatabaseMetaData* interface. 2015. [http://www-01.ibm.com/support/knowledgecenter/ssw\\_ibm\\_i\\_71/rzaha/datameta.htm?lang=es](http://www-01.ibm.com/support/knowledgecenter/ssw_ibm_i_71/rzaha/datameta.htm?lang=es).

- Ilarri, S. Transacciones, recuperación y control de concurrencia. Diseño de Bases de Datos Relacionales, 2012. [http://webdiis.unizar.es/asignaturas/BD/transparenciasBD/PDFs\\_1x1/TransaccionesSilarri.pdf](http://webdiis.unizar.es/asignaturas/BD/transparenciasBD/PDFs_1x1/TransaccionesSilarri.pdf).
- ISO/IEC, "International Standard ISO/IEC 25012", Suiza, 2008. <https://www.iso.org/standard/35736.html>
- Jelínek, P. Logical Replication in PostgreSQL. 2016. <https://wiki.postgresql.org/images/7/7b/Pglogical-tallinn-2016-petr-jelinek.pdf>.
- Jiménez, R. y Patiño, M. Coordinación y Acuerdo. Microsoft Research Cambridge. Universidad Politécnica de Madrid, 2015. [http://resist.isti.cnr.it/free\\_slides/distributed/jimenez/coordinacion-bn.pdf](http://resist.isti.cnr.it/free_slides/distributed/jimenez/coordinacion-bn.pdf).
- Juneau, J. Java Message Service. Berkeley, CA, Apress, Berkeley, CA, 2013. [https://link.springer.com/chapter/10.1007%2F978-1-4302-4426-4\\_13](https://link.springer.com/chapter/10.1007%2F978-1-4302-4426-4_13).
- Justi, R. La enseñanza de ciencias basada en la elaboración de modelos. *Enseñanza de las Ciencias* 24(2), 2006, p. 173-184.
- Kemme, B. y Alonso, G. Don't be lazy, be consistent: Postgres-R, a new way to implement database replication. In Proceedings of the 26th VLDB Conference, Cairo, Egypt, 2000. <http://www.vldb.org/conf/2000/P134.pdf>.
- Kendall, K. E. Análisis y diseño de sistemas. 6ta ed. México, Pearson Educación, 2005 [13 marzo 2018]. <https://luiscastellanos.files.wordpress.com/2014/02/analisis-y-disenio-de-sistemas-kendall-kendall.pdf>.
- Lapedra, R.; Devence, C. y Guiral, J. Introducción a la gestión de sistemas de información en la empresa. 1ra ed. Universitat Jaume, Castelló de la Plana, 2011.
- Lapiente, M. J. Hipertexto: el nuevo concepto de documento en la cultura de la imagen. Tesis de doctorado, Universidad Complutense de Madrid, España, 2013, p. 25-46. <http://www.hipertexto.info/>.
- Len, S. J. Base de Datos Distribuidas Estudio de Actualización de Réplicas. Facultad de Informática, Argentina, 2001.
- Long, E. Sync Schema (DDL) Changes. *SymmetricDS* 2012. <http://www.symmetricds.org/docs/how-to/sync-schema-ddl-changes>.
- Long, E.; Henson, Ch.; Hanes, M.; Wilmer, G. y Brouger, A. SymmetricDS Pro QuickStart Guide 2013. <http://www.jumpmind.com/downloads/symmetricds/doc/SymmetricDSPro-QuickStart-3.5.pdf>.
- Marqués, M. Bases de datos. 1ra ed. Castelló de la Plana, 2011. <http://libros.metabiblioteca.org/handle/001/353>.
- Martínez, J. Sistema Multi-Agente para monitorizar el control de concurrencia de Transacciones Animadas sobre grupos de dispositivos móviles. Tesis de maestría, Instituto Politécnico Nacional, Centro de Investigación en Computación, México, 2007, p. 56-58p.
- Martínez, L. y Cuba, O. Módulo para la resolución de conflictos generado durante el proceso de réplica de estructura.doc. Tesis de grado, Universidad de las Ciencias Informáticas, Cuba, 2015, p. 26-29.

