

SISTEMA DE REPLICACIÓN DE DATOS BASADO EN MENSAJES XML Y COLAS FIFO
PARA LA EMPRESAS CARGA S.A.S.

TOMO METODOLÓGICO

JOHN JAIRO BOTERO RIVERA
JULIÁN ANDRÉS SANDOVAL GALLEGO

FUNDACIÓN UNIVERSITARIA SAN MARTÍN
FACULTAD DE UNIVERSIDAD ABIERTA Y A DISTANCIA
“EDUCACIÓN A TRAVÉS DE ESCENARIOS MÚLTIPLES”
PROGRAMA DE INGENIERÍA DE SISTEMAS
SABANETA

2013

SISTEMA DE REPLICACIÓN DE DATOS BASADO EN MENSAJES XML Y COLAS FIFO
PARA LA EMPRESAS CARGA S.A.S.

JOHN JAIRO BOTERO RIVERA
JULIÁN ANDRÉS SANDOVAL GALLEGO

Trabajo de grado para optar al título de Ingeniero de Sistemas

Asesor

GUILLERMO LEÓN RESTREPO HURTADO

Ingeniero de Sistemas

FUNDACIÓN UNIVERSITARIA SAN MARTÍN
FACULTAD DE UNIVERSIDAD ABIERTA Y A DISTANCIA
“EDUCACIÓN A TRAVÉS DE ESCENARIOS MÚLTIPLES”
PROGRAMA DE INGENIERÍA DE SISTEMAS
SABANETA

2013

NOTA DE ACEPTACIÓN

Firma del Presidente del Jurado

Firma del Jurado 1

Firma del Jurado 2

RESUMEN

En un medio cada vez más competitivo, como lo es el sector de transporte de mercancía, las empresas transportadoras, buscan tener presencia en cualquier lugar que el generador de carga así lo requiera, motivo por el cual, la información disponible y actualizada en las diferentes sucursales, juega un papel muy importante. En aquellos casos en donde un aplicativo web no puede cumplir con todos los requisitos funcionales, se recurre a otros métodos, los cuales, en algunas ocasiones, conllevan a trabajar con bases de datos descentralizadas.

Por medio del presente proyecto, se muestra cómo la aplicabilidad de diferentes conceptos, adquiridos en el pregrado de Ingeniería de Sistemas, permiten desarrollar e implementar al interior de la empresa Carga S.A.S., un sistema de replicación de datos, cuya finalidad es crear un canal que permita la sincronización de los mismos en cada una las sedes separadas geográficamente, ofreciendo solución a una necesidad detectada en la organización, en el cual se busca que las sucursales operen de forma independiente, pero al mismo tiempo, el resultado de su actividad, se vea reflejada en el resto de las oficinas que conforman el Grupo Organizacional.

Palabras clave: Transporte, sincronización, bases de datos, ingeniería, logística de transporte, generador de carga, aplicación, protocolo de comunicación, servicios web

ABSTRACT

In an increasingly competitive environment, such as the merchandise freight transport, the transport logistic companies, looks for make presence in any place that the merchandise freight's generator requires, reason why, the information available and updated in every single office, plays a very important role. On those situations in where a web app does not accomplish with the whole functional requirements, other techniques are used, wich in some situations, involve working with databases decentralized.

Through in this project, shown how applicability of differents concepts, acquired in the undergraduate of System's Engineering, allow to develop and carry through within of the organization Carga S.A.S., a data replication System, whose ultimate goal is to create a channel that allow the sync between every sites of the organization, geographically separated, offering answer to a weakness detected in the Organization, in wich seeks that the remote sites work independently, but at the same time, the result of its operation, is shown in the rest of the offices that conform the Organizational Group.

Key words: Transport, sync, databases, engineering, transport logistic companies, merchandise freight's generator, app, protocol of communication, web service

TÍTULO

**SISTEMA DE REPLICACIÓN DE DATOS BASADO EN MENSAJES XML Y COLAS
FIFO PARA LA EMPRESAS CARGA S.A.S.**

CONTENIDO

APÉNDICE	12
INTRODUCCIÓN.....	13
PROBLEMA.....	15
PLANTEAMIENTO	15
FORMULACIÓN.....	16
SISTEMATIZACIÓN.....	16
JUSTIFICACIÓN	18
TEÓRICA	18
METODOLÓGICA	20
PRÁCTICA.....	20
OBJETIVOS	22
OBJETIVO GENERAL	22
OBJETIVOS ESPECÍFICOS	22
MARCO REFERENCIAL	23
MARCO TEÓRICO / CONCEPTUAL	23
Conceptos de la Organización.....	23
Replicación, Definición Básica	25

El Dato	26
Sistema	26
Sistema de Información	27
Almacenamiento y Estructura de los Datos	29
Entidades y Atributos	30
Base de Datos Relacionales	31
DBMS (Database Management System)	32
Sistema de Comunicación	33
Protocolo TCP/IP (Transmission Control Protocol/Internet Protocol)	34
IPSec (Internet Protocol Security) y VPN (Virtual Private Network)	35
Datagrama IP	36
Modelo TCP/IP y Modelo OSI	39
XML (eXtensible Markup Language)	40
Colas FIFO	43
Replicación, un concepto más profundo	44
MARCO HISTÓRICO	49
MARCO LEGAL	55
DISEÑO METODOLÓGICO	57
REFERENTE METODOLÓGICO	57
Tipo de Estudio	57

Método de Investigación	58
Fuentes de Información	58
REFERENTE PRÁCTICO	59
RUP (Rational Unified Process - Proceso Racional Unificado)	59
Inicio	60
Elaboración	60
Construcción	60
Transición	61
ESTRUCTURA DEL TRABAJO	62
PRESUPUESTO	67
Recursos Financieros	67
Recurso Humano	67
Recursos Tangibles e Intangibles.....	68
CRONOGRAMA	69
CONCLUSIONES.....	70
RECOMENDACIONES.....	71
BIBLIOGRAFÍA.....	72
CIBERGRAFÍA	74

LISTA DE TABLAS

Tabla 1 – Recurso Humano	67
Tabla 2 – Recursos Tangibles e Intangibles.....	68
Tabla 3 - Cronograma de Actividades.....	69

TABLA DE ILUSTRACIONES

Ilustración 1 – Componentes de un Sistema	27
Ilustración 2 – Sistemas de Información (en la informática)	28
Ilustración 3 – Ejemplo de Entidad y Atributos.....	31
Ilustración 4 – Sistema de Base de Datos	33
Ilustración 5 – Sistema de Comunicación Básico	35
Ilustración 6 – Transporte de datos con la utilización de claves asimétricas	36
Ilustración 7 – Datagrama IP y campo de Datos.....	39
Ilustración 8 – Modelo TCP/IP vs. Modelo OSI.....	40
Ilustración 9 – Ejemplo XML Bien Formado (Well Formed).....	42
Ilustración 10 – Cola FIFO.....	44
Ilustración 11 – Replicación Transaccional	47
Ilustración 12 – Replicación en Mezcla	48
Ilustración 13 – Replicación de Instantáneas	48

APÉNDICE

El presente apéndice se desarrolla con los conceptos comprendidos por parte de los autores del proyecto con los términos que apropiaran al lector durante la aplicación del trabajo escrito:

Colas: Son secuencias de datos, en las cuales existe la oportunidad de hacer inserción y eliminación de los mismos.

FIFO: Es la traducción de primero en entrar primero en salir, término utilizado para transacciones con datos.

Mensajes: Los mensajes son un técnica empleada de programación para informar el estado de situaciones específicas.

Replicación: Es la copia y actualización de la información en varias bases de datos, ya que es una necesidad crítica para muchas organizaciones.

Servidor: Es un nodo que hace parte de una red, este proporciona servicios a otros nodos llamados clientes.

Transacción: Es un conjunto de ordenes ejecutadas bajo un plan de trabajo.

INTRODUCCIÓN

El sector de transporte de mercancía en Colombia, como muchos otros, no ha obtenido la suficiente inversión en tecnología, dado que la gran mayoría de sus procesos, tienen su origen en entes privados, bien sean generadores de carga o administradores logísticos de transporte, los cuales son los que determinan qué cantidad de esfuerzo, tanto económico como tecnológico, se debe aplicar, dicho esfuerzo, generalmente se enfoca en simplemente dar cumplimiento a los requerimientos técnicos impuestos por el Ministerio de Transporte¹, como por ejemplo el Manifiesto Carga², el cual cuenta con su versión electrónica por medio de una herramienta denominada RNDC³.

Dicha limitante de inversión de recursos por parte de los administradores logísticos de transporte, ha hecho que a nivel general, la administración para los propietarios de vehículos sea cada vez más compleja, enfrentando varias problemáticas, las cuales afectan también a los conductores, entre ellas, el monopolio en los turnos para la adjudicación de mercancía.

Carga S.A.S., es una empresa que está fuertemente comprometida con realizar aportes que contribuyan al desarrollo del sector transporte, y por lo tanto, haciendo alusión a su política de

¹ Organismo del Gobierno Nacional encargado de formular y adoptar las políticas, planes, programas, proyectos y regulación económica del transporte, el tránsito y la infraestructura, en los modos carretero, marítimo, fluvial, férreo y aéreo del país, <https://www.mintransporte.gov.co/publicaciones.php?id=33>

² Documento que ampara el transporte de mercancías ante las distintas autoridades, cuando estas se movilizan en vehículos de servicio público, mediante contratación a través de empresas de transporte de carga legalmente constituidas y debidamente habilitadas por el Ministerio de Transporte, <http://web.mintransporte.gov.co/consultas/mercapeli/documentos/manifiesto.htm>

³ Registro Nacional de Despacho de Carga (RNDC), es un sistema de información que permite recibir, validar y transmitir la información generada en las operaciones del Servicio Público de Transporte de Carga por Carretera, <http://rndc.mintransporte.gov.co/LinkClick.aspx?fileticket=G1pa0WNWbEo%3d&tabid=80&language=es-MX>

transparencia, ofrece tanto a conductores como propietarios, nuevas opciones que les permitan estar siempre informados sobre las mercancías disponibles para cargar, estado de los turnos de conductores, disponibilidad del pago de fletes⁴ a nivel nacional, entre otros, pero debido a que dicha información debe ser transmitida a las diferentes sucursales, se debe contar con un sistema que tenga la capacidad de replicar los datos de forma confiable y oportuna

Dentro del siguiente proyecto, se presenta el desarrollo de una propuesta tecnológica, cuya función principal es la de servir como una plataforma de integración a nivel de software, y adicionalmente, permite cerrar la brecha con respecto a los problemas de replicación de información detectados en la empresa Carga S.A.S.

La solución se basa en un patrón de Arquitectura Orientada a Servicios (SOA por sus siglas en inglés, Service Oriented Architecture) y un software intermedio (o middleware), el cual se encarga del transporte de los datos utilizando un sistema de colas, dicho middleware es también conocido como ESB (Enterprise Service Bus).

La combinación de todos estos componentes busca satisfacer todas las necesidades del cliente con respecto a la replicación de información entre las diferentes sedes, ajustándose de manera adecuada tanto en tiempo como en presupuesto, y poder de esta forma cumplir con los lineamientos estratégicos y la política de transparencia y calidad, trazados por la organización.

⁴ Precio que pactan el remitente o destinatario de la carga con la empresa de transporte por concepto del contrato de transporte terrestre automotor de carga, <https://www.mintransporte.gov.co/descargar.php?idFile=2185>

PROBLEMA

PLANTEAMIENTO

Carga S.A.S. es una empresa de transporte de carga masiva, constituida en 1984, la cual cuenta con varios reconocimientos importantes como lo son: Certificación ISO 9001:2008, Certificación BASC (Business Alliance for Secure Commerce), primer empresa del sector servicios en el transporte en recibir la Mención a la excelencia BASC, premio a la transparencia otorgado por Bavaria S.A., entre otros. Está conformada por varias agencias a nivel nacional, cada una de las cuales opera de manera independiente de la Sede Principal, ubicada en el Municipio de Sabaneta - Antioquia.

En la actualidad Carga S.A.S. y las demás empresas del sector cuentan con grandes inconvenientes al momento de mantener la información disponible y actualizada (bases de datos, información resultante de la operación del negocio, informes, turnos, entre otros), debido a los costos elevados que implica tener un sistema en el cual la información se encuentre disponible en todas y cada una de las sedes y que esta sirva para mantener a flote el modelo de negocio.

