INFOTEC CENTRO DE INVESTIGACIÓN E INNOVACIÓN EN TECNOLOGÍAS DE LA INFORMACIÓN Y COMUNICACIÓN

DIRECCIÓN ADJUNTA DE INNOVACIÓN Y CONOCIMIENTO
GERENCIA DE CAPITAL HUMANO
POSGRADOS

“ESTRATEGIAS ALTERNAS PARA EL ALMACENAMIENTO Y RECUPERACIÓN DE DATOS EN NACIONAL FINANCIERA”

Tipo de Proyecto: Solución Estratégica Empresarial

Que para obtener el grado de MAESTRO en Dirección Estratégica de las Tecnologías de Información y Comunicación

Presenta:

Ing. Joaquín Chaparro Cortés

Asesor:

Dr. Ricardo Marcelín Jiménez

México D.F., a 30 de enero de 2016

NÚMERO DE FÓLIO: 025

AUTORIZACIÓN PARA IMPRESIÓN DE PROYECTO INTEGRADO

La Coordinación académica del área de Posgrados INFOTEC hace constar que la investigación titulada:

"ESTRATEGIAS ALTERNAS PARA EL ALMACENAMIENTO Y RECUPERACIÓN DE DATOS EN NACIONAL FINANCIERA"

Desarrollada por el alumno

Nombre: JOAQUÍN

Apellido paterno: CHAPARRO

Apellido materno: CORTÉS

Con número de matrícula: 169-2013-0112

Alumno de la Maestría en dirección estratégica en tecnologías de información y comunicación

Desarrollado bajo la asesoría de

Nombre del Tutor: Dr. Ricardo Marcelín Jiménez

Ha sido revisada y aprobada por los profesores:

Dr. Ricardo Marcelín Jiménez

Dr. Ramón Reyes Carrión

Quienes han depositado en esta coordinación en su oportunidad sus reflexiones y comentarios que han sido atendidos e integrados **en su totalidad** por el alumno a la nueva versión escrita de su proyecto integrado, siendo corroborados por los mismos revisores.

Por lo cual esta coordinación expide la presente **autorización para la impresión del proyecto integrado** al que se ha hecho mención.

Vo. Bo.

Mtro. Héctor David Berriolope Galván

Coordinador Académico del área de Posgrados de INFOTEC

* Anexar la presente autorización al inicio de la versión impresa del proyecto integrado que ampara la misma.

DIRECCIÓN ADJUNTA DE INNOVACIÓN Y CONOCIMIENTO

GERENCIA DE CAPITAL HUMANO

Nubia, muchas gracias por incentivar a continuar mis estudios, de no haber sido por ti seguiría como proyecto y no como una realidad.

Tabla de Contenido

Índice de Figuras

Índice de Tablas

Siglas y abreviaturas

Banxico	Banco de México
BCP	Business Continuity Plan
BIA	Businness Impact Analysis
CCC	Centro de Cómputo de Contingencia
CCP	Centro de Cómputo Principal
CNVB	Comisión Nacional Bancaria y de Valores
DRP	Disaster Recovery Plan
IAAS	Infrastructure As A Service
IDA	Information Dispersal Algorithm
NAFIN	Nacional Financiera
PYMES	Micros, Pequeñas y Medianas Empresas
RPO	Recovery Point Objective
RTO	Recovery Time Objective
SAN	Storage Area Network
SFP	Secretaria de la Función Pública
SG	Secretaria de Gobernación
UMA	Unidad Máxima de Almacenamiento

Resumen

Como parte de los procesos continuos de mejora, reducción de costos e innovación, en Nacional Financiera se revisan regularmente todas las operaciones que dan sustento a la organización. El presente proyecto se inscribe en este contexto y tiene por objetivo revisar las operaciones de almacenamiento y recuperación de la información, plantear soluciones alternativas y aportar elementos para valorar su adopción.

Para contar con una referencia, se presenta la arquitectura actual de almacenamiento y recuperación de datos procesados en el centro de cómputo principal en Nacional Financiera, así como una breve descripción del plan de continuidad de negocio y del plan de recuperación ante desastres, relacionado con los procedimientos de respaldo de estos datos.

Como parte de la investigación también se considera el marco normativo, que incluye a las entidades del gobierno federal, como la Comisión Nacional Bancaria y de Valores, o el Banco de México, que regulan los temas de almacenamiento, respaldo, recuperación de datos y planes de continuidad de negocio en el sector financiero, así como su reglamentación asociada con estos temas. Se consideran también los artículos que detallan las obligaciones que los bancos deben cumplir en esta materia, sean estos privados o públicos, como es el caso de Nacional Financiera. Adicionalmente, se reconoce que en la misma materia existen regulaciones para todas las entidades federales, cuya supervisión queda a cargo de la Secretaría de Gobernación y de la Secretaría de la Función Pública.

Para los fines de este trabajo, se considera que el Sistema Babel, basado en la tecnología de almacenamiento distribuido, puede representar un punto de partida con el que se pueda plantear una alternativa para las necesidades que NAFIN tiene en el tema de almacenamiento de información. Con la finalidad de evaluar la viabilidad del cómputo en la nube que incorpore los principios de operación de Babel, en este trabajo se describe la arquitectura de un sistema adhoc y se presentan los resultados de un conjunto de pruebas de desempeño, que aportan elementos de decisión para evaluar la adopción de esta tecnología.

Introducción

Una parte estratégica de la operación de los bancos está soportada por los servicios basados en tecnologías de la información, sobre los que se implementan los planes para enfrentar contingencias técnicas u operativas. Por ello, las entidades reguladoras de los bancos, como la Comisión Nacional Bancaria y de Valores (CNBV) y el Banco de México (Banxico) principalmente, han formulado varios requerimientos y políticas que todo banco debe cumplir. Entre los principales aspectos a considerar está la necesidad de un plan de continuidad de negocio y acciones para el almacenamiento, recuperación y respaldo de los datos procesados en los centros de cómputo de cada banco.

Nacional Financiera (NAFIN) forma parte de la banca de desarrollo dentro del sector hacendario del gobierno federal. Como banco debe cumplir los requerimientos y políticas dictadas por la CNBV y Banxico, así como con las políticas señaladas en la Estrategia Digital Nacional, reguladas por las Secretarías de Gobernación y de la Función Pública.

Actualmente, NAFIN cuenta con una estrategia de almacenamiento y recuperación de datos, basada en arquitecturas tradicionales con equipos de red de área de almacenamiento, conocidos como SAN por sus siglas en inglés (Storage Area Network) Por otra parte, el respaldo de los datos se realiza a través de medios magnéticos como cintas alojadas en equipos específicos para su grabación y lectura, llamadas librería robótica, con unidades de cintas para respaldos. Para cumplir las políticas de la CNBV, como la de garantizar la continuidad de negocio, NAFIN cuenta con un Centro de Cómputo de Contingencia o de respaldo (CCC), en el cual también se tiene una arquitectura similar para el almacenamiento y recuperación de datos, pero con una menor capacidad que la infraestructura instalada en el centro principal (CCP), ambos centros de cómputo están conectados de forma permanente a través de enlaces de datos dedicados con la finalidad de transmitir los datos de los servicios considerados más críticos, desde los servidores instalados en el centro de cómputo principal hacia la infraestructura instalada en el centro de contingencia. Para los demás servicios, al término del plan de producción y después de realizados los respaldos en cintas, se genera una segunda cinta o copia, la cual se envía por medio de una compañía especializada en transportes, hacia una bóveda fuera de sitio en el CCC.

Los datos trasferidos en línea y copias de las cintas en el centro de cómputo de contingencia forman parte de la estrategia institucional de Continuidad de Negocio o BCP por sus siglas en inglés (Business Continuity Plan), la cual fue diseñada bajo los requerimientos de negocio de hace 10 años. Si bien, esta estrategia cumple las normativas y políticas requeridas, es necesario evaluar alternativas que reduzcan los tiempos de recuperación comprometidos, tanto en el BCP, como en el Plan de Recuperación ante Desastre DRP por sus siglas en inglés (Disaster Recovery Plan), así como reducir los costos involucrados en la estrategia actual manteniendo los niveles de seguridad informática.

Una alternativa tecnológica que se ha evaluado sobre todo por la reducción de costos, es utilizar los servicio de cómputo en la nube o “cloud computing”, pero el hecho de que los datos estarían hospedados en infraestructura ubicada en centros de cómputo externos a la institución, incluso fuera del territorio nacional, y con la posibilidad por mínima que sea de que personal ajeno a la entidad pudiera tener acceso a estos datos, han hecho que esta alternativa no sea vista por la alta dirección como factible hasta el momento.

Por otra parte, el almacenamiento distribuido parece ofrecer principios que garantizan la disponibilidad de la información, sin comprometer su integridad y confidencialidad. En los sistemas basados en esta tecnología, cada archivo que se recibe es transformado en varios archivos más pequeños, llamados dispersos, usando técnicas de procesamiento relacionadas con los códigos para tratamiento de errores y el secreto compartido. Luego, cada uno de estos dispersos puede almacenarse en un dispositivo diferente. Estos esquemas de codificación tienen un conjunto de propiedades que pueden ser aprovechadas en beneficio de la disponibilidad y confidencialidad: i) el archivo original puede recuperarse sólo si se dispone de un subconjunto del total de dispersos, llamado umbral, cuyo número es configurable, ii) un subconjunto de dispersos menor al umbral no puede revelar el contenido del archivo original. En combinación, estas propiedades nos dicen que podemos guardar los dispersos en diferentes emplazamientos y que, aun si algunos de estos se vieran comprometidos, el archivo

original no podrá recuperarse, en tanto no se dispongan de un número suficiente (que puede configurarse).

Supongamos ahora que usamos una combinación de almacenamiento distribuido y cómputo en la nube. Esto es, procesamos un archivo conteniendo, por ejemplo, la imagen de una base de datos y generamos sus dispersos, que luego emplazamos en diferentes sitios en la nube, ajenos y operados de forma independiente entre sí. Supongamos también que usamos un esquema de codificación tal que cada archivo se transforma, por ejemplo, en 5 dispersos y se puede recuperar a partir de cualesquiera 3 de ellos. Un usuario no autorizado tendría que conseguir el acceso a por lo menos 3 de estos sitios, para poder recuperar la información almacenada bajo esta nueva solución. Si esta alternativa no fuera suficiente, podría reconfigurarse el esquema de codificación para hacer más difícil la captura del umbral de recuperación.

Por lo anterior, la presente investigación tiene como objetivo principal diseñar una estrategia alterna para el almacenamiento y recuperación de datos en la plataforma distribuida de Nacional Financiera, con la finalidad de optimizar recursos en la recuperación de los datos, en caso de alguna incidencia que impida la operación en el Centro de Cómputo Principal.

CAPITULO I.- Almacenamiento y Recuperación de Datos en Nacional Financiera.

Marco Normativo para Instituciones Financieras sobre procedimientos de respaldo de Información procesada en sus Centros de Cómputo y de continuidad de negocio.

Todos los bancos públicos o privados que operan en territorio mexicano deben cumplir varios lineamientos y políticas, vinculadas a su operación, administración, así como lo relacionado con aspectos técnicos y de continuidad de negocio, entre otras varias regulaciones.

Entre las principales normativas dispuestas a los bancos están las impuestas tanto por la Comisión Nacional Bancaria y de Valores (CNBV), como por el de Banco de México (Banxico), y en últimas fechas la Secretaria de Gobernación y la Secretaria de la Función Pública, en el marco de la Estrategia Digital Nacional.

Para lo relacionado con los procedimientos de respaldo de Información procesada en sus Centros de Cómputo, así como los procedimientos de continuidad de negocio que se debe cumplir, están las normas que a continuación se describen.

La CNBV a través de las “Disposiciones de carácter general aplicables a las instituciones de crédito” (1) norma lo relacionado con el plan de continuidad de negocio, en sus artículos 71, 78, 154, 156, 164 y 164 Bis, lo relacionado con los procesos de respaldos se encuentra en los artículos 75, 164 y 316 Bis 11, mientras que el denominado “Anexo 67” (2) norma los requerimientos mínimos que debe contener un Plan de Continuidad de Negocio en sus puntos I y II, y el “Anexo 52” (3) en sus puntos I, II y III señalan los lineamientos mínimos de operación y seguridad para la contratación de servicios de apoyo tecnológico.

Por otra parte el Banco de México (Banxico) a través de la circular 4/2012 (4), indica los 31 puntos que deben cumplirse para que las instituciones de crédito mantengan el permiso de prestar servicios de operaciones de derivados, en esta circular en el punto 11 hace referencia al cumplimiento de requerimientos de su Anexo y en particular los puntos 4 y 19 de este anexo, son las disposiciones que se deben cumplir relacionadas con el plan de continuidad y respaldo y recuperación de datos respectivamente.

Por último, la Secretaria de Gobernación (SG) y la Secretaria de la Función Pública (SFP), como parte del Plan Nacional de Política Digital 2012 – 2014 emitieron el "acuerdo que tiene por objeto emitir las políticas y disposiciones para la Estrategia Digital Nacional, en materia de tecnologías de la información y comunicaciones, y en la de seguridad de la información, así como establecer el Manual Administrativo de Aplicación General en dichas materias" (5), en el artículo 13 señalan lineamientos para servicios de Centros de Datos y en el artículo 27 indica lo relacionado con los procesos de respaldos.

En el Anexo "I", y como parte de este documento, se detallan los artículos señalados en los anteriores párrafos.

Nacional Financiera

Nacional Financiera (NAFIN) es el banco de desarrollo del Gobierno Federal Mexicano para atender las necesidades de financiamiento, capacitación y asistencia técnica a las micros, pequeñas y medianas empresas mexicana (PYMES). Fue creada en 1934 para otorgar la liquidez y flexibilidad que requerían los bancos mexicanos debido a la crisis de la Gran Depresión de 1929 en los Estados Unidos y las dificultades financieras del México post revolucionario. Desde 1989, NAFIN otorga recursos financieros y garantías, principalmente como Banco Nacional, además de ser agente financiero del Gobierno Federal con relación a la negociación, contratación y manejo de créditos del interior, cuyo objetivo sea fomentar el desarrollo económico; así como ofrecer servicios fiduciarios a los sectores privados (6).

A partir del año 2001, NAFIN experimentó un profundo cambio estructural con la clara visión de partir de las necesidades de los clientes, aplicando las mejores prácticas internacionales (innovación, alineación a procesos, apoyándose en una sólida plataforma tecnológica y certificación de calidad) para cumplir con sus objetivos. Siendo NAFIN el principal detonador del financiamiento y servicios de apoyo destinados a las PYMES (7).

Actualmente la misión de NAFIN es la de promover el acceso de las PYMES a los servicios financieros; impulsar el desarrollo de proyectos sustentables y estratégicos para el país; promover el desarrollo del mercado de valores y fungir como Agente Financiero del Gobierno Federal, con el fin de contribuir al crecimiento regional y a la

creación de empleos, a través de innovación y calidad, con gente comprometida y guiada por valores compartidos [8].

Los principales Productos y Servicios que ofrece NAFIN son [8]:

- Intermediarios financieros No Bancarios
- Compras del Gobierno Federal
- Programas Empresariales
- Cadenas Productivas
- Capacitación empresarial
- Piso Financiero

Como parte del gobierno federal y siendo a la vez un banco, NAFIN está obligada a cumplir con los lineamientos y políticas mencionadas en la sección anterior, la forma en que se da cumplimiento se describe en las siguientes secciones.

Situación actual de la estrategia de almacenamiento y recuperación de datos

Como parte del cambio estructural en el 2001, la dirección general a través de una arquitectura empresarial se focalizó en las necesidades de los clientes (las PYMES en este caso) y se planteó el objetivo de contar con las mejores prácticas en las diferentes áreas de la institución, entre ellas la dirección de informática, la cual además de diseñar para cada dominio tecnológico una arquitectura, también procedió a definir y homologar estándares tecnológicos.

Como parte de esta definición y con la finalidad de priorizar los recursos técnico y económicos asignados, se inició la tarea de definir y clasificar la cartera de servicios que proporcionaba conforme a un análisis de impacto al negocio o BIA por sus siglas en inglés (Businness Impact Analysis), definiéndose cuatro niveles de criticidad en donde los servicios aplicativos clasificados con el número “1” serían los de mayor criticidad y los definidos como número “4” los de menor criticidad.

