



**Inteligencia de negocios basada en Bases de Datos In-Memory**

**Mario Raúl Morales Morales <sup>1</sup>, Santiago Leonardo Morales Cardoso <sup>2</sup>**

**1 Universidad Central del Ecuador, mmorales@uce.edu.ec**

**2 Universidad Central del Ecuador, smorales@uce.edu.ec**

**RESUMEN.**

La tecnología In-Memory ha sido propiciada por la necesidad de procesamiento de grandes volúmenes de datos de manera muy rápida y fundamentalmente por el desarrollo de los procesadores y el incremento en la capacidad de memoria basada en la arquitectura de 64-bits. Esto ha hecho posible el procesamiento paralelo masivo de las operaciones de base de datos, albergando todos los datos relevantes en memoria. Una base de datos In-Memory (IMDB) usa la memoria como el principal soporte de almacenamiento, y para su procesamiento en memoria no requieren su paso desde el disco duro hacia ella, lo que reduce el tiempo de respuesta de la base de datos dramáticamente. Las bases de datos tradicionales almacenan la data en disco y las operaciones de I/O son muy lentas comparadas con aquellas hechas en memoria RAM.

Las IMDBs poseen una técnica de almacenamiento columnar lo que posibilita el acceso a la data a grandes velocidades y capacidades de analítica en tiempo-real. Al organizar los valores en la forma de un vector de atributos (columnar) permite una fácil compresión de datos y también permite una alta velocidad de escaneo y filtraje. La velocidad es en efecto tan alta que se puede dejar de lado la idea de pre-agregación de la data transaccional, la base de los sistemas de información en las décadas pasadas.

**PALABRAS CLAVES.**

In Memory, In Memory Data Base, Almacenamiento Columnar



**Business intelligence based on In-Memory Databases**

**ABSTRACT.**

In-memory technology has been driven by the need to process huge volumes of data very fastly and mainly due to the development of processors and the increase in memory capacity based on 64-bit architecture. This has made possible the massive parallel processing of the database operations, storing all relevant data in memory. An In-Memory database (IMDB) uses memory as the main storage medium, and for processing in memory does not require its passage from the hard disk to memory, which dramatically reduces database response time. Traditional databases store data on disk and I / O operations are very slow compared to those made in RAM. IMDB has a columnar storage foundation, which enables access to data at high speeds and real-time analytical capabilities. By organizing the values in the form of an attribute vector (columnar) allows for easy data compression and also allows a high speed of scanning and filtering. Speed is indeed so high that one can leave out the idea of pre- aggregation of transactional data, the basis of information systems in the past decades.

**KEYWORDS.**

In Memory, In Memory Data Base, Columnar Storage



## **1. INTRODUCCIÓN.-**

### **1.1 La evolución computacional.-**

La tecnología In-Memory ha sido propiciada por el desarrollo de los procesadores y la memoria en los últimos años (Mihaela-Laura, 2014). En este tiempo, entre los más importantes desarrollos en la tecnología computacional está la construcción de procesadores (CPU) multi-core (capaces de albergar hasta 15 unidades de procesamiento completamente funcionales) y el incremento en la capacidad de memoria basada en la arquitectura de 64-bits, la cual permite soportar terabytes de espacio directamente direccionable (Plattner & Leukert, 2015). Asociado a estos avances, se suma el abaratamiento de los costos de memoria, la capacidad de la arquitectura multi-core para el procesamiento paralelo, y el desarrollo de técnicas de compresión para mantener más data en menos memoria (Mihaela-Laura, 2014).

La arquitectura multi-core permite el procesamiento paralelo masivo de las operaciones de base de datos, y ya que todos los datos relevantes permanecerían en memoria, el procesamiento toma lugar a la más alta velocidad posible. Las operaciones de lectura son completamente independientes de cualquier acceso más lento a dispositivos de almacenamiento en disco; a su vez, las operaciones de escritura también toman lugar en memoria, pero tienen que ser grabadas en un almacenamiento no volátil para garantizar persistencia de la data (Plattner & Leukert, 2015). Bajo este esquema, los procesos In-Memory son soportados por CPU's multicore que permiten el procesamiento paralelo de data en memoria (Mihaela-Laura, 2014).

La necesidad de rendimiento en el dominio de la Tecnologías de Información (TI) combinado con las ventajas de la computación In-Memory son factores importantes que han influenciado la aparición de bases de datos In-Memory. Una base de datos In-Memory usa la memoria como el principal soporte de almacenamiento, comparado con los clásicos sistemas de gestión de base de datos (DBMS por sus siglas en Inglés) que usan el disco como lugar de almacenamiento principal (Babeanu & Ciobanu, 2015). En suma, estas innovaciones técnicas forman la base para una nueva manera de trabajo de aplicaciones de negocio, particularmente las bases de datos. En contraste a las bases de datos relacionales, las IMDB ya tienen la data a ser procesada en memoria y no requiere su paso desde el disco duro hacia ella; esto reduce el tiempo de respuesta de la base de



datos dramáticamente y actúa como un impulso para las aplicaciones empresariales (Mihaela-Laura, 2014).