Específicamente para Carga S.A.S., el no contar con información oportuna y adecuada, dificulta el prestar un servicio óptimo a los clientes, también genera retrasos en el proceso de facturación, y aumenta los riesgos asociados al transporte de carga como lo son el hurto, el

gemeleo⁵, y el fraude, y por consiguiente dificultad la selección de personal adecuado para el transporte de mercancía y retrasa el pago del servicio.

Esta situación se vería mejorada al ser implementado un sistema de replicación de datos que permitiera tener acceso a la información desde cualquier ubicación de forma eficiente y veraz (actualizada y permanentemente). Dicho sistema podría llegar a soportar diferentes motores de bases de datos, lo cual permite mantener la integridad de los datos para que la información sea utilizada de manera eficiente, generando valor para sus clientes, oportunidad de carga, eficiencia en el servicio y mejoramiento del modelo del negocio.

FORMULACIÓN

¿Cómo se lograría optimizar la disponibilidad de la información al interior de Carga S.A.S., a fin de mejorar sus procesos internos y la percepción de un buen servicio por parte del cliente?

SISTEMATIZACIÓN

¿Cómo garantizar la disponibilidad, la seguridad y la confiabilidad permanente de la plataforma de la tecnología informática con respecto a los equipos, aplicaciones y servicios relacionados con el sistema de replicación de datos?

⁵ Gemeleo: Palabra utilizada en el sector del transporte para hacer alusión a un vehículo que ha sido clonado. Concepto de dos vehículos con una misma placa.

¿Cómo medir la rentabilidad de la inversión en el mediano y largo plazo de un sistema de replicación de datos?

¿Por qué se da la creación de valor al cliente externo y al cliente interno con un sistema de información que entregue la información en el tiempo justo?

¿Cuál es la situación de la competencia con relación a los sistemas de información y sus métodos de disponibilidad de la información en cada una de sus sucursales?

¿Por qué asesorar a la alta gerencia y al departamento administrativo en la adquisición de sistemas de información que apoyen el modelo de replicación de datos?

JUSTIFICACIÓN

TEÓRICA

Actualmente, Internet ha cambiado la dinámica de los negocios y es muy común encontrar en estos tiempos, varios conceptos ligados fuertemente a la conectividad, como lo son la inteligencia colectiva, la web semántica, la web social, la web de las cosas, aplicaciones o servicios en la nube, entre otros, los cuales tienen como objetivo principal, poder tener un acercamiento y/o interacción con el usuario (cliente interno o externo de la organización) sin importar la ubicación geográfica del mismo.

El concepto de replicación de bases de datos no es nuevo, y más que un concepto, es un conjunto de técnicas orientadas a la arquitectura de software, la implementación de dichas técnicas, ofrecen una herramienta muy poderosa para aquellas organizaciones en las cuales, no se puede mantener una comunicación constante con la base de datos principal, o simplemente se necesita trabajar con réplicas para aumentar la eficiencia en la transaccionalidad de los datos.

Carga S.A.S., de acuerdo con sus políticas de calidad y servicio, se mantiene siempre a la vanguardia de la tecnología, y de esta manera puede ofrecer un servicio diferenciador y con valor agregado tanto al generador de carga como a los propietarios, conductores y demás personas que directa o indirectamente están relacionadas con la organización.

La implementación de un nuevo sistema de replicación de datos, permite a la organización obtener una ventaja competitiva muy importante, dado que prepara todo su entorno tecnológico, tanto a nivel de hardware como de software, para afrontar los retos que supone la nueva era de la informática, puesto que permite trabajar los nuevos desarrollos de software de manera sincrónica o asincrónica, adicionalmente, se obtiene una mejora significativa en cuanto a tiempos de entrega con respecto al envío y recepción de la información generada en las distintas sucursales, obtener un mayor control sobre qué oficinas se encuentran conectadas con el sincronizador, adicionar capas de seguridad en la transmisión, escalar fácilmente tanto de forma vertical como horizontal y contar con sitios de redundancia completa de datos para la implementación de un esquema de recuperación de desastres, minimizando los posibles riesgos que afecten la operación normal del modelo del negocio.

Para el desarrollo del sistema de replicación, se utilizan como base un conjunto de técnicas aplicadas en diferentes motores de bases de datos, como es el caso de Oracle MySQL, el cual se detalla en su manual de referencia, capítulo 6: Replicación en MySQL, igualmente documentado de forma amplia por Charles Dye, arquitecto de bases de datos y autor del libro Oracle Distributed Systems. Similares esquemas de replicación que se utilizaron como apoyo para este proyecto, es el utilizado por el motor de base de datos SQL Server, de la empresa Microsoft, y el cual se detalla de forma clara en su sitio web technet.microsoft.com, en el apartado de Replicación de SQL Server.

El sistema de colas de mensajería, utiliza como base un producto llamado ActiveMQ, el cual es un proyecto de código abierto, de la fundación de software Apache y que ha sido

implementado en diferentes soluciones que requieren arquitecturas complejas basadas en un sistema de mensajes.

METODOLÓGICA

A lo largo del desarrollo del proyecto se puede evidenciar el uso de un tipo de investigación aplicada, dado que se adaptan y utilizan conceptos y teorías ampliamente conocidas, para obtener como resultado final, un producto cuyo objetivo, es mejorar de manera sustancial el antiguo esquema de comunicaciones de la organización. Adicionalmente, implementando una adecuada arquitectura de software, se crea una potente capa de integración, la cual fortalece el corazón tecnológico de la empresa para afrontar de manera exitosa los cambios que se tienen planteados en el corto plazo.

No obstante, es necesario la utilización de una investigación de tipo básica, con el fin de conocer con un alto nivel de detalle, todas aquellas causas que rodean la problemática inicial, las cuales, son primordiales para marcar un punto de partida y de esta manera sustentar todas las decisiones que se tomen con respecto a la arquitectura o modelo a utilizar en un momento determinado.

PRÁCTICA

Para afrontar de manera exitosa el proyecto, se debe contar con una metodología apropiada, la cual permita establecer los lineamientos del trabajo, asignando roles y responsabilidades e identificando todos los actores del proceso en un momento determinado. Al

ser éste un proyecto fundamentado en la integración, se debe prestar especial atención a las etapas iniciales, las cuales se basan en requerimientos, diseño y arquitectura del software, adicionalmente, el esquema planteado como solución, debe tener la capacidad, de adaptarse exitosamente al entorno analizado y anticiparse a requerimientos futuros.

De acuerdo a lo anterior, se emplea como proceso de desarrollo de software, la metodología RUP (Rational Unified Process - Proceso Racional Unificado), permitiendo asegurar las diferentes etapas concernientes al mismo (el desarrollo de software).

Dentro del ciclo de vida de todo el proyecto, se inicia con el modelamiento del negocio, el cual permite obtener una visión más amplia todos los procesos de la organización, identificar sus actores y el área (o las áreas) de aplicación del producto resultante. El siguiente paso, es definir todos los requerimientos del sistema, tanto a nivel funcional como no funcional, estableciendo las posibles restricciones y delimitando su alcance, luego, se realiza el análisis y diseño del sistema, utilizando herramientas que cumplan las especificaciones de Lenguaje Unificado de Modelado o UML (Unified Modeling Language).

Una vez cubiertas las fases anteriores, se puede comenzar con la etapa de implementación, la cual incluye, entre otros, el uso de lenguajes de programación, plasmando todo el diseño lógico en un diseño a nivel físico, dando forma al producto final, el cual una vez culminado, se coloca en un ambiente de pruebas, evaluando la funcionalidad de todos los requerimientos, para por último, proceder a realizar la elaboración de los instaladores, manuales y puesta en producción de la solución.

OBJETIVOS

OBJETIVO GENERAL

Desarrollar un aplicativo al interior de la empresa CARGA S.A.S., como puente entre las estaciones remotas y el servidor principal de la organización, que permita un flujo constante de los datos con los cuales opera el negocio, apoyados de las herramientas ActiveMQ, C# .NET y SQL en las versiones Compact Edition y Server Express.

OBJETIVOS ESPECÍFICOS

- Identificar el estado actual de las condiciones tecnológicas, a nivel de equipos de cómputo y conectividad, de cada una de las agencias que se encuentran distribuidas a nivel nacional y que hacen parte de la empresa CARGA S.A.S.
- Establecer el proceso que va a intervenir en el sistema de replicación.
- Diagramar casos de uso que permitan evidenciar la funcionalidad del cliente con el desarrollo.
- Construir la aplicación utilizando las herramientas ActiveMQ, C# .NET y SQL en las versiones Compact Edition y Server Express.
- Realizar pruebas al sistema para establecer la pertinencia entre el desarrollo con los requisitos establecidos.
- Instalar la aplicación en un ambiente de producción.
- Capacitar y entregar documentación a la organización.

MARCO REFERENCIAL

MARCO TEÓRICO / CONCEPTUAL

Conceptos de la Organización

Carga S.A.S.

Constituida en 1984. Inicia su expansión a nivel nacional en 1991. En el 2002 recibe las certificaciones en la norma ISO 9001 versión 2000 y BASC (Business Alliance for Secure Commerce). En el 2005, el BASC Capitulo Antioquia le otorga el reconocimiento Mención a la excelencia BASC, siendo CARGA S.A.S., la primera empresa del sector servicios en recibirlo.

En el 2009, Bavaria S.A. y la Corporación Transparencia por Colombia, otorgan el reconocimiento a la gestión integral en la implantación del programa Rumbo Empresas Integras y Transparentes.

Pensamiento Organizacional

Los actos diarios de las personas que integran la organización estén siempre enmarcados en la honradez, la rectitud y la verdad.

El actuar diario de la organización esté enmarcado en el cumplimiento de los compromisos adquiridos, en forma oportuna.

La igualdad y el respeto a la dignidad de las personas, son independientes de la labor que desempeñan.

El éxito que se pueda obtener a través del tiempo, permita que conlleve a una organización dispuesta a aprender y reconocer los errores.

Generar un ambiente laboral de armonía y trabajo en equipo, la cual ayude a mejorar cada día el potencial humano y sea la base para el crecimiento personal y empresarial.

Reafirmar el compromiso con la sociedad y el país, respetando sus valores, cumpliendo con rectitud sus normas y acatando a cabalidad las autoridades legalmente constituidas.

Misión

Satisfacer con seriedad y responsabilidad las necesidades de los clientes, suministrando un servicio confiable.

Contribuir al desarrollo de los proveedores en general y los transportadores de carga en particular, a través de un trato digno, respetuoso y profesional, con unas condiciones de contratación acordes con su servicio y la realidad del mercado, siempre enmarcado en un pago oportuno.

Mantener a los accionistas un rendimiento satisfactorio, acorde con su inversión.

Visión

Ser la mejor opción en la prestación de servicios logísticos.

Replicación, Definición Básica

Existen varios métodos a emplear al momento de crear un sistema de replicación, el concepto no es nuevo, lo fundamental es entender qué es y cómo funcionan cada uno de los componentes que participan en este ecosistema.

Según la vigésima segunda edición del Diccionario de la Lengua Española de la RAE (Real Academia Española), replicar significa, (en una de sus varias definiciones): “*Repetir lo que se ha dicho*”. Tomando este concepto y haciendo una abstracción⁶ al mundo de los sistemas, replicar la información, es crear una copia de la misma en otro lugar.

La replicación cumple un papel muy importante en aquellas situaciones en las cuales, no es posible disponer de la información de forma centralizada, y según su implementación, se puede lograr ésta meta de manera económica, cumpliendo con el objetivo de contar con los datos actualizados y que éstos sean fácilmente accesibles para las estaciones remotas.

⁶ Crear una síntesis mental de las cualidades básicas de un objeto con el fin de comprenderlo en su esencia, <http://es.wiktionary.org/wiki/abstraer>

El Dato

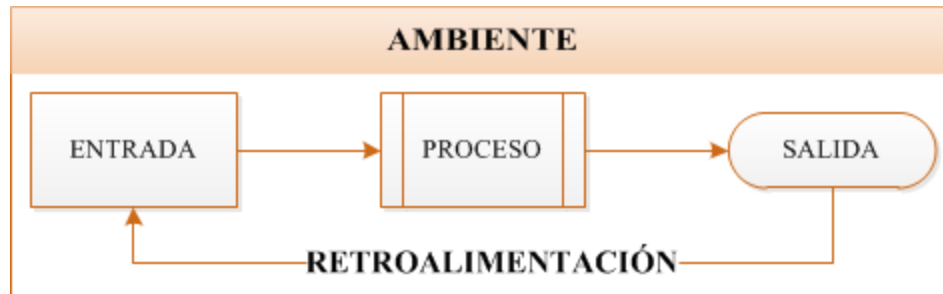
Los datos son, según la definición de Jeff Maynard, autor del libro Diccionario del Procesamiento de Datos: *“Un término general para denotar alguno o todos los hechos, letras, símbolos y números referidos a, o que describen un objeto, idea, situación, condición u otro factor”*, el conjunto de datos de manera organizada y comprensible se denomina información, siendo este uno de los activos más valiosos para las organizaciones, es aquí donde intervienen los sistemas de información, pero antes de definir lo que es un sistema de información, se debe entender que es un sistema en su conjunto. Existen muchos conceptos y definiciones de lo que es un sistema, como lo expresa Ludwig von Bertalanffy, filósofo y biólogo austríaco, en su libro Teoría General de los Sistemas: *“Si alguien se pusiera a analizar las nociones y muletillas de moda hoy por hoy, en la lista aparecería sistemas entre los primeros lugares. El concepto ha invadido todos los campos de la ciencia y penetrado en el pensamiento y el habla populares en los medios de comunicación de masas.”*. Bertalanffy define un sistema como: *“Un conjunto de unidades recíprocamente relacionadas”*, lo cual quiere decir, que todos sus elementos interactúan entre sí, y cuya finalidad es el logro de un objetivo.

Sistema

Los elementos más básicos que tiene un sistema son las entradas, los procesos y la salida, aunque también son igualmente importantes el ambiente y la retroalimentación, dado que del primero reciben las entradas y allí mismo retornan las salidas, y del segundo reciben la información necesaria para el continuo mejoramiento del sistema.

Ilustración 1 – Componentes de un Sistema

Fuente: Desarrollo de Sistemas de Información, Luis Castellanos, 2011



Sistema de Información

Teniendo en cuenta los conceptos anteriores, dato, información y sistema, se puede decir que un Sistema de Información, en informática⁷, es todo aquel, que utiliza la computación⁸, para el procesamiento de los datos, los cuales representan las entradas, y cuyo objetivo es satisfacer una necesidad de información, ésta puede ser almacenada, manipulada o transmitida. Dicho sistema, está compuesto por hardware, los cuales son todos aquellos componentes físicos, internos o externos, de una computadora, software (el componente lógico - programas -) y un recurso humano, también denominado humanware.

Éste tipo de sistemas de información, consta, como mínimo, de las siguientes etapas:

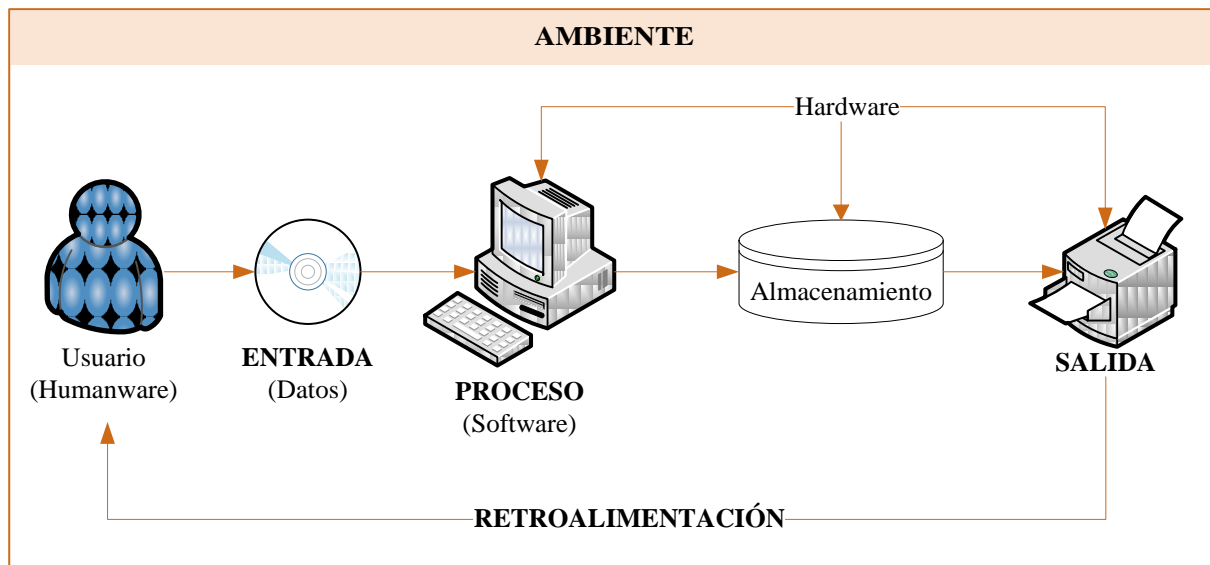
⁷ Conjunto de conocimientos científicos y técnicas que hacen posible el tratamiento automático de la información por medio de ordenadores, <http://lema.rae.es/drae/?val=informática>

⁸ Conjunto de conocimientos científicos y de técnicas que hacen posible el tratamiento automático de la información por medio de computadoras, <http://www.ecured.cu/index.php/Computación>

- Entrada, la cual consiste en ingresar al sistema los datos a procesar por medio de dispositivos tales como teclados, medios de almacenamiento extraíbles, (por ejemplo CD, DVD, USB), otros sistemas de información, entre otros.
- Proceso, paso en el cual los datos son convertidos en información propiamente dicha.
- Almacenamiento, permite la recuperación de la información para ser modificada o consultada.
- Salida, es el medio de comunicación del sistema de información con el exterior, el cual también incluye otros sistemas.

Ilustración 2 – Sistemas de Información (en la informática)

Fuente: Elaboración por parte de los autores del proyecto



Almacenamiento y Estructura de los Datos

En los sistemas de información, se mantienen las características propias de un sistema, (entradas, procesos y salida), y adicionalmente se reconoce uno nuevo, denominado almacenamiento, el cual se encarga de guardar los datos (generalmente en CD, DVD, Discos Duros, USB, entre otros) que pertenecen a un contexto en particular y de manera estructurada. Cuando se habla de datos estructurados, quiere decir que éstos poseen un tipo, el cual no es más que una particularidad que muestra cómo va a ser su procesamiento en los sistemas e indicando el rango de valores que puede contener.

Entre los distintos tipos de datos, se encuentran los de tipo Carácter, Entero, Flotante y Lógicos o Booleanos, denominados tipos de datos primitivos, ya que son comunes para la gran mayoría de lenguajes de programación.

Los tipos de datos carácter, generalmente, aceptan valores que van desde los números “0” al “9”, las letras de la “A” a la “Z” (mayúsculas y minúsculas) y caracteres especiales tales como “@”, “#”, “\$”, entre otros. La cantidad de caracteres almacenados para un dato en particular, como por ejemplo el nombre de una persona, está dado por su longitud, lo que quiere decir que para éste tipo de dato, se debe especificar un número, que generalmente va entre 0 (cero) y 255 ($2^8 - 1$), pero para los sistemas de información modernos, se han creado otros tipos de datos que permiten superar dicho límite.

Los tipos de datos enteros, aceptan números enteros (no poseen punto decimal), o números reales, (los cuales poseen una parte entera y otra parte decimal), y tanto los enteros como los reales pueden ser positivos o negativos.

El rango de valores, para los enteros pueden variar entre 0 y 4,294,967,295 (si sólo va a tomar valores positivos) ó entre -2,147,483,648 y 2,147,483,647 (si toma valores positivos y negativos). Lo anterior estaría dado por un entero de 32 bits (2^{32}), lo cual indica que puede tomar 2^{32} posibles valores, existen también enteros de 8, 16 y 64 bits, denominados generalmente TinyInt, SmallInt y BigInt respectivamente.

Los tipos de datos lógicos o booleanos almacenan sólo dos valores TRUE (verdadero) o FALSE (falso), lo cual correspondería, a nivel de bits, 1(uno) para TRUE (verdadero), ó 0 (cero) para FALSE (falso).

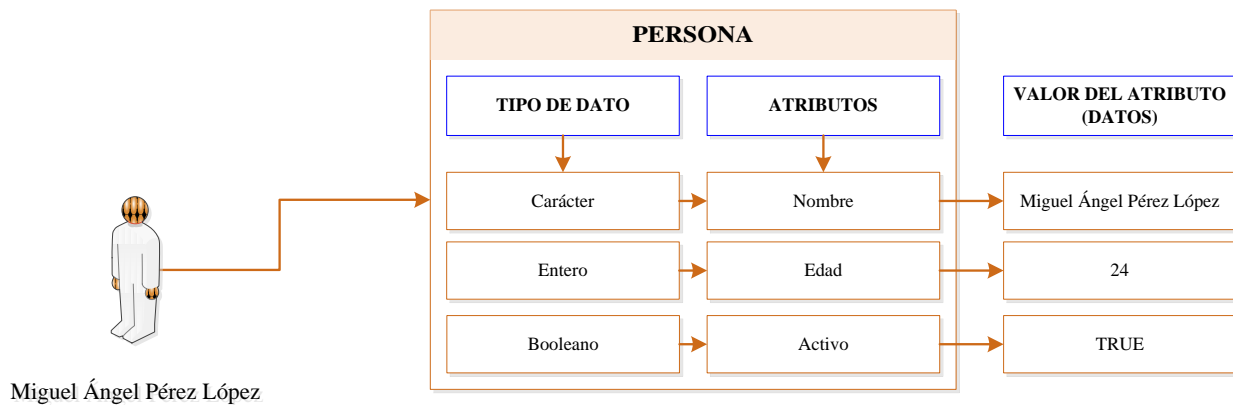
Entidades y Atributos

Dado que la información que se almacena, es una percepción del mundo real, es necesario crear una distinción entre cada agrupación de datos comunes, por ejemplo, se debe diferenciar entre una persona y un vehículo. A este tipo de elementos se les denomina Entidades, los cuales son objetos distinguibles de otros por el conjunto de sus atributos⁹, por otro lado, el valor de los mismos brinda la información necesaria para identificar luego dicho objeto en el mundo real.

⁹ Cualidad o característica propia de una persona o una cosa, especialmente algo que es parte esencial de su naturaleza, <http://es.thefreedictionary.com/atributo>

Ilustración 3 – Ejemplo de Entidad y Atributos

Fuente: Elaboración por parte de los autores del proyecto



Base de Datos Relacionales

Al momento de almacenar sistemáticamente cualquier cantidad de entidades con sus respectivos datos, se van construyendo uno o varios contenedores, denominados Bases de Datos, que no son más que colecciones de objetos que contienen información. Todas estas entidades se relacionan unas con otras para manipular e interpretar la información de manera adecuada, pero para disponer de todo este flujo de datos, se debe contar con un componente denominado Base de Datos Relacional, que es aquel que implementa el concepto Modelo Relacional, postulado en 1970 por Edgar Frank Codd, el cual se basa en la lógica de predicados¹⁰ y en la teoría de conjuntos, muy utilizado actualmente para la administración de los datos. En el modelo relacional, se tienen entidades totalmente estructuradas y relacionadas entre sí, para dar coherencia a la información, por algo son el método preferido para el almacenamiento en aplicaciones multiusuario, dispositivos móviles, agendas, entre otros, además, facilitan la labor tanto de los usuarios como de los desarrolladores. Las bases de datos sólo almacenan

¹⁰ Es aquella que distingue qué se afirma (predicado) y de quién se afirma (objeto).

información, pero para su administración, es necesario contar un Gestor (o administrador) de Bases de Datos o DBMS por sus siglas en inglés (Database Management System), siendo éste el encargado de realizar operaciones tales como inserciones, actualizaciones y borrados, entre otros, apoyándose en un lenguaje de consultas estructurado o SQL (Structured Query Language), siendo éste el más reconocido DML (Data Manipulation Language – Lenguaje de Manipulación de Datos).

SQL es también utilizado como DDL (Data Definition Language - Lenguaje de Definición de Datos), encargado de administrar las instrucciones para creación, modificación o eliminación de objetos (o entidades), los cuales ya en términos de Bases de Datos se denominan Tablas.