En paralelo y para cumplir con la normativa y regulación requerida por la CNBV, se diseñó e implantó el plan de recuperación ante desastre o DRP por sus siglas en inglés, donde tomando como base la clasificación de servicios aplicativos se definieron los tiempos objetivos de recuperación o RTO por sus siglas en inglés (Recovery Time

Objective) y el punto objetivo de recuperación o RPO por sus siglas en inglés (Recovery Point Objective), para el caso de que ocurriera algún incidente que impidiera continuar operando desde el Centro de Cómputo Principal (CCP), y se requiriera iniciar operación desde el Centro de Cómputo de Contingencia (CCC), estos tiempos están señalados en la tabla 1:

Tabla 1.- RTO y RPO en caso de DRP

Criticidad de servicio	(*) Tiempo objetivo de recuperación RTO	(**) Punto objetivo de recuperación RPO
1	24 horas	10 minutos
2	24 horas	1 hora
3	1 semana	último respaldo realizado
4	1 mes	último respaldo realizado

(*) Tiempo máximo en restaurar operación posterior a la declaración de contingencia

(**) Datos recuperados anteriores al momento del desastre

Los datos de la tabla anterior quedaron acordados y comprometidos con las diferentes áreas usuarias responsables de cada servicio aplicativo a través de acuerdos de niveles de servicio, los cuales fueron firmados tanto por las áreas usuarias como por el staff de la Dirección de informática. Cabe señalar que cada año se realizan por lo menos dos simulacros donde participa las áreas técnicas junto con las áreas usuarias para mantener capacitado tanto al personal técnico, como al operativo de negocio y, al término del simulacro se evalúa y califica el cumplimiento de los niveles de servicio acordados.

Arquitectura de la Infraestructura Distribuida, Almacenamiento y Respaldo

Para cumplir con los acuerdos de niveles de servicio, se diseñó e implemento la arquitectura de servidores y de telecomunicaciones mostradas en la siguiente figura 1, en esta se esquematiza las diferentes plataformas utilizadas y los canales de comunicación tanto para la replicación de datos, como el acceso y navegación por Internet:

Figura 1.- arquitectura de servidores, almacenamiento, respaldo y telecomunicaciones

Se pueden observar plataformas duplicadas en ambos centros de cómputo, pero con diferentes capacidades de procesamiento y almacenamiento, las cuales están descritas en la tabla 2. El objetivo de esta arquitectura es contar con una infraestructura instalada y disponible para utilizarse un par de horas después de declarada la contingencia. Sin embargo, por razones económicas, la infraestructura instalada de forma permanente en el Centro de Cómputo de Contingencia será solo aquella que permita iniciar la operación productiva en máximo 24 horas después de la contingencia y recuperar los servicios aplicativos clasificados con criticidad 1 y 2. En tanto, para los servicios clasificados como 3 y 4, el plan sería contratar como servicio o arrendar la infraestructura complementaria para que en un máximo de 20 días hábiles se tenga la capacidad total de operar todos los servicios.

Tabla 2.- volumetría de la infraestructura distribuida

Infraestructura	C. Cómputo Principal	C. Cómputo Contingencia
Servidores físicos	19	4
Servidores virtuales	64	11
SAN	17.5 TB	3.3 TB
Librería	45 cintas	45 cintas
Internet	32.7 Mbps	2 Mbps
Enlace de datos	10.2 Mbps	

Como resumen de la arquitectura instalada y conforme a cada plataforma podemos señalar:

(1) **Servidores.-** todos los servidores instalados en el CCP están bajo un esquema de alta disponibilidad, es decir redundantes, y más del 80% de estos servidores son virtuales, en el CCC los servidores instalados no tienen equipos redundantes aunque el diseño de cada servidor sea de alta disponibilidad contando con tarjetas y fuentes redundantes, sin embargo existen equipos críticos activos para el inicio inmediato de operaciones como un directorio activo el cual esta sincronizado en línea con los del CCP y servidores de seguridad como firewalls también disponibles de inmediato. En lo que respecta al tema de virtualización, también se cuenta de forma permanente con equipos y licenciamiento actualizado con servidores de servicios criticidad 1 y 2 de forma pasiva, y el resto de los servidores están configurados para que, en caso de contingencia, serían creados de forma automática.

(2) **Almacenamiento.**- se cuenta con una SAN por sus siglas en inglés (Storage Area Network) en cada centro de cómputo, en la del CCP se aloja el 100% de los servicios y bases de datos de la plataforma distribuida, esta plataforma ofrece alta disponibilidad de tarjetas, discos, accesos, fuentes, etc., la SAN instalada en el CCC es de las mismas características que en el centro principal, pero con una capacidad de almacenamiento menor, solo está recibiendo los datos de los servicios críticos clasificados como 1 y 2 con un retardo máximo de 10 minutos. No obstante, en caso de declararse una contingencia dispone de slots libres para adquirir los discos faltantes y en plazo máximo de un mes alcanzar el 100% la capacidad del CCP.

(3) **Respaldo. -** como parte de un proceso continuo y del plan de producción, el área operativa de informática, al término de los procesos de cierre de línea y Batch (si aplicara), procede de forma manual o automática a realizar el respaldo en cintas magnéticas de cada servicio y cada base de datos, sin importar la clasificación de criticidad. Este proceso se hace a través de una infraestructura conocida como librería robótica con unidades de cintas, con la capacidad de almacenar todas las

cintas que se utilicen a diario y conservando para consulta o una posible restauración las realizadas semanas e incluso meses o años anteriores, esta librería cuenta con dos brazos mecánicos y cuatro drivers de lectura/escritura, con lo que se pueden efectuar varios respaldos o restauraciones de forma simultánea. Al término de la generación de las cintas que se denominan “normales” de forma automática, la librería inicia el proceso de generar una segunda cinta o copia fiel de cada una de las cintas generadas en el día, esta segunda cinta se le denominará de “respaldo”, todo este proceso se debe iniciar y concluir por la noche, ya que por la mañana se retiran de la librería las cintas de respaldo y se empaquetan para trasladarse a través una empresa de seguridad especializada en traslado de bienes, a una bóveda de cintas ubicada en el CCC. Este mismo lugar también cuenta con una librería de similares características que la del CCP, con un brazo mecánico y dos drivers de lectura/escritura, prácticamente sin cintas ya que en caso de declararse una contingencia, se llenaría con las cintas de respaldo bajo custodia de la bóveda fuera de sitio para iniciar la restauración de los servicios aplicativos clasificados con criticidad 3 y 4.

(4) **Telecomunicaciones.**- para garantizar que los datos de los servicio críticos clasificados como 1 y 2 estén replicados en ambos centros de cómputo, se cuenta con una infraestructura de red y telecomunicaciones redundante, en ruteadores y enlaces dedicados, adicionalmente se cuenta con equipos redundantes también de optimizadores de anchos de banda, los cuales además de aprovechar mejor el canal de datos, manejan prioridades en el tipo de tráfico y asigna calidad de servicio diferenciado, adicionalmente se cuenta en cada centro de cómputo con un enlace o conectividad hacia Internet, el del CCC con menor ancho de banda que el del CCP, pero bajo contrato con la factibilidad de crecerlo en un tiempo máximo de un mes.

En lo relacionado con la seguridad tanto física como lógica, los accesos a los centros de cómputo descritos, están monitorizados, supervisados y auditados por el área de seguridad física de NAFIN, y en lo que corresponde a la seguridad informática el acceso es exclusivamente a los usuarios operativos permitidos. Cabe señalar que el acceso es

desde los equipos de cómputo personal instalados en las instalaciones de NAFIN, es decir no hay acceso remoto a las consolas de administración de los servidores o aplicativos desde equipos externos a la red de datos NAFIN, ya sea por enlaces dedicados o Internet. Adicionalmente, todo el personal que tengan claves de usuario con acceso a las infraestructuras debe firmar una carta responsiva. Por otra parte, a través del área de seguridad informática se auditan los logs y bitácoras de accesos para identificar si algún usuario intento acceder a algún sitio no permitido por su perfil o hizo buen uso de su usuario administrador.

Con esta arquitectura se garantiza no solo los niveles de servicio acordados con las áreas usuarias, sino que se cumple con las disposiciones y políticas requeridas por la CNBV, el marco normativo de la SG y la SFP, así como con las constantes auditorias de otras entidades como el Banco de México. De esta forma NAFIN mantiene la operación de algunos de sus servicios como son mercados de valores, piso financiero, derivados, futuros, etc.

Por último, el tema del precio por esta infraestructura queda señalado en la siguiente tabla:

Tabla 3.- Precios anuales de la infraestructura distribuida (USD)

Infraestructura	C. Cómputo Principal	C. Cómputo Contingencia
Servidores físicos	$ 39,650.00	$ 5,660.00
Servidores virtuales	$ 39,050.00	$ 10,520.00
SAN	$ 38,615.00	$ 24,690.00
Librería	$ 33,970.00	$ 33,070.00
Internet	$ 95,945.00	$ 8,742.00
Enlace de datos	$ 260,983.00	

Las anteriores consideraciones, tanto del precio, como de la seguridad, servirán como criterios para evaluar la factibilidad de una estrategia alterna de respaldo y recuperación de datos, que será descrita y analizada en los siguientes capítulos.

CAPITULO II.- El sistema de archivos BABEL

Definiciones y conceptos básicos

Antes de iniciar con la definición del Sistema BABEL, y si bien este trabajo no pretende realizar un análisis e investigación a detalle sobre conceptos y definiciones relacionadas con almacenamiento, tolerancia a fallas, redundancia y métodos de codificación de datos, con la finalidad de contar con un marco de referencia y conceptos básico de estos temas, y apoyados en la tesis realizada por Becerril y Pichardo [9], se describe el siguiente resumen:

Almacenamiento de Información

Por almacenamiento de la información se entiende el mantenimiento de la información en un periodo de tiempo para posterior recuperación, dada una cierta capacidad a disposición, siendo las principales tendencias tecnológicas de almacenamiento:

- Arreglos de discos a unidades de disco duro (HDD); esta tecnología guarda los datos al magnetizar partículas microscópicas en la superficie de un disco o una cinta. Las partículas conservan su orientación magnética hasta que se modifica la orientación, lo cual hace a los discos y a las cintas medios de almacenamiento bastante permanentes, pero modificables [10].
- De estado sólido (SSS); utiliza memoria no volátil, como la memoria flash para almacenar datos, en lugar de los discos magnéticos de las unidades de discos duros convencionales.
- En la Nube (cloud storage); basado en redes de computadoras los datos están alojados en espacios de almacenamiento virtualizado, por lo general aportados por terceros.
- Masivo o “Big Data”; es el conjunto de datos que exceden los límites y tamaños de capacidades de procesamiento normales, cualquier conjunto de datos que rompe los límites y capacidades convencionales de TI diseñadas para apoyar las operaciones del día a día.
- Para la virtualización; en donde se unen múltiples dispositivos de almacenamiento en red, en lo que aparenta ser una única unidad de almacenamiento.

Almacenamiento distribuido

Sin embargo, las tendencias señaladas evidencian el espacio o dispositivo de almacenamiento, pero no resuelven el posible problema de fallas en ellos o de la pérdida de información. En razón de ello, se han desarrollo técnicas para almacenar los datos de forma fiable, entre los cuales se encuentran los sistemas basados en arquitectura de redes de computadoras, que agrupa diversos dispositivos de almacenamiento en una red a fin de aumentar la capacidad de los sistemas así como distribuir la información sobre las computadoras disponibles dentro de una red, para que ante alguna falla en sus componentes (dispositivos de almacenamiento), se mantengan los contenidos almacenados.

Siendo las dos principales ventajas de los sistemas de almacenamiento distribuido frente a los métodos tradicionales, el aumento de la capacidad total del sistema debido a que reúne un conjunto de computadoras compartiendo sus recursos de almacenamiento y su confiabilidad.

Como ejemplo de este almacenamiento son las Redes de Área de Almacenamiento o Storage Area Networks (SAN) y las Redes Peer-to-Peer (P2P).

La SAN emplea tecnología de canal de fibra cuya tasa de transferencia es muy alta y la probabilidad de error en las transmisiones es muy baja (del orden de 10-9), evidenciando un alto grado de confiabilidad. Estas redes se comunican mediante diferentes protocolos, siendo el más común el protocolo de canal de fibra o Fibre Channel Protocol (FCP).

En cuanto a las Redes P2P, se basa en la conectividad de diversos usuarios de forma remota o local, los cuales comparten sus recursos de almacenamiento. Cada computadora conectada a una red de este tipo cumple funciones de cliente y también de servidor, por ello, los usuarios pueden acceder a la información solicitada.

En el almacenamiento distribuido por la forma en que almacenan y distribuye la carga de diferentes archivos se podrían agrupar en:

- Archivos de Datos basados en la Replicación (replication-based data archives); se distribuyen réplicas de datos o erasure resilient fragments (borrado resistente de

fragmentos) a un conjunto diverso de nodos para asegurar la disponibilidad de datos y la supervivencia a largo plazo.

- Intercambio de archivos y Sistemas editores (File Sharing and Publishing Systems); son sistemas de intercambio de archivos peer-to-peer, en los que muchos nodos contribuyen con los recursos de almacenamiento y de red para construir un ambiente de alta disponibilidad de los archivos. Corresponde a cada nodo individual la decisión acerca de qué archivos se deben guardar y desalojar a lo largo del tiempo.

Tolerancia a fallas y redundancia

Por otra parte, es necesario garantizar la fiabilidad del procedimiento de almacenamiento distribuido, y por consiguiente la fiabilidad se podría clasificar en:

i. Disponibilidad (availability); es la capacidad del sistema para que los datos sean accesibles ante su petición por parte de un cliente.
ii. Confiabilidad (reliability); es la capacidad de un sistema para funcionar en presencia de fallas durante un determinado tiempo.
iii. Seguridad (safety); es la probabilidad de que no ocurra ningún suceso que provoque una falla o accidente.
iv. Capacidad de Mantenimiento (maintainability); es la capacidad del sistema para restaurarse en un determinado tiempo después de presentarse una falla.

También se pueden presentar fallas y/o averías en los componentes de un sistema, las cuales interrumpen el funcionamiento de éste evitando el cumplimiento de algunas o todas las propiedades descritas anteriormente.

Mientras que una falla es la desviación de un sistema sobre su funcionamiento normal, una avería es la causa de un error en el funcionamiento interno de un sistema e indirectamente la causa de una falla. Las fallas se clasifican de acuerdo a diferentes modelos, los cuales describen sus causas, siendo las fallas más comunes [11]:

i. Paro; corresponde a la situación en la que un componente detiene su funcionamiento y permanece en este estado. Los demás componentes pueden detectar este tipo de fallas.
ii. Caída; al igual que en la falla de paro, se detiene el funcionamiento, sin embargo los demás componentes no pueden detectar esta falla.

iii. Caída y enlace; una falla de este tipo se presenta cuando un componente detiene su funcionamiento, permanece en ese estado y además un enlace falla perdiendo mensajes sin retrasar, corromper o duplicarlos.
iv. Omisión de recepción; un componente incurre en este tipo de falla cuando recibe sólo un subconjunto de mensajes de los que se le enviaron.
v. Omisión de transmisión; esta falla ocurre cuando un componente envía un subconjunto de los mensajes que debería enviar.
vi. Omisión general; este tipo de falla incluye a las dos anteriores, es decir, de forma simultanea tanto en la recepción como en la transmisión.
vii. Fallas Bizantinas; este tipo de fallas surgen cuando un componente presenta un comportamiento arbitrario.

Un sistema tolerante a fallas debe ser capaz de detectar y/o corregir los errores antes de que se produzcan las averías y el usuario detecte su existencia. El factor clave para lograr este objetivo es integrar redundancia en el sistema.

Finalmente, la redundancia consiste en repetir aquellos datos o hardware de carácter crítico que se quiere asegurar ante los posibles fallos que puedan surgir por su uso continuo. La redundancia es otro elemento que se debe considerar en el almacenamiento, consiste en proveer al sistema de mecanismos para detectar y enmascarar los errores internos que se producen, evitando que el usuario pueda observar los efectos de las fallas. Se define la redundancia como los recursos adicionales que no serían necesarios en un sistema ideal.

Existen tres tipos de redundancia:

i. Redundancia de recursos físicos; consiste en la duplicación de componentes físicos en el sistema. En el caso del almacenamiento distribuido, se deben tener duplicados los datos almacenados en diversos dispositivos. Si existe un error en alguno de los dispositivos, se puede obtener la información de alguna copia. La ventaja principal de este tipo de redundancia es que permite un acceso rápido para almacenar y recuperar los datos, sin embargo, representa un costo mayor debido a que se duplican componentes físicos para una misma función.

ii. Redundancia en tiempo; consiste en la repetición de operaciones de cómputo o repetición de transmisión de datos en un lapso de tiempo determinado. Por ejemplo, cuando se transmite un archivo, si el receptor no notifica que se ha recibido en un determinado tiempo, el emisor debe enviarlo de nuevo hasta que se le notifique una recepción exitosa o hasta que se intente un número predefinido de eventos. La ventaja de emplear este tipo de redundancia consiste en que se asegura el resultado que se obtendrá, aunque la desventaja es que aumenta el número de operaciones a ejecutar o aumentan la transferencia de archivos introduciendo retardos en el sistema.

iii. Redundancia en información; consiste en técnicas de codificación específicas (información adicional al mensaje original). Por ejemplo, al agregar un bit de paridad a los datos que se envían a través de un canal se permite detectar y corregir errores.

Los tres tipos de redundancia mencionados se emplean en el diseño de sistemas tolerantes a fallas. Cada tipo se emplea en diferentes casos y son complementarios en una arquitectura tecnológica integral. La redundancia de recursos físicos se emplea para tolerar fallas permanentes tanto en software como en hardware. La redundancia en tiempo se emplea para tolerar fallas temporales en la transferencia o transmisión de archivos entre dispositivos y por último la redundancia en información se emplea para proteger la integridad de la información.

Métodos de codificación

Algoritmo de dispersión de la información (Information Dispersal Algorithm: IDA)

Este método de codificación genera información redundante. Produce fragmentos resistentes al borrado, que luego pueden almacenarse sobre diferente infraestructura, de modo que si se comprometen los datos originales en alguno de estas plataformas, se podrá recuperar siempre que sobreviva un mínimo de fragmentos, y se ha visto como una alternativa al cifrado de datos por secreto compartido [(12)].

Una instancia del algoritmo de dispersión de información (IDA) transforma un archivo de datos de tamaño | F | se transforma en “n” nuevos archivos llamados dispersos. Cada disperso es del tamaño | F |/ m donde n > m. Estos dispersos se almacenan en “n”

distintos nodos de almacenamiento que hacen parte de la zona de servicio de almacenamiento de información. Como parte de las propiedades del algoritmo IDA, se puede reconstruir el archivo original, siempre que cualesquiera “m” dispersos estén disponibles. La combinación de parámetros (n, m) tolera la falta de hasta (n − m) dispersos, evitando un incremento en el precio por un exceso de redundancia igual a (n − m) / m [(13)].

La tabla 4 muestra la cantidad de redundancia adicional para “n” servidores cuando requieren (al menos) “m” dispersos bajo un esquema tolerante a fallas [(13)]:

Tabla 4.- Combinación de parámetros de IDA y su redundancia asociada

n (servidores)	m=1	m=2	m=3	m=4	m=5	m=6
n=2	100%					
n=3	200%	50%				
n=4	300%	100%	33.3%			
n=5	400%	150%	66.70%	25%		
n=6	500%	200%	100%	50%	20%	
n=7	600%	250%	133.3%	75%	40%	16.7%

Por ejemplo, en implementaciones en IDA con parámetros (5, 3), esto significa que cada archivo se transforma en n = 5 dispersos y el archivo original estará disponibles siempre que cualesquiera m = 3 dispersos estén a la mano. Esto implica que el sistema hará uso de un 66.7% adicional por redundancia. Esta combinación (5,3) representa un buen balance entre fiabilidad (2 nodos pueden fallar), un espacio adicional para la recuperación (66.7%) y rendimiento (el sistema distribuye en cinco nodos el almacenamiento). La combinación (4, 3) es mejor que el anterior en cuanto a capacidad puesto que requiere menos espacio, pero sólo permite el sistema soportar la falta de un solo nodo. La combinación (6, 4) ofrece una fiabilidad similar a (5, 3) así como un mejor trato en cuanto a capacidad, sin embargo, en esta combinación el rendimiento se ve afectado por los costos de la distribución y nodos de cómputo adicionales.

Secreto compartido (Secret Sharing or Secret Splitting)

En este esquema, un valor secreto se distribuye en fragmentos entre un conjunto de participantes, de tal manera que sólo algunas coaliciones calificadas de los participantes pueden recuperar el valor secreto de sus acciones.

Un Esquema de Secreto Compartido, es un método que permite cifrar un archivo “S” en “n” piezas llamadas sombras S1, S2,… Sn, donde cada sombra por sí sola no revela información de “S”, pero “S” puede ser reconstruido totalmente si se reúne un cierto mínimo de sombras.

Por ejemplo, considere un umbral (k,n) de un secreto compartido, donde “k” es número aleatorio de sombras y “n” el número total de sombras, entonces con cualesquiera “k” de “n” sombras puede descifrarse la información secreta “S”. Por lo consiguiente, incluso si n-k sombras son destruidas por algún intruso, la información secreta “S” puede ser recuperada de las “k” sombras restantes. Sin embargo, si un intruso roba k-1 sombras, ningún tipo de información acerca de “S” puede ser obtenida. Esto significa que, el secreto compartido es incondicionalmente seguro, porque no está basado en suposiciones de deficiencias computacionales, como la factorización de enteros o el cálculo de algoritmos discretos, siendo por lo tanto apropiado para un almacenamiento de datos.

Así también, el concepto de secreto compartido puede ser entendido de otra forma como en el caso de una estructura de acceso con un umbral (k,n), se asume que cada sombra es igualmente importante, pero hay casos en los que se requiere que algunas sombras sean más importantes que otras. Por ejemplo: El presidente de una compañía quiere distribuir las sombras del secreto “S” a los directores de la compañía, de tal forma que “S” pueda ser descifrado si y sólo si dos vicepresidentes o más de cinco directores (exceptuando vicepresidentes) cooperan entre ellos. En tales casos, las sombras de los directores son menos importantes que la de los vicepresidentes. Desde esta forma, la decodificación no puede ser realizada mediante un umbral (k,n) del secreto compartido, una estructura general de acceso deberá ser introducida para lograr la seguridad deseada. Una estructura general de acceso consiste en grupos cualificados y grupos prohibidos. Un grupo calificado es aquel grupo de sombras que puede descifrar la información secreta o el secreto mientras que un grupo prohibido es aquel grupo de sombras que no pueden obtener ningún tipo de información del secreto.

Por otra parte, para realizar un eficiente y seguro almacén de datos, es necesario hacer el tamaño o medida de cada sombra, tan pequeña como sea posible.

Finalmente, es importante enfatizar que en el caso del secreto compartido, si no se reúne el mínimo de sombras es imposible reconstruir parcial o totalmente el archivo original. En tanto, con el IDA, es posible reconstruir parcialmente el archivo original, aun si no se reúne el mínimo especificado por el algoritmo.

Sistema de archivos BABEL

Conforme al trabajo realizado por el doctor Miguel López Guerrero sobre la construcción de una solución de almacenamiento de información denominado BABEL [(14)], este sistema en resumen se conforma básicamente de:

- Un Proxy el cual sirve para autenticar a los usuarios que quieren hacer uso del sistema Babel y es la única entrada/salida con los nodos de almacenamiento
- Cinco nodos de almacenamiento, en esta infraestructura será donde se almacene de forma distribuida los diferentes fragmentos o dispersos de los datos originales.
- Un firewall, para mantener por seguridad separada la zona de almacenamiento distribuido de segmentos o entidades externas a la institución, de acuerdo al tipo y procedimientos de almacenamiento que se utilizará y a la arquitectura de seguridad con la que se cuente podría no ser necesario utilizar este componente.
- Un sistema de monitoreo o consola, desde este equipo es donde se monitorea, supervisa y da seguimiento a los procesos de almacenamiento y recuperación de archivos.

Para el caso de Nacional Financiera dado que el almacenamiento será dentro de las instalaciones, así como la programación y ejecución de los procesos de almacenamiento y recuperación, no será necesario contar con un firewall.

En términos generales, los datos originales que están en el equipo del usuario son fragmentados y distribuidos en bloques entre los diferentes nodos de almacenamiento, siendo por lo tanto un almacenamiento virtual y de información redundante, donde en caso de falla o problemas con uno de estos nodos, el sistema Babel con por lo menos tres de los cinco nodos es capaz de restaurar la información o datos originales, haciendo de Babel un sistema de almacenamiento de información confiable (disponibilidad, tolerancia a fallas y escalabilidad). Adicionalmente Babel hace este almacenamiento

distribuido independientemente del sistema operativo e infraestructura tecnológica de almacenamiento que se esté o requiera utilizar.

En la siguiente figura se muestra el diseño conceptual del sistema Babel.

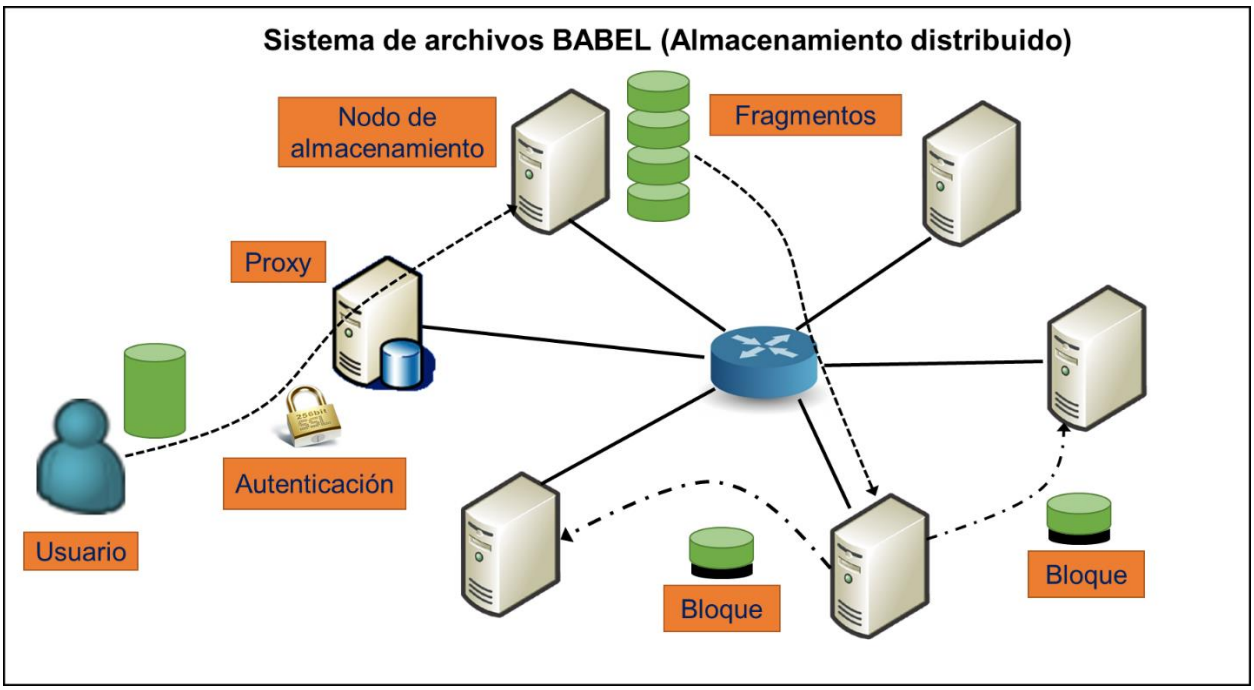


Figura 2.- esquema sistema de archivos Babel

Por último y de acuerdo a la publicación realizada por sus diseñadores del sistema Babel y conforme a la figura 2, señalamos el procedimiento de almacenamiento distribuido [(14)]:

1) El usuario conecta con la interfaz de la célula.
2) El servidor Proxy valida si el usuario está autorizado para ejecutar el almacenamiento.
3) Cuando el usuario presenta su archivo al proxy, éste crea un flujo entre su computadora y uno de los cinco nodos de almacenamiento. Este nodo es seleccionado por el proxy conforme un procedimiento de selección al azar, también graba esta operación en el registro actual y crea una nueva entrada en su base de datos (metadatos), con el fin de soportar la futura recuperación.

4) El nodo de almacenamiento empieza a recibir el flujo o secuencia del archivo. Para fomentar el balanceo de procesamiento, el sistema define un parámetro llamado unidad máxima de almacenamiento (UMA o MSU por sus siglas en ingles). Los archivos se dividen en tantas unidades de procesamiento, o fragmentos, según sea necesario para garantizar que la longitud de cada fragmento no exceda la UMA dada.

5) Cada nodo de almacenamiento mantiene una lista rotativa o cíclica llamada anillo o ring, con las identidades de los nodos de almacenamiento de información que colaboran con él. De acuerdo al orden que aparecen en el anillo, el nodo a cargo asigna a cada uno de los fragmentos resultantes a los otros nodos, a partir del último lugar señalado.

6) Para lograr la tolerancia a fallas y alta disponibilidad, los fragmentos se generan de forma redundante. El sistema soporta dos procedimientos diferentes; replicación simple o a través de un algoritmo de dispersión de información (IDA [15] por sus siglas en ingles). La replicación crea dos copias idénticas del fragmento llamado *bloques, para e*n el caso de IDA, se crean "n" unidades de almacenamiento diferentes, también llamados *bloques*, estos bloques se almacenan invocando un procedimiento denominado *oráculo*, el cual define el sitio final donde se guarda cada bloque. El oráculo está pensado para garantizar el balance de carga de almacenamiento.

7) Un nodo que ha sido requerido para procesar o almacenar una unidad de información (archivo, fragmento o bloque), confirma la realización de esta tarea a la fuente inmediata que ha requerido su capacidad.

8) Tanto el proxy como los nodos de almacenamiento mantienen una réplica de sus metadatos en un par de nodos que integran su anillo. Siendo la consistencia de los metadatos un aspecto muy importante que cuida el sistema Babel.

CAPITULO III.- Buscando una estrategia alterna

Como se mencionó en el capítulo I, la arquitectura y procedimientos actuales en NAFIN cumplen con la normatividad aplicable al almacenamiento, recuperación y planes de continuidad, sin embargo, en la búsqueda de hacer más eficiente los recursos técnicos y económicos, ha llevado a investigar sobre la posibilidad de utilizar otras alternativas tecnologías como servicios de almacenamiento en la nube.

Los servicios de almacenamiento en la nube desde el punto de vista económico los hace más eficientes que el almacenamiento tradicional ya que una de sus funciones es pagar por la cantidad real de datos almacenados o procesamiento utilizado, mientras que en la forma tradicional se requiere adquirir o contar con una infraestructura tanto de procesamiento como de capacidad de almacenamiento mayor a la utilizada, esto debido principalmente para soportar un uso mayor en determinados horarios (picos). Otra variable importante a considerar es que la disponibilidad del servicio de almacenamiento en la nube es 7 X 24, es decir los siete días de la semana durante las 24 horas del día, para igualar este soporte se requeriría contar con recursos humanos durante varios turnos para cubrir esta misma ventana, y por lo menos en el caso de NAFIN las prestaciones laborales hacen que se encarezca esta alternativa.

Sin embargo, un punto muy importante también a tomar en cuenta es lo relacionado con la seguridad de los datos almacenados, mientras que los datos almacenados en equipos instalados en las instalaciones de NAFIN se perciben con una mayor seguridad de custodia y acceso, que si estuvieran en equipos instalados en centros de cómputo del proveedor del servicio en la nube, la custodia de estos datos ya no estarían a la “vista” del personal responsable de NAFIN, estarían fuera de nuestras instalaciones e incluso podrían estar fuera del territorio nacional, y aunque estuvieran firmados contratos o acuerdos de no acceso o confidencialidad de la información, así como bitácoras de acceso de usuarios permitidos o no autorizados, el riesgo de intrusiones no permitidas es mayor que el tener los equipos y datos en las instalaciones propias.

Hasta el momento, este factor de riesgo ha pesado más sobre las ventajas económicas, por esta razón la alta dirección de NAFIN ha decidido, por el momento, no contratar servicios de almacenamiento en la nube.

Utilizando la tecnología de almacenamiento distribuido y en especial el sistema BABEL, los datos originales se fragmentarían en diferentes partes, cada una de las partes estarían alojadas en diferentes dispositivos de almacenamiento, y si uno o varios de los dispositivos estuvieran bajo el concepto de “Nube” se reducirían los costos en lugar de tenerlos bajo una plataforma tradicional, pero además, manteniendo acotados los niveles de riesgo en la seguridad, ya que en caso de que personal no autorizado tuviera acceso a este almacenamiento para robar, alterar o borrar la base, no le serviría de mucho ya que solo estaría accediendo a un fragmento de la base original, tendría que acceder a la vez a las otros plataformas distribuidas en diferentes sitios, por lo que la probabilidad de robar o alterar la información se reduce considerablemente.

El esquema de almacenamiento distribuido haciendo uso del sistema Babel esta esquematizado en la figura 3, donde se podrían contratar empresas que actualmente ofrecen ya servicios en la nube en especial infraestructura, conocido como IaaS (Infrastructure as a Service).