- Mato, R. M. *Sistemas de Bases de Datos*. La Habana, Pueblo y Educación, 2005.
- Mazilu, M. C. Database Replication. *Database Systems Journal I* (2), 2010, p. 33-38.
- Miaton, I. C. Simulador para la evaluación de tiempos de respuesta en transacciones distribuidas y para el estudio de recuperación de errores. 2001. <http://sedici.unlp.edu.ar/handle/10915/3878>.
- Microsoft. Prácticas recomendadas para la Administración de replicación. 2014. <http://technet.microsoft.com>.
- Microsoft Corporation. Sincronizar y resolver conflictos en un conjunto de réplicas (MDB). 2014. <http://office.microsoft.com/es-mx/access-help/sincronizar-y-resolver-conflictos-en-un-conjunto-de-replicas-mdb-HP003069834.aspx>.
- Ocampo, F. Protocolo de control de copias en bases de datos distribuidas. Tesis de maestría. Instituto Politécnico Nacional, Centro de Investigación en Computación, México, 2001, p. 20-30. <https://www.cs.cinvestav.mx/TesisGraduados/1998/TesisFabiolaOcampo.pdf>.
- Ocampo, J. A. Registros Administrativos, Calidad de los Datos y Credibilidad Pública. CEPAL, Presentación y Debate de los Temas Sustantivos de la Segunda Reunión de la Conferencia Estadística de las Américas de la CEPAL, 2004.
- Oracle Corporation. Data Replication and Integration. 2007. <http://www.oracle.com/technetwork/database/information-management/streams-fov-11g-134280.pdf>.
- Oracle Help Center. Streams Concepts and Administration. 2018. [https://docs.oracle.com/cd/E11882\\_01/server.112/e17069.pdf](https://docs.oracle.com/cd/E11882_01/server.112/e17069.pdf).
- Orellana, A. Modelo para la detección de variabilidad en procesos hospitalarios utilizando técnicas de Minería de Procesos. Tesis de doctorado, Universidad de las Ciencias Informáticas, Cuba, 2016, 89-91p. <http://repositorio.uci.cu/jspui/handle/123456789/7159>.
- Oxford University Press. Dialect. *Diccionario Oxford* 2013. <http://oxforddictionaries.com/definition/english/dialect>.
- Panamsource, Inc. Spring Tool Suite. *Spring Latinoamérica* 2015. <https://www.springla.io/spring/spring-tool-suite/>.
- Pérez, D. y Mena, L. E. Módulo de Reko para la replicación entre bases de datos con diferentes estructuras. *Serie Científica de la Universidad de las Ciencias Informáticas* 5(4), Cuba, 37-48p. [https://publicaciones.uci.cu/?journal=SC&page=article&op=viewFile&path\[\]=748&path\[\]=524](https://publicaciones.uci.cu/?journal=SC&page=article&op=viewFile&path[]=748&path[]=524)
- Pérez, J. Solución integrada de réplica de información. Tesis de maestría, Universidad de las Ciencias Informáticas, Cuba, 2008, p. 18-21. [http://repositorio\\_institucional.uci.cu/jspui/handle/ident/TD\\_2810\\_08](http://repositorio_institucional.uci.cu/jspui/handle/ident/TD_2810_08).
- Pérez, J. L. Modelo de replicación de datos en sistemas distribuidos de bases de datos relacionales. Tesis de grado, Universitas Major Pacensis Divi Andreae, 40-89p, Bolivia. 2006. <http://repositorio.umsa.bo/handle/123456789/387>.
- Peter, J.; Freund, A.; Rijkers, E.; Ringer, C. y Sawada, M. Logical Replication in PostgreSQL. 2016. [https://www.pgcon.org/2017/schedule/attachments/473\\_Logical%20Replication%20in%20PostgreSQL%2010.pdf](https://www.pgcon.org/2017/schedule/attachments/473_Logical%20Replication%20in%20PostgreSQL%2010.pdf).
- PostgreSQL. Announcing the release of pglogical 2.0. *Oxford, United Kingdom* 2017. <https://www.postgresql.org/about/news/1744/>.