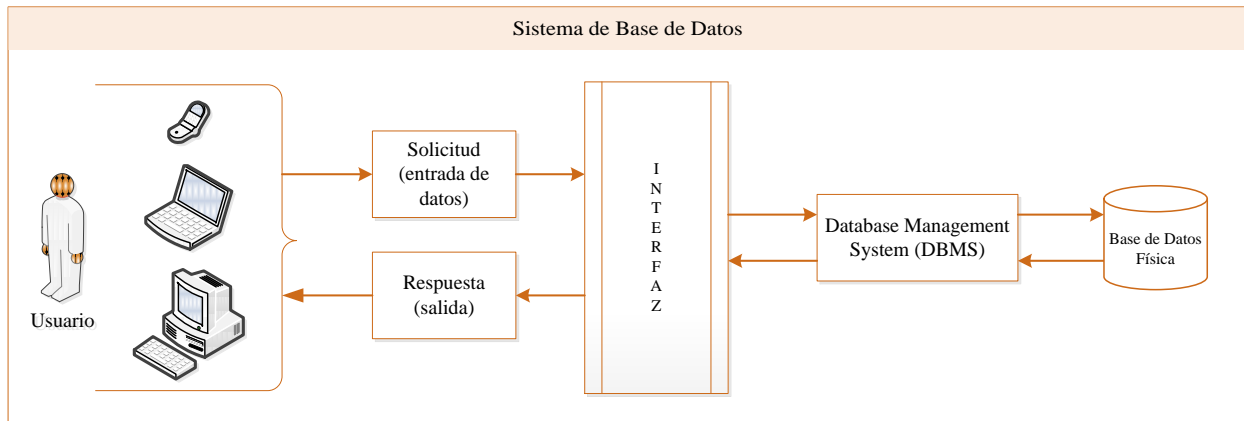
DBMS (Database Management System)

El DBMS (Database Management System – Sistema de Administración de Bases de Datos), utiliza también como herramientas el SDL (Store Definition Language – Lenguaje de Definición de Almacenamiento) para especificar el cómo se almacenara la estructura de la base de datos y el VDL (View Definition Language – Lenguaje de Definición de Vistas), el cual, según la definición del libro Database Systems, Concepts, Design and Applications (Sistemas de Bases de Datos, Conceptos, Diseño y Aplicaciones), explica: *“VDL es utilizado para especificar las vistas de los usuarios (esquema externo) y sus correspondencias con el esquema conceptual. Sin embargo, en la mayoría de los DBMS’s, se utiliza el DDL para especificar tanto el esquema conceptual como el esquema externo. Hay dos vistas de los datos, una, la vista lógica, la cual es la forma en que el programador percibe estar en. La otra es la vista física, la cual refleja la*

manera en como el dato está guardado actualmente en el disco (o cualquier otro dispositivo de almacenamiento)”

Ilustración 4 –Sistema de Base de Datos

Fuente: Database Systems, Concepts, Design and Applications, S. K. Singh, 2009



Sistema de Comunicación

Ahora bien, la manera como se logra que varios usuarios o equipos, actualicen o consulten las bases de datos, es por medio de un sistema de comunicaciones, el cual consta de un emisor, un medio de transmisión, por el cual transita la información, y un receptor. Dentro de los medios de transmisión, se pueden encontrar los Guiados y los No Guiados, el primero utiliza un cable que se encarga de conducir las señales de un extremo a otro, y el segundo aprovecha el uso de antenas (radio, microondas o de luz), tanto para la transmisión como para la recepción de las señales. Pero para que los sistemas sean capaces de entender las señales que reciben fue necesario

crear un estándar en las comunicaciones, y uno de los más difundidos y que dio paso a la que Internet alcanzara su masificación, es el protocolo TCP/IP.

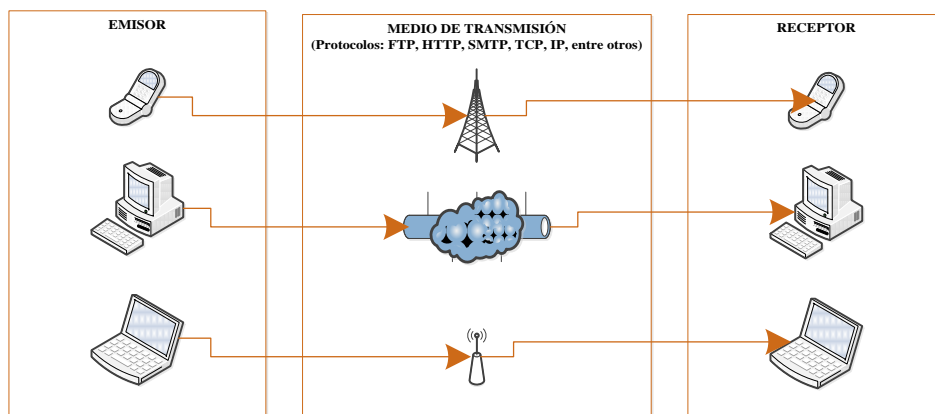
Protocolo TCP/IP (Transmission Control Protocol/Internet Protocol)

Douglas E. Comer, en su libro Redes Globales de Información con Internet TCP/IP, define protocolo como: *“Descripción formal de formatos de mensajes y reglas que dos o más máquinas deben seguir para intercambiar mensajes”* y adicionalmente presenta el siguiente concepto de lo que es TCP/IP: *“(Transmission Control Protocol) Protocolo de nivel de transporte TCP/IP estándar que proporciona el servicio de flujo confiable...permite que el proceso en una máquina envíe un flujo de datos hacia el proceso de otra...está orientado a la conexión en el sentido de que, antes de transmitir datos, los participantes deben establecer la conexión...”*.

TCP/IP, son realmente una colección de protocolos, pero debido a que TCP (Transmission Control Protocol – Protocolo de Control de Transmisión) e IP (Internet Protocol – Protocolo de Internet), son los dos protocolos más importantes del grupo, el conjunto de protocolos completo se denomina más frecuentemente como TCP/IP.

Ilustración 5 – Sistema de Comunicación Básico

Fuente: Elaboración por parte de los autores del proyecto



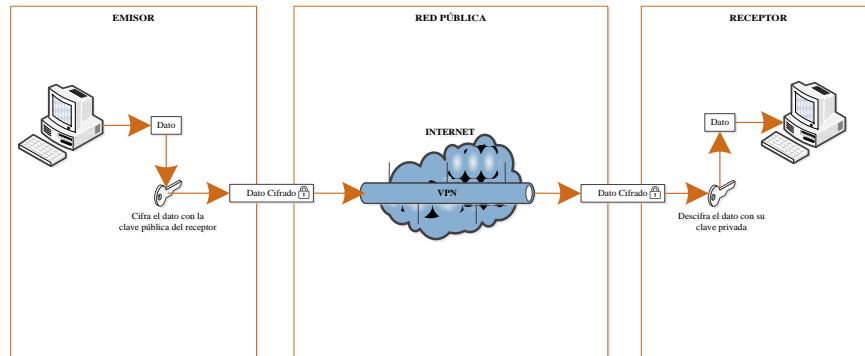
IPSec (Internet Protocol Security) y VPN (Virtual Private Network)

Se debe tener en cuenta que la información que viaja por cualquier medio de transmisión, puede ser interceptada, poniendo en peligro la integridad de la misma y su privacidad. Para evitar dicho inconveniente, se cuenta con un protocolo de seguridad IP, más conocido como IPSec (Internet Protocol Security), el cual trabaja en conjunto con las Redes Privadas Virtuales o VPN (Virtual Private Network), ésta, permite enviar los datos a través de Internet de manera cifrada (o codificada) y sólo aquellos equipos que se encuentran dentro de la red VPN, tienen la capacidad de descifrar dichos mensajes. El cifrado de datos, se genera por medio de la utilización de claves asimétricas, el cual consiste en utilizar un par de claves, (una pública y una privada), para el envío de mensajes. Cada usuario (o equipo) puede poseer la clave pública de cualquiera, por lo tanto, cuando se requiera transmitir un mensaje, simplemente se cifra con la clave pública correspondiente, y luego el destinatario lo descifra con su clave privada (la cual no es conocida

por nadie sino por él mismo). Gracias a la complejidad que presentan los sistemas criptográficos para la generación de claves, se garantiza que un par de claves puede crearse sólo una única vez.

Ilustración 6 – Transporte de datos con la utilización de claves asimétricas

Fuente: Elaboración por parte de los autores del proyecto



Datagrama IP

Para transportar los datos (encriptados o no) a lo largo de toda la red TCP/IP, se utilizan una serie de paquetes, denominados datagramas ip, que según la definición de Douglas E. Comer, son: “Unidad básica de información que pasa a través de una red de redes TCP/IP. Un datagrama IP es a una red de redes lo que un paquete de hardware es a una red física. Contiene las direcciones de fuente y destino junto así como los datos”. Actualmente se pueden encontrar las versiones 4 (IPv4) y 6 (IPv6). IPv4 presenta ciertos limitantes e inconvenientes, los cuales son solucionados con la implementación de IPv6, pero a pesar de ello, IPv4 aún sigue siendo muy utilizado en Internet y en soluciones de Redes de Áreas Locales (LAN). Los datagramas ip, en IPv4, poseen un tamaño o longitud, el cual, generalmente, para redes IP, se encuentra entre 576 bytes (como mínimo) y 1500 bytes (en las redes más comunes), dicha longitud se denomina MTU (Maximum Transport Unit – Unidad Máxima de Transporte). Si el dato a transmitir es

mayor al MTU permitido en la red, puesto que realmente es la red quien determina el tamaño del máximo MTU, éstos se dividen en varios datagramas (fragmentación), que luego se reensamblan en el destino, para realizar este proceso, los dispositivos de red cuentan con la capacidad de controlar dicha fragmentación.

El datagrama IP consta de los siguientes campos:

- Versión: Versión del Protocolo IP.
- Longitud del Encabezado (o cabecera): La cabecera IP debe contener como mínimo 20 bytes y máximo 60 bytes.
- Tipo de Servicio: Establece la prioridad del paquete en algunas redes.
- Longitud Total: Contiene el tamaño total del datagrama ip (cabecera + datos). Si existe fragmentación de datagramas, este campo contiene el tamaño del fragmento, no del tamaño total del datagrama original.
- Identificador: Único por cada datagrama, utilizado en caso de que haya fragmentación de los mismos.
- Flags: Se utiliza para saber cómo tratar un paquete, tales como, si es divisible/no divisible, si es el último fragmento o si existen más fragmentos.
- Posición del Fragmento: Indica qué posición ocupa el fragmento dentro de un datagrama fragmentado

- Tiempo de Vida (TTL): Indica por cuantos enrutadores¹¹ puede pasar un paquete antes de ser descartado. Su valor se va disminuyendo en uno (1), cuando éste llegue a cero (0), el paquete se descarta.
- Protocolo: Indica el protocolo a utilizar.
- Suma de Control de Cabecera: Si no corresponde con el contenido, el datagrama se desecha.
- Dirección IP¹² de Origen: Identifica quien envía el datagrama.
- Dirección IP de Destino: Identifica el destino del datagrama.
- Opciones: En ocasiones utilizado para pruebas o depuración.
- Relleno: Si se utiliza el campo opciones, se marca éste campo con ceros (0), hasta completar un total de 32 bits.
- Datos: Información transmitida.

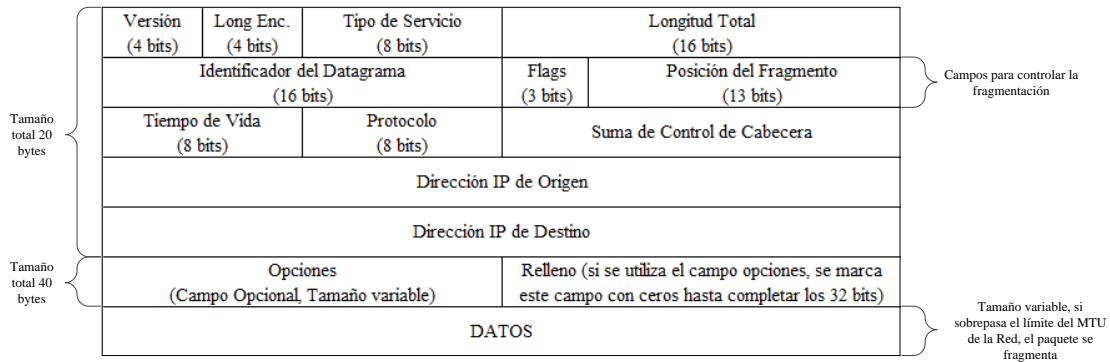
En teoría un datagrama no puede tener un tamaño mayor a 65,536 bytes, pero como se mencionó anteriormente, dicho tamaño es establecido por el MTU (Maximum Transport Unit – Unidad Máxima de Transporte), aunque lo importante, es que todos los puntos de la red que necesiten interactuar con el datagrama, estén en condiciones de manejar el tamaño del objeto enviado, de no ser así, simplemente la comunicación entre ellos no sería posible.

¹¹ Dispositivo utilizado para el envío de paquetes entre las redes.

¹² Dirección de 32 bits asignada a cada anfitrión que participa en una red de redes TCP/IP. Una dirección IP es una abstracción de la dirección de hardware físico. Para hacer el ruteo eficiente, cada dirección IP se divide en parte en red y en parte en anfitrión. Douglas E. Comer, Redes Globales de Información con Internet y TCP/IP.

Ilustración 7 – Datagrama IP y campo de Datos

Fuente: Elaboración por parte de los autores del proyecto



Modelo TCP/IP y Modelo OSI

Para comprender mejor como se transportan los datos de un lugar a otro por medio de las comunicaciones en red, es necesario conocer los modelos que han servido como referencia.

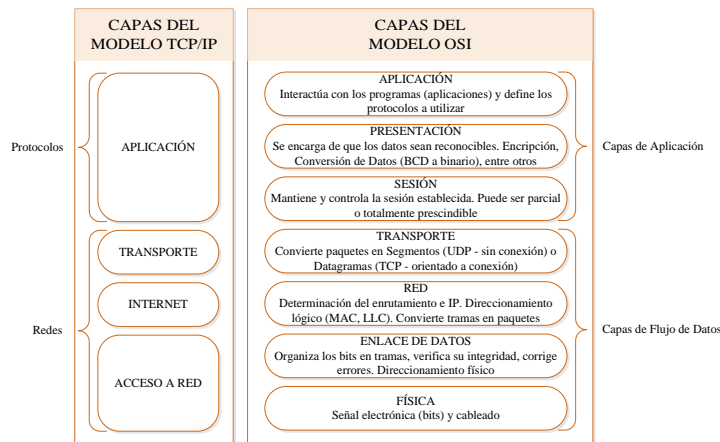
Tanto el modelo TCP/IP (Transmission Control Protocol/Internet Protocol) como el modelo OSI (Open Systems Interconnection), se utilizan como base para definir la arquitectura que deben seguir las comunicaciones. El modelo TCP/IP, involucra sólo un protocolo de red, el IP (Internet Protocol), por otro lado, el modelo OSI fue creado como un estándar mundial por la ISO (International Organization for Standardization - Organización Internacional para la Estandarización), en 1984, motivo por el cual, es más genérico (independiente del protocolo).

Aunque existen diferencias entre un modelo y otro, y cada uno presenta ventajas en un momento determinado, pero generalmente, se suele hacer más referencia al comportamiento de las comunicaciones en red con el modelo OSI.

Para los desarrolladores de software, el modelo TCP/IP, es más flexible en su capa de aplicación, con respecto al modelo OSI, puesto que agrupa las capas de sesión presentación y aplicación en una sola. Utiliza únicamente dos protocolos en la capa de transporte, UDP (User Datagram Protocol – Protocolo de Datagrama de Usuario), que no es orientado a conexión y TCP (Transmission Control Protocol – Protocolo de Control de Transmisión) que es orientado a conexión, adicionalmente, involucra sólo un protocolo en la capa de red, el protocolo IP (Internet Protocol).

Ilustración 8 – Modelo TCP/IP vs. Modelo OSI

Fuente: <http://www.zak.com.mx/tuto/tcpip.pdf>,



XML (eXtensible Markup Language)

Cómo se mencionó anteriormente, replicar es crear una copia de la información, en uno, o varios lugares, por lo tanto, así como existen protocolos de comunicación para los sistemas en la capa de aplicación del modelo OSI (Open Systems Interconnection), como por ejemplo SMTP (Simple Mail Transfer Protocol – Protocolo Simple para Transferencia de Correo), FTP (File

Transfer Protocol – Protocolo de Transferencia de Archivos), entre otros, es conveniente que exista también un “formato”, el cual permita a los equipos remotos interpretar, cómo deben procesar la información que constantemente está fluyendo en el sistema de replicación de datos. Cada sistema de replicación define su propio esquema de mensajes para realizar dicha tarea.

Dado que el método de replicación a utilizar, se basa en una comunicación constante por medio de mensajes, es necesario utilizar un lenguaje que permita flexibilidad y escalabilidad, características que encajan perfectamente con XML (eXtensible Markup Language – Lenguaje de Marcas Extensible), dicho lenguaje, permite definir el formato del mensaje, estructurar documentos de gran tamaño y es totalmente abierto para integración con otros sistemas o aplicaciones, dado que permite el intercambio de información, adicionalmente es una tecnología relativamente sencilla, simple y eficaz.

Un documento XML, se puede clasificar en:

- Bien Formados (Well Formed): Lo cual quiere decir que cumple reglas básicas de sintaxis. Dichas reglas se pueden resumir en lo siguiente:
 - El documento inicia con una declaración, la cual proporciona como máximo tres tipos de información como son: la versión (obligatorio), la codificación de caracteres (opcional), que es la que indica cómo se deben interpretar internamente y si es, también opcional, un documento independiente (standalone)
 - Debe tener sólo un elemento raíz.
 - Debe estar correctamente escrito (sintácticamente hablando), es decir, los elementos deben estar correctamente delimitados, los valores de los atributos

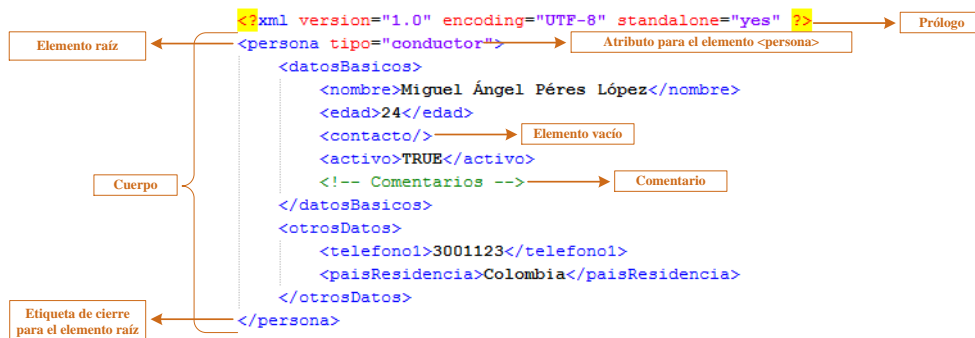
deben ir entre comillas (simples o dobles), se deben cerrar correctamente los elementos, los nombres de las etiquetas deben empezar con una letra, entre otros.

- Válidos: Deben estar bien formados y cumplir con las especificaciones de un documento adicional, el cual sirve de validador, como lo son los DTD (Document Type Definition), Schemas, o cualquier otro

Para el caso en particular de envío de mensajes por medio de un sistema de replicación, es más recomendable utilizar XML Bien Formados simplemente, dado que no hay que realizar una tarea de validación contra un elemento validador, evitando de esta manera, un procesamiento adicional, que para el caso, puede resultar innecesario e impactar directamente en el rendimiento del sistema de replicación.

Ilustración 9 – Ejemplo XML Bien Formado (Well Formed)

Fuente: Elaboración por parte de los autores del proyecto



Colas FIFO

Es muy importante, al momento de pensar en replicación de datos, controlar el orden en que los mensajes serán entregados a todos los equipos cliente, para este tipo de situaciones, se puede utilizar una estructura de datos¹³ básica denominada colas FIFO (First In First Out – Primero en Entrar Primero en Salir), la idea básica de la misma, es que cada uno de los datos que allí se encolan, se entreguen en el mismo orden en que fueron despachados.

De acuerdo al libro de la Universidad Nacional del Litoral de Argentina, titulado Algoritmos y Estructuras de Datos, una cola se define como: *“Un objeto muchas veces usado como buffer o pulmón, es decir un contenedor donde almacenar una serie de objetos que deben ser procesados, manteniendo el orden en el que ingresaron”*

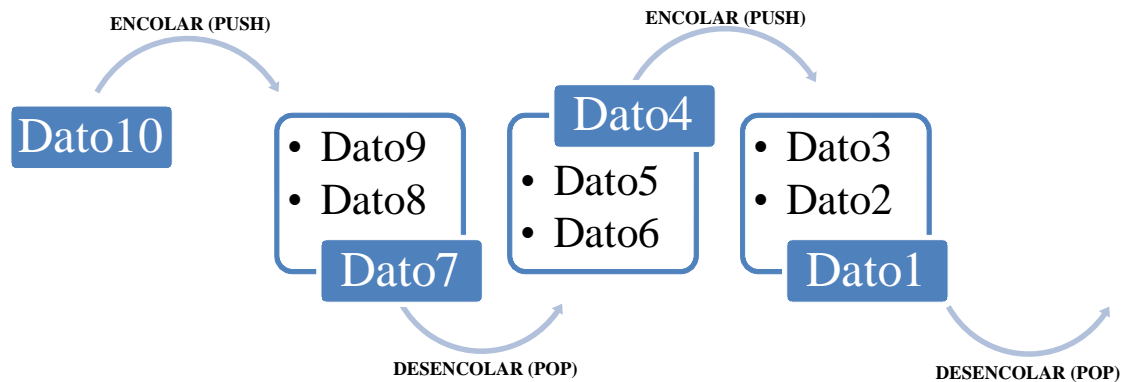
Dentro de las operaciones básicas, (aparte de la creación - create -), que se pueden realizar con una cola son: encolar (push), desencolar (pop) y consultar.

Cabe anotar, que pese a que el funcionamiento a nivel general del sistema de replicación de datos se base en colas FIFO, los mensajes XML pueden tener un atributo que indique la prioridad del mismo, similar a como se implementan las prioridades en las cabeceras de un datagrama IP en el campo Tipo de Servicio (TOS - Type of Service) , lo cual quiere decir que al llegar éste mensaje a su destino, el equipo cliente, será éste quien determine si se procesará inmediatamente o si se procesa de manera cronológica (de acuerdo al orden de llegada), tal cual es su funcionamiento predeterminado.

¹³ Recurso para organizar un grupo de datos el cual busca facilitar el manejo de los mismos.

Ilustración 10 – Cola FIFO

Fuente: Elaboración por parte de los autores del proyecto



Replicación, un concepto más profundo

Se han esbozado una serie de elementos, que de una u otra forma tienen relevancia con la replicación de datos, por lo cual, se debe tener en cuenta, que la replicación, va mucho más del concepto en sí, es una técnica, que apoyada de varias herramientas tecnológicas, logra su objetivo.

Se puede definir la replicación de datos como el proceso de copiar y/o distribuir todas o algunas de las entidades o elementos que constituyan información, pertenecientes a uno o varios contenedores de datos, con el fin de mantenerla dispuesta y actualizada en los diferentes puntos pertenecientes a un sistema de bases de datos distribuida, para su posterior manipulación y/o consulta, o simplemente con fines de respaldo y/o contingencia ante posibles fallos en alguno de los sistemas de información.

Por otro lado, Microsoft, para su motor base de datos SQL Server, define replicación como *“Un conjunto de tecnologías destinadas a la copia y distribución de datos y objetos de base de datos desde una base de datos a otra, para luego sincronizar ambas bases de datos y mantener su coherencia. La replicación permite distribuir datos entre diferentes ubicaciones y entre usuarios remotos o móviles mediante redes locales y de área extensa, conexiones de acceso telefónico, conexiones inalámbricas e Internet”*.

Existen varios tipos de replicación, los cuales se detallan a continuación:

- **Replicación Transaccional:** Consiste en definir un contenedor de datos principal (también denominado Master o Publicador), quien envía, por medio de un agente, las inserciones, borrados o modificaciones de las tablas pertenecientes a la base de datos a cada uno de los equipos clientes (denominados Suscriptores). En éste esquema, los datos entregados a los suscriptores, no pueden ser modificados en ninguno de los nodos (o equipos clientes), sólo están disponibles para efectos de consultas. Las modificaciones, deben hacerse directamente en el Publicador. En algunos sistemas, este tipo de replicación permite la modificación en los suscriptores, pero sus modificaciones no se transmiten hacia el Publicador.
- **Replicación en Mezcla:** Permite el intercambio de movimientos entre el Publicador y el Suscriptor en el momento en que ambos estén accesibles por medio de la red. En este esquema los cambios realizados en el Publicador, viajan hacia el Suscriptor y viceversa. Se debe tener un sistema de Resolución de Conflictos para evitar la incoherencia entre los datos.