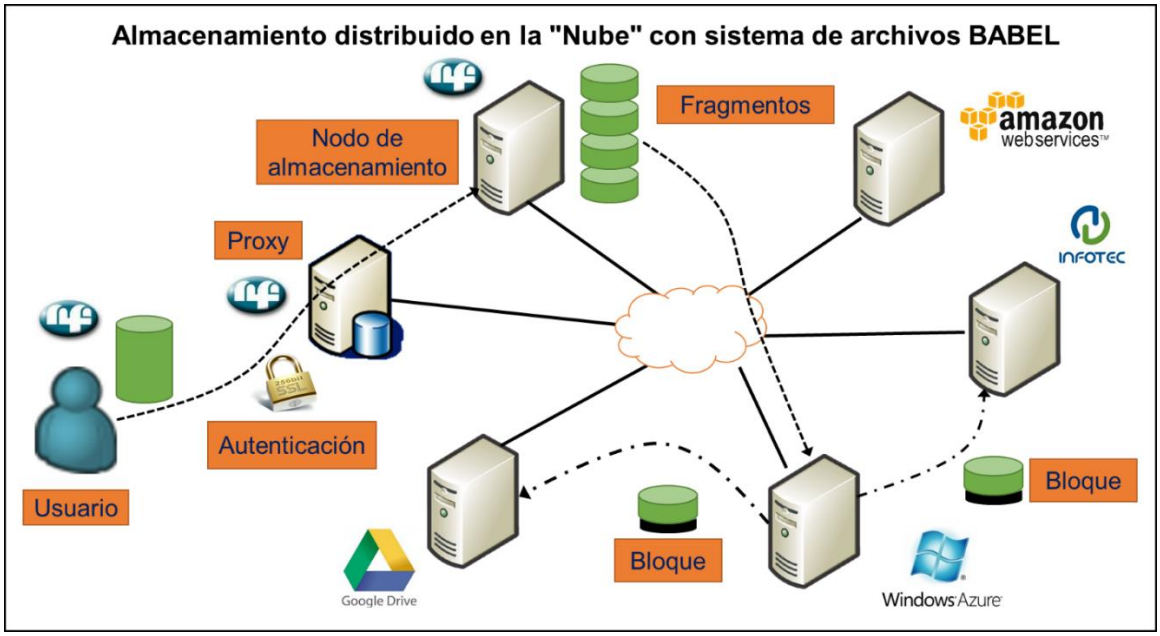


Figura 3.- esquema almacenamiento distribuido en la “nube” con sistema de archivos Babel

Por otra parte, suponiendo que el intruso afectó o dañó un fragmento, el sistema Babel requiere por lo menos tres de los bloques para recuperar la base original, por lo tanto, con los bloques restantes se podría recuperar los datos sin mayores dificultades.

En los siguientes capítulos se describe la prueba de concepto y los resultados obtenidos del sistema Babel con la arquitectura e infraestructura de Nacional Financiera, así como el análisis costo - beneficio de esta alternativa.

CAPITULO IV.- Prueba de concepto del sistema BABEL con la arquitectura e infraestructura de Nacional Financiera.

La prueba de concepto se realizó en diferentes etapas, partiendo de la arquitectura 100% interna y en una misma red local, hacia una distribución de los nodos en diferentes segmentos de red local e incluso en diferentes dominios.

Primera etapa

La arquitectura de la primera etapa se muestra en la figura 4, con cinco equipos físicos de cómputo personal como nodos de almacenamiento, un equipo de Proxy y un equipo de consola, la red local se configuró a 1 Gbps.

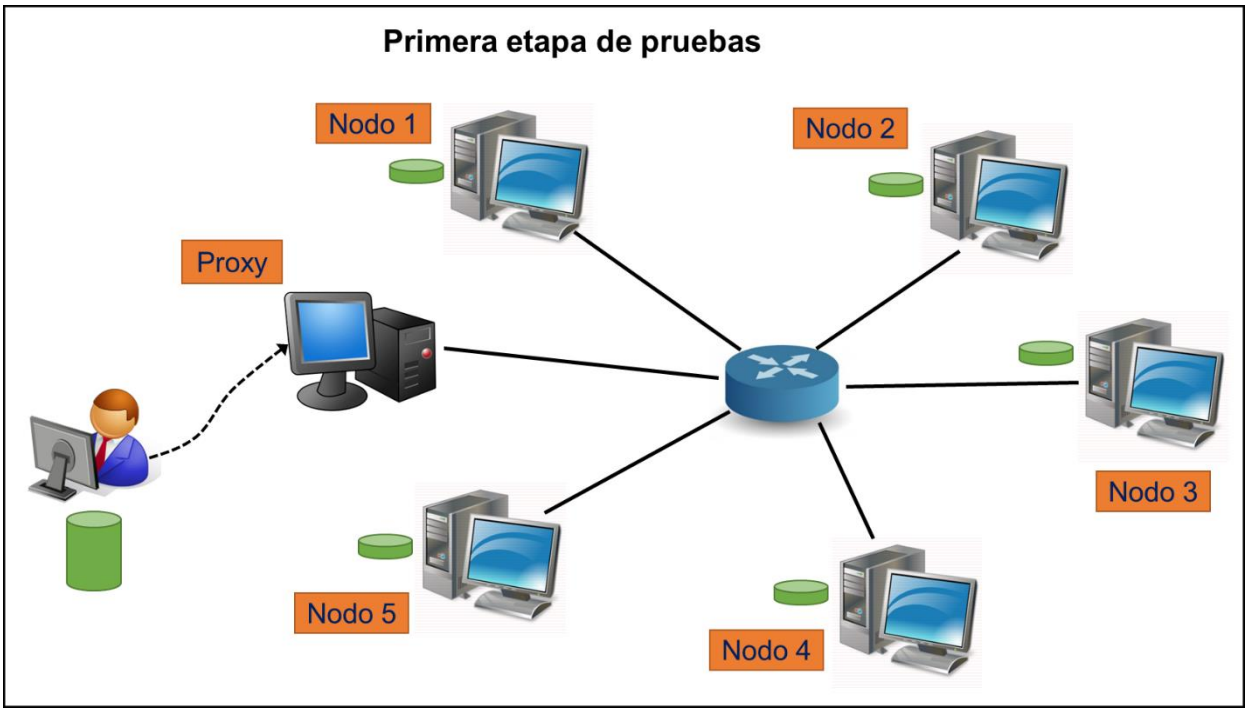


Figura 4.- primera etapa de pruebas con equipos dedicados y en una misma LAN

Para esta primera etapa fue muy importante primero definir el tamaño máximo de archivos a respaldar y distribuir, así como también definir el tamaño óptimo de la UMA de acuerdo al ambiente y desempeño de los equipos e infraestructura involucrada.

Para estas pruebas se utilizó la herramienta de MD5Summer, la cual compara el archivo original contra el respaldado, distribuido y recuperado, proporcionando la

validación si el archivo recuperado es exactamente igual al originalmente respaldado, en otras palabras, evalúa la integridad de los datos recuperados.

De acuerdo a las pruebas realizadas se obtuvieron las siguientes conclusiones (el detalle de cada prueba está en el Anexo “II”:

a) **Definición tamaño máximo de archivo:** se realizaron diferentes pruebas con archivos de 800, 1,000 y 2,000 Mb, donde el archivo recuperado en el 100% de las muestras, correspondieron al original y por lo tanto garantiza la integridad de datos, fue el de 800 Mb, por lo tanto, el archivo a utilizar durante las pruebas será máximo de **800 Mb**.

b) **Definición del tamaño de UMA óptimo:** haciendo uso del archivo de 800 Mb, se realizaron pruebas de respaldo, distribución y recuperación con UMAs de 128, 256 y 512, de acuerdo a las pruebas se pudo observar que conforme se incrementaba el tamaño de UMA, el tiempo de los procesos eran mayores y el número de pruebas fallidas al comparar el archivo original contra el recuperado se incrementaba, por lo tanto la UMA más consistente y que asegura la integridad de los datos correspondían a la de **128 unidades**, en la tabla 5 se muestran los tiempos promedio y número total de fallas de las 10 pruebas realizadas con cada UMA:

Tabla 5.- selección UMA de acuerdo al tiempo de proceso y pruebas fallidas

UMA	Tiempo proceso	Pruebas fallidas
128	**00:54.09**	**NINGUNA**
256	01:02.42	UNA
512	02:22.61	DOS

c) **Comparación con ambientes de 5, 4 y 3 nodos:** Durante esta etapa se confirmó que para la recuperación de los datos almacenados de forma distribuida se requiere contar con un mínimo de tres nodos activos, ya que en caso de tener uno o dos nodos activos simplemente no es factible la recuperación de los datos, ya que no solo no se efectúa la distribución sino también provoca inconsistencia en algunos de los nodos teniendo que reinstalar y sincronizar estos con el sistema integral.

La siguiente tabla muestra los tiempos promedios con 5, 4 y 3 nodos:

Tabla 6.- tiempos promedios de distribución y recuperación con diferentes nodos de distribución

Nodos	Almacenamiento	Recuperación
5	54.09	1:06.99
4	3:10.26	2:40.22
3	3:27.13	2:09.72

Durante todas estas pruebas tanto la distribución como la recuperación con el sistema Babel fueron exitosas, concluyendo que a menor cantidad de nodos de distribución los tiempos de los diferentes procesos de almacenamiento y recuperación se incrementan.

Segunda etapa

La arquitectura de la segunda etapa se muestra en la figura 5, con cuatro equipos físicos de cómputo personal y un equipo virtual en el centro de cómputo principal como nodos de almacenamiento, un equipo de Proxy y un equipo de consola, la red local se configuró en los equipos dedicados a 1 Gbps y el servidor físico para máquinas virtuales a 10 Gbps.

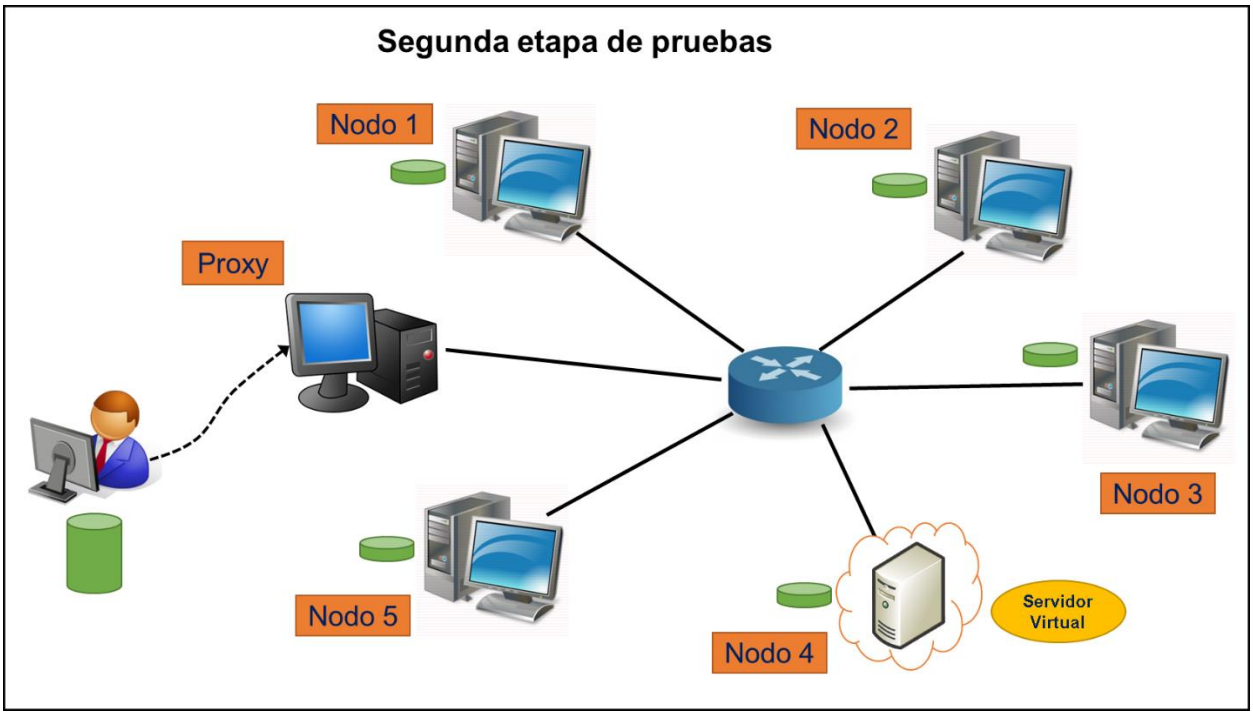


Figura 5.- segunda etapa de pruebas con equipos dedicados y virtuales en una misma LAN

En la ejecución de las pruebas de esta etapa se mantuvo constante los valores de tamaño de archivo y UMA utilizados en la primera etapa, es decir 800 Mb y 128 unidades respectivamente, así como el ambiente de prueba fue exclusivamente con cinco nodos

El detalle de los resultados de esta prueba se señala en el “Anexo II”.

La siguiente tabla muestra los tiempos promedios obtenidos en la primera y segunda etapa:

Tabla 7.- tiempos promedios de distribución y recuperación obtenidos en la primera y segunda etapa de pruebas

Etapa	Almacenamiento	Recuperación
Primera	54.09	1:06.99
Segunda	50.44	1:08:80

De acuerdo a los resultados de esta prueba se puede concluir que es factible contar con ambientes tanto físicos como virtuales, ya que los tiempos de almacenamiento y recuperación son muy similares. Cabe señalar que para esta prueba tanto los nodos físicos como el virtual estuvieron en la misma ubicación física y su conectividad fue a través de una red local.

Tercera etapa

La arquitectura de la tercera etapa se muestra en la figura 6, con tres equipos físicos de cómputo personal, un equipo virtual en el centro de cómputo principal y un equipo virtual en el centro de cómputo de contingencia como nodos de almacenamiento, un equipo de Proxy y un equipo de consola, la red local se configuró en los equipos dedicados a 1 Gbps, el servidor físico para máquinas virtuales en el CCP a 10 Gbps, el servidor físico para máquinas virtuales en el CCC a 1 Gbps y el enlace de comunicación entre centros de cómputo a 10.2 Mbps.

Cabe señalar que con esta prueba se estaría simulando el contar con servidores virtuales de almacenamiento en “la nube” a través del servidor virtual creado en el CCC, si bien los medios y protocolos de comunicación no serían iguales, ya que por ejemplo en el caso de un servicio en la nube se utiliza como medio de comunicación el Internet y en la prueba fue a través de la red privada virtual MPLS de Nacional Financiera, lo

importante era validar la funcionalidad, integridad e incremento del tiempo de almacenamiento y recuperación.

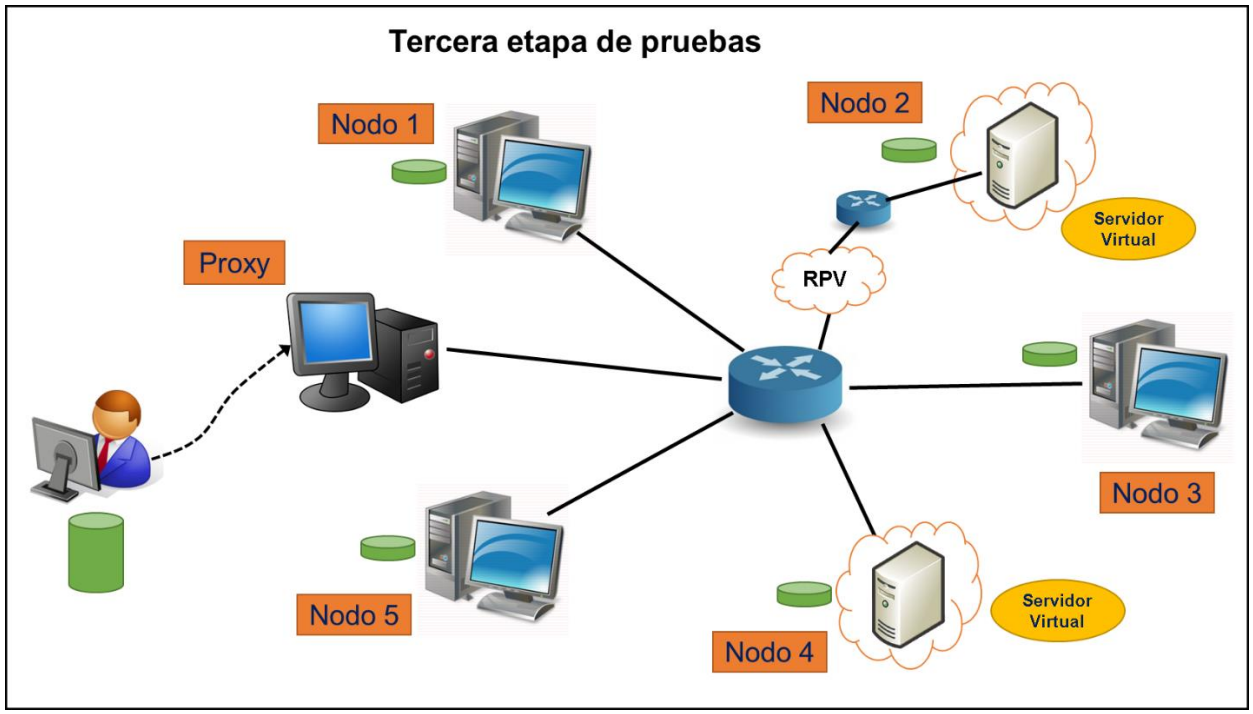


Figura 6.- tercera etapa de pruebas con equipos dedicados y virtuales en diferentes LAN

El detalle del resultado de esta prueba se describe en el “Anexo II”.

La siguiente tabla muestra los tiempos promedios obtenidos en la primera, segunda y tercera etapa:

Tabla 8.- tiempos promedios de distribución y recuperación obtenidos en la primera, segunda y tercera etapa de pruebas

Etapa	Almacenamiento	Recuperación
Primera	54.09	1:06.99
Segunda	50.44	1:08:80
Tercera	1:06.80	3:33.91

Conclusión de la etapa de pruebas

De acuerdo con las pruebas realizadas se concluye que:

- El concepto de almacenamiento distribuido fue funcional en cualquiera de las tres etapas.