- PowerData. ¿Qué se entiende por integridad de los datos? *PowerData. Especialistas en Gestión de datos*, 2013. <https://blog.powerdata.es/el-valor-de-la-gestion-de-datos/bid/348870>.
- Quero, M. Confiabilidad y coeficiente Alpha de Cronbach. *Revista de Estudios Interdisciplinarios en Ciencias Sociales* 12, 2010, 248-252.
- Ramírez, M. S. Modelos y estrategias de enseñanza para ambientes innovadores. Editorial Digital. Tecnológico de Monterrey, México, 2012, p. 1-3. [https://www.editorialdigitaltec.com/materialadicional/ID254\\_RamirezMontoya\\_Modelosyestrategiasdeensenanza.cap1.pdf](https://www.editorialdigitaltec.com/materialadicional/ID254_RamirezMontoya_Modelosyestrategiasdeensenanza.cap1.pdf).
- Rojas, D. Propuesta Metodológica para el Desarrollo y la Elaboración de Estadísticas ambientales para los países de América Latina y el Caribe. United Nations Publications, 2009.
- Russo, P. Gestión documental en las organizaciones. Editorial UOC, Barcelona, España, 2014.
- Sampieri, R. H.; Collado, C. F. y Lucio, P. B. Metodología de la investigación Quinta edición. 5ta ed. México, McGraw-Hill, 2010.
- Shull, F.; Singer, J. y Sjøberg, D. Guide to Advanced Empirical Software Engineering. London, Springer Verlag, 2008. <http://read.pudn.com/downloads159/ebook/712887/2008-Guide%20to%20Advanced%20Empirical%20Software%20Engineering.pdf>.
- Silberschatz, A.; Korth, H. F.; y Sudarshan, S. Fundamentos de Bases de Datos. 4ta ed. España, McGraw-Hill, 2002.
- SL, M. Solución de Replicación. 2014. Disponible en la Web: <http://www.solucionesit.com.ve>.
- Snyder, Bruce; Bosnanac, Dejan y Davies, Rob. ActiveMQ in action. Greenwich, Conn., Manning, 2011.
- SQL Server 2012. Replicación de SQL Server. 2012. <http://msdn.microsoft.com/es-es/library/ms151198.aspx>.
- SQL Server 2014. Sincronizar datos. 2014. <http://msdn.microsoft.com/es-es/library/ms151793.aspx>.
- Urbano, R. Advanced Replication. U.S. Government, Oracle Corporation, 2013.
- Vila, Joan. Replicación en sistemas distribuidos. DISCA/UPV Departamento de Informática de Sistemas y Computadoras Universidad Politécnica de Valencia, 2017.
- Weitzenfeld, A. Ingeniería de software orientada a objetos con UML, Java e Internet. Thomson, 2005. <https://books.google.com/cu/books?id=MOviEp0ApQcC&pg=PP6&lpg=PP6&dq=weitzenfeld+alfredo.+ingenier%C3%ADA+de+software+orientada+a+objetos&source=bl&ots=OWzHheQv9x&sig=PTmThWW1GaLbcSvn-vGzdqxJOY&hl=es&sa=X&ei=vykUUfS6NMnC0QGSzoCQAQ>.
- Wheeler, W. y White, J. Spring in practice: covers Spring 3. Shelter Island, NY, Manning, 2013.

## ANEXOS

### Anexo 1 Composición de los expertos involucrados en el diagnóstico

Tabla 4 Composición de los expertos involucrados en el diagnóstico

<b>-No-</b>	<b>-Perfil de los expertos-</b>	<b>-Cantidad-</b>
<b>1</b>	<b>Perfil de trabajo</b>	
	Jefe de proyecto	2
	Desarrollador	3
	Analista	2
	Arquitecto	1
	Diseñador de BD	2
	Profesor	1
	Directivo con conocimientos en el tema	1
<b>2</b>	<b>Centros involucrados, Entidad</b>	
	CDAE, UCI	7
	FORTEX, UCI	1
	DATEC, UCI	4
<b>3</b>	<b>Años de experiencia</b>	
	Entre 5 y 10	6
	Menos de 5	6

**Anexo 2 Cuestionario para determinar los criterios para realizar el análisis comparativo entre modelos y soluciones de replicación de datos**

Cuestionario para determinar los atributos para realizar un análisis comparativo entre modelos y soluciones de replicación de datos				
Entidad:				
Área:				
Datos del encuestado				
Nombre y apellidos:				
Cargo o rol:				
Nivel escolar:		Técnico medio <input type="checkbox"/>	Universitario <input type="checkbox"/>	
Categoría docente:		Instructor <input type="checkbox"/>	Asistente <input type="checkbox"/>	Auxiliar <input type="checkbox"/> Titular <input type="checkbox"/>
Categoría científica:		Especialista <input type="checkbox"/>	Máster <input type="checkbox"/>	Doctor <input type="checkbox"/>
Años de experiencia:				
No	Preguntas	Repuestas		
1	¿Ha realizado usted alguna investigación sobre temas de replicación? (en proyecto, investigación, despliegue).	Sí ____ No ____ No recuerdo ____		
2	¿En el análisis documental de soluciones homólogas, cuáles tópicos le han sido útiles para comprender el objeto de estudio?	__ Escenarios de replicación en los que se basa la solución. __ Descripción de los procesos __ Herramientas y tecnologías empleadas __ Variable que experimentan (según objetivo y validación) __ Otros: _____		
3	¿Por qué usted considera que parte del conocimiento documentado no ha podido aplicarse?	__ No he comprendido la solución propuesta. __ La solución no se ha correspondido con el escenario a implementar. __ No hay correspondencia entre el objetivo, propuesta y validación. __ Las soluciones fundamentadas son para tecnologías específicas que no han formado parte de mi entorno. __ Carecen de métodos y algoritmos a seguir. Otras consideraciones: _____		
4	Sobre los escenarios de replicación: ¿Qué características considera útiles en su comprensión y aplicación?	__ La clasificación de la réplica en todos sus aspectos. __ Conocer si la herramienta está accesible y no bajo licencia. __ Costo económico para emplear la herramienta. __ Beneficios de optar por la selección de los escenarios descritos. __ Reflejar en las descripciones el proceder tecnológico. Otros: _____		
5	Sobre la descripción de los procesos: ¿Qué elementos considera importantes para favorecer su aplicación?	__ Describir todos los procesos asociados a la réplica. __ Estandarización de los procesos. __ RR. HH necesarios para el desarrollo. __ Orientación en cada proceso hacia su implementación (qué y cómo).		