- **Replicación de Instantáneas:** Básicamente consta de la transmisión (programada), por parte del Publicador, de una copia completa de la base de datos principal hacía los Suscriptores. Al igual que en la replicación básica (o transaccional), tampoco existe replicación por parte de los Suscriptores hacía el Publicador, la diferencia entre estas radica, en que para la replicación transaccional, se envía cada modificación que se realice, bien sea a un campo de una tabla, o al registro completo (o fila) de la misma, mientras que en la replicación de instantáneas se transmiten tablas completas o incluso toda la base de datos.

Cabe anotar que los tipos de replicación anteriormente mencionados, son los más comunes, pero existen sistemas más avanzados, los cuales poseen la capacidad de combinar las tres técnicas de replicación, aportando más potencia al modelo de replicación.

Algo común a todos los sistemas de replicación de datos es lo siguiente:

- Siempre debe existir una copia inicial total de la base de datos (o instantánea), y a partir de allí se comienzan a generar los datos a replicar.
- La replicación se realiza de manera programada (replicación por lotes o batch) o con la ayuda de disparadores¹⁴ (o triggers), los cuales son una funcionalidad propia de la gran mayoría de los motores de bases de datos.
- La replicación, en la gran mayoría de los casos, es asincrónica¹⁵.

¹⁴ Procedimiento que se ejecuta al cumplirse una condición particular en una base de datos. Normalmente se programan para que se ejecuten en el momento de realizar operaciones de inserción, modificación o borrado.

¹⁵ Es aquella comunicación que se establece entre dos o más personas de manera diferida en el tiempo, es decir, cuando no existe coincidencia temporal. Actualmente es un tipo de la comunicación desarrollada mediante

El replicar la información puede traer consigo una serie de inconvenientes, los cuales deben ser tratados por medio de un mecanismo de resolución de conflictos, el cual debe tener la capacidad de solucionar o eliminar las incoherencias presentadas al momento de actualizar la información. Uno de los conflictos más comunes son los de violación de clave primaria, lo cual quiere decir que se está intentando ingresar un registro que actualmente existe en la base de datos de destino.

Ilustración 11 – Replicación Transaccional

Fuente: Elaboración por parte de los autores del proyecto

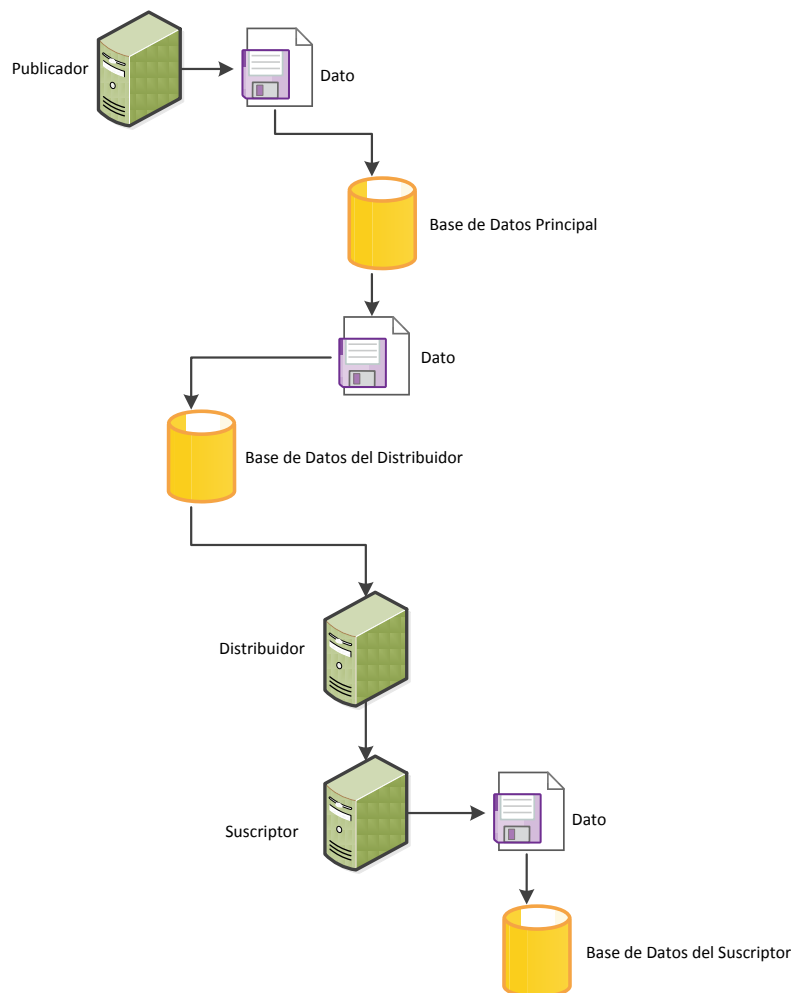


Ilustración 12 – Replicación en Mezcla

Fuente: Elaboración por parte de los autores del proyecto

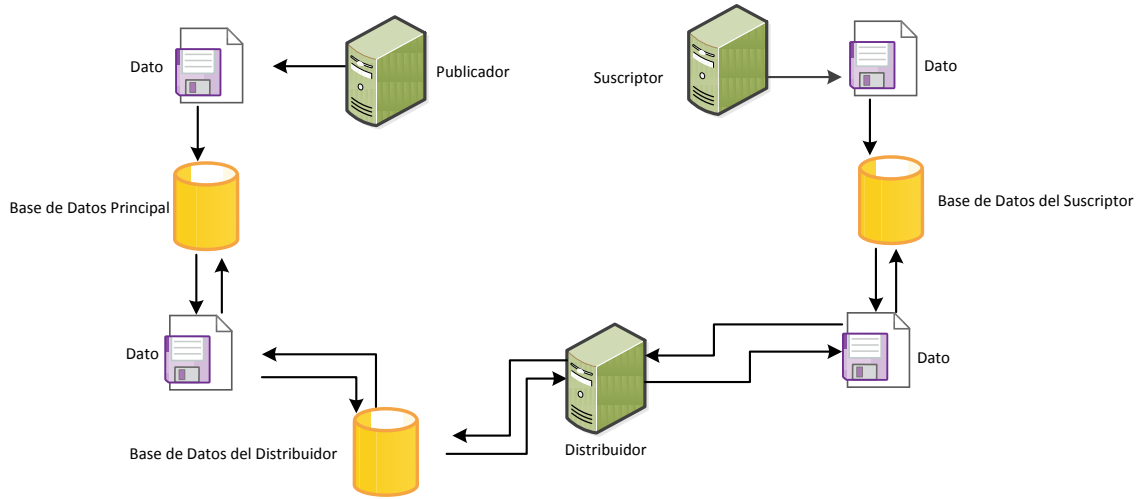
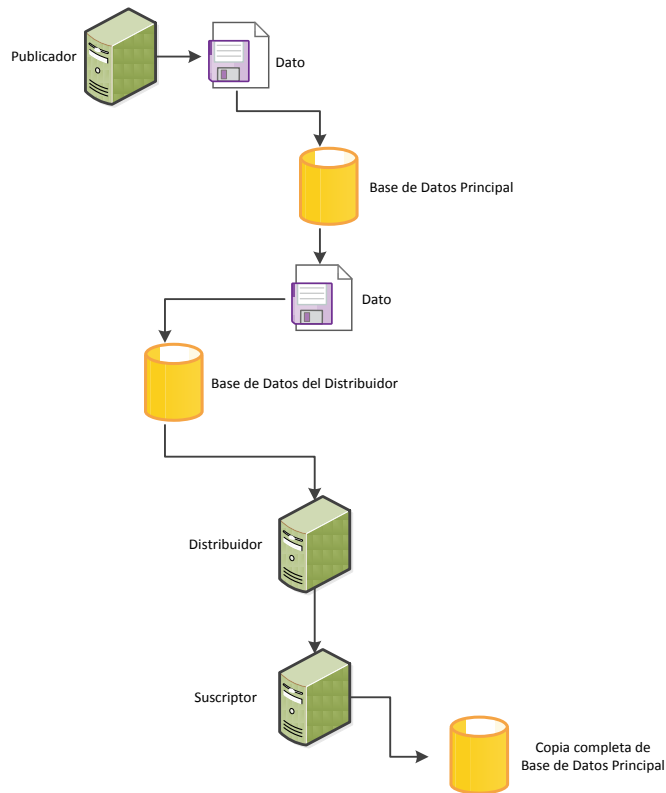


Ilustración 13 – Replicación de Instantáneas

Fuente: Elaboración por parte de los autores del proyecto



MARCO HISTÓRICO

Mucho tiempo ha transcurrido, desde que las computadoras de primera generación, aproximadamente entre 1951 y 1958, eran simples máquinas, cuya función consistía en realizar operaciones matemáticas básicas. Para realizar dicho proceso, se contaba con tarjetas perforadas, las cuales contenían los datos a procesar y su respectivo conjunto de instrucciones (programa), estos, se cargaban en un tambor magnético, el cual era considerado la memoria principal de aquella época y es el precursor de los discos duros de platos, que aún, al día de hoy son muy utilizados, luego, se grababa el resultado en cintas de papel o tarjetas perforadas.

El inconveniente del tambor magnético, era que no podía ser removido de la computadora que lo albergaba, por lo tanto, el traslado de la información resultante se realizaba por medio de tarjetas perforadas, hechas en lámina de cartulina, o similares, como las cintas de papel, y debido a lo delicado de las mismas, resultaba un tanto riesgoso su manipulación constante.

La primera generación de computadores, da inicio a la creación e implementación de los medios magnéticos de almacenamiento, a medida que se va madurando la tecnología de los computadores, se va reemplazando paulatinamente el uso de las tarjetas perforadas, y aparecen como opciones alternativas, la cinta magnética y los discos magnéticos.

Los nuevos medios magnéticos, brindaban capacidades de almacenamiento cada vez mayores, y al mismo tiempo surgían nuevos requerimientos, prestando especial atención en la implementación de políticas de respaldo de la información.

La aparición de los medios magnéticos permitió la sistematización de los datos, siendo el EDVAC (Electronic Discrete Variable Automatic Computer) en 1949, el primer computador en utilizar una cinta magnética para almacenar información, luego en 1955, el UNIVAC (Universal Automatic Computer), almacena la información proveniente de tarjetas perforadas en una cinta magnética, la desventaja que presentaban las cintas magnéticas, radicaba en la forma como se leían los datos para su recuperación, puesto que debía hacerse de manera secuencial, solución que vino de la mano con los discos magnéticos, los cuales acceden a la información de manera aleatoria, es decir, no es necesario recorrer todo el disco para recuperar un dato.

En la década de los 60's, comienzan las primeras generaciones de bases de datos y en 1970, Edgar Frank Codd publica un trabajo llamado A Relational Model of Data for Large Shared Data Banks (Un modelo relacional de datos para grandes bancos de datos compartidos), el cual sirvió como base para el desarrollo de un sistema de gestión de bases de datos relacionales llamado Oracle.

Se plantea el interrogante de si es posible compartir información entre dos equipos ubicados en sitios geográficamente diferentes, el cual, da paso al nacimiento la primera red de comunicación WAN (Wide Area Network – Red de Área Amplia), cuya solución se basa, en transmitir dicha información por medio de líneas telefónicas de baja velocidad (aproximadamente 1200 bps para aquel entonces), siendo Lawrence G. Roberts el precursor de este nuevo modelo de comunicación, con el cual se diseñó, en el año de 1965, la primer red de computadores, compuesta por un TX-2, construido en el laboratorio Lincoln del MIT (Massachusetts Institute Technology) y un IBM AN/FSQ-32.

A partir de ese momento comienzan una serie de desarrollos enfocados a la interconexión por medio de la conmutación de paquetes, y aparece un primer acercamiento a lo que sentaría las bases de la red de redes, Internet.

Sería entonces en 1969 cuando se conectarían cuatro computadores, a una velocidad de 50 Kbps, entre la UCLA (University of California, Los Angeles), UCSB (University of California, Santa Barbara), SRI (Stanford Research Institute) y U of U (University of Utah), dicha red se denominó ARPANET (Advanced Research Projects Agency Network – Red de la Agencia de Investigación de Proyectos Avanzados), propiedad del Departamento de Defensa de los Estados Unidos.