En el campo empresarial, las primeras IMDB aparecieron a inicios de los 90. En 1992 White Cross Systems Limited Co. ofreció la primera versión de este tipo de bases de datos la cual fue entregada en hardware propietario. En la actualidad las grandes marcas como Oracle, IBM, SAP y Microsoft tienen versiones de bases de datos In-Memory (Babeanu & Ciobanu, 2015)

## **1.2 El advenimiento de Big Data.-**

Según un estudio publicado por (Wixom, y otros, 2014) basado en el 3er. Congreso de la AIS (*Association for Information Systems*), indica que cada vez más las nuevas tecnologías se están usando para direccionar y habilitar *Big Data*, y que ellas son significativamente diferentes de las tecnologías de datos heredadas (antiguas o tradicionales). Señala a la analítica In-Memory como una opción cada vez más viable, por lo que crecientemente las compañías la están usando para aplicaciones analíticas específicas.

*Big Data* está creando un cambio de paradigma con el movimiento de datos. Las organizaciones se están moviendo de traer los datos hacia el procesamiento (entiéndase como servidores) a empujar el procesamiento (distribuido) hacia la data. El análisis de datos In-Memory, con las arquitecturas distribuidas y el procesamiento paralelo masivo (MPP) está ganando cada vez mayor “momentum” (Wixom, y otros, 2014).

## **2 TECNOLOGÍA IN-MEMORY.-**

Previo al inicio del estudio más detallado de lo que es la tecnología In-Memory, es interesante notar tres elementos que están en la cima de la agenda de la gestión tecnológica según la investigación propuesta por (Hahn & Packowski, 2015): *Big Data*, analítica avanzada y bases de datos en-memoria. Ellos son vistos como habilitadores clave para una reforzada toma de decisiones en los negocios. Su investigación se enfoca principalmente bajo una perspectiva de las aplicaciones de analítica In-Memory en el campo de la gestión de la cadena de suministro. Los conceptos relacionados que describen el análisis de data de negocios para objetivos de toma-de-decisiones han recibido una amplia atención tanto en la comunidad académica como de negocios. Los principales directivos ven a la analítica de negocio como un factor diferenciador para tomar ventaja competitiva y por ello están cada vez más interesados en capturar este potencial valor. Algunos estudios que cita el documento reportan un incremento de 5% a 6% en la



productividad de las compañías que colocan en el tercio superior de sus industrias al uso de la analítica de negocios. Este impacto positivo de las capacidades de la analítica de negocio en el rendimiento de la cadena de suministro ha sido también confirmado en varios estudios empíricos. Otros estudios indican el potencial disruptivo de la ‘toma de decisiones basada en datos’ para los negocios y para la gestión de operaciones en particular (Hahn & Packowski, 2015).

## **2.1 Bases de Datos In-Memory.-**

Muchas de las compañías, aunque tienen implementados soluciones de *Data Warehouse* y *Business Intelligence*, tienen grandes problemas en obtener la data a tiempo. La obtención de información en el momento correcto se torna en un trabajo extremadamente duro y en muchos de los casos debido a una inapropiada combinación de soporte de hardware y software. Las bases de datos tradicionales almacenan la data en disco y las operaciones de I/O son muy lentas comparadas con aquellas hechas en memoria RAM (Babeanu & Ciobanu, 2015).

La necesidad de procesamiento de grandes volúmenes de datos de manera muy rápida fue la principal razón que condujo al desarrollo de bases de datos In-Memory (IMDB). Estas bases de datos revolucionaron íntegramente el área de *Business Intelligence* y consiguieron resultados excepcionales exactamente donde las bases de datos clásicas habían fallado (Babeanu & Ciobanu, 2015).

Las IMDB constituyen un sistema de gestión de base de datos diseñadas para un alto rendimiento con la condición de que exista suficiente memoria para mantener la data necesaria en ella. Poseen una técnica de almacenamiento columnar lo que posibilita el acceso a la data a grandes velocidades y capacidades de analítica en tiempo-real. En comparación con *Cloud Computing*, el beneficio para el usuario es inmediatamente entendible, porque viene de un rápido análisis de grandes volúmenes de datos (Mihaela-Laura, 2014).

### **2.1.1 Características de diseño de una base de datos columnar In-Memory.-**

Para manejar la eficiencia en los sistemas empresariales era necesario separar la carga inserción-actualización (carga transaccional) de los trabajos analíticos. Esto obligaba a separar los sistemas, lo que resultaba en un alto costo por almacenamiento redundante, transferencia y sincronización entre sistemas. La tecnología In-Memory permite juntar los sistemas OLTP y OLAP. Por primera vez el tiempo, velocidad, costo y simplicidad apuntan en una misma dirección favorable (Plattner & Leukert, 2015).



En formato de filas, todos los atributos de entrada de una tabla son almacenados uno junto a otro como una secuencia, usando uno o múltiples bloques de memoria. En formato columnar, todos los valores de un atributo de la tabla son almacenados como un vector usando múltiples bloques de memoria y todos los vectores de atributos de una tabla son almacenados secuencialmente. Matemáticamente ambos formatos son comparables, pero su implementación técnica en memoria principal tiene cualidades muy diferentes. Al organizar los valores en la forma de un vector de atributos permite una fácil compresión de datos y también permite una alta velocidad de escaneo y filtraje. Esto resulta en mucho procesamiento secuencial donde el formato columnar tiene una enorme ventaja comparada con la tradicional base de datos de disco orientado a filas. Junto con la opción de procesamiento paralelo, se puede alcanzar una muy alta velocidad para filtraje o cualquier tipo de agregación (que constituyen algunas de las principales cargas en procesamiento analítico). La velocidad es en efecto tan alta que se puede dejar de lado la idea de pre-agregación de la data transaccional, la base de los sistemas de información en las décadas pasadas. Además ya no se requieren índices adicionales para acceso más rápido a la data (Plattner & Leukert, 2015).

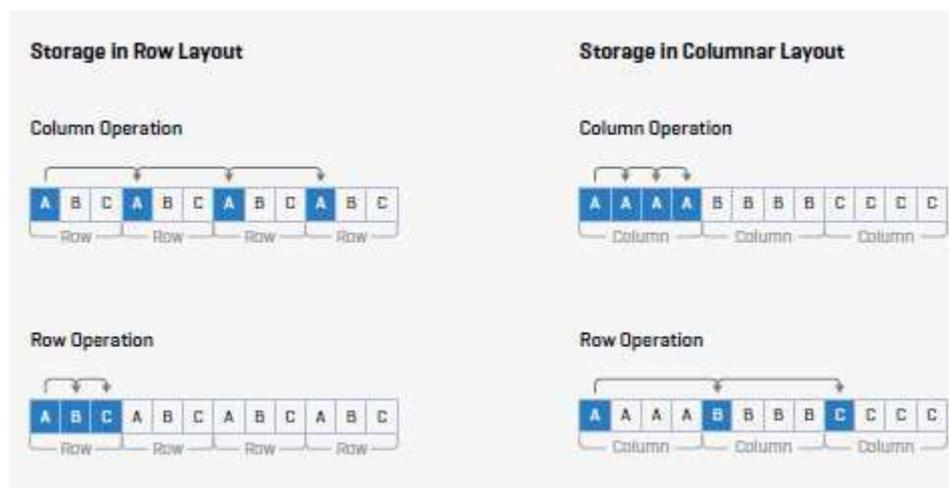


Figura 1. Operaciones de filas y columnas sobre un diseño de datos de filas y columnas

Fuente: (Plattner & Leukert, 2015)

### 2.1.2 Funcionalidades clave de un sistema IMDB.-

Algunas de las funcionalidades clave de un sistema IMDB son los siguientes (Sakulsorn,



2011):

- **Caching.**- Debido a la razón de un bajo rendimiento, todos los sistemas de gestión de bases de datos convencionales usualmente combinan la técnica de *caching* para mantener la data más recientemente usada en memoria. El método de *caching* también incluye el concepto de sincronización del caché, el cual asegura que una imagen de la data en caché sea consistente con la base de datos física y el *cache lookup* para justificar si la data requerida de la aplicación está en el cache o no. Sin embargo, con un sistema IMDB, estos procesos de *caching* han sido removidos; esto significa que la sobrecarga de rendimiento y los procesos complejos entre RAM y CPU son eliminados.
- **Proceso de transferencia de datos.**- En un sistema de base de datos convencional, el proceso de transferencia de datos es requerido antes de que cualquier aplicación pueda leer y modificar la data desde el almacenamiento de disco. El proceso puede ser ejemplificado en la siguiente figura:

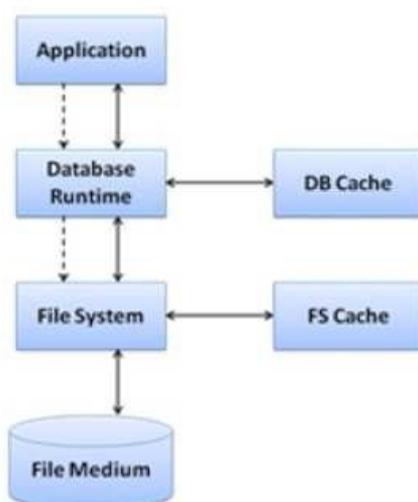


Figura 2. Flujo de datos en un sistema de gestión de base de datos convencional.

Fuente: (Sakulsorn, 2011)

Con las flechas continuas el gráfico ilustra la transferencia de datos, mientras que las flechas discontinuas representan la trayectoria del mensaje. El flujo es el siguiente:



- Una aplicación envía un requerimiento para el ítem de data solicitada hacia el motor de la base de datos.
  - El motor de la base de datos envía una instrucción al sistema de archivos para recuperación de data desde el almacenamiento físico.
  - El sistema de archivos genera una copia de la data para el  *caching*  y pasa otra copia al motor de la base de datos.
  - El motor de la base de datos mantiene una copia para su  *caching*  y pasa otra copia a la aplicación.
  - La aplicación hace cambios a su copia y pasa de regreso al motor de la base de datos.
  - El motor de la base de datos copia la data cambiada de regreso a su caché.
  - La data copiada en el caché del motor de base de datos es escrita en el sistema de archivos, y actualiza en el caché del sistema de archivos
  - Finalmente, la data modificada es escrita de regreso al almacenamiento físico.
- **Procesamiento de transacciones.**- En los sistemas convencionales de bases de datos existen mecanismos para prevenir pérdidas (como las originadas por la pérdida de poder eléctrico) tales como  *logs*  de transacciones y  *backups* . Igualmente, el sistema IMDB también debe gestionar la integridad de las transacciones. El proceso se hace manteniendo la imagen original de la data modificada y una lista de conteo de páginas durante la transacción. Cuando la transacción es confirmada por la aplicación, la imagen almacenada y las páginas referenciadas serán enviadas de regreso a la pila de memoria. En caso de que la transacción sea abortada, la imagen original será recuperada de la base de datos y las páginas referenciadas recientemente añadidas serán traídas de regreso a la memoria.

Una gran diferencia entre la IMDB y los sistemas tradicionales basados en disco es que la imagen en un IMDB se perderá si el sistema se enfrenta a una situación catastrófica (Sakulsorn, 2011).

### **2.1.3 Principios nucleares en el formato columnar de bases de datos In-Memory.-**

El formato columnar es una característica fundamental para llegar a obtener mayor velocidad de acceso. Está íntimamente ligado a la cercanía (o secuencialidad) de bloques en memoria y el desperdicio que significa esto cuando se tiene formato de filas (un caché de memoria perdido cuesta 100 nanosegundos). Con una codificación adecuada del



diccionario se puede reducir el almacenamiento por un factor de cinco. Además, esto permite que los procesos (agregación, escaneo, filtraje) puedan subdividirse en múltiples subprocesos y éstos puedan ejecutarse en procesamiento paralelo en diferentes *cores* del CPU. Tablas de gran volumen puede dividirse en particiones y residir en nodos de servidor separados. Luego con un proceso Map/Reduce se pueden trabajar tales particiones en paralelo (las operaciones de base de datos pueden llegar a correr entre 1,000 y 10,000 veces más rápido). Una base de datos columnar IMDB no necesita índices específicos, lo que a su vez requiere menor administración de la misma (Plattner & Leukert, 2015).

Un gestor importante para el almacenamiento de sistemas IMDB es la eficiencia en entrada/salida (I/O) bajo una variedad de condiciones. En razón de la continua disminución de precios y el incremento de la capacidad de memoria, ahora es posible llegar a asentar íntegramente bases de datos OLTP de terabytes en memoria. Las bases de datos IMDB pueden ejecutar las transacciones en orden de magnitud más rápido que las convencionales. Las bases de datos IMDB pueden reducir en forma muy alta la latencia de lectura, y aunque usualmente realizan menos escrituras I/O debido a la ausencia de árboles-B-residentes en disco, el procesamiento de transacciones de alto performance genera una correspondiente alta tasa de actualización I/O (Kim, Salem, & Khuzaima, 2015).

Las bases de datos In-Memory dejan de almacenar la data a nivel de disco, moviéndola en su lugar a un almacenamiento en la memoria. Estos cambios traen grandes ventajas, aunque también conllevan riesgos. Considerando que las principales características de estos sistemas están en el almacenamiento en memoria, lo cual puede conducir a una sorprendente velocidad de procesamiento. Las ventajas traídas por una base de datos In-Memory son, primeramente, debido al hecho que dejan de existir operaciones lentas de I/O en disco, tales como (Babeanu & Ciobanu, 2015):Muy rápido procesamiento de grandes volúmenes de datos

- Reducido tiempo de bloqueo de registros, debido a que todas las operaciones de *insert*, *update* y *delete* son hechas en tiempo real
- Consultas de datos muy rápidas

Los sistemas IMDB generalmente usan uno de los dos paradigmas para actualización persistente. El método *Update-In-Place* (UIP) escribe la data in-memory hacia el disco usando un mapeo directo (orden aleatorio); una variante es usar *snapshotting* para



optimizar el rendimiento través de la conversión de escritura de objetos de una manera secuencial (UIP-S). En el otro paradigma todos las actualizaciones (operaciones *update*) son añadidos a una serie de archivos de control (*checkpoints*) sobre el disco, se lo llama *Copy-On-Write* (COW-D) y tiene la ventaja que convierte I/O aleatorios en I/O secuenciales. Una variante es el COW-M, que escribe en memoria y también en almacenamiento persistente, lo que mejora el rendimiento (Kim, Salem, & Khuzaima, 2015). Estos autores han planeado un estudio para desarrollar un modelo de optimización considerando la eficiencia del I/O cuando se hacen actualizaciones a la IMDB. Para ello desarrollan un valor denominado “*write amplification*” que mide la relación entre I/O y las actualizaciones a la base de datos. La investigación también indica que existen estudios orientados a mejorar significativamente la gestión de almacenamiento persistente para sistemas IMDB, lo cual lleva a pensar en la escala de madurez que están teniendo estos sistemas (Kim, Salem, & Khuzaima, 2015).

Respecto a los métodos de acceso de datos, hay diferentes métodos de indexamiento, ya sea basados en funciones hash o en varios tipos de árboles. Los árboles-T son un tipo especial, creados particularmente para esas clases de bases de datos. Es un árbol balanceado y es especialmente optimizado para indexamiento de la data almacenada exclusivamente en memoria. La principal característica de este árbol es que el índice no almacena ningún dato, solo almacena un puntero al dato, debido a que todo permanece en memoria. Como se indicó, las bases de datos IMDB usan para la representación de la data el modelo de almacenamiento columnar. Este lleva a un eficiente almacenamiento de los datos, porque para la construcción de una tupla relacional se usan punteros. De esta manera el valor efectivo es almacenado solamente una vez en un diccionario y, en lugar de almacenar el mismo valor muchas veces en la memoria, las tuplas serán construidas almacenando solamente el puntero al valor. Si se tiene que tratar con grandes valores que aparecen muchas veces en la base de datos, el espacio de almacenamiento será significativamente mejorado. El rendimiento es la característica clave para las bases de datos In-Memory, y esto ya no es influenciado por las operaciones lentas de I/O de la memoria al disco y de regreso, ahora es solo basada en la velocidad de procesamiento (Babeanu & Ciobanu, 2015).

#### **2.1.4 Limitaciones de un sistema IMDB.-**

Con el fin de garantizar la confiabilidad para las transacciones de la base de datos, se necesita definir un conjunto de propiedades deseables las cuales son llamadas ACID:



atomicidad, consistencia, aislamiento y durabilidad. Las tres primeras características son comúnmente soportadas por un IMDB. Sin embargo, en la forma más simple de un sistema IMDB, la característica de durabilidad desaparece. La RAM o memoria principal, es un almacenamiento volátil; por tanto, requiere el suministro de energía para mantener la información almacenada de otro modo se perderá su contenido durante un *reset* o cuando haya un corte de energía. Entonces, el almacenamiento volátil podría causar vulnerabilidad y falta de durabilidad para el sistema BI. Sin embargo, para alcanzar la durabilidad del sistema, hay varias soluciones planteadas para soportar el almacenamiento de data In-Memory (Sakulsorn, 2011):

- **On-line Backup:** servicio de *backup* automático para la base de datos. Es la solución más simple, y al mismo tiempo provee durabilidad en mínimo grado.
- **Log de Transacciones:** mantiene la historia de las acciones ocurridas después de un *checkpoint* periódico. Estos archivos log deberían ser grabados en un dispositivo no-volátil, tal como un disco duro, con el propósito de garantizar la propiedad de durabilidad del sistema.
- **Replicación de la base de datos:** mantener copias replicadas de los datos se lo conoce como implementación de alta disponibilidad. El sistema de replicación de base de datos maneja la data distribuida sobre diferentes y múltiples nodos para asegurar que ésta no se perderá aún en el caso de la falla de un nodo.
- **NVRAM:** dispositivo de memoria no-volátil la cual es llamada RAM estática con un *backup* de suministro de energía. NVRAM es capaz de recuperar la base de datos incluso sobre un re-inicio. Diferente al *Transaction Logging* y la Replicación de Base de Datos, esta solución no involucra ninguna latencia I/O de disco ni sobrecarga de comunicación. Sin embargo, los fabricantes de base de datos comerciales raramente proveen esta tecnología NVRAM debido al hecho de que su costo es bastante caro aún.

## 2.2 Inteligencia de Negocios basada en bases de datos In-Memory.-

Aunque los sistemas de base de datos basados en disco (DRDB) pueden cachear la data en memoria principal, de todas maneras necesitan ser procesadas y almacenadas en varias capas en un almacenamiento primario residente en disco. En su lugar, un IMDB mantiene toda la data íntegra en memoria. La memoria principal es directamente accesada por el CPU(s) y la rapidez resulta en órdenes de magnitud más rápido. Los sistemas IMDB permiten una alta velocidad de procesamiento y un análisis de grandes volúmenes de



datos almacenando la data directamente en memoria principal, evitando consumo de tiempo en operaciones con disco (Knabke & Olbrich, 2015), (Hahn & Packowski, 2015). Se debe tomar en cuenta la constante reducción de precios para la memoria principal y el uso de técnicas de compresión que posibilitan mantener toda la data en memoria incluso de empresas de gran tamaño (Knabke & Olbrich, 2015). También, la emergencia de sistemas IMDB ha estado promovida por una mejora en los procedimientos de gestión de datos y una arquitectura multi-core que ha llegado a estar disponible recientemente (Hahn & Packowski, 2015).

La infraestructura de Inteligencia de Negocios basada en IMDB usa almacenamiento de datos orientado-a-columnas para soportar óptimamente las aplicaciones basadas en OLAP. Esta forma de almacenamiento además permite una mejor aplicación de técnicas de compresión y una enorme ganancia en rendimiento –sobre un factor de 1000 con datos reales (Knabke & Olbrich, 2015).

Las dificultades en proyectos de Inteligencia de Negocios (BI por sus siglas en Inglés) se derivan de varias causas. Aspectos tecnológicos respecto del uso de herramientas de BI descubren los problemas de herramientas demasiado complicadas para los usuarios operacionales, la escasez de escalabilidad en los sistemas, la insatisfacción con el rendimiento del software, y la codificación dura (quemada) para el cumplimiento de requerimientos de negocios. Las pruebas resultantes revelan que *Business Intelligence In-Memory* amplía el rendimiento de los sistemas BI convencionales mejorando el tiempo de carga del software y el tiempo de respuesta (Sakulsorn, 2011).

El enfoque del BI convencional es en su mayoría basado en OLAP y en herramientas de consulta a la base de datos (*query tools*) los cuales facilitan la extracción, clasificación, sumarización y despliegue de la data. La tecnología OLAP es importante para los negocios empresariales desde que permite funciones a través de vistas multi-dimensionales de los datos para tener un rápido, interactivo y consistente acceso a las consultas. La arquitectura OLAP y las herramientas basadas en consultas (*queries*) descansan en pre-cálculos de todas las medidas sobre la asociación de varias tablas separadas. Diferentes vistas dimensionales para la toma de decisiones gerenciales deben definirse previamente y todos los datos necesarios deben ser calculados y cargados en las tablas antes de cualquier requerimiento de los usuarios. En la actualidad, con las posibles tecnologías habilitadoras de rápida velocidad y el incremento de la cantidad de memoria en los procesadores, las herramientas de BI recientemente se han movido a una nueva



tecnología basada en arquitectura ‘In-Memory’ (Sakulsorn, 2011).

El *Business Intelligence* In-Memory es un método novedoso para el análisis de data que ha tomado fuerza en los últimos años aunque la tecnología IMDB ha existido desde hace aproximadamente 30 años. No obstante, la computación de 64-bits recién ha llegado a ser práctica lo que ha impulsado el BI In-Memory. En este escenario la data está simplemente mantenida y analizada dentro de la memoria principal (RAM), diferente al BI convencional el cual mantiene la data en almacenamiento en disco. Generalmente, no se pre-calculan todas las medidas en la forma en la que se hace en un BI convencional, en su lugar descansan en la velocidad de la memoria para dejar que los valores medidos sean calculados al momento (*‘on the fly’*) en cuanto sean necesarios. Sin embargo, no solamente la manera de almacenar la data en RAM es la diferencia, también el estilo de consulta varía de acuerdo a los diferentes fabricantes de software. Actualmente los enfoques BI In-Memory ofrecidos por varios fabricantes consisten de tres principales soluciones – motor de consulta de alta velocidad, cálculos *on-demand*, e interface de usuario visualmente interactiva (Sakulsorn, 2011).

Las herramientas de Inteligencia de Negocios In-Memory incorporan nuevas capacidades de alto rendimiento y provee de varios beneficios entre los que se cuenta la analítica acelerada, el procesamiento de cálculos en memoria, la factibilidad de una extensa y más amplia explotación de datos, la reducción de tiempos de espera para reportes; facilitando un mayor soporte para la toma de acciones y decisiones en el negocio. Además, el incremento en el desempeño de las herramientas de Inteligencia de Negocios permite analizar en forma rápida grandes volúmenes de datos, dando paso al análisis en tiempo real (Mihaela-Laura, 2014).

Una de las principales consideraciones de la tecnología In-Memory es que posibilita un mayor rendimiento, no solo en analítica y *reporting*, sino también en procesamiento transaccional con una estimación de cuadruplicar su velocidad para estas aplicaciones. Estas soluciones basadas en tecnología IM son capaces de mantener la data completa de sus sistemas en memoria, sin limitaciones, incluso para las más grandes empresas del mundo (Plattner & Leukert, 2015).

El motor de consultas rápido permite al área de Tecnología de la Información y a los usuarios operacionales consultar la data y presentarla inmediatamente. El motor categorizará la data en dos tipos, data relevante e irrelevante, de acuerdo a cada selección, mientras los cálculos *on-demand* se refieren al cálculo de los valores medidos los cuales



serán realizados, cuando hay un requerimiento de usuario. Finalmente, la interface de usuario visualmente interactiva provee un pre-construido modelo asociativo para la interacción de la data tal como tablas, gráficos, cuadros, *dashboards* los cuales usualmente cubren mucho de las mediciones principales del negocio. Los enfoques de gestión de bases de datos convencionales suelen ser demasiado lentos para procesar la información y de alto coste en términos de rendimiento del sistema. Con el fin de ayudar a extender las características de las aplicaciones de Inteligencia de Negocios, el software debe ser capaz de hacer frente a una gran cantidad de datos complejos de manera eficiente. Por lo tanto, sin la comunicación de dispositivos I/O para el procesamiento de información en disco, los sistemas de bases de datos en memoria proporcionan una mayor velocidad de acceso a los datos y tienen menos procesos de interacción que los sistemas de bases de datos convencionales (Sakulsorn, 2011).

Los sistemas clásicos de Inteligencia de Negocios usan tecnología OLAP para la agregación de la data las cuales son incrementalmente cargadas de varios sistemas OLTP en los denominados cubos y a través de intervalos definidos de tiempo. Con la tecnología In-Memory (IMDB), la separación artificial de OLAP y OLTP llega a ser obsoleta y la data transaccional está inmediatamente disponible para análisis en tiempo real (Hahn & Packowski, 2015).

### **2.2.1 Implicaciones para la investigación y la práctica industrial.-**

El estudio confirma la hipótesis que el análisis en tiempo real usando data transaccional –en contraste al convencional sistema de Inteligencia de Negocios basado en OLAP usando agregaciones– representa una novedad y una promesa en el campo de aplicaciones (empresariales) que están habilitadas por tecnología IMDB (Hahn & Packowski, 2015).

En la actualidad existe una gran disponibilidad de fabricantes de software In-Memory, muchos de ellos se han enfocado extensivamente en resolver los problemas de rendimiento en las herramientas convencionales de BI. Algunos fabricantes son: Qlik Inc., MS Power BI, Spotfire, SAP HANA (Sakulsorn, 2011).

### **2.2.2 Problemas en los BI convencionales.-**

Entre algunos de los obstáculos tecnológicos incluyen la dificultad en el uso de las herramientas del BI convencional, la necesidad de grandes cantidades de inversión inicial y los costos de mantener en marcha, caros costos de entrenamiento, y una complejidad para personalizar las herramientas para tipos particulares de usuarios. La revisión de la literatura indica que las limitaciones más críticas están relacionados con el problema de



integración. Varios tipos de problemas de integración sobre sistemas BI se pueden clasificar en problemas tecnológicos y gerenciales (Sakulsorn, 2011):

Obstáculos tecnológicos:

- Capacidad de inteligencia de datos insuficiente
- Inconveniente interacción de datos, arquitectura y soporte de flujo de trabajo

Obstáculos gerenciales:

- Insuficiente compromiso de negocio
- Escasez de cooperación entre usuarios TI y usuarios operacionales

Además de los problemas anteriores, muchas fuentes indican un significativo problema de calidad de data en las aplicaciones de BI. Una encuesta hecha por Gartner Group y BARC revela que la pobre calidad de data es el más común problema para las implementaciones BI.

### **3 RESULTADOS Y DISCUSIÓN.-**

La tecnología IM es significativamente diferentes de las tecnologías de análisis de datos tradicionales (Wixom, y otros, 2014). Los sistemas IMDB se están enfocando en la eficiencia de atributos tales como el I/O. Con ello se destaca los beneficios de la tecnología IMDB incluso en la actualización para mantener la persistencia de la data. No obstante, es claro notar que esta tecnología permite ventajas en la actualización de los datos en el disco por sobre los sistemas nativos de bases de datos tradicionales (en disco). Esto revela la potencial madurez a la que están llegando los sistemas IMDB en varios de sus aspectos estructurales (Kim, Salem, & Khuzaima, 2015).

La tecnología IM ha sido posible y ha podido expandirse y fortalecerse en base a dos grandes avances tecnológicos en la línea de hardware: la expansión de la arquitectura de procesadores multi-core (que permiten procesamiento paralelo) y el incremento de la capacidad de memoria en 64 bits. Ambas, favorecidas con una notable disminución de precios que sigue ocurriendo hasta el momento (Plattner & Leukert, 2015), (Mihaela-Laura, 2014).

Una características de mucho interés está data por la posibilidad de evitar la separación de conjuntos de datos transaccionales de los datos analíticos, lo que conllevaría a desechar prácticas casi impuestas en la implementación de proyectos de Inteligencia de negocios (Plattner & Leukert, 2015).

La posibilidad (efectuada ya en la práctica de aplicaciones empresariales) de mantener grandes volúmenes de datos transaccionales en forma completa e íntegra en memoria. Se



puede observar que en el campo de aplicaciones empresariales, esta tecnología IM ha sido completamente disruptiva y de un impensable impacto en su funcionamiento y en la potencialidad para el futuro (Plattner & Leukert, 2015). La velocidad que esta tecnología habilita, posibilitará el análisis de información en tiempo real, incluso en grandes volúmenes de datos. También permitirá una mejor integración de modelos de datos transaccionales a la analítica de procesos de negocio, lo que está siendo cada vez más adoptado por el segmento empresarial (Plattner & Leukert, 2015), (Hahn & Packowski, 2015), (Wixom, y otros, 2014).

Las bases de datos IMDB están orientadas a una estructura columnar para de esta manera asegurar un alto nivel de compresión de datos y velocidad de acceso. Adicionalmente, para indexamiento se usa un algoritmo especial, principalmente basado en árboles-T, para ofrecer un alto rendimiento (Babeanu & Ciobanu, 2015).

La tecnología In-Memory elimina los procesos convencionales lógicos, por ejemplo, tecnología ETL, creación de cubos de datos, y creación de modelos dimensionales, los cuales son muy conocidos y se han usado durante décadas. Además, el In-Memory BI parece ser apropiado para reaccionar de manera flexible con los constantes cambios y volatilidad en el medio de negocios tal como un método de cálculo al momento (*on the fly*). Como resultado, está en línea con métodos ágiles en los procesos de desarrollo de software que se enfocan en rápidos y flexibles respuestas al cambio, tanto como contribuir a la creación de cambio (Sakulsorn, 2011).

Estas ventajas tecnológicas que trae IM han sido adecuadamente aprovechadas por plataformas de Inteligencia de Negocios, en base a ello varias aplicaciones empresariales están teniendo notorios resultados (Mihaela-Laura, 2014).

La tecnología IMDB provee la plataforma con un modelo de datos integrado y con librerías funcionales que construyen las capacidades analíticas predictivas disponibles para un amplio grupo de usuarios no expertos (Hahn & Packowski, 2015).

#### **4 REFERENCIAS BIBLIOGRÁFICAS.-**

- Babeanu, R., & Ciobanu, M. (2015). In-memory databases and innovations in Business Intelligence. *Database Systems Journal*, VI(1), 9. Recuperado el 28 de Marzo de 2017
- Hahn, G. J., & Packowski, J. (Agosto de 2015). A perspective on application of In-Memory analytics in supply chain management. *Decision Support Systems*, 76, 8. Recuperado el 30 de Marzo de 2017, de <http://doi.org/10.1016/j.dss.2015.01.003>



- Kim, J., Salem, K., & Khuzaima, D. (2015). Write Amplification: An Analysis of In-Memory Database Durability Techniques. *IMDM '15 Proceedings of the 3rd VLDB Workshop on In-Memory Data Management and Analytics* (pág. 7). Kohala Coast, HI, USA: ACM Digital Library. doi:10.1145/2803140.2803141
- Knabke, T., & Olbrich, S. (2015). In Memory Technology and the Agility of Business Intelligence - A Case Study at a German Sportswear Company. *ResearchGate*, 18. Recuperado el 27 de Marzo de 2017, de <https://www.researchgate.net/publication/283794267>
- Mihaela-Laura, I. (2014). Characteristics of In-Memory Business Intelligence. *Informatica Economica*, 18(No. 3), 9. Recuperado el 27 de Marzo de 2017, de <https://www.researchgate.net/publication/284368743>
- Plattner, H., & Leukert, B. (2015). *The In-Memory Revolution*. Walldorf, Alemania: Springer. doi:10.1007/978-3-319
- Sakulsorn, P. (2011). *In-memory Business Intelligence*. KTH Royal Institute Of Technology, Department of Computer and Systems Sciences. Stockholm: KTH Royal Institute Of Technology. Recuperado el 04 de Abril de 2017
- Wixom, B., Ariyachandra, T., David, D., Goul, M., Gupta, B., & Others. (2014). The Current State of Business Intelligence in Academia: The Arrival of Big Data. *Communications of the Association for Information Systems*, 34(Article 1), 16. Recuperado el 25 de Marzo de 2017, de <http://aise1.aisnet.org/cais/vol34/iss1/1>