- Es factible contar con células en máquinas virtuales, los tiempos de almacenamiento y recuperación en una misma infraestructura de red local fueron muy similares.
- La prueba con una célula remota (simulando servicio en la nube) y en ambiente virtual fue satisfactoria, como era de esperarse al estar en diferentes ubicaciones físicas se incrementó el tiempo de almacenamiento en un 27.8% y en 215.3% el tiempo de recuperación.
- Se debe considerar un incremento en los tamaños de bases de datos a ser almacenadas de forma distribuida en un 67 %.
- Se requiere hacer ajustes al sistema Babel para que pueda procesar bases de datos de mayor tamaño, por lo menos de 20 GB, dado que en Nacional Financiera el 80% de los datos almacenados se encuentran en esta capacidad, en la siguiente tabla se muestra la cantidad de servidores y su capacidad:

Tabla 9.- cantidad de servidores instalados y su capacidad almacenada

Capacidad almacenada (GB)	Cantidad de servidores
1 a 10	41
11 a 20	10
21 a 30	4
31 a 40	6
41 a 50	1
51 a 60	0
61 a 70	0
71 a 80	1

CAPITULO V.- Diseño conceptual de la arquitectura de infraestructura distribuida en Nacional Financiera con almacenamiento distribuido.

Con base en las pruebas de concepto detalladas en el Capítulo IV, donde se comprobó la funcionalidad y viabilidad de utilizar sistemas de almacenamiento distribuido como una alternativa al almacenamiento tradicional, sobre todo utilizando sitios físicos alternos a donde se encuentre la base de datos original, se propondría la arquitectura mostrada en la figura 7.

Cabe señalar que, por la complejidad de contratación temporal de servicios de almacenamiento en la nube, simulamos este concepto a través de poner una célula en el centro de cómputo alterno de NAFIN, sin embargo para fines de este estudio y de las pruebas conceptuales se asumirán similares resultados.

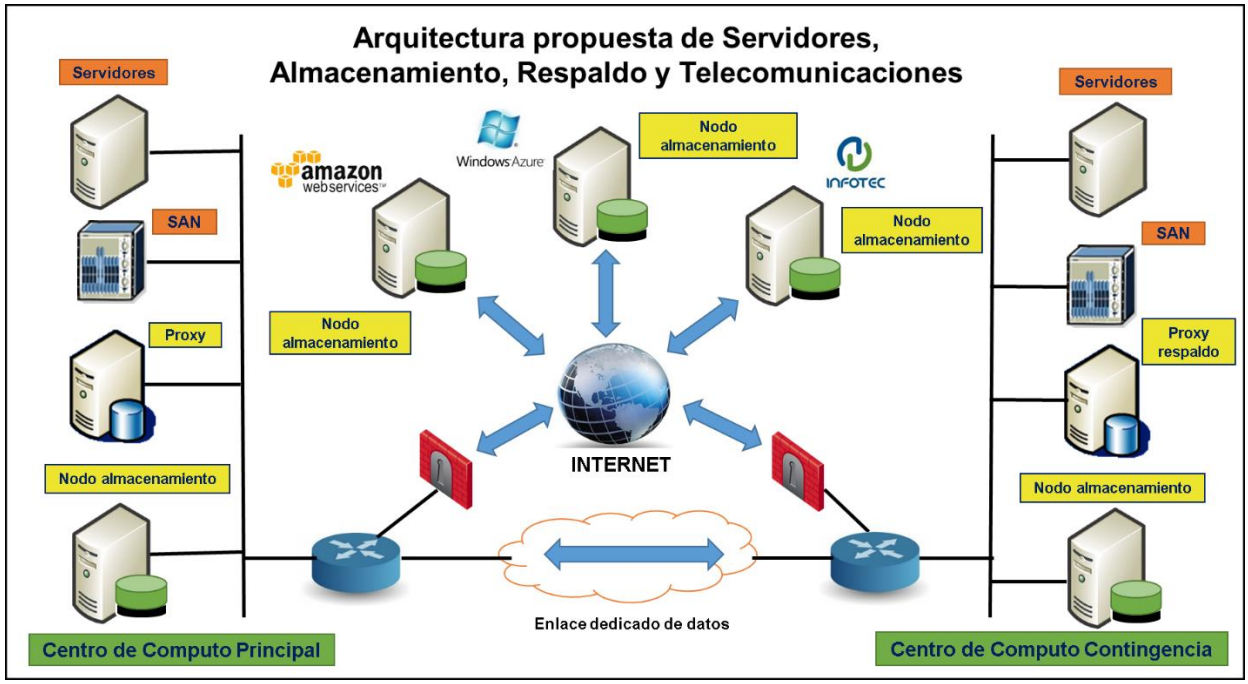


Figura 7.- arquitectura propuesta de servidores, almacenamiento, respaldo y telecomunicaciones

Cambios a la arquitectura actual

En la configuración propuesta se mantiene el concepto de plataformas duplicadas en ambos centros de cómputo para garantizar los procesos, procedimientos, capacidades y tiempos de recuperación en caso de un DRP del CCP señalados en el capítulo I.

En relación a la implantación del sistema de almacenamiento distribuido Babel, se propone instalar un Proxy en el CCP para realizar el almacenamiento distribuido y un segundo Proxy de respaldo, para en caso de un DRP del CCP poder recuperar los datos distribuidos en las células de almacenamiento aun activas, la distribución de estas células seria, una en el CCP, otra en el CCC y tres más (para hacer el total de cinco) con diferentes proveedores de servicios de almacenamiento en la Nube.

Por lo que los cambios a la arquitectura actual quedarían:

(1) **Servidores. -** se mantienen los servidores instalados (físicos y virtuales) bajo un esquema de alta disponibilidad en el CCP y en el CCC con equipos no redundantes.

(2) **Almacenamiento.** - con la implementación del almacenamiento distribuido con Babel se incrementaría en 67% la capacidad de las células distribuidas instaladas en los centros de cómputo principal y de contingencia, destinadas para respaldar bases de datos en la SAN.

(3) **Respaldo. -** los respaldos en cinta que actualmente se realizan serían sustituidos por el sistema de almacenamiento distribuido, es decir, al término de los procesos de cierre de línea y Batch, se ejecutarían los procedimientos de almacenamiento distribuido hacia las cinco células definidas, por lo tanto la infraestructura de librerías estarían temporalmente solo para recuperar información de las cintas realizadas hasta la fecha de liberación del sistema de almacenamiento distribuido.

(4) **Telecomunicaciones.**- se mantendría el enlace de datos dedicado entre los centros de cómputo, para continuar con el espejeo de datos y ahora para la generación de una de las células del almacenamiento distribuido, en lo que respecta a los enlaces o conectividad hacia Internet, estos tampoco habría algún cambio, ya que si bien la generación de las otras tres células del almacenamiento distribuido sería a través de esta comunicación con los proveedores externos, su ocupación estaría siendo en un horario diferente (por la noche) al del servicio de navegación y consulta por los clientes de NAFIN y usuarios internos.

(5) **Almacenamiento distribuido.** - se contaría con dos Proxies uno en cada centro de cómputo, el del CCP sería el que lleve el control del almacenamiento distribuido y el del CCC solo entraría en funcionamiento en el caso de activarse el DRP. Las células de almacenamiento estarían en diferentes ubicaciones físicas, uno el CCP, otra en el CCC y cada una de las tres restantes con diferentes proveedores de almacenamiento en la nube.

Por último, con respecto a la seguridad, se mantendrían las políticas de acceso a los sistemas y perfiles de usuarios con los que se cuenta actualmente, y como se comentó en el capítulo III, aunque algunos datos estarían almacenados en sitios diferentes a NAFIN, dada la funcionalidad de Babel, un intruso o personal no autorizado tendría que tener acceso y alterar de forma simultánea en por lo menos tres de las cinco células de almacenamiento para provocar algún daño, lo cual es menos probable que como estamos hoy en día, ya que los datos los tenemos íntegros en dos ubicaciones, y si algún intruso accediera a uno de estos equipos podría causar daño a la institución.

Aspectos económicos

Como lo comentamos, la alternativa propuesta además de ser funcional y mantener por lo menos los aspectos de seguridad de información actual, también debería proporcionar benéficos económicos, por lo que realizamos el siguiente análisis de costos al implementar la solución de almacenamiento distribuido Babel contra los costos actuales.

Precios actuales

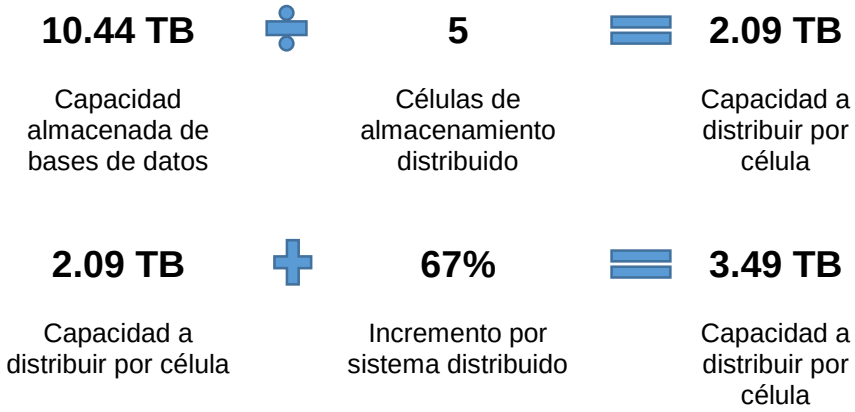
De los precios anuales señalados en la tabla 3 y como lo señalamos en este capítulo, se mantienen los precios de servidores tanto físicos como virtuales, Internet y el del enlace de datos, por lo tanto, la comparación la haremos con los precios de la SAN y Librerías.

Tabla 10.- Precios anuales de la infraestructura almacenamiento y respaldo actual (USD)

Infraestructura	C. C. P.	C. C. C.	Total
SAN	$ 38,615.00	$ 24,690.00	$ 63,305.00
Librería	$ 33,970.00	$ 33,070.00	$ 67,040.00
Total	**$ 72,585.00**	**$ 57,760.00**	**$ 130,345.00**

Precios con almacenamiento distribuido

Almacenamiento.- como se ha señalado, el uso de almacenamiento distribuido a través de Babel implica un incremento del 67% del tamaño original de las bases a ser distribuidas, en el caso de NAFIN los servidores y bases de datos actuales ocupan 12 TB del almacenamiento en las SAN del CCP, sin embargo el alcance para almacenar de forma distribuida sería de 10.44 TB, ya que los servidores y base de datos correspondiente a correo electrónico seguiría la estrategia actual de espejeo de un centro de cómputo al otro, por lo tanto para la creación de las células de almacenamiento distribuido se requiere:



Para cada célula que estará en los centros de cómputo de NAFIN se requerirá de 3.49 TB, de acuerdo al contrato actual de servicios, el precio anual por incremento en la SAN es de $ 561.44 dólares por cada 500 Gb adicionales, por lo tanto, se requeriría un gasto anual de $ 3,930.08 dólares por centro de cómputo.

Almacenamiento en la nube.- conforme a la nueva arquitectura propuesta se tendrán cinco células de almacenamiento distribuido, dos en los centros de cómputo de NAFIN y las tres restantes con diferentes proveedores de servicio de almacenamiento en la nube, para fines de costeo, cotizamos en la página de internet de la empresa Microsoft el servicio AZURE [16], el cual tomaremos como referencia para el cálculo de los tres servicios, por lo tanto el precio anual estimado en cada célula distribuida en la nube sería:

a) Por almacenamiento. - se cobra por GB almacenado, para este análisis será la capacidad de cada célula almacenamiento distribuido, siendo el cálculo:

$ 0.024		3,487		$ 82.29	$ 987.52
Precio GB/mes		GB almacenado por célula		Precio mensual	Precio anual

b) Por transmisión de datos. - se cobra cuando se recuperan los datos almacenados, actualmente solo en los simulacros de BCP o DRP se restablecen los datos respaldados, para fines de costeo asumiremos una recuperación trimestral del total de datos almacenados:

$ 0.083		3,487		$ 289.42	$ 868.26
Precio GB transmitido		GB almacenado por célula		Precio por evento	Precio anual (tres eventos)

c) Por servidor Linux.- se cobra por las horas de uso del tipo de servidor creado, es decir cuantas horas es utilizado el número de núcleos de procesamiento, la cantidad de GB de memoria y el espacio en disco duro asignados, para este análisis se formarían servidores de 4 núcleos, 14 GB de memoria y 200 GB de almacenamiento (configuración similar a la utilizada en las pruebas), asumiendo que se utilizarán estos recursos durante el proceso de almacenamiento distribuido (de lunes a viernes durante 12 horas), siendo el cálculo:

$ 0.376		240		$ 90.24	$ 1,082.88
Precio hora		Horas al mes de uso		Precio mensual	Precio anual

Por lo tanto, el costo de almacenamiento distribuido a través de un proveedor de este servicio en la nube es de $ 2,938.66 dólares anuales

Por último, y como lo señalamos en este capítulo, asumiremos similares costos de los otros dos proveedores de almacenamiento en la nube, por lo tanto, el costo anual por el almacenamiento en la nube de las tres células distribuidas sería de $ 8,815.98 dólares.

Soporte a BABEL.- el sistema Babel hace uso, entre otros, de CentOS como sistema operativo y Python como lenguaje de programación, los cuales en NAFIN no contamos con la experiencia suficiente para darle el soporte adecuado, por lo tanto contrariamos un recursos técnico que conozca de estos sistemas y complementarlo con

la capacitación necesaria, con lo que cualquier problema podrá ser atendido con oportunidad, para este análisis consideraríamos un sueldo integral aproximado de $1,000.00 dólares mensuales, por lo tanto, el costo anual estimado sería de $ 12,000.00 dólares.

Presupuesto final. - de acuerdo a los costos por almacenamiento tradicional, por el almacenamiento en la nube y por el soporte al sistema Babel, tendríamos finalmente los costos finales por la nueva arquitectura propuesta, en la tabla 11, se muestra el cálculo final.

Tabla 11.- Precios anuales de la infraestructura almacenamiento y almacenamiento distribuido

Infraestructura	Precio anual (USD)
SAN C.C.P.	$ 38,615.00
SAN C.C.C.	$ 24,690.00
Incremento SAN por célula en CCP y CCC	$ 7,860.16
Almacenamiento distribuido en la NUBE	$ 8,815.98
Soporte sistema Babel	$ 12,000.00
Total arquitectura propuesta	**$ 91,981.14**

Comparación de alternativas

De acuerdo a las pruebas funcionales, las cuales fueron satisfactorias para ambas alternativas (Actual y Almacenamiento distribuido con Babel) y de acuerdo a la siguiente comparación de precios anuales señalados en la tabla 12:

Tabla 12.- Precios anuales estimados para cada alternativa

Arquitectura	Precio anual (USD)
Actual	$ 130,345.00
Almacenamiento distribuido	$ 91,981.14
Diferencia ($)	**$ 38,363.86**
Diferencia (%)	**29.4 %**

Por lo tanto, el hacer uso del almacenamiento distribuido con Babel mantendría la funcionalidad de respaldo y recuperación esperada, cumpliendo con las políticas y lineamentos de las diferentes entidades reguladoras, así como mantener por lo menos los niveles de seguridad de información actual, pero con un presupuesto anual 29.4 % menor que actualmente, es decir $ 38,363.86 dólares menos.

Incluso, si se mantuviera temporalmente el servicio de la librería más económica para en caso de recuperar alguna de las cintas previamente grabadas, la solución de almacenamiento distribuido seguiría siendo más económica en un 4.1 %, como se muestra en la tabla 13:

Tabla 13.- Precios anuales estimados para cada alternativa con recuperación de cintas

Arquitectura	**Precio anual (USD)**
Actual	$ 130,345.00
Almacenamiento distribuido	$ 91,981.14
+ librería	$ 33,070.00
Diferencia ($)	**$ 5,293.86**
Diferencia (%)	**4.1 %**

Conclusión

En relación a las pruebas en laboratorio y comparando la arquitectura actual con la de almacenamiento distribuido, concluyo que es muy viable diseñar una arquitectura alterna para el almacenamiento, recuperación y respaldos de datos utilizando el almacenamiento distribuido con el sistema de archivos Babel, siendo soportada por servicios de infraestructura en la nube, con lo que se tendrían menores niveles de riesgo en temas de seguridad en el acceso de la información, pero siendo una solución más eficiente, es decir con menores costos de infraestructura.

La evaluación costo beneficio resultó satisfactoria, con una reducción inicial en relación al presupuesto actual del 4% con una temporalidad entre dos a tres años, para aumentar en 29% después de este periodo, con esto se hace eficiente el presupuesto asignado en Nacional Financiera para esta temática y por otra parte también apoyaría al poder ejecutivo del gobierno federal en relación a su Estrategia Digital Nacional en materia de tecnologías de la información y comunicaciones, y en la de seguridad de la información, en el cual entre sus objetivos está el "*favorecer el uso del cómputo en la nube para el aprovechamiento de la economía de escala, eficiencia en la gestión gubernamental y estandarización de las TIC, teniendo en consideración la seguridad de la información*" [(5)].

A través de esta arquitectura alterna se aprovecharía la tendencia tecnológica de servicios en la nube, manteniendo los niveles de seguridad de información, ya que se tendría que vulnerar por lo menos en tres sitios diferentes para provocar un problema crítico con la información almacenada, menos probable que ahora que todos los datos están concentrados en un solo lugar.

Adicionalmente desde el punto de vista de riesgo operativo este se reduciría, mejorando la seguridad con esta arquitectura alterna, dado que actualmente tenemos todos los datos en dos infraestructuras (normal y respaldo), y en caso de que una persona quisiera vulnerar nuestra información tiene que acceder una de las dos instancias (50%) mientras que con la arquitectura alterna tendría que vulnerar tres de las cinco instancias (60%), haciéndolo menos probable y por lo tanto más seguro.

Por último, agradecer todo el apoyo al Dr. Ricardo Marcelin y su equipo, sin el cual no hubiera sido posible realizar las pruebas de laboratorio con las que se pudo validar la factibilidad técnica y económica de esta arquitectura alterna, así como al Dr. Ramón Reyes por sus comentarios en la segunda revisión.

Bibliografía

(1) México, Normatividad de la Comisión Nacional Bancaria y de Valores “Disposiciones de carácter general aplicables a las instituciones de crédito (Circular Única de Bancos)”, publicada en el diario oficial de la federación el 2 de diciembre del 2005.

(2) México, Normatividad de la Comisión Nacional Bancaria y de Valores “Anexo 67, requerimientos mínimos del plan de continuidad de negocio, de la Circular Única de bancos”, publicada en el diario oficial de la federación el 2 de diciembre del 2005.

(3) México, Normatividad de la Comisión Nacional Bancaria y de Valores “Anexo 52, lineamientos mínimos de operación y seguridad para la contratación de servicios de apoyo tecnológico, de la Circular Única de bancos”, publicada en el diario oficial de la federación el 2 de diciembre del 2005.

(4) México, Normatividad del Banco de México “Circular 4/2012 aplicable a las Instituciones de Crédito, Casas de Bolsa, Sociedades de Inversión, Sociedades Financieras de Objeto Limitado y a la Financiera Rural (Operación Derivados)”, publicada en el diario oficial de la federación el 15 de junio del 2012.

(5) México, Normativa de la Secretaria de Gobernación y de la Secretaria de la Función Pública “ACUERDO que tiene por objeto emitir las políticas y disposiciones para la Estrategia Digital Nacional, en materia de tecnologías de la información y comunicaciones, y en la de seguridad de la información, así como establecer el Manual Administrativo de Aplicación General en dichas materias”, publicado en el diario oficial de la federación el 8 de mayo del 2014.

(6) México, Nacional Financiera 1934-1984 / Medio siglo de banca de desarrollo / Testimonio de sus directores generales. Nacional Financiera (1984).

(7) México, La transformación lograda por Nacional Financiera. Nacional Financiera (2008).

(8) Portal de Nacional Financiera, http://www.nafin.com/portalnf/content/home/home.html.

(9) Becerril Sierra Israel y Pichardo Flores Lorena, “Modelo de Sistema de Gestión de Archivos Electrónicos en la nube como medida técnica de seguridad para la protección de información clasificada como reservada y confidencial”, Tesis presentada para obtener la Maestría en Gestión de Innovación de las TIC en INFOTEC, 2014.

(10) Pichardo Lorena, “Informática I: Serie basada en competencias y valores” para Bachillerato de la SEP, 2011, IURE editores, México (http://www.gandhi.com.mx/index.cfm/id/Producto/dept/libros/pid/466806),

(http://www.elsotano.com/libro-informatica-i-serie-basada-en-competencias-y-valores-bachillerato-10369409).

(11) Quezada Naquid Moises, “Evaluación de desempeño de un Sistema de Almacenamiento Distribuido”, 2007.

(12) Portal de Techtarget, http://searchnetworking.techtarget.com/Information-dispersal-algorithms-Data-parsing-for-network-security

(13) J.L.Gonzalez, Jesus Carretero Perez, Victor Sosa Sosa, Juan F. Rodriguez Cardoso, Ricardo Marcelín Jiménez, “An approach for constructing private stprage services as a unified faul-toulerant system”, publicado 16 marzo 2013 en Journal of Systems and Software.

(14) López Guerrero Miguel, Marcelín Jiménez Ricardo y Quezada Naquid Moisés, “The Babel File System”, en proceso de publicación.

(15) M. O. Rabin, "dispersión eficiente de información para la seguridad, balanceo de carga y tolerancia de fallas" diario de la ACM, vol. 36, págs. 335-348, abril de 1989, doi:10.1145/62044.62050.

(16) Portal del servicio AZURE de Microsoft, http://azure.microsoft.com/es-es/pricing/

ANEXO “I”. - Normatividad aplicable a bancos relacionada con plan de continuidad de negocio y respaldo de información.

Normativa de la CNBV relacionada con el plan de continuidad de negocio

Para la CNBV se entenderá como Plan de Continuidad de Negocio y Contingencia Operativa como [(1)]:

- ***Contingencia Operativa:*** *a cualquier evento fortuito que dificulte o inhabilite a una Institución a prestar sus servicios o realizar sus procesos, cuya actualización derive en daño o pérdida para sus clientes, para el público en general, para sus contrapartes o para la Institución misma.*
- ***Plan de Continuidad de Negocio:*** *al conjunto de estrategias, procedimientos y acciones a que hace referencia el Artículo 164 Bis de estas disposiciones que permitan, ante la verificación de Contingencias Operativas, la continuidad en la prestación de los servicios o en la realización de los procesos críticos de las Instituciones, o bien su restablecimiento oportuno, así como la mitigación de las afectaciones producto de dichas Contingencias.*

Los artículos e inciso relacionados con la creación, autorización, responsabilidades, operación y evaluación del plan de continuidad de negocio son [(1)]:

- ***Artículo 71.-*** *El comité de riesgos, para el desarrollo de su objeto, desempeñará las funciones siguientes:*

  VII. Informar al Consejo, cuando menos una vez al año, sobre el resultado de las pruebas de efectividad del Plan de Continuidad de Negocio.

- ***Artículo 78.-*** *Las Instituciones deberán contemplar en el Marco para la Administración Integral de Riesgos, cuando menos, los aspectos siguientes:*

  IX. Los planes de acción y de contingencia para restablecer la operación de la Institución en los procesos de negocio clasificados como críticos de acuerdo con el Análisis de Impacto al Negocio al que hace referencia la fracción I del Anexo 67 de estas disposiciones, en caso de presentarse eventos fortuitos o de fuerza mayor, quedando incluidos el Plan de Continuidad de Negocio así como el Plan de Financiamiento de Contingencia.

- ***Artículo 154.-*** *El Comité de Auditoría deberá proponer para aprobación del Consejo, el Sistema de Control Interno que la propia Institución requiera para su adecuado funcionamiento, así como sus actualizaciones.*

  IV. El Plan de Continuidad de Negocio, el cual deberá ser sometido regularmente a pruebas de funcionamiento y hacerse del conocimiento del personal.

- ***Artículo 156.-*** *El Comité de Auditoría, en el desarrollo de sus funciones, deberá, por lo menos, desempeñar las actividades siguientes:*

  VI. Informar al Consejo, cuando menos una vez al año, sobre la situación que guarda el Sistema de Control Interno de la Institución. El informe deberá contener, como mínimo, lo siguiente:

  f. Una evaluación del alcance y efectividad del Plan de Continuidad de Negocio, su divulgación entre las áreas pertinentes y la identificación, en su caso, de los ajustes necesarios para su actualización y fortalecimiento.

- ***Artículo 164.-*** *La Dirección General será la responsable de la debida implementación del Sistema de Control Interno; lo anterior, en el ámbito de las funciones que correspondan a dicha dirección.*

  En la implementación deberá procurarse que su funcionamiento sea acorde con las estrategias y fines de la Institución, aplicando las medidas preventivas y correctivas necesarias para subsanar cualquier deficiencia detectada.

  Al efecto, a la Dirección General, en adición a lo señalado en estas disposiciones, le corresponderá llevar a cabo las actividades siguientes:

  I. Elaborar, revisar y, en su caso, actualizar o proponer la actualización, para someter a la consideración del Comité de Auditoría y posterior presentación al Consejo, por lo menos una vez al año o con frecuencia mayor de acuerdo a lo determinado al efecto por el propio Consejo, los objetivos y lineamientos del Sistema de Control Interno, el código de conducta de la Institución, así como el Plan de Continuidad de Negocio.

- ***Artículo 164 Bis. -*** *La Dirección General deberá elaborar el Plan de Continuidad de Negocio observando al efecto lo establecido en el Anexo 67 de las presentes disposiciones; dicho plan será presentado para aprobación del Consejo de Administración a través del Comité de Auditoría.*

  El Director General será responsable de:

  I. La implementación, continúa actualización y difusión del plan al interior de la Institución. Al efecto, deberá establecer un programa de capacitación que responda a la participación del personal tanto en los procesos que al efecto se identifiquen como críticos, como en el desarrollo del propio plan.

  III. Prever lo necesario para hacer del conocimiento de la Comisión, las Contingencias Operativas que se presenten en cualquiera de sus canales de atención al público tales como sus Oficinas Bancarias, Medios Electrónicos o Comisionistas a que hace referencia el Artículo 319 de las presentes disposiciones; lo anterior, siempre que estas interrupciones registren una duración de al menos sesenta minutos y generen una afectación en al menos treinta por ciento de cualquiera de los canales de atención disponibles en una región, de acuerdo con la división geográfica que para efectos operativos o de negocio mantengan las propias Instituciones.

  En todo caso, la notificación señalada deberá efectuarse dentro de los 60 minutos siguientes a la verificación de los criterios antes mencionados.

  Asimismo, el director general deberá enviar a esta Comisión, en un plazo no mayor a quince días naturales posteriores a la conclusión de la Contingencia Operativa, un análisis de las causas que la motivaron, la afectación causada en términos cualitativos y cuantitativos que incluya el impacto monetario, temporal y en los canales de atención al público, así como la indicación de las acciones que se implementarán para minimizar el daño en situaciones similares subsecuentes.

  Por otra parte, el mismo día de la publicación de la circular única también la CNBV publicó el que denomino “Anexo 67”, el cual contiene los requerimientos mínimos

que debe contener un Plan de Continuidad de Negocio, del cual señalo los siguientes puntos más importantes y relacionados con este estudio [(2)]:

I. *Las Instituciones, previo al desarrollo del Plan de Continuidad de Negocio deberán llevar a cabo un análisis de impacto al negocio que:*

   a) *Identifique los procesos críticos que se consideran indispensables para la continuidad de las operaciones.*

   b) *Determine los recursos (humanos, logísticos, materiales, de infraestructura tecnológica y de cualquier otra naturaleza) mínimos necesarios para mantener y restablecer los servicios y procesos de la Institución ante la ocurrencia de una Contingencia Operativa, así como al término de ésta.*

   e) *Defina la prioridad de recuperación para cada uno de los procesos identificados como críticos.*

   f) *Determine el tiempo objetivo de recuperación (conocido como RTO, por sus siglas en inglés), para cada uno de los procesos críticos.*

   g) *Establezca, en su caso, el punto objetivo de recuperación (conocido como RPO, por sus siglas en inglés) entendido como la máxima pérdida de datos tolerable para cada uno de los procesos críticos.*

   h) *Identifique y evalúe los riesgos relacionados con los procesos operativos y servicios de procesamiento y transmisión de datos contratados con proveedores, así como los relacionados con custodia y resguardo de información de la Institución o de sus clientes.*

II. *En el desarrollo del Plan de Continuidad de Negocio, las Instituciones deberán incorporar las siguientes estrategias:*

   a) *De prevención, que comprenderá al menos la determinación, con base en el Análisis de Impacto al Negocio, de las acciones y procedimientos relativas a:*

iii. *El establecimiento de un programa de pruebas al funcionamiento y suficiencia del Plan de Continuidad de Negocios que contemple la actualización al menos anual, o antes si ocurre un cambio significativo en la infraestructura tecnológica, procesos, productos y servicios, u organización interna de la institución, y que evalúen todas las etapas y componentes del Plan de Continuidad de Negocios.*

Normativa de la CNBV relacionada con los procesos de respaldo de datos

Los siguientes son los artículos e incisos que se debe cumplir en relación a respaldos y recuperación de los datos, publicados en la Circular Única [(1)].

- ***Artículo 75.-*** *La unidad para la Administración Integral de Riesgos, para llevar a cabo la medición, vigilancia y control de los diversos tipos de riesgos discrecionales y la valuación de las posiciones de la Institución, deberá:*

  II. *Contar con modelos y sistemas de medición de riesgos que reflejen en forma precisa el valor de las posiciones y su sensibilidad a diversos Factores de Riesgo, asegurando que dichos modelos y sistemas estén adecuadamente elaborados y calibrados, e incorporando información proveniente de fuentes confiables para tales efectos. Dichos sistemas deberán:*

     d. *Contar con adecuados mecanismos de respaldo y control que permitan la recuperación de datos, de los sistemas de procesamiento de información empleados en la administración de riesgos y de modelos de valuación.*

- ***Artículo 164.-*** *La Dirección General será la responsable de la debida implementación del Sistema de Control Interno; lo anterior, en el ámbito de las funciones que correspondan a dicha dirección.*

  En la implementación deberá procurarse que su funcionamiento sea acorde con las estrategias y fines de la Institución, aplicando las medidas preventivas y correctivas necesarias para subsanar cualquier deficiencia detectada.

Al efecto, a la Dirección General, en adición a lo señalado en estas disposiciones, le corresponderá llevar a cabo las actividades siguientes:

IV. Prever las medidas que se estimen necesarias para que las transacciones u operaciones de la Institución y el Sistema de Control Interno, sean congruentes entre sí, adoptando, entre otras, las medidas siguientes:

f. Proteger la integridad y adecuado mantenimiento de los sistemas informáticos, incluidos los sistemas automatizados de procesamiento de datos y redes de telecomunicaciones a que se refiere el Artículo 52 de la Ley, así como la inalterabilidad, confidencialidad y disponibilidad de la información procesada, almacenada y transmitida por los mismos, determinando los mecanismos de respaldo de la información en caso fortuito o de fuerza mayor, así como los planes de contingencia que permitan asegurar la capacidad y continuidad de los sistemas. Adicionalmente, dichas medidas deberán establecer procedimientos para que los clientes puedan reportar el robo o extravío de sus Factores de Autenticación, incluso cuando las Instituciones operen a través de sus comisionistas.

V. Prever las medidas que se estimen necesarias a fin de que los sistemas informáticos que utilicen las Instituciones para realizar sus operaciones y para la prestación de servicios al público, cumplan con lo siguiente:

e. Cuenten con controles tanto de seguridad que protejan la confidencialidad de la información, como de acceso para garantizar la integridad de los sistemas y de la información generada, almacenada y transmitida por éstos. Dichas medidas serán acordes con el grado de criticidad de la información.

f. Minimicen el riesgo de interrupción de la operación con base en mecanismos de respaldo y procedimientos de recuperación de la información, así como de la infraestructura tecnológica para su procesamiento.

- ***Artículo 316 Bis 11.-*** *Las Instituciones deberán contar con controles para el acceso a las bases de datos y archivos correspondientes a las operaciones y servicios efectuados a través de Medios Electrónicos, aun cuando dichas bases de datos y archivos residan en medios de almacenamiento de respaldo. Para efectos de lo anterior, las Instituciones deberán ajustarse a lo siguiente:*

    I. El acceso a las bases de datos y archivos estará permitido exclusivamente a las personas expresamente autorizadas por la Institución en función de las actividades que realizan. Al otorgarse dichos accesos, deberá dejarse constancia de tal circunstancia y señalar los propósitos y el periodo al que se limitan los accesos.

    II. Tratándose de accesos que se realicen en forma remota, deberán utilizarse mecanismos de Cifrado en las comunicaciones.

    III. Deberán contar con procedimientos seguros de destrucción de los medios de almacenamiento de las bases de datos y archivos que contengan Información Sensible de sus Usuarios, que prevengan su restauración a través de cualquier mecanismo o dispositivo.

    IV. Deberán desarrollar políticas relacionadas con el uso y almacenamiento de información que se transmita y reciba por los Medios Electrónicos, estando obligadas a verificar el cumplimiento de sus políticas por parte de sus proveedores y afiliados.

Normativa de la SG y de la SFP relacionada con los procesos de respaldo de datos

El siguiente artículo e inciso se debe cumplir en relación a respaldos y recuperación de los datos, publicados por la SG y la SFP en materia de políticas y disposiciones para la Estrategia Digital Nacional en materia de Tecnologías de Información y Comunicaciones y en la seguridad de la Información [(5)].