En 1972 se envía el primer correo electrónico, en 1989 se crea la World Wide Web (WWW), con alrededor de 100,000 equipos conectados entre varios países y ya en 1991 la Web es presentada como un servicio público en Internet, la cual trabaja con un conjunto de protocolos denominados TCP/IP (Transmission Control Protocol/Internet Protocol), compuestos por un subconjunto mucho más grande, dentro de los cuales se encuentran los protocolos necesarios para el acceso a las páginas web, envío de correos, transferencia de archivos...entre otros.

El nuevo protocolo, TCP/IP, se declaró como estándar para las comunicaciones militares en 1982, reemplazando al protocolo NCP (Network Control Program) utilizado por ARPANET. La gran aceptación y adaptación de TCP/IP por parte de los fabricantes y su alto grado de fiabilidad, hizo posible la Internet de nuestros tiempos.

Los sistemas operativos comenzaron a implementar el nuevo protocolo, y empieza el desplazamiento del sistema operativo Novell Netware, desarrollado por Novell, el cual utilizaba IPX/SPX como protocolo de red, dominante en el mercado durante los años 90.

BSD (Berkeley Software Distribution) en su versión 4.1a, ya había hecho un primer acercamiento en 1981, y mejorándola en la versión 4.3 en 1986, Linux por su parte, debió escribir el código para TCP de manera independiente debido a una incompatibilidad, (hasta antes de 1999), entre las licencias GPL (usada por Linux) y la Licencia BSD , Microsoft en 1990 incluyó TCP/IP, junto con NetBEUI (NetBIOS Extended User Interface), en su versión Windows NT, Apple Computer con su Mac OS X del año 2001, adopta el protocolo, basándose en una implementación de BSD 4.4, aunque ya existían protocolos diseñados por Apple, y compatibles (en su momento), como por ejemplo MacTCP, reemplazada luego por Open Transport y esta a su vez reemplazada por sockets Berkley. En este punto los sistemas operativos ya contaban con todo lo necesario, tanto a nivel de hardware y software, para poder comunicarse por medio de una LAN (Local Area Network – Red de Área Local) o una WAN (Wide Area Network – Red de Área Amplia).

La posibilidad de poder contar con la información en línea, abre una gran oportunidad de expansión para todas aquellas empresas que estuviesen en condiciones de afrontar el costo de dicha inversión. La pregunta ahora, es el tipo de modelo a utilizar con respecto a la forma de acceder y mantener actualizada la información, dentro de los cuales, el más común es una arquitectura Cliente/Servidor, proveniente de los sistemas distribuidos (en este caso, a nivel de aplicaciones), cuya primer definición se dio a conocer aproximadamente en 1978, en el cual, existe un proveedor de recursos y/o servicios (servidor) y un solicitante de los mismos (cliente).

El reto de compartir la información, remotamente, trae consigo varias situaciones por resolver, las cuales son inherentes al modelo de un sistema distribuido, como lo son la disponibilidad de la información, la integridad de los datos, una administración total del sistema mucho más compleja, retrasos en la entrega de la información, más puntos de fallas en la arquitectura, mayores requerimientos de hardware en cuanto a procesamiento, costos de implementación...entre otros. Uno de los inconvenientes más críticos en el modelo Cliente/Servidor, es mantener la alta disponibilidad del servicio, dado que si estos no se encuentran disponibles, sencillamente las estaciones remotas no tendrían manera de operar.

Para poder recuperar la información, ésta debe estar almacenada en un contenedor lógico. Una de las técnicas utilizadas para sopesar la indisponibilidad del servicio en sedes remotas, es contar con un esquema de replicación de datos, ubicando en este punto un Sistema de Bases de Datos Distribuidas (SBDD), para la cual, los nuevos estándares de comunicación por redes, favoreció mucho dicho aspecto, apoyándose además en un lenguaje de consulta estructurado, o SQL (Structured Query Language) para afrontar el reto de obtener un proceso de replicación confiable.

SQL es un lenguaje que surge prácticamente a la par con los conceptos de Bases de Datos relacionales, el cual es estandarizado en el año de 1986 y ampliado en 1992. Actualmente la gran mayoría de los Sistemas de Gestión de Bases de Datos incluyen un estándar de SQL. Los motores de bases de datos más conocidos han implementado de forma nativa la replicación de sus bases de datos, como por ejemplo MySQL, motor de base de datos, desarrollado por Sun Microsystems, Sql Server, desarrollado por Microsoft, PostgreSQL, desarrollado por PostgreSQL Global Development Group, Oracle, desarrollado por Oracle Corporation...entre otros.

Todo lo anterior muestra el recorrido que ha tenido la estructuración y tránsito de los datos por los diferentes sistemas.

MARCO LEGAL

El sistema de replicación de datos, conlleva una copia de los mismos en equipos de cómputo diferentes, para luego ser consultados, modificados o simplemente para tener un respaldo de la información.

Dichos equipos de cómputo, pueden estar distribuidos a lo largo del territorio Nacional Colombiano, motivo por el cual las personas que se encuentran registradas en las bases de datos, deben ser informadas de dicha situación, dando cumplimiento a la Ley Estatutaria 1581 de Octubre 17 de 2012, la cual se refiere a la disposición general para la protección de los datos personales, decretando, en su Título I, su objeto, ámbito de aplicación y definiciones, adicionalmente el Título II, Artículo 4, numeral B, expone el Principio de Finalidad, en su numeral C, el Principio de libertad y en los numerales G y H Principio de Seguridad y Confidencialidad respectivamente. El incumplimiento de la Ley Estatutaria 1581 trae consigo sanciones hasta por dos mil salarios mínimos mensuales legales vigentes, suspensión de las actividades, cierres temporales o definitivos de las operaciones relacionadas con el tratamiento.

Se debe tener en cuenta que el titular de la información, puede ejercer los derechos establecidos en el Título II, de la Ley Estatutaria 1266 de Diciembre 31 de 2008, más conocida como Ley de Habeas Data, y los operadores, fuentes y usuarios de información deben garantizar lo expuesto en el Título III de la misma Ley.

La Ley 1273 de Enero 5 de 2009, modifica el Código Penal y crea un nuevo bien jurídico tutelado, el cual denomina “Protección de la Información y de los Datos”, en el cual se debe

prestar especial cuidado al Capítulo III, Artículo 29F, que trata sobre la violación de los Datos Personales.

Teniendo en cuenta que la replicación de los datos, utiliza como base para su funcionamiento la transmisión por medio de redes, y en la mayoría de los casos dichas redes son diferentes de la LAN (Local Area Network – Red de Área Local), se debe verificar primero si el destino de alguno de los equipos que van a almacenar la información, está ubicado por fuera del territorio Colombiano, de ser así, es necesario consultar las Leyes o Normas que aplican en el país destino, pues son estas las que entrarían a regir en un momento determinado.

DISEÑO METODOLÓGICO

REFERENTE METODOLÓGICO

Tipo de Estudio

El tipo de estudio a aplicar, se puede catalogar como exploratorio, dado que antes de abordar de manera directa la situación evidenciada en el antiguo esquema de sincronización de datos de la empresa Carga S.A.S., se realiza un análisis a nivel general de la problemática, para luego poder obtener partes mucho más pequeñas, denominadas sub-problemas, las cuales permiten identificar los puntos más críticos, éstos, se priorizan, y se presta especial atención a aquellos cuyo nivel de prioridad sea más alto. El conjunto de sub-problemas, también permiten que se alcance un alto grado de familiaridad con el problema inicial, aportando claridad al momento de formular, más adelante, todas las posibles soluciones, tanto a nivel de diseño como de desarrollo.

Adicionalmente, con el conocimiento adquirido por el análisis de la problemática, se utiliza un tipo de estudio descriptivo, empleando técnicas puntuales para recolectar información que permiten identificar aquellas variables que ayudan a exponer de forma precisa el problema tal y como es, para luego determinar el modelo arquitectónico que mejor se adapte a la solución del mismo.

Método de Investigación

Como se comentaba anteriormente, uno de los tipos de estudio permite dividir el problema en partes mucho más pequeñas, brindando mayor claridad del mismo, igualmente ocurre con el método de investigación a aplicar, en este caso, el analítico, que por definición permite descomponer una situación global y observar todas aquellas causas que lo provocan, su naturaleza y su efecto, por lo tanto, al comprender la esencia misma del problema como su comportamiento, se pueden proponer soluciones puntuales y acordes a la situación. También se utiliza un modelo deductivo, dado que se adoptan conceptos probados y validados, lo cual los convierte en normas generales (desde el punto de vista tecnológico), extrayendo conclusiones que permiten entregar resultados enfocados siempre en encontrar un producto resultante que cumpla con las expectativas de la organización.

Fuentes de Información

En búsqueda de plantear siempre las mejores prácticas, metodologías y diseños arquitectónicos más adecuados, se recurren a dos tipos de fuentes de información, las cuales se clasifican en primarias y secundarias.

Dentro de las fuentes de información primaria se destacan las entrevistas con el personal que conforman el Departamento de T.I. (Tecnologías de la Información) de Carga S.A.S., entrevista con Ingenieros de la empresa Sistemas I.A., desarrollador del actual sistema de comunicaciones en producción en la empresa Carga S.A.S., asesoría en la implementación de

listeners y dispatchers en lenguaje C# por empresas con experiencia certificada, manuales de referencia del sistema de colas de mensajería ActiveMQ de la fundación se software Apache.

Dentro de las fuentes de información secundaria, se encuentran textos con información concerniente al desarrollo de colas de mensajería, conceptos de bróker, formato xml, arquitecturas basadas en servicios, implementaciones de sistemas integrados con un bus de datos, referencias bibliográficas, artículos y revistas especializadas.

REFERENTE PRÁCTICO

RUP (Rational Unified Process - Proceso Racional Unificado)

El producto final tiene como objetivo implementar un sistema de replicación de datos basado en una arquitectura orientada a servicios (SOA - Service Oriented Architecture). Para poder consumir y aprovechar dichas funcionalidades, es necesario desarrollar (a nivel de programación) varios componentes, motivo por el cual se debe cumplir con una serie de buenas prácticas, técnicas y demás, que permitan asegurar la calidad dentro de todo el ciclo de vida de desarrollo del software. Para tal efecto, la metodología a seguir es RUP (Rational Unified Process - Proceso Racional Unificado), el cual es ampliamente conocido y utilizado por grandes compañías de desarrollo.

El ciclo de vida del desarrollo, dentro de la metodología RUP, se compone de cuatro fases, al final de las mismas, se realiza una revisión, la cual permite validar el cumplimiento de los objetivos trazados.

A continuación se detallan cada una de las fases de la metodología:

Inicio

Permite obtener una visión de las capacidades del producto resultante, describe el modelo del negocio, define los objetivos del proyecto, su alcance, identifica los casos de uso y los posibles riesgos asociados.

Elaboración

Define una arquitectura a nivel básico, basándose tanto en los casos funcionales como no funcionales, el producto resultante, debe responder satisfactoriamente a las necesidades identificadas y dentro del alcance propuesto.

Construcción

Complementa el desarrollo del sistema por medio de tareas de análisis, diseño e implementación, se establece una arquitectura mucho más firme, adicionalmente, cada una de las iteraciones (o conjunto de tareas) es sometida a una etapa de pruebas, se documenta la construcción del sistema y su manejo.

Transición

Asegura que el producto esté disponible y totalmente funcional para el usuario (o propósito) final. Soluciona inconvenientes detectados en las pruebas de aceptación, se realizan ajustes finales para luego ejecutar la liberación del producto en un ambiente de producción. En este punto, también se complementan los manuales de usuario.

ESTRUCTURA DEL TRABAJO

La información suministrada por parte de Carga S.A.S., permitió analizar de manera detallada los inconvenientes presentados con el actual sistema de replicación de datos, y adicionalmente, realizar conclusiones, tanto cualitativas como cuantitativas, las cuales apoyan y/o sustentan, la mejora o desarrollo e implementación de un sistema más robusto de replicación de la información, el cual proporciona mayor estabilidad en toda la solución tecnológica de replicación y tiempos de respuesta más oportunos.

El primer paso es corroborar si es necesario, mantener los datos actualizados, en cada una de las estaciones remotas, y si dicho objeto, aporta valor a la organización, bien sea a nivel competitivo u operativo, de ser así, el planteamiento de una solución de replicación de datos, que sustituya, mejore o se implemente por primera vez, como apoyo a los sistemas de información que sostienen la operación de la empresa, es totalmente recomendada.