		Otros: _____
6	Sobre las herramientas y tecnologías: ¿Considera importante orientarlos técnicamente?	Sí ____ No ____ No sé ____
	6.1 ¿Cuáles son las razones por las que considera que sí debe aplicarse?	<input type="checkbox"/> Para valorar su utilización en los proyectos. <input type="checkbox"/> Para comparar su accesibilidad a la solución y mejorar costos de desarrollo. <input type="checkbox"/> Para contribuir a la reutilización y disminuir tiempo de desarrollo. <input type="checkbox"/> Para valorar competencia en el mercado. Otros: _____
	6.2 ¿Por qué usted considera que no es necesario?	Criterio: _____
7	¿Considera usted que es necesario que la solución examinada refleje las relaciones entre la variable a experimentar y los procesos comprendidos?	Sí ____ No ____ No sé ____
8	¿Considera usted importante establecer la relación entre los beneficios del escenario propuesto, el alcance definido para la solución tecnológica y descripciones técnicas que se requieren?	Sí ____ No ____ No sé ____
9	¿Qué indicadores se han evidenciado en el análisis documental o cuáles han sido cubiertos por las soluciones desde los objetivos, alcance hasta desarrollo y validación?	<input type="checkbox"/> Performance <input type="checkbox"/> Velocidad de transferencia <input type="checkbox"/> Consistencia <input type="checkbox"/> Balanceo de carga <input type="checkbox"/> Tolerancia a fallo <input type="checkbox"/> Otros: _____
10	¿Considera importante valorar el grado de consistencia ofrecido por las soluciones?	Sí ____ No ____ No sé ____
11	¿Usted ha examinado alguna solución que se enfoque en la consistencia de los datos?	<input type="checkbox"/> Sí <input type="checkbox"/> No, pero implícitamente lo han tratado <input type="checkbox"/> No <input type="checkbox"/> No, pero se menciona <input type="checkbox"/> No recuerdo
12	¿Considera usted que es importante gestionar la consistencia en los procesos asociados a la réplica de datos?	<input type="checkbox"/> Para mantener la coherencia y homogeneidad de los datos. <input type="checkbox"/> Para que se puedan tomar las decisiones correctas. <input type="checkbox"/> Mantener la integridad referencial y completitud de los datos. <input type="checkbox"/> Actualizar de manera correcta los cambios DML y DDL. Otras consideraciones: _____

### Anexo 3 Encuesta y procedimiento empleado para determinar el coeficiente de competencia de los candidatos a expertos y resultados obtenidos

Encuesta para determinar el coeficiente de competencias de los expertos

Compañero (a): \_\_\_\_\_

Usted ha sido seleccionado como posible experto para ser consultado respecto a temas relacionados con la replicación de datos en ambientes distribuidos, con vistas a la investigación que se está llevando a cabo. Agradecemos sinceramente su valiosa cooperación.

Gracias.

1. Marque con una cruz (X) en la tabla siguiente el valor que se corresponde con el grado de conocimiento que usted posee sobre "replicación de datos". (Escala ascendente).

0	1	2	3	4	5	6	7	8	9	10

2. Realice una autoevaluación del grado de influencia que cada una de las fuentes que le presentamos a continuación ha tenido en su conocimiento y criterio sobre "análisis de los procesos de replicación de datos". Marque con una cruz (X) según corresponda en A (alto), M (medio) o B (bajo).

No	Fuente de argumento	Grado de influencia de cada una de las fuentes		
		A (alto)	M (medio)	B (bajo)
1	Estudios teóricos realizados por usted.			
2	Experiencia adquirida durante su vida profesional.			
3	Conocimiento de investigaciones y/o publicaciones nacionales e internacionales.			
4	Conocimiento propio sobre el estado del tema de investigación.			
5	Actualización en cursos de posgrado, diplomados, maestrías, doctorado, etc.			
6	Intuición			

#### Cálculo del coeficiente de competencia de los expertos que evaluaron el modelo desarrollado

El cálculo de dicho coeficiente se realiza de la forma siguiente:

$K_{comp} = \frac{1}{2} (K_c + K_a)$ , donde:

**K<sub>comp</sub>**: coeficiente de competencia.

**K<sub>c</sub>**: coeficiente de conocimiento o información que tiene el experto acerca del problema, calculado sobre la valoración del propio experto en una escala de 0 a 10 y multiplicado por 0,1.

**K<sub>a</sub>**: coeficiente de argumentación o fundamentación de los criterios del experto, obtenido como resultado de la suma de los puntos de acuerdo con la siguiente tabla patrón:

Tabla 5 Fuentes de argumentación del conocimiento de los expertos

-No-	-Fuentes de argumentación-	-A (alto)-	-M (medio)-	-B (bajo)-
1	Estudios teóricos realizados por usted.	0,30	0,20	0,10
2	Experiencia adquirida durante su vida profesional.	0,50	0,37	0,30



3	Conocimiento de investigaciones y/o publicaciones nacionales e internacionales.	0,05	0,04	0,03
4	Conocimiento propio sobre el estado del tema de investigación.	0,05	0,04	0,03
5	Actualización en cursos de posgrado, diplomados, maestrías, doctorado, etc.	0,05	0,04	0,03
6	Intuición	0,05	0,03	0,02
	<b>Total</b>	1,00	0,70	0,50

Se plantea entonces que:

La competencia del experto es de Alta (A): Si  $K_{comp} > 0,7$

La competencia del experto es Media (M): Si  $0,5 < K_{comp} = < 0,7$

La competencia del experto es Baja (B): Si  $K_{comp} = < 0,5$

**RESULTADOS:** Competencia de los expertos que participaron en la valoración del modelo elaborado.

Tabla 6 Resultados de la encuesta para determinar nivel de competencia

-Experto-	-Kc-	1	2	3	4	5	6	-Ka-	-Kcomp-	-Valor-
1	0,9	0,30	0,50	0,04	0,05	0,04	0,05	0,99	0,95	alto
2	0,7	0,20	0,37	0,05	0,05	0,03	0,05	0,75	0,73	alto
3	0,6	0,20	0,50	0,03	0,03	0,03	0,05	0,84	0,72	alto
4	0,8	0,30	0,37	0,04	0,04	0,05	0,05	0,85	0,83	alto
5	0,7	0,20	0,37	0,04	0,04	0,05	0,03	0,73	0,72	alto
6	0,9	0,30	0,50	0,04	0,05	0,05	0,05	0,84	0,87	alto
7	0,7	0,20	0,37	0,04	0,05	0,04	0,05	0,75	0,73	alto
8	0,6	0,10	0,37	0,04	0,05	0,03	0,05	0,64	0,62	medio
9	0,6	0,30	0,50	0,05	0,05	0,05	0,05	1,00	0,80	alto
10	0,9	0,30	0,37	0,05	0,05	0,05	0,05	0,87	0,89	alto
11	0,7	0,30	0,37	0,04	0,04	0,05	0,05	0,85	0,78	alto
12	0,6	0,10	0,37	0,03	0,05	0,05	0,05	0,65	0,63	medio
13	0,8	0,20	0,50	0,03	0,05	0,03	0,05	0,86	0,83	alto
14	0,6	0,20	0,50	0,03	0,03	0,03	0,05	0,84	0,72	alto
15	0,4	0,10	0,30	0,03	0,04	0,03	0,03	0,55	0,47	bajo
16	0,4	0,10	0,37	0,03	0,03	0,04	0,03	0,60	0,50	bajo
17	0,2	0,10	0,37	0,03	0,03	0,03	0,05	0,61	0,41	bajo
18	0,6	0,20	0,37	0,03	0,04	0,04	0,05	0,72	0,67	medio
19	0,6	0,10	0,37	0,04	0,05	0,03	0,05	0,64	0,62	medio
20	0,4	0,10	0,30	0,03	0,04	0,03	0,03	0,53	0,47	bajo

#### Anexo 4 Cuestionario para evaluar la aceptación del modelo por los expertos

Estimado experto (a): la presente encuesta forma parte de una investigación que está dirigida al desarrollo de un modelo para la replicación de datos en ambientes distribuidos, que contribuya a mantener la consistencia de los datos replicados. Por cuanto, sus valoraciones acerca de los asuntos que se someten a su consideración servirán de ayuda.

MODELO PARA LA RÉPLICA DE DATOS EN SISTEMAS DE BASES DE DATOS DISTRIBUIDAS				
<b>Datos del encuestado</b>				
Entidad:				
Área:				
Nombre y apellidos:				
Cargo o rol:				
Nivel escolar:		Técnico medio <input type="checkbox"/>	Universitario <input type="checkbox"/>	
Categoría docente:		Instructor <input type="checkbox"/>	Asistente <input type="checkbox"/>	Auxiliar <input type="checkbox"/> Titular <input type="checkbox"/>
Categoría científica:		Especialista <input type="checkbox"/>	Máster <input type="checkbox"/>	Doctor <input type="checkbox"/>
Años de experiencia:				
No	Preguntas	Repuestas		
1	Los principios teóricos establecidos sustentan un modelo que contribuye a mantener la consistencia de los datos en ambientes distribuidos.	MA <input type="checkbox"/>	ED <input type="checkbox"/>	
		DA <input type="checkbox"/>	CD <input type="checkbox"/>	
		Sí-No <input type="checkbox"/>		
2	La estructura del modelo articula adecuadamente los principios establecidos, representando apropiadamente los componentes para el desarrollo de soluciones de replicación de datos consistentes.	MA <input type="checkbox"/>	ED <input type="checkbox"/>	
		DA <input type="checkbox"/>	CD <input type="checkbox"/>	
		Sí -No <input type="checkbox"/>		
3	La forma en que el modelo propuesto integra los principios y componentes establecidos en su funcionamiento, contribuye a su aplicación en entornos reales y agiliza el desarrollo de soluciones de replicación de datos.	MA <input type="checkbox"/>	ED <input type="checkbox"/>	
		DA <input type="checkbox"/>	CD <input type="checkbox"/>	
		Sí-No <input type="checkbox"/>		
4	Los componentes que conforman el modelo y las fuentes teóricas que los sustentan, cubren aspectos relevantes para mantener la consistencia de los datos en ambientes distribuidos.	MA <input type="checkbox"/>	ED <input type="checkbox"/>	
		DA <input type="checkbox"/>	CD <input type="checkbox"/>	
		Sí-No <input type="checkbox"/>		
5	Los procesos vinculados a cada componente son modelados de manera adecuada y describen correctamente su funcionamiento.	MA <input type="checkbox"/>	ED <input type="checkbox"/>	
		DA <input type="checkbox"/>	CD <input type="checkbox"/>	
		Sí-No <input type="checkbox"/>		
6	Los componentes son descritos detalladamente y de forma genérica, favoreciendo la comprensión tanto para los analistas de sistema como para los programadores de una solución de réplica de datos.	MA <input type="checkbox"/>	ED <input type="checkbox"/>	
		DA <input type="checkbox"/>	CD <input type="checkbox"/>	
		Sí-No <input type="checkbox"/>		
7	El modelo desarrollado contribuye a mantener la consistencia de los datos replicados en ambientes distribuidos.	MA <input type="checkbox"/>	ED <input type="checkbox"/>	
		DA <input type="checkbox"/>	CD <input type="checkbox"/>	
		Sí-No <input type="checkbox"/>		
8	El modelo propuesto otorga beneficios sobre importantes funciones de integración de los procesos que se llevan a cabo en los sistemas distribuidos y para diseminar la información entre múltiples localizaciones, procesos sustanciales en los sistemas hospitalarios, gubernamentales, bancarios, entre otros como: empresariales, con alto flujo de datos.	MA <input type="checkbox"/>	ED <input type="checkbox"/>	
		DA <input type="checkbox"/>	CD <input type="checkbox"/>	
		Sí-No <input type="checkbox"/>		

¡MUCHAS GRACIAS POR SU COLABORACIÓN!

**Respuestas emitidas por los expertos para cada indicador**

5: Muy De Acuerdo (MA)

4: De Acuerdo (DA)

3: Ni de Acuerdo Ni en Desacuerdo (Sí-No)

2: En Desacuerdo (ED)

1: Completamente en Desacuerdo (CD)

Tabla 7 Respuesta emitida por los expertos para cada indicador

<b>-Experto-</b>	<b>-Indicador-</b>							
	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>
<b>1</b>	5	5	5	5	5	5	5	5
<b>2</b>	5	5	5	5	5	5	5	5
<b>3</b>	4	5	5	4	5	5	4	5
<b>4</b>	5	5	5	4	5	5	5	5
<b>5</b>	5	5	5	4	5	5	5	5
<b>6</b>	4	4	4	4	4	4	4	4
<b>7</b>	4	5	4	4	5	4	4	5
<b>8</b>	5	5	4	5	5	5	5	5
<b>9</b>	5	4	5	5	5	5	5	5
<b>10</b>	5	4	5	5	5	5	5	5
<b>11</b>	4	4	5	5	5	4	5	5
<b>12</b>	5	5	5	5	5	5	5	4
<b>13</b>	5	5	5	5	5	5	5	5
<b>14</b>	5	5	5	5	5	5	5	5
<b>15</b>	5	5	5	5	5	5	5	5
<b>16</b>	5	5	5	5	5	5	5	5

### Anexo 5 Cálculo del coeficiente de confiabilidad de los expertos en el cuestionario

$$\alpha = \frac{k}{k-1} \left[ 1 - \frac{\sum S_i^2}{S_T^2} \right]$$

**k:** Número de ítems.

$\sum S_i^2$ : Sumatoria de Varianzas de los elementos (ítems).

$S_T^2$ : Varianza de la suma de los elementos (ítems).

**$\alpha$ :** Coeficiente Alfa-Cronbach.

Las respuestas fueron codificadas a los valores 1, 2, 3, 4 y 5 respectivamente. Se procedió a calcular la varianza de la población. Los valores obtenidos para cada ítem fueron:

Tabla 8 Varianza de los resultados de la encuesta por preguntas

-Ítems-	-Varianza de la población-
Pregunta 1	0,19
Pregunta 2	0,19
Pregunta 3	0,15
Pregunta 4	0,21
Pregunta 5	0,06
Pregunta 6	0,15
Pregunta 7	0,15
Pregunta 8	0,11

La sumatoria de varianzas de los ítems ( $\sum S_i^2$ ) arrojó como resultado 1,21 y la varianza de la suma de los ítems ( $S_T^2$ ) resultó 4,62. Al aplicar la fórmula anterior y conociendo que el valor del número de ítems (K) es 8, el coeficiente Alpha de Cronbach, como se muestra a continuación, resultó 0,84:

$$\alpha = \frac{8}{8-1} \left[ 1 - \frac{1,21}{4,62} \right]$$

$$\alpha = 1,14 [1 - 0,26]$$

$$\alpha = 0,84$$

## Anexo 6 Cuestionario para evaluar la satisfacción de los usuarios con respecto al modelo

MODELO PARA LA RÉPLICA DE DATOS EN SISTEMAS DE BASES DE DATOS DISTRIBUIDAS				
<b>Datos del encuestado</b>				
Entidad:				
Área:				
Nombre y apellidos:				
Cargo o rol:				
Nivel escolar:	Técnico medio <input type="checkbox"/>	Universitario <input type="checkbox"/>		
Categoría docente:	Instructor <input type="checkbox"/>	Asistente <input type="checkbox"/>	Auxiliar <input type="checkbox"/>	Titular <input type="checkbox"/>
Categoría científica:	Especialista <input type="checkbox"/>	Máster <input type="checkbox"/>	Doctor <input type="checkbox"/>	
Años de experiencia:				
No	Preguntas	Repuestas		
1	¿Ha analizado procesos de replicación de datos en ambientes distribuidos en algún momento?	Sí <input type="checkbox"/>	No <input type="checkbox"/>	
2	¿Conoce usted de algún modelo que le permita hacer estos análisis?	Sí <input type="checkbox"/>	No <input type="checkbox"/>	
3	¿Considera usted que sin mantener la consistencia de los datos replicados, se pueda realizar efectivamente la replicación de los datos en ambientes distribuidos?	Sí <input type="checkbox"/>	No <input type="checkbox"/>	
4	¿Considera usted que es aconsejable analizar la consistencia de los datos replicados en la ejecución de los procesos asociados a la réplica?	Sí <input type="checkbox"/>	No <input type="checkbox"/>	
5	¿Si usted necesitara mantener la consistencia de los datos en la ejecución de los procesos de replicación usaría el modelo propuesto?	Sí <input type="checkbox"/>	No <input type="checkbox"/>	
6	¿Considera que la utilización de este modelo estimulará el desarrollo de soluciones de replicación de datos consistentes?	Sí <input type="checkbox"/>	No <input type="checkbox"/>	
7	¿Le satisface el modelo propuesto para mantener la consistencia de los datos en ambientes distribuidos?	Me satisface mucho No me satisface tanto Me da lo mismo Me insatisface más de lo que me satisface No me satisface nada No sé qué decir		
8	¿Qué relevancia le concede usted a la vinculación del análisis de procesos de replicación con un modelo para mantener la consistencia de los datos replicados?	Entre 0 y 10 __		
9	Valore el impacto que produciría el uso de un modelo para mantener la consistencia de los datos en la ejecución de procesos de replicación de cambios.	Entre 0 y 10 __		

10- ¿Qué elementos considera positivos del modelo propuesto?

11- ¿Qué elementos considera negativos del modelo propuesto?

12- ¿Qué sugeriría mejorar del desarrollo e implantación del modelo propuesto?

### Procedimiento para aplicar la técnica ladov

La técnica se basa en el análisis del cuestionario anterior que tiene una estructura interna determinada que sigue una relación entre tres cuestionamientos cerrados y el análisis posterior de dos cuestiones abiertas. La relación entre las preguntas cerradas se establece a través del denominado "Cuadro Lógico de ladov".

Tabla 9 Cuadro Lógico de ladov modificado por la autora

	¿Considera usted que sin mantener la consistencia de los datos replicados, se pueda realizar efectivamente la replicación de los datos en ambientes distribuidos?								
	No			No sé			Sí		
	¿Si usted necesitara mantener la consistencia de los datos en la ejecución de los procesos de replicación usaría el modelo propuesto?								
¿Le satisface la representación de este modelo para mantener la consistencia de los datos en la ejecución de los procesos de replicación?	Sí	No sé	No	Sí	No sé	No	Sí	No sé	No
Me satisface mucho	1	2	6	2	2	6	6	6	6
No me satisface tanto	2	2	3	2	3	3	6	3	6
Me da lo mismo	3	3	3	3	3	3	3	3	3
Me insatisface más de lo que me satisface	6	3	6	3	4	4	3	4	4
No me satisface nada	6	6	6	6	4	4	6	4	5
No sé qué decir	2	3	6	3	3	3	6	3	4

Luego de aplicado el cuestionario y haber triangulado las preguntas cerradas en el Cuadro Lógico de ladov, el número resultante de la interrelación de las tres preguntas indica la posición de cada experto en la escala de satisfacción siguiente:

1. Clara satisfacción
2. Más satisfecho que insatisfecho
3. No definida
4. Más insatisfecho que satisfecho
5. Clara insatisfacción
6. Contradictoria

Para medir este grado de satisfacción se tomó una muestra de 11 usuarios pertenecientes a los centros CDAE, DATEC, DIN y FORTES, teniendo en cuenta los años de experiencia en el desarrollo y/o despliegue de soluciones de replicación y/o gestión de datos, el rol que desempeñan, entre otros aspectos.

Tabla 10 Distribución de los usuarios por escalas de años de experiencia

-Años de experiencia-	-Cantidad de usuarios-
Entre 0 y 5 años	1
Entre 5 y 10 años	5
Más de 10 años	5

Tabla 11 Resultados de la satisfacción individual

-Escala de satisfacción-	-Cantidad de usuarios-	-Porcentaje-
Clara satisfacción	10	91 %
Más satisfecho que insatisfecho	1	9 %
No definida	-	-
Más insatisfecho que satisfecho	-	-
Clara insatisfacción	-	-
Contradictoria	-	-

## Anexo 7 Composición de los participantes en el grupo focal

Tabla 12 Composición de los participantes en el grupo focal

-No-	-Ocupación de los participantes-	-Procedencia-	-Nivel académico o científico-	-Años de experiencia-
1	Directora de centro	CDAE	Doctor en Ciencias Técnicas	10
2	Jefa del proyecto Replicador de datos (REKO)	CDAE	Ingeniero	5
3	Desarrollador Replicador de datos (REKO)	CDAE	Ingeniero	6
4	Planificador y Analista del Sistema de Auditoría de datos para la Contraloría General de la República (AUDATv2.0)	DATEC	Máster en Calidad de Software	6
5	Desarrollador	DATEC	Ingeniero	4
6	Miembro del Subcomité 7	CALISOF	Máster en Calidad de Software	9
7	Miembro del Subcomité 7	CALISOF	Ingeniero	8
8	Auditora líder	CALISOF	Ingeniero	8

## Anexo 8 Guía de desarrollo del Grupo focal

**Número de participantes:** 8

**Fecha:** 28 de mayo del 2018

**Lugar:** Departamento de Soluciones Empresariales del centro CDAE

**Hora:** 9:00 AM

### Apertura

- Describir lo que constituye un grupo focal.
- Explicar el objetivo de la reunión.
- Explicar procedimiento.
- Presentar el modelo.

### Objetivos

**Objetivo de la investigación:** desarrollar un modelo para la réplica de datos en sistemas de bases de datos distribuidas que contribuya a mantener la consistencia de los datos replicados.

**Objetivo del grupo focal:** definir las medidas adecuadas para evaluar los indicadores propuestos en la operacionalización de la variable dependiente.

**Nombre el moderador:** Ing. Mirnerys Delgado del Solar

**Nombre del observador:** Ing. Reiman Alfonso Azcuy

### Guía de preguntas

¿Considera útil el modelo para mantener la consistencia de los datos replicados?

¿Considera importante definir indicadores para evaluar la consistencia de los datos en ambientes distribuidos?

¿Qué indicadores considera que deban aplicarse para evaluar la consistencia de los datos en los sistemas de replicación? Para evaluar la consistencia de los datos en la replicación debe considerarse la medida base:

- a) Acciones construidas
- b) Acciones persistidas
- c) Acciones replicadas o enviadas
- d) Acciones aplicadas
- e) Conflictos generados
- f) Conflictos resueltos
- g) Acciones importadas
- h) Acciones exportadas
- i) Otras. ¿Cuáles?

¿Qué indicadores considera que son relevantes para emitir un resultado final en la evaluación de la consistencia?

¿La herramienta informática que soporta el modelo propuesto y las indicaciones metodológicas elaboradas facilitan su aplicación?