- ***Artículo 27.-*** *Las Instituciones mantendrán los componentes de software y de seguridad de los dominios tecnológicos actualizados para evitar vulnerabilidades, de acuerdo a lo que se establece en el MAAGTICSI, para lo cual implementarán, entre otros, elementos de seguridad de la información, los siguientes:*

VII. Implementar medidas y procedimientos para el respaldo de información.

Normativa de la CNBV relacionada en prestación de servicios de infraestructura hospedada en centros de cómputo de terceros

Adicionalmente la CNBV publicó el Anexo 52 con los "LINEAMIENTOS MÍNIMOS DE OPERACIÓN Y SEGURIDAD PARA LA CONTRATACIÓN DE SERVICIOS DE APOYO TECNOLÓGICO", es decir, normativa que se debe cumplir en caso de que una institución crediticia contrate servicios de hospedaje e infraestructura en Centros de Cómputo de terceros, sean esto públicos o privados, señalo los puntos más relevantes de este anexo, ya que en caso de que contratar una o varias bases de datos que resulten del proceso de almacenamiento distribuido y se alojen en centros de cómputo diferentes a los de NAFIN, deberán apegarse a estos lineamientos [(3)]:

Las Instituciones deberán considerar los aspectos siguientes:

I. Aspectos en materia de operación.

- *a. Esquemas de redundancia o mecanismos alternos en las telecomunicaciones de punto a punto que permitan contar con enlaces de comunicación que minimicen el riesgo de interrupción en el servicio de telecomunicaciones.*
- *b. Estrategia de continuidad en los servicios informáticos que proporcionen a la Institución la capacidad de procesar y operar los sistemas en caso de contingencia, fallas o interrupciones en las telecomunicaciones o de los equipos de cómputo centrales y otros que estén involucrados en el servicio de procesamiento de información de operaciones o servicios.*

II. Aspectos en materia de seguridad.

- *a. Medidas para asegurar la transmisión de la Información Sensible del Usuario en forma cifrada punto a punto y elementos o controles de seguridad en cada uno de los nodos involucrados en el envío y recepción de datos.*

III. Auditoría y Supervisión.

a. *Políticas y procedimientos relativos a la realización de auditorías internas o externas sobre la infraestructura, controles y operación del centro de cómputo del tercero, relacionado con el ambiente de producción para la institución de crédito, al menos una vez cada dos años con el fin de evaluar el cumplimiento de lo mencionado en el presente anexo.*

   Tratándose de las operaciones a que se refiere la fracción X del Artículo 319 de las disposiciones que se realicen a través de Administradores de Comisionistas, la auditoría a que se refiere el párrafo anterior, deberá realizarse por la propia Institución, al menos una vez al año.

b. *Mecanismos de acceso al ambiente tecnológico, incluyendo información, bases de datos y configuraciones de seguridad, desde las instalaciones de la Institución en territorio nacional.*

Normativa de la SG y de la SFP relacionada con prestación de servicios de en centros de cómputo de terceros

El siguiente artículo e incisos se debe cumplir en relación a que servicios se podrían contratar en centros de cómputo de terceros, publicados por la SG y la SFP en materia de políticas y disposiciones para la Estrategia Digital Nacional en materia de Tecnologías de Información y Comunicaciones y en la seguridad de la Información [(5)].

- ***Artículo 13.-*** *En el caso de servicios de Centros de Datos, las Instituciones, deberán observar lo siguiente:*

  III. Evaluar la conveniencia de contratar servicios de Centro de Datos, tomando en cuenta el beneficio económico, eficiencia, privacidad, seguridad de los datos y de la información, en comparación con la de utilizar un Centro de Datos propio o compartido con otra Institución;

  VI. Almacenar y administrar en los Centros de Datos que se encuentren en las instalaciones de las Instituciones, los datos considerados de seguridad nacional, seguridad pública e información reservada y confidencial, conforme a la normatividad aplicable.

Normativa de Banxico relacionada con los puntos que deben cubrir las instituciones bancarias que presten servicios de DERIVADOS, conocidos como los 31 puntos de Banco de México

Otra normativa importante que requiere se cumplan planes de contingencia y respaldos y recuperación es el Banco de México (Banxico), a través de la Circular 4/2012, estas disposiciones se conocen como los 31 puntos de Banco de México y son de carácter obligatorio para que Banxico mantenga a las instituciones financieras el permiso de prestar servicios de operaciones de derivados.

En referencia a este estudio señalo los puntos que se requieren cumplir [(4)]:

11. LÍMITE, SUSPENSIÓN O REVOCACIÓN DE OPERACIONES

El Banco de México podrá limitar, suspender o revocar las autorizaciones otorgadas a las Entidades en términos de las presentes Reglas para realizar Operaciones Derivadas cuando:

*b) Dejen de reunir cualquier requerimiento del **Anexo** de estas Reglas;*

***Anexo;** Requerimientos para las entidades que pretendan realizar operaciones derivadas:*

4. La Dirección General deberá tener un procedimiento de acción contingente que le permita actuar cuando se detecte que son deficientes las políticas, procedimientos, controles internos, el sistema de información gerencial o los niveles de tolerancia de riesgo o cuando ocurran violaciones a las leyes, normas o circulares aplicables.

Adicionalmente, deberá contarse con un plan de contingencia operativo que garantice la continuidad de la operación ante eventos inesperados.

19. Los sistemas de procesamiento de datos, de administración de riesgos y de los modelos de valuación, deberán tener un adecuado respaldo y control que incluya la recuperación de datos.

ANEXO “II”. - Pruebas de laboratorio.

Primera etapa

Se utilizaron siete computadoras personales con las siguientes características técnicas:

- HP Compaq 8200 Elite CMT PC
- Procesador Intel Core i7-2600 - 3.4 GHz, 4 cores
- 16 GB de memoria RAM
- 500 GB de Disco Duro
- Tarjeta de Red Interna Gigabit Intel 82579LM

La velocidad de transmisión en la red local se configuro a 1 Gbps.

Se utilizó la herramienta MD5Summer para validar la integridad de los archivos recuperados contra el archivo originalmente respaldado y distribuido.

Selección del tamaño máximo de archivo

a) Se utilizaron tres diferentes archivos como muestras para las pruebas, de 800, 1,000 y 2,000 Mb, con una UMA (Unidad Máxima de Almacenamiento) de 128 unidades, realizándose diez pruebas con cada archivo tanto para el almacenamiento distribuido (carga) como para la recuperación (descarga) de la muestra original, obteniéndose los siguientes datos:

UMA	Tamaño de Archivo (MB)	Nombre del archivo	Tipo de respaldo	Almacenamiento			Recuperado			Integridad
				Inicio	Termino	Tiempo	Inicio	Termino	Tiempo	
128	800	miArchivo0.Naf	IDA	13:32:13.163	13:32:59.236	00:00:46.073	13:49:49.360	13:51:02.973	00:01:13.613	OK
128	800	miArchivo1.Naf	IDA	13:33:07.682	13:34:02.350	00:00:54.668	13:51:10.585	13:52:09.467	00:00:58.882	OK
128	800	miArchivo2.Naf	IDA	13:34:13.802	13:35:07.362	00:00:53.560	13:52:28.584	13:53:30.714	00:01:02.130	OK
128	800	miArchivo3.Naf	IDA	13:35:38.313	13:36:25.368	00:00:47.055	13:53:38.230	13:54:45.074	00:01:06.844	OK
128	800	miArchivo4.Naf	IDA	13:36:35.322	13:37:32.436	00:00:57.114	13:54:52.121	13:56:12.873	00:01:20.752	OK
128	800	miArchivo5.Naf	IDA	13:38:09.514	13:39:09.149	00:00:59.635	13:56:19.917	13:57:34.535	00:01:14.618	OK
128	800	miArchivo6.Naf	IDA	13:39:28.605	13:40:17.188	00:00:48.583	13:57:41.069	13:58:37.739	00:00:56.670	OK
128	800	miArchivo7.Naf	IDA	13:40:26.735	13:41:40.645	00:01:13.910	13:58:44.336	13:59:49.939	00:01:05.603	OK
128	800	miArchivo8.Naf	IDA	13:41:50.734	13:42:42.406	00:00:51.672	13:59:57.041	14:00:59.773	00:01:02.732	OK
128	800	miArchivo9.Naf	IDA	13:43:26.395	13:44:14.972	00:00:48.577	14:01:06.307	14:02:14.348	00:01:08.041	OK
PROMEDIO						00:00:54.085			00:01:06.988	

UMA	Tamaño de Archivo (MB)	Nombre del archivo	Tipo de respaldo	Almacenamiento			Recuperado			Integridad
				Inicio	Termino	Tiempo	Inicio	Termino	Tiempo	
128	1,000	miArchivo0.Naf	IDA	09:58:24.761	09:59:33.800	00:01:09.039	10:48:07.141	10:49:21.487	00:01:14.346	Failed
128	1,000	miArchivo1.Naf	IDA	09:45:09.245	09:46:12.787	00:01:03.542	10:49:32.445	10:50:48.221	00:01:15.776	OK
128	1,000	miArchivo2.Naf	IDA	09:43:50.743	09:44:51.293	00:01:00.550	10:51:06.814	10:52:37.845	00:01:31.031	OK
128	1,000	miArchivo3.Naf	IDA	10:24:02.802	10:25:10.348	00:01:07.546	10:52:45.500	10:54:01.812	00:01:16.312	OK
128	1,000	miArchivo4.Naf	IDA	10:22:11.291	10:23:35.353	00:01:24.062	10:54:09.694	10:55:34.869	00:01:25.175	OK
128	1,000	miArchivo5.Naf	IDA	10:21:01.293	10:21:55.343	00:00:54.050	10:55:45.702	10:57:03.496	00:01:17.794	Failed
128	1,000	miArchivo6.Naf	IDA	10:19:50.782	10:20:45.331	00:00:54.549	11:03:03.389	11:04:27.415	00:01:24.026	OK
128	1,000	miArchivo7.Naf	IDA	10:18:35.292	10:19:34.842	00:00:59.550	11:05:05.097	11:06:28.878	00:01:23.781	OK
128	1,000	miArchivo8.Naf	IDA	10:14:18.273	10:15:21.334	00:01:03.061	11:06:39.747	11:07:56.735	00:01:16.988	OK
128	1,000	miArchivo9.Naf	IDA	10:06:51.271	10:08:08.325	00:01:17.054	11:10:02.589	11:11:31.918	00:01:29.329	OK
PROMEDIO						00:01:05.300			00:01:21.456	

UMA	Tamaño de Archivo (MB)	Nombre del archivo	Tipo de respaldo	Almacenamiento			Recuperado			Integridad
				Inicio	Termino	Tiempo	Inicio	Termino	Tiempo	
128	2,000	miArchivo0.Naf	IDA	10:34:25.596	10:36:22.135	00:01:56.539	12:56:51.387	12:59:42.491	00:02:51.104	Failed
128	2,000	miArchivo1.Naf	IDA	10:39:48.806	10:41:59.956	00:02:11.150	13:02:46.699	13:05:27.771	00:02:41.072	OK
128	2,000	miArchivo2.Naf	IDA	10:42:09.660	10:44:31.033	00:02:21.373	13:05:41.007	13:08:26.932	00:02:45.925	Failed
128	2,000	miArchivo3.Naf	IDA	10:44:39.627	10:49:26.303	00:04:46.676	13:08:31.917	13:11:18.802	00:02:46.885	Failed
128	2,000	miArchivo4.Naf	IDA	11:20:39.192	11:22:57.263	00:02:18.071	13:11:53.983	13:14:50.697	00:02:56.714	Failed
128	2,000	miArchivo5.Naf	IDA	11:23:19.000	11:25:31.896	00:02:12.896	13:14:55.721	13:17:44.450	00:02:48.729	Failed
128	2,000	miArchivo6.Naf	IDA	11:29:29.867	11:31:48.471	00:02:18.604	13:19:05.260	13:21:44.004	00:02:38.744	Failed
128	2,000	miArchivo7.Naf	IDA	11:32:13.239	11:39:29.671	00:07:16.432	13:22:01.941	13:24:23.934	00:02:21.993	Failed
128	2,000	miArchivo8.Naf	IDA	11:39:46.517	11:44:24.975	00:04:38.458	13:24:32.441	13:27:14.029	00:02:41.588	OK
128	2,000	miArchivo9.Naf	IDA	11:45:10.732	11:47:33.897	00:02:23.165	13:27:21.489	13:30:31.032	00:03:09.543	OK
PROMEDIO						00:03:14.336			00:02:46.230	

De acuerdo a estas mediciones, con archivos de 1 y 2 Gb el número de pruebas no exitosas o de falta de integridad de los datos se incrementa, mientras que con el archivo de 800 Mb en todos los casos los datos se mantuvieron íntegros, por lo tanto, el tamaño máximo de archivos o base de datos conforme a la arquitectura utilizada y los equipos involucrados será de 800 Mb.

b) Como una prueba complementaria se cambió el tamaño de la UMA para verificar si con un tamaño mayor de esta se garantizaba la integridad de los datos, sin embargo, los resultados fueron similares, es decir en archivos superiores de 800 Mb, no se podría garantizar la integridad de datos, como lo muestra siguiente tabla:

UMA	Tamaño de Archivo (MB)	Nombre del archivo	Tipo de respaldo	Almacenamiento			Recuperado			Integridad
				Inicio	Termino	Tiempo	Inicio	Termino	Tiempo	
256	2,000	miArchivo0.Naf	IDA	17:31:26.141	18:01:36.460	00:30:10.319	11:20:33.387	11:23:38.506	00:03:05.119	Failed
256	2,000	miArchivo1.Naf	IDA	08:47:17.503	10:13:57.791	01:26:40.288	11:23:46.006	11:26:54.377	00:03:08.371	Failed
256	2,000	miArchivo2.Naf	IDA	11:36:11.315	11:38:27.854	00:02:16.539	11:28:24.970	11:31:57.638	00:03:32.668	OK
256	2,000	miArchivo3.Naf	IDA	14:29:12.574	15:01:10.602	00:31:58.028	11:32:12.686	11:35:13.385	00:03:00.699	Failed
256	2,000	miArchivo4.Naf	IDA	16:00:30.422	16:07:43.633	00:07:13.211	13:14:04.376	13:17:07.058	00:03:02.682	OK
256	2,000	miArchivo5.Naf	IDA	16:32:57.382	16:41:10.034	00:08:12.652	13:17:30.762	13:20:07.631	00:02:36.869	Failed
256	2,000	miArchivo6.Naf	IDA	16:43:16.379	16:45:35.261	00:02:18.882	13:21:10.504	13:24:13.388	00:03:02.884	Failed
256	2,000	miArchivo7.Naf	IDA	16:46:15.383	17:05:29.494	00:19:14.111	13:24:29.090	13:27:04.629	00:02:35.539	Failed
256	2,000	miArchivo8.Naf	IDA	17:06:10.132	17:11:10.774	00:05:00.642	13:27:51.265	13:30:28.289	00:02:37.024	OK
256	2,000	miArchivo9.Naf	IDA	08:43:17.885	09:01:48.982	00:18:31.097	13:31:42.381	13:34:35.015	00:02:52.634	Failed
PROMEDIO						00:21:09.577			00:02:57.449	

Selección del tamaño óptimo de la Unidad Máxima de Almacenamiento (UMA)

Se realizaron pruebas con tres diferentes tamaños de UMA, 128, 256 y 512, para cada UMA se efectuaron 10 pruebas con el archivo de 800 Mb obteniéndose los siguientes datos:

UMA	Tamaño de Archivo (MB)	Nombre del archivo	Tipo de respaldo	Almacenamiento			Recuperado			Integridad
				Inicio	Termino	Tiempo	Inicio	Termino	Tiempo	
128	800	miArchivo0.Naf	IDA	13:32:13.163	13:32:59.236	00:00:46.073	13:49:49.360	13:51:02.973	00:01:13.613	OK
128	800	miArchivo1.Naf	IDA	13:33:07.682	13:34:02.350	00:00:54.668	13:51:10.585	13:52:09.467	00:00:58.882	OK
128	800	miArchivo2.Naf	IDA	13:34:13.802	13:35:07.362	00:00:53.560	13:52:28.584	13:53:30.714	00:01:02.130	OK
128	800	miArchivo3.Naf	IDA	13:35:38.313	13:36:25.368	00:00:47.055	13:53:38.230	13:54:45.074	00:01:06.844	OK
128	800	miArchivo4.Naf	IDA	13:36:35.322	13:37:32.436	00:00:57.114	13:54:52.121	13:56:12.873	00:01:20.752	OK
128	800	miArchivo5.Naf	IDA	13:38:09.514	13:39:09.149	00:00:59.635	13:56:19.917	13:57:34.535	00:01:14.618	OK
128	800	miArchivo6.Naf	IDA	13:39:28.605	13:40:17.188	00:00:48.583	13:57:41.069	13:58:37.739	00:00:56.670	OK
128	800	miArchivo7.Naf	IDA	13:40:26.735	13:41:40.645	00:01:13.910	13:58:44.336	13:59:49.939	00:01:05.603	OK
128	800	miArchivo8.Naf	IDA	13:41:50.734	13:42:42.406	00:00:51.672	13:59:57.041	14:00:59.773	00:01:02.732	OK
128	800	miArchivo9.Naf	IDA	13:43:26.395	13:44:14.972	00:00:48.577	14:01:06.307	14:02:14.348	00:01:08.041	OK
PROMEDIO						00:00:54.085			00:01:06.988	