En la siguiente etapa, se realiza la fase de elaboración del plan de trabajo, el cual permite conocer la situación actual, y cómo el nuevo producto brinda soluciones y respuestas a todas las necesidades de la organización relacionadas con el sistema de replicación. En este punto, se obtiene un enfoque global de los objetivos que se deben cumplir, y se adquiere un conocimiento general de cuáles van a ser las áreas en que va a impactar (positivamente) el proyecto y quienes son las personas involucradas en el mismo, tanto del proveedor como del cliente.

Es necesario conocer a fondo, cómo interactúan las bases de datos, con los distintos sistemas de información que van a ser incluidos en el sistema de replicación, cuál es el volumen transaccional de los mismos y la cantidad de usuarios concurrentes en sesiones normales de trabajo. Para lograr obtener dicha información, se realizan monitoreos (no invasivos), como por ejemplo revisando archivos de registros (logs), en aquellos servidores y contenedores de bases de datos, que participan de manera activa en el procesamiento de la información que constituyen o pueden llegar a constituir, objeto de replicación.

Al mismo tiempo, se realiza un procedimiento de levantamiento de requisitos, que permite conocer información relevante para el cumplimiento exitoso de los objetivos propuestos para el sistema de replicación de información. En este punto también se detectan requisitos adicionales que pudieron no tenerse en cuenta en la etapa inicial y que no impactan de manera dramática al proyecto en general. La etapa de levantamiento de requerimientos, ayuda también a detectar los requisitos no funcionales, bien sea a nivel de comportamiento, estabilidad, usabilidad o seguridad, en los componentes que conforman el sistema de replicación en general.

Por medio de la información de rendimiento, (performance), obtenida de los servidores y contenedores de bases de datos, se crea una matriz que almacene los tiempos de respuesta de los equipos y bases de datos, operando sin el sistema de replicación, versus los mismos, con el sistema de replicación trabajando, de esta forma se mide el impacto que se genera en los tiempos de respuesta para las aplicaciones que los utilicen y así, poder realizar las acciones de mejora o correcciones pertinentes (tuning o puesta a punto).

Adicionalmente, se miden los tiempos de tránsito de los datos en la red, desde el servidor de replicación, hacía los equipos clientes y viceversa. El comparativo con los tiempos de tránsito en la red, generados por el sistema de replicación actual, evidencian la efectividad del nuevo sistema en la entrega de la información en los diferentes puntos.

El diseño del proyecto cuenta con tecnologías acordes a la época, las cuales ofrecen funcionalidad, fiabilidad, eficiencia, mantenibilidad y portabilidad.

La replicación de la información consiste, en mantener los datos actualizados, en cada uno de los puntos o nodos, representados en este caso por equipos de cómputo (denominados clientes), los cuales, deben pasar primero por un proceso de registro contra los servidores principales.

Uno de los elementos principales en el presente proyecto de replicación, es la utilización de un sistema de colas FIFO (First in - First Out – Primero en Entrar – Primero en Salir), el cual es el encargado de mantener ordenados cronológicamente la entrega de mensajes y enviarlos a todos los clientes. Dicha cola es administrada por medio de un patrón arquitectónico denominado Bróker de mensajes (Message Broker - agente de mensajes), cuya funcionalidad consiste en gestionar la entrega de los mismos en cada uno de los nodos, en este punto, se utiliza, un patrón llamado Publisher/Subscriber (Publicador/Subscriber), en el cual existen nodos que publican datos (envían información) y nodos que consumen los datos publicados (reciben información), creando de esta forma una red en la cual, los cambios en la información que se origina en cualquiera de los puntos, es replicada hacía los demás nodos. Los clientes nunca encolan mensajes directamente en el Bróker de Mensajes, esto garantiza una capa de seguridad en el

modelo, el encargado de entregar los mensajes a distribuir al Bróker de Mensajes es administrado por un único proceso denominado Broker Server Watcher. La comunicación entre el bróker de mensajería y los nodos, se establece por medio de una red privada virtual o VPN (Virtual Private Network), la cual garantiza, por medio de un protocolo de seguridad IP, IPSec, la privacidad de la información, dado que ésta viaja de manera encriptada (característica propia de una red VPN trabajando en conjunto con IPSec), dando la posibilidad de utilizar internet como medio de transporte, evitando incurrir en costos adicionales como lo son por ejemplo los canales dedicados. Como se puede evidenciar, en este sistema de intercambio de mensajes, intervienen varios actores, unos funcionan como servidores y otros como clientes.

Los servidores son los encargados de centralizar los procesos principales de distribución y almacenamiento de mensajes, dentro de estos, encontramos al Bróker de Mensajería (mencionado anteriormente), el servidor de Base de Datos principal, encargado de almacenar los mensajes de todos los nodos y su estado (si ha sido o no entregado al Bróker de Mensajes), y por último, el Broker Server Watcher, cuya función consiste en recibir los mensajes de los clientes y almacenarlos en el servidor de base de datos, además se encarga de entregar los mensajes directamente al Bróker de Mensajería.

Los clientes cuentan con una serie de servicios, unos que permanecen a la espera de nuevos mensajes, denominados Listeners, encargados de recibir nuevos mensajes y otros que se denominan Watchers, encargados de enviar mensajes. Los Listeners trabajan de la mano con un complemento adicional denominado Dispatchers (o despachadores), cuya función consiste en procesar de manera correcta cada uno de los nuevos mensajes que recibe el cliente.

El Bróker de Mensajería cuenta con la capacidad de suministrar cierto tipo de información que permite determinar el estado de los mensajes, bien sea porque se encuentran en espera o la totalidad de los procesados por cada uno de los clientes, permitiendo de esta manera detectar (y detener) comportamientos anómalos que se puedan presentar en los procesos de replicación, bien sea en uno o varios nodos.

PRESUPUESTO

Recursos Financieros

El presente anteproyecto no tiene costo financiero, dado que los autores del proyecto tuvieron la oportunidad de acceder a la organización, recopilar datos de ella y estructurar una posible solución, a cambio del ofrecimiento de una herramienta funcional, la cual será el producto del proyecto.

Recurso Humano

Para el desarrollo del proyecto, inicialmente se cuenta con los autores del mismo.

Tabla 1 – Recurso Humano

Fuente: Elaboración por parte de los autores del proyecto

ECURSO	UNIDAD HORA	VALOR	VALOR
Autores de proyecto	(944 horas) x (2)	\$8.333	\$15.732.704
Total Recurso Humano	1.888		\$15.732.704

Recursos Tangibles e Intangibles

Adicionalmente también se contaron con los siguientes gastos operacionales.

Tabla 2 – Recursos Tangibles e Intangibles

Fuente: Elaboración por parte de los autores del proyecto

ELEMENTO	VALOR
Portátil DELL	\$1'100.000
Portátil SAMSUNG	\$1'200.000
Multifuncional Canon	\$200.000
Combustible	\$180.000
Papelería	\$35.000
Cartuchos Impresora	\$100.000
Almuerzos	\$500.000
Total	\$3'315.000

CRONOGRAMA

Tabla 3 - Cronograma de Actividades

Fuente: Elaboración por parte de los autores del proyecto

Actividad	SEMANAS																		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17		
Identificar el estado actual de la organización.																			
Establecer el sistema de replicación.																			
Diagramar casos de uso																			
Construir la aplicación																			
Realizar pruebas al sistema																			
Instalar la aplicación en un ambiente de producción.																			
Capacitar y entregar documentación a la organización.																			
Sustentación del proyecto																			

Semana 1: Julio 29 a Agosto 2

Semana 17: Noviembre 25 al 29

CONCLUSIONES

El seleccionar un buen diseño arquitectónico, permite alcanzar una integración exitosa entre las distintas aplicaciones que conforman el ecosistema transaccional o productivo de la Organización.

Los sistemas de replicación de datos, son productos complejos que requieren un alto nivel de conocimiento en diferentes áreas, y su éxito en la implementación, como en cualquier otro modelo de integración, depende de que exista una correlación muy precisa entre todos sus elementos.

Las organizaciones hoy día, aún no están dispuestas a apostar por soluciones que para su funcionamiento, dependan de una conexión permanente a internet, y mucho menos cuando estas aplicaciones forman parte esencial del núcleo del negocio.

Grandes aplicaciones modernas basadas en cloud computing (computación en la nube), utilizan esquemas de trabajo fuera de línea (offline), para luego ser sincronizadas contra un servidor principal, lo que ratifica que los esquemas de sincronización de información juegan un papel muy importante en el mundo de la tecnología.

RECOMENDACIONES

El modelo arquitectónico de la solución permite que la implementación no sea simplemente un sistema de replicación de datos, existe un alto potencial, el cual es responsabilidad de la organización llevarlo a su máxima expresión, un ejemplo de ello, es utilizarlo como monitoreo de sucursales en línea, sistema de apoyo en el proceso continuidad del negocio en eventos de recuperación de desastres, esquema de actualización de aplicaciones, central de reportes del estado de servidores o estaciones de trabajo, unificación de diferentes fuentes de información, entre otros.

Las organizaciones que deseen enfrentar el reto de implementar soluciones de integración similares a la planteada en este proyecto, deben tener una visión clara de cuáles son las metas a corto, mediano y largo plazo, para que la base arquitectónica inicial sea totalmente estable y funcional, y de esta manera, los cambios a aplicaciones existentes y los requerimientos de las futuras puedan adaptarse fácilmente al modelo establecido.

Por parte de los autores del proyecto, se alienta a la universidad a incentivar en los estudiantes el explorar y aplicar los diferentes modelos de arquitectura de software, dado que son una herramienta de alto nivel que en muchas ocasiones, brindan soluciones de calidad y altamente escalables, a esquemas complejos.

BIBLIOGRAFÍA

Castellanos, Luis (2011). *Desarrollo de Sistemas de Información bajo un enfoque incremental*.

Maracaibo: UNIVERSIDAD NACIONAL EXPERIMENTAL DE LA FUERZA ARMADA.

Comer, Douglas E. (1996). *Redes Globales de Información con Internet y TCP/IP: Principios*

Básicos, Protocolos y Arquitectura (3 ed.). México: PRENTICE HALL HISPANOAMERICANA.

Date, C.J. (2003). *An Introduction to Database Systems (8 ed.)*. Canadá: ADDISON-WESLEY

PROFESSIONAL.

Kurose, James F. y Ross Keith W (2010). *Redes de Computadores: Un enfoque descendente (5*

Ed.). Ribera del Loira, Madrid: PEARSON EDUCACIÓN S.A.

Peña Ayala, Alejandro (2006). *Ingeniería de Software: Una Guía para Crear Sistemas de*

Información. México: INSTITUTO POLITÉCNICO NACIONAL.

Perlman, Radia. (2000). *Interconnections: Bridges, Routers, Switches and Internetworking*

Protocols (2 ed.). Canadá: ADDISON-WESLEY PROFESSIONAL.

Ramos, María Jesús, Ramos, Alicia y Montero, Fernando: *Sistemas Gestores de Bases de Datos*.

España: MCGRAW HILL

CIBERGRAFÍA

Moreno León, Jesús y Molina Coballes, Alberto (2013). *IPSec VPNs*. Obtenido el 16 de marzo de 2013 en:

http://informatica.gonzalonazareno.org/plataforma/pluginfile.php/2487/mod_resource/content/0/tema5/IPSEC_VPN.pdf?forcedownload=1

Mur Siles, Pedro Antonio (2013). *Seguridad a Nivel de IP: IPSec*. Obtenido el 16 de marzo de 2013 en <http://upcommons.upc.edu/revistes/bitstream/2099/9944/1/Article005.pdf>

Piedrahita, Alejandro (2013). *Introducción a los Computadores: Estructura del Computador*. Obtenido el 16 de marzo de 2013 en

<http://ciencias.udea.edu.co/programas/pregrado/CNM-130/docs/clase1sinp.pdf>

UNIVERSIDAD DE OVIEDO (2013). *Desarrollo de aplicaciones distribuidas utilizando el estándar CORBA*. Obtenido el 16 de marzo de 2013 en

<http://usuarios.multimania.es/tralqb/Apuntes de CORBA.pdf>

Vásquez, César y Herrera, Nadia (2013). *Base de Datos Distribuidas: Replicación*. Obtenido el 16 de marzo de 2013 en

<http://www.itescam.edu.mx/principal/sylabus/fpdb/recursos/r11294.PDF>

Santamaría Carmona, Rafael (2013). *Copias de Seguridad: Estudio y Metodología*. Obtenido el

16 de marzo de 2013 en

<http://openaccess.uoc.edu/webapps/o2/bitstream/10609/14001/6/rsantamariaTFC0612memoria>