



Universidad Cenfotec

Maestría en Tecnologías de Bases de Datos

Documento final de Proyecto de Investigación Aplicada 2

Evaluación comparativa del modelo HTAP con el modelo OLAP para su utilización
en el área de inteligencia de negocios

Zelaya Rivas, Óscar Andrés

Mayo 2020

© 2020

Zelaya Rivas, Oscar Andrés

Se autoriza la reproducción total o parcial, con fines académicos, por cualquier medio o procedimiento, incluyendo la cita bibliográfica del documento.

Agradecimiento

Primeramente, agradezco a Dios por permitirme alcanzar una meta más ya que es mi principal fortaleza para superar los retos que cada día se presentan, así como la principal razón de las cosas buenas que pasan en mi vida.

Así mismo agradezco a mi familia por siempre estar ahí para mí y apoyarme en todo lo que me propongo y dando todo el apoyo y amor posible. De igual forma a mis amigos por tantos años de estar ahí impulsándonos entre nosotros para ser mejores.

Agradezco a la universidad y profesores por todo el conocimiento que me dieron, así como a José Cabezas por su ayuda como tutor de este trabajo. También a Sughey Carrillo por su gran ayuda como lectora del mismo.

TRIBUNAL EXAMINADOR

Este proyecto fue aprobado por el Tribunal Examinador de la carrera: **Maestría en Tecnología de Bases de Datos**, requisito para optar por el título de grado de **Maestría**, para el estudiante: **Oscar Andrés Zelaya Rivas**.

JOSE ALBERTO
CABEZAS
JAIKEL (FIRMA)

Digitally signed by JOSE
ALBERTO CABEZAS JAIKEL
(FIRMA)
Date: 2020.07.13 08:30:03
-06'00'

MBD. José A. Cabezas Jaikel
Tutor



MAP. Sugey Carrillo Contreras
Lector 1

IGNACIO
TREJOS ZELAYA
(FIRMA)

Digitally signed by
IGNACIO TREJOS ZELAYA
(FIRMA)
Date: 2020.07.13 21:16:46
-06'00'

M. Sc. Ignacio Trejos Zelaya
Lector 2

San José, Costa Rica, 10 de julio de 2020

Firmado digitalmente, de conformidad con la Ley de Certificados, Firmas Digitales y Documentos Electrónicos N° 8454, destacando el artículo 9°

TABLA DE CONTENIDO

Capítulo I. Introducción	1
1.1. Generalidades.....	1
1.2. Antecedentes del Problema	1
1.3. Definición y Descripción del Problema.....	1
1.4. Justificación.....	2
1.5. Viabilidad.....	3
1.5.1. Punto de Vista Técnico	3
1.5.2. Punto de Vista Operativo	3
1.5.3. Punto de Vista Económico.....	4
1.6. Objetivos	4
1.6.1. Objetivo General	5
1.6.2. Objetivos Específicos.....	5
1.7. Alcances y Limitaciones.....	5
1.7.1. Alcances	5
1.7.2. Limitaciones	5
1.8. Estado de la Cuestión	7
1.8.1. Formulación de la Pregunta	7
1.8.2. Selección de Fuentes	9
1.8.3. Selección de Estudios.....	10

1.8.4. Ejecución de la Selección	10
1.8.5. Extracción de Información.....	13
Capítulo II. Marco Conceptual	14
Capítulo III. Marco Metodológico.....	18
3.1. Tipo de Investigación	18
3.2. Alcance Investigativo	18
3.3. Enfoque.....	19
3.4. Diseño.....	19
3.5. Instrumentos de Recolección de Datos.....	20
3.5.1. Creación de Consultas para Obtener Medidas de Rendimiento	20
3.5.2. Iteraciones de Ejecuciones de Consultas	22
3.5.3. Herramientas Utilizadas para la Recolección de Medidas de Rendimiento.....	23
3.6. Técnicas de Análisis de Información.....	24
Capítulo IV. Análisis del Diagnóstico.....	26
4.1. Resultados Referentes a Métricas de Tiempo	26
4.2. Resultados Referentes a Operaciones de Entrada y Salida de Datos	27
4.3. Resultados Referentes al Porcentaje de Tiempo de Procesador Usado ...	30
4.4. Resultados Referentes al Uso de Memoria.....	31
4.5. Análisis de Implementaciones.....	32

Capítulo V. Propuesta de Solución.....	34
5.1. Diseño de la Base de Datos Transaccional.....	35
5.2. Carga de Datos Simulados	36
5.3. Diseño de Modelos de Analítica.....	39
5.3.1. Diseño OLAP	39
5.3.2. Diseño HTAP	41
Capítulo VI. Conclusiones y Recomendaciones.....	43
6.1. Conclusiones.....	43
6.2. Recomendaciones	46
6.2.1. Trabajos a Futuro (Líneas Futuras de Investigación).....	47
Referencias	48

LISTA DE FIGURAS

Figura 1 Código Powershell para la iteración de las consultas	23
Figura 2 Gráfico para comparación de medidas de duración.....	26
Figura 3 Gráfico para comparación de volúmenes de bytes por segundo	28
Figura 4 Gráfico para comparación de operaciones de datos por segundo	29
Figura 5 Gráfico para comparación de fallos de página por segundo	30
Figura 6 Gráfico para comparación de porcentaje de tiempo de procesador utilizado	31
Figura 7 Gráfico para comparación de conjuntos de páginas de memoria utilizados	32
Figura 8 Modelo relacional de la base de datos transaccional	36
Figura 9 Modelo relacional del Data Warehouse desarrollado	40

LISTA DE TABLAS

Tabla 1 Costos de investigación.....	4
Tabla 2 Artículos relevantes de ACM Digital Library	10
Tabla 3 Artículos relevantes de Science Direct.....	12
Tabla 4 Artículos relevantes de Sage Journals	12
Tabla 5 IEEE Xplore Digital Library	12
Tabla 6 Consultas de procesamiento analítico.....	20
Tabla 7 Valores límite de duración de las consultas	27
Tabla 8 Valores límite de utilización de conjuntos de páginas de memoria	32
Tabla 9 Descripción y número de registros de las tablas transaccionales	38

RESUMEN

El presente proyecto busca dictar el beneficio que se puede obtener al optar por la opción de Hybrid Transactional/Analytical Processing en lugar de los sistemas transaccionales y analíticos que se implantan por separado. Al considerarse una alternativa tan reciente, no existe información clara sobre aspectos de rendimiento o funcionalidades que se pueden obtener de este modelo. Por eso, se decidió llevar a cabo una evaluación comparativa que permita aclarar lo que este modelo ofrece en el aspecto analítico, en contraste con el tradicional On-line Analytical Processing.

Para esta comparación se implementaron sistemas que cumplieran el mismo propósito, pero con modelos de procesamiento analítico distintos. Cada implementación utilizó uno de los modelos mencionados y, en el caso de On-line Analytical Processing, se optó por dos de sus alternativas: la relacional y la multidimensional.

Palabras clave: HTAP, OLAP, MOLAP, ROLAP, OLTP, bases de datos, procesamiento analítico, inteligencia de negocios, in-memory.

Capítulo I. Introducción

1.1. Generalidades

Hybrid Transaction/Analytical Processing (HTAP) es una solución reciente para el manejo de sistemas transaccionales y analíticos en un único sistema, se introdujo como concepto en el 2014. Para el desarrollo del proyecto se trabajó con un sistema simulado que no pertenece a una entidad real, pero se procuró contar con diseños y volúmenes de datos que se aproximen a lo que se puede encontrar en una organización real que maneje ambos tipos de sistemas de bases de datos. El propósito del proyecto fue analizar y comparar el rendimiento y características que ofrecen ambas soluciones, pero sin dejar de lado la etapa de diseño e implementación que se necesita para las dos.