UMA	Tamaño de Archivo (MB)	Nombre del archivo	Tipo de respaldo	Almacenamiento			Recuperado			Integridad
				Inicio	Termino	Tiempo	Inicio	Termino	Tiempo	
256	800	miArchivo0.Naf	IDA	10:10:49.793	10:11:53.838	00:01:04.045	13:55:18.693	13:56:18.186	00:00:59.493	Failed
256	800	miArchivo1.Naf	IDA	10:12:01.788	10:13:10.345	00:01:08.557	14:28:30.710	14:29:46.206	00:01:15.496	OK
256	800	miArchivo2.Naf	IDA	10:14:35.845	10:15:25.899	00:00:50.054	14:29:54.204	14:31:09.706	00:01:15.502	OK
256	800	miArchivo3.Naf	IDA	10:15:44.868	10:16:52.431	00:01:07.563	14:31:48.710	14:33:19.203	00:01:30.493	OK
256	800	miArchivo4.Naf	IDA	10:16:59.399	10:18:05.549	00:01:06.150	14:33:28.707	14:34:32.213	00:01:03.506	OK
256	800	miArchivo5.Naf	IDA	10:18:42.493	10:19:51.585	00:01:09.092	14:34:47.216	14:36:01.211	00:01:13.995	OK
256	800	miArchivo6.Naf	IDA	10:21:19.046	10:22:25.594	00:01:06.548	14:40:52.219	14:42:07.208	00:01:14.989	OK
256	800	miArchivo7.Naf	IDA	10:24:23.062	10:25:17.136	00:00:54.074	14:42:16.722	14:43:31.214	00:01:14.492	OK
256	800	miArchivo8.Naf	IDA	10:39:24.466	10:40:16.037	00:00:51.571	14:43:45.715	14:44:58.209	00:01:12.494	OK
256	800	miArchivo9.Naf	IDA	10:42:51.489	10:43:58.033	00:01:06.544	14:45:06.715	14:46:10.213	00:01:03.498	OK
PROMEDIO						00:01:02.420			00:01:12.396	

UMA	Tamaño de Archivo (MB)	Nombre del archivo	Tipo de respaldo	Almacenamiento			Recuperado			Integridad
				Inicio	Termino	Tiempo	Inicio	Termino	Tiempo	
512	800	miArchivo0.Naf	IDA	10:59:02.281	11:01:41.824	00:02:39.543	12:34:34.414	12:36:41.905	00:02:07.491	Failed
512	800	miArchivo1.Naf	IDA	11:05:49.270	11:08:00.832	00:02:11.562	12:42:16.415	12:44:23.408	00:02:06.993	OK
512	800	miArchivo2.Naf	IDA	11:08:16.784	11:10:31.829	00:02:15.045	12:45:13.912	12:47:21.415	00:02:07.503	Failed
512	800	miArchivo3.Naf	IDA	11:28:07.329	11:30:37.882	00:02:30.553	12:47:52.414	12:50:15.915	00:02:23.501	OK
512	800	miArchivo4.Naf	IDA	11:30:46.328	11:32:54.894	00:02:08.566	13:14:07.943	13:16:12.934	00:02:04.991	OK
512	800	miArchivo5.Naf	IDA	11:33:07.328	11:35:19.383	00:02:12.055	13:20:23.486	13:22:24.490	00:02:01.004	OK
512	800	miArchivo6.Naf	IDA	11:35:31.834	11:37:51.387	00:02:19.553	13:47:37.303	13:49:42.804	00:02:05.501	OK
512	800	miArchivo7.Naf	IDA	11:49:46.331	11:52:27.387	00:02:41.056	13:51:35.802	13:53:57.301	00:02:21.499	OK
512	800	miArchivo8.Naf	IDA	11:53:23.333	11:55:54.391	00:02:31.058	15:19:58.394	15:22:18.976	00:02:20.582	OK
512	800	miArchivo9.Naf	IDA	11:56:03.343	11:58:20.409	00:02:17.066	15:22:28.977	15:24:33.490	00:02:04.513	OK
PROMEDIO						00:02:22.606			00:02:10.358	

De acuerdo a estas pruebas, con la UMA a 128 unidades, es con la que obtuvimos un 100% de integridad, e incluso un tiempo promedio menor tanto en almacenamiento como en recuperación en comparación con las UMAs de 256 y 512, por lo tanto, la UMA óptima para nuestra arquitectura de pruebas y equipos involucrados será de 128.

Pruebas de almacenamiento y recuperación con cinco, cuatro y tres nodos o células, obtenido la siguiente información:

a) Con 5 nodos

UMA	Tamaño de Archivo (MB)	Nombre del archivo	Tipo de respaldo	Almacenamiento			Recuperado			Integridad
				Inicio	Termino	Tiempo	Inicio	Termino	Tiempo	
128	800	miArchivo0.Naf	IDA	13:32:13.163	13:32:59.236	00:00:46.073	13:49:49.360	13:51:02.973	00:01:13.613	OK
128	800	miArchivo1.Naf	IDA	13:33:07.682	13:34:02.350	00:00:54.668	13:51:10.585	13:52:09.467	00:00:58.882	OK
128	800	miArchivo2.Naf	IDA	13:34:13.802	13:35:07.362	00:00:53.560	13:52:28.584	13:53:30.714	00:01:02.130	OK
128	800	miArchivo3.Naf	IDA	13:35:38.313	13:36:25.368	00:00:47.055	13:53:38.230	13:54:45.074	00:01:06.844	OK
128	800	miArchivo4.Naf	IDA	13:36:35.322	13:37:32.436	00:00:57.114	13:54:52.121	13:56:12.873	00:01:20.752	OK
128	800	miArchivo5.Naf	IDA	13:38:09.514	13:39:09.149	00:00:59.635	13:56:19.917	13:57:34.535	00:01:14.618	OK
128	800	miArchivo6.Naf	IDA	13:39:28.605	13:40:17.188	00:00:48.583	13:57:41.069	13:58:37.739	00:00:56.670	OK
128	800	miArchivo7.Naf	IDA	13:40:26.735	13:41:40.645	00:01:13.910	13:58:44.336	13:59:49.939	00:01:05.603	OK
128	800	miArchivo8.Naf	IDA	13:41:50.734	13:42:42.406	00:00:51.672	13:59:57.041	14:00:59.773	00:01:02.732	OK
128	800	miArchivo9.Naf	IDA	13:43:26.395	13:44:14.972	00:00:48.577	14:01:06.307	14:02:14.348	00:01:08.041	OK
PROMEDIO						00:00:54.085			00:01:06.988	

b) Con 4 nodos

UMA	Tamaño de Archivo (MB)	Nombre del archivo	Tipo de respaldo	Almacenamiento			Recuperado			Integridad
				Inicio	Termino	Tiempo	Inicio	Termino	Tiempo	
128	800	miArchivo0.Naf	IDA	15:50:57.885	15:52:49.454	00:01:51.569	16:25:04.911	16:26:59.907	00:01:54.996	OK
128	800	miArchivo1.Naf	IDA	15:52:58.889	15:55:52.946	00:02:54.057	16:27:15.934	16:29:56.924	00:02:40.990	OK
128	800	miArchivo2.Naf	IDA	15:56:05.884	15:58:57.937	00:02:52.053	16:30:05.435	16:32:48.966	00:02:43.531	OK
128	800	miArchivo3.Naf	IDA	15:59:10.889	16:02:04.422	00:02:53.533	16:33:51.965	16:36:42.983	00:02:51.018	OK
128	800	miArchivo4.Naf	IDA	16:02:11.885	16:04:59.945	00:02:48.060	16:36:51.495	16:39:45.053	00:02:53.558	OK
128	800	miArchivo5.Naf	IDA	16:05:06.885	16:08:03.957	00:02:57.072	16:40:26.559	16:43:15.554	00:02:48.995	OK
128	800	miArchivo6.Naf	IDA	16:08:11.392	16:11:15.455	00:03:04.063	16:43:26.056	16:46:20.555	00:02:54.499	OK
128	800	miArchivo7.Naf	IDA	16:11:22.388	16:14:14.442	00:02:52.054	16:46:27.059	16:49:13.065	00:02:46.006	OK
128	800	miArchivo8.Naf	IDA	16:14:20.889	16:20:54.943	00:06:34.054	16:49:21.081	16:51:26.169	00:02:05.088	OK
128	800	miArchivo9.Naf	IDA	16:21:21.897	16:24:17.927	00:02:56.030	16:52:57.708	16:56:01.264	00:03:03.556	OK
PROMEDIO						00:03:10.255			00:02:40.224	

c) Con 3 nodos

UMA	Tamaño de Archivo (MB)	Nombre del archivo	Tipo de respaldo	Almacenamiento			Recuperado			Integridad
				Inicio	Termino	Tiempo	Inicio	Termino	Tiempo	
128	800	miArchivo0.Naf	IDA	11:10:31.047	11:16:42.621	00:06:11.574	13:59:35.844	14:00:56.349	00:01:20.505	OK
128	800	miArchivo1.Naf	IDA	11:16:53.087	11:19:18.165	00:02:25.078	14:01:11.350	14:05:55.404	00:04:44.054	OK
128	800	miArchivo2.Naf	IDA	12:01:50.495	12:03:45.048	00:01:54.553	14:07:53.443	14:10:33.472	00:02:40.029	OK
128	800	miArchivo3.Naf	IDA	12:04:52.995	12:09:41.615	00:04:48.620	18:12:12.549	18:14:13.529	00:02:00.980	OK
128	800	miArchivo4.Naf	IDA	12:09:49.565	12:11:47.610	00:01:58.045	18:14:58.026	18:17:13.500	00:02:15.474	OK
128	800	miArchivo5.Naf	IDA	12:36:29.262	12:40:42.803	00:04:13.541	18:19:42.746	18:21:32.878	00:01:50.132	OK
128	800	miArchivo6.Naf	IDA	13:12:51.764	13:16:16.862	00:03:25.098	18:21:42.497	18:23:04.747	00:01:22.250	OK
128	800	miArchivo7.Naf	IDA	13:21:24.378	13:26:58.457	00:05:34.079	18:23:13.246	18:24:38.254	00:01:25.008	OK
128	800	miArchivo8.Naf	IDA	13:27:11.429	13:28:25.489	00:01:14.060	18:24:43.245	18:26:47.750	00:02:04.505	OK
128	800	miArchivo9.Naf	IDA	13:28:33.453	13:31:20.126	00:02:46.673	18:27:13.247	18:29:07.512	00:01:54.265	OK
PROMEDIO						00:03:27.132			00:02:09.720	

Concluyendo que en todos los casos fue exitoso tanto la distribución como la recuperación y que a memos número de nodos tiende a ser mayor el tiempo de proceso.

Segunda etapa

Se ocuparon seis de las computadoras personales utilizadas en la primera etapa y se creó un servidor virtual con las siguientes características:

- HP Proliant DL 380p G8
- Procesador XEON E5-2640 – 2.5 GHz con 2 procesadores, 6 cores por procesador, a la máquina virtual se le asignaron 4 cores
- 64 GB de memoria RAM, a la máquina virtual se le asignaron 8 GB
- 100 GB de Disco Duro
- Tarjeta de Red Intel 3350 GB - 1000

La velocidad de transmisión en la red local del servidor físico se configuro a 10 Gbps.

El formato de medición de tiempos utilizado fue, *horas: minutos: segundos*

Se continuó utilizando la herramienta MD5Summer para validar la integridad de los archivos recuperados contra el archivo originalmente respaldado y distribuido.

Se mantuvo constante el tamaño del archivo de prueba de 800 Mb, así como la UMA de 128.

Infraestructura de servidores con hipervisor VMWare

Obteniendo las siguientes mediciones:

UMA	Tamaño de Archivo (MB)	Nombre del archivo	Tipo de respaldo	Almacenamiento			Recuperado			Integridad
				Inicio	Termino	Tiempo	Inicio	Termino	Tiempo	
128	800	miArchivo1.Naf	IDA	14:40:56.768	14:41:45.815	00:00:49.047	14:53:13.770	14:54:27.774	00:01:14.004	Ok
128	800	miArchivo4.Naf	IDA	14:43:51.767	14:44:45.820	00:00:54.053	14:58:54.776	15:00:13.773	00:01:18.997	Ok
128	800	miArchivo2.Naf	IDA	14:12:36.374	14:13:20.435	00:00:44.061	14:13:29.871	14:14:30.372	00:01:00.501	Ok
128	800	miArchivo3.Naf	IDA	13:41:03.845	13:41:55.380	00:00:51.535	14:14:51.883	14:16:00.906	00:01:09.023	Ok
128	800	miArchivo5.Naf	IDA	14:44:55.766	14:45:45.795	00:00:50.029	15:01:05.779	15:02:09.275	00:01:03.496	Ok
128	800	miArchivo5.Naf	IDA	13:54:07.371	13:55:01.898	00:00:54.527	14:19:10.908	14:20:14.410	00:01:03.502	Ok
128	800	miArchivo6.Naf	IDA	14:45:52.771	14:46:41.323	00:00:48.552	15:05:50.780	15:06:52.777	00:01:01.997	Ok
128	800	miArchivo7.Naf	IDA	13:58:59.866	13:59:48.903	00:00:49.037	14:21:45.914	14:23:04.405	00:01:18.491	Ok
128	800	miArchivo8.Naf	IDA	14:06:15.873	14:07:05.908	00:00:50.035	14:23:53.416	14:25:00.913	00:01:07.497	Ok
128	800	miArchivo9.Naf	IDA	14:07:11.376	14:08:04.901	00:00:53.525	14:25:08.413	14:26:18.909	00:01:10.496	Ok
PROMEDIO						00:00:50.440			00:01:08.800	

Tercera etapa

Se ocuparon cinco de las computadoras personales utilizadas en la primera etapa, el servidor virtual creado para la segunda etapa de pruebas y se creó un servidor virtual con las siguientes características técnicas en el centro de cómputo alterno:

- HP Proliant DL 380p G8
- Procesador XEON E5-2640 – 2.5 GHz con 2 procesadores, 6 cores por procesador, a la máquina virtual se le asignaron 4 cores
- 64 GB de memoria RAM, a la máquina virtual se le asignaron 8 GB
- 100 GB de Disco Duro
- Tarjeta de Red Intel 3350 GB - 1000

La velocidad de transmisión en la red local del CCC se configuro a 1 Gbps

La velocidad de transmisión entre los centros de cómputo es de 10.2 Mbps.

El formato de medición de tiempos utilizado fue, *horas: minutos: segundos*

Se continuó utilizando la herramienta MD5Summer para validar la integridad de los archivos recuperados contra el archivo originalmente respaldado y distribuido.

Se mantuvo constante el tamaño del archivo de prueba de 800 Mb, así como la UMA de 128.

Infraestructura de servidores con hipervisor VMWare

Obteniendo las siguientes mediciones:

UMA	Tamaño de Archivo (MB)	Nombre del archivo	Tipo de respaldo	Almacenamiento			Recuperado			Integridad
				Inicio	Termino	Tiempo	Inicio	Termino	Tiempo	
128	800	miArchivo0.Naf	IDA	15:13:22.132	15:14:06.191	00:00:44.059	15:29:29.641	15:31:23.636	00:01:53.995	Ok
128	800	miArchivo1.Naf	IDA	15:14:12.137	15:15:39.192	00:01:27.055	15:56:28.213	16:02:43.710	00:06:15.497	Ok
128	800	miArchivo2.Naf	IDA	15:16:01.634	15:17:34.677	00:01:33.043	16:03:04.716	16:04:17.716	00:01:13.000	Ok
128	800	miArchivo3.Naf	IDA	15:18:24.632	15:19:13.692	00:00:49.060	16:04:39.215	16:07:09.712	00:02:30.497	Ok
128	800	miArchivo4.Naf	IDA	15:20:45.138	15:21:34.632	00:00:49.494	08:20:43.386	08:24:27.879	00:03:44.493	Ok
128	800	miArchivo5.Naf	IDA	15:21:52.639	15:22:59.681	00:01:07.042	08:25:23.883	08:33:51.944	00:08:28.061	Ok
128	800	miArchivo6.Naf	IDA	15:23:05.636	15:24:24.682	00:01:19.046	09:05:18.210	09:07:12.223	00:01:54.013	Ok
128	800	miArchivo7.Naf	IDA	15:24:33.139	15:25:48.703	00:01:15.564	09:13:50.296	09:17:55.292	00:04:04.996	Ok
128	800	miArchivo8.Naf	IDA	15:26:24.631	15:27:23.169	00:00:58.538	09:18:53.797	09:20:32.291	00:01:38.494	Ok
128	800	miArchivo9.Naf	IDA	15:27:47.640	15:28:52.693	00:01:05.053	09:21:05.798	09:25:01.795	00:03:55.997	Ok
PROMEDIO						00:01:06.795			00:03:33.904	