1.2. Antecedentes del Problema

Debido a lo reciente que es el modelo HTAP en comparación con la trayectoria que han tenido los sistemas On-Line Analytical Processing (OLAP), existen muy pocas evaluaciones sobre enfocadas en el rendimiento que este modelo puede ofrecer. Aunque se puede encontrar una cantidad considerable de literatura con fundamento académico sobre los mismos, la mayoría se enfoca en el modelo y su implementación sin valorar lo que puede ofrecer o los aspectos en los que puede quedar debiendo.

El crecimiento acelerado que presentan los volúmenes de datos en la actualidad ha sacado a relucir la necesidad de mejorar su procesamiento analítico. Por eso, se destaca la importancia de alternativas como la que se propone con HTAP.

1.3. Definición y Descripción del Problema

Cuando se habla de inteligencia de negocios y procesamiento analítico de datos es común que de inmediato se asocie el tema con un diseño OLAP. Desde 1970, cuando se desarrolló la primera herramienta multidimensional, se creó el

concepto OLAP, pero fue nombrado hasta inicios de los años 90. Esta antigüedad ha permitido que el modelo llegue a considerarse la base para el procesamiento analítico de datos en el mercado.

Hasta hace pocos años, la compañía de investigación Gartner propuso el modelo HTAP como una alternativa para el procesamiento analítico, presenta una opción que unifica los tradicionales sistemas transaccionales con los sistemas analíticos en uno solo. Se busca que el procesamiento analítico sea factible en vivo y sin la latencia que toma la transferencia completa de datos, así como la falta de sincronización que implica un sistema OLAP tradicional que requiere de un trabajo Extraction-Transformation-Load (ETL) intermedio entre ambos sistemas. Todo esto sin afectar al procesamiento transaccional y con un procesamiento analítico eficiente (Özcan, Tian y Tözün, 2017).

1.4. Justificación

OLAP ha sido una de las principales opciones para el procesamiento analítico de datos en organizaciones por muchos años. La historia y evolución que le ha acompañado por más de 40 años le da esta posición como solución de inteligencia de negocios. Sin embargo, el crecimiento acelerado y mejora que se tiene a nivel de *software* y de *hardware* en la actualidad ha generado la interrogante de si OLAP es la opción más acertada para este propósito.

Como una alternativa que busca sacar provecho de las opciones tecnológicas que se tienen en el mercado actualmente, surgió el concepto de HTAP como un híbrido entre los sistemas transaccionales On-line Transactional Processing (OLTP) y los analíticos OLAP, orientado al manejo de analítica utilizando datos *en vivo* y para mantener la calidad de rendimiento que pueden ofrecer estos sistemas por separado. Por lo reciente del modelo, las organizaciones no se atreven a migrar sus sistemas actuales por la escasez de información y la incertidumbre. Por esto, se planteó una comparación entre funcionalidades y aspectos de rendimiento que puede ofrecer el modelo HTAP al contrastarse con OLAP, este último es un punto de vista más familiar para la

comunidad informática lo que facilita la comparación.

Con esta evaluación comparativa se desea presentar una base más tangible que permitiera considerar la implementación de sistemas HTAP, de acuerdo con los resultados que se obtengan y evaluando aspectos que se puedan utilizar como referencia a partir de la necesidad y requerimientos que pueda tener cada organización.

1.5. Viabilidad

1.5.1. Punto de Vista Técnico

Para implementar los sistemas que se utilizaron en la comparación entre modelos se contó con el conocimiento básico de manejo de bases de datos necesarias. Además, para el sistema OLAP, se contó con experiencia en ambientes reales, así como conocimiento académico previo, tanto para el diseño como mantenimiento y manejo de este sistema, lo que facilitó esta implementación.

Por otro lado, para la implementación del sistema HTAP, se contó con cierta base del manejo de bases de datos en memoria, este es uno de los pilares para la implementación. Además, se llevó a cabo una investigación profunda sobre este modelo para obtener un sistema funcional y en búsqueda de que este fuera un punto de referencia para la comparación con el modelo OLAP implementado.

1.5.2. Punto de Vista Operativo

Al tratarse de un proyecto de investigación que utilizó una simulación de los dos tipos de sistemas, no existió algún impacto a nivel operativo debido a que no se manejaba un ambiente en producción, sino que los sistemas se implementaron desde cero, con simulación de datos y transacciones de acuerdo con el sistema implementado.

1.5.3. Punto de Vista Económico

En el aspecto económico, no existió ningún pago de personal o servicios debido a la naturaleza investigativa del proyecto y la forma en que se llevó a cabo la comparación entre ambos sistemas basándose en el desarrollo de una simulación. Asimismo, el *software* utilizado fue gratuito. De forma resumida, en la *Tabla 1* se presenta un resumen de costos que podría implicar esta investigación:

- Consultor para el diseño e implementación de los sistemas, así como de la preparación y ejecución de las consultas utilizadas para comparar los modelos: \$13 la hora, para un aproximado de 160 horas, para un total de \$2080.
- Alquiler del computador para el desarrollo del proyecto: \$300 por un mes de alquiler.
- Licencias: no es necesario ningún gasto en licenciamiento con el *software* que se plantea.

Tabla 1 *Costos de investigación*

	Costo	Lapso	Total
Consultor	\$13/hora	160 horas	\$2080
Alquiler de equipo	\$300	1 mes	\$300
Licencias de software	-	-	-

Fuente: elaboración propia.

1.6. Objetivos

La taxonomía cognitiva utilizada en el presente proyecto fue la original de Bloom, debido a que se buscó comparar por medio de un ejemplo de uso y el análisis de sus resultados y los modelos de bases de datos enfocados en inteligencia de negocio propuestos. Esto mediante la expansión de conocimiento, lo cual se logró de manera natural con esta taxonomía jerárquica.

1.6.1. Objetivo General

Valorar las características y medidas de rendimiento que ofrece el modelo de bases de datos HTAP en su implementación para inteligencia de negocios comparado con el tradicional modelo OLAP.

1.6.2. Objetivos Específicos

1. Describir los aspectos que caracterizan ambos modelos como soluciones de bases de datos para inteligencia de negocios.
2. Proponer un escenario con datos ficticios como ejemplo de uso para la implementación de ambos modelos.
3. Desarrollar una simulación de procesamiento analítico basado en iteraciones que contemple el estado real en el que podría encontrarse cada uno de los sistemas propuestos.
4. Cuantificar medidas de rendimiento que presenta cada uno de los modelos respecto al uso de recursos del sistema.
5. Comparar los resultados que ofrece cada uno de los modelos en términos de métricas de rendimiento y características de implementación y manejo.

1.7. Alcances y Limitaciones

1.7.1. Alcances

Como resultado de la investigación e implementación de los sistemas se presentan los resultados de los indicadores de rendimiento medidos, así como su interpretación, y de la comparación descriptiva respecto a implementación y manejo de ambos sistemas.

1.7.2. Limitaciones

Por otra parte, entre las limitaciones del proyecto, la implementación de los sistemas se limitó a lo necesario para el procesamiento analítico y funcionamiento

regular, tanto del sistema OLAP como del sistema HTAP. No se pudo contar con ambientes completamente en producción por el alcance evaluativo de este proyecto.

1.8. Estado de la Cuestión

1.8.1. *Formulación de la Pregunta*

1.8.1.1. Foco de la Pregunta. El propósito de esta revisión fue encontrar estudios e investigaciones que fueron realizados referentes a los modelos HTAP y OLAP en términos de funcionalidades y rendimiento de ambos, así como posibles comparaciones que se había hecho entre ambos sobre estos aspectos.

1.8.1.2. **Amplitud y Calidad de la Pregunta.**

1.8.1.2.1. Definición del Problema. En la actualidad, los datos son considerados un activo sumamente importante para las organizaciones y esta importancia se relaciona con el crecimiento acelerado que poseen los volúmenes de datos. Por esto, se presenta la solución HTAP como una alternativa al modelo OLAP, ya que la misma elimina la necesidad de un sistema aparte para la analítica de negocio, juntando este sistema con el transaccional en uno solo. Debido al poco tiempo que tiene HTAP en el mercado, se buscó tener una percepción más detallada sobre la comparación entre ambos modelos para tener clara la viabilidad de un sistema HTAP y su comparación de utilidad con uno OLAP

1.8.1.2.2. Pregunta de Investigación. ¿Cuáles son los posibles beneficios o desventajas del uso de un sistema HTAP al compararse con un sistema OLAP en aspectos de funcionalidades y rendimiento?

1.8.1.2.3. Palabras Clave. A continuación, se presenta una lista de palabras clave utilizadas para la búsqueda en los repositorios seleccionados, tanto en inglés como en español:

- Hybrid Transactional Analytical Processing
- HTAP
- Online Analytical Processing
- OLAP
- Análisis de datos
- Analytics Database
- Columnstore
- In-memory Database
- Base de datos híbrida

1.8.1.2.4. Cadenas de Búsqueda Utilizadas. A continuación se presentan las cadenas de texto utilizadas en las fuentes utilizadas para esta investigación.

- “HTAP”+“OLAP”.
- “HTAP”.
- “Hybrid transactional analytical processing”.
- “OLAP”.
- “Online Analytical Processing”.
- “Data Analysis”.
- “Análisis de datos”.
- “Base de datos híbrida”.
- “Hybrid database”.
- “In-memory Database”.
- “Columnstore”+ “database”.
- “Data Analytics”.

1.8.2. Selección de Fuentes

1.8.2.1. Definición del Criterio de Selección de Fuentes. Para la selección de fuentes, los repositorios a utilizarse se tomaron en consideración de acuerdo con la importancia que poseen los mismos en el ambiente informático y para la comunidad.

1.8.2.1.1. Lenguajes de los Estudios. Para contar con un mayor volumen de fuentes y debido a que muchas de las investigaciones y los textos informáticos son escritos en inglés, se utilizaron tanto palabras en inglés como en español para la investigación.

1.8.2.1.2. Identificación de Fuentes. Como punto de partida para la identificación de fuentes, se tomaron las palabras clave mencionadas con el estudio del contenido de las fuentes y sus descripciones.

1.8.2.1.3. Método de Búsqueda en las Fuentes. Mediante los repositorios clave mencionados y las palabras clave, se procedió con la búsqueda de las fuentes de esta investigación, asegurando la confiabilidad de los repositorios, así como la calidad de los escritos.

1.8.2.1.4. Lista de Fuentes. A continuación se presentan los repositorios seleccionados para la obtención de fuentes:

- ACM Digital Library.
- Science Direct.
- Sage Journals.
- IEEE Xplore Digital Library.

1.8.3. Selección de Estudios

1.8.3.1. Procedimiento para la Selección de Estudios. Las búsquedas se basaron en las palabras clave mencionadas para la obtención de fuentes relacionadas con este estudio.

1.8.3.2. Definición de Criterios de Inclusión y Exclusión de Estudios. Los criterios de inclusión iniciales se realizaron basados en el título y el resumen presentados, seleccionando los que se ajusten a los propósitos del proyecto. Posteriormente, se estudió más a fondo el contenido de las fuentes incluidas para terminar de descartar aquellas que no fueron útiles para el proyecto.

1.8.4. Ejecución de la Selección

A continuación se presenta la selección inicial de fuentes realizada en los repositorios listados. En la *Tabla 2* se encuentra la selección de ACM Digital Library, prosiguiendo con Science Direct (*Tabla 3*), Sage Journals (*Tabla 4*) y Pro Quest Ebook Central (*Tabla 5*).

Tabla 2 Artículos relevantes de ACM Digital Library

Artículo

Título: HTAPBench: Hybrid Transactional and Analytical Processing *Benchmark*.

Autor: Fábio Coelho, João Paulo, Ricardo Vilaça, José Pereira, Rui Oliveira.

Cita: Coelho, F., Paulo, J., Vilaça, R., Pereira, J. y Oliveira, R. (2017). HTAPBench. *Proceedings Of The 8Th ACM/SPEC On International Conference On Performance Engineering-ICPE '17*. doi: 10.1145/3030207.3030228

Link: <https://dl.acm.org/citation.cfm?doid=3030207.3030228>

Título: Bridging the Archipelago between Row-Stores and Column-Stores for Hybrid Workloads

Autor: Joy Arulraj, Andrew Pavlo, Prashanth Menon.

Cita: Arulraj, J., Pavlo, A. y Menon, P. (2016). Bridging the Archipelago between Row-Stores and Column-Stores for Hybrid Workloads. *Proceedings Of The 2016 International Conference On Management Of Data-SIGMOD '16*. doi: 10.1145/2882903.2915231

Link: <https://dl.acm.org/citation.cfm?doid=2882903.2915231>

Título: Columnstore and B+ tree-Are Hybrid Physical Designs Important?

Autor: Adam Dziedzic, Jingjing Wang, Sudipto Das, Bolin Ding, Vivek R. Narasayya, Manoj Syamala.

Cita: Dziedzic, A., Wang, J., Das, S., Ding, B., Narasayya, V. & Syamala, M (2018). Columnstore and B+ tree-Are Hybrid Physical Designs Important? *Proceedings Of The 2018 International Conference On Management Of Data-SIGMOD '18*. doi: 10.1145/3183713.3190660

Link: <https://dl.acm.org/citation.cfm?doid=3183713.3190660>

Título: BatchDB: Efficient Isolated Execution of Hybrid OLTP+OLAP Workloads for Interactive Applications

Autor: Darko Makreshanski, Jana Giceva, Claude Barthels, Gustavo Alonso.

Cita: Makreshanski, D., Giceva, J., Barthels, C. & Alonso, G (2017). BatchDB. *Proceedings Of The 2017 ACM International Conference On Management Of Data-SIGMOD '17*. doi: 10.1145/3035918.3035959

Link: <https://dl.acm.org/citation.cfm?doid=3035918.3035959>

Título: Hybrid Transactional/Analytical Processing: A Survey

Autor: Fatma Özcan, Yuanyuan Tian, Pinar Tözün.

Cita: Özcan, F., Tian, Y. y Tözün, P. (2017). Hybrid Transactional/Analytical Processing. *Proceedings Of The 2017 ACM International Conference On Management Of Data-SIGMOD '17*. doi: 10.1145/3035918.3054784

Link: <https://dl.acm.org/citation.cfm?doid=3035918.3054784>

Título: Benchmarking database design for mixed OLTP and OLAP workloads

Autor: Anja Bog, Kai Sachs, Alexander Zeier

Cita: Bog, A., Sachs, K. y Zeier, A. (2011). *Benchmarking Database Design for Mixed OLTP and OLAP Workloads*. *Proceedings Of The 2Nd ACM/SPEC International Conference On Performance Engineering*. doi: 10.1145/1958746.1958806

Link: <https://dl.acm.org/citation.cfm?doid=1958746.1958806>

Título: Data warehousing and OLAP over big data: current challenges and future research directions

Autor: Alfredo Cuzzocrea, Ladjel Bellatreche, Il-Yeol Song

Cita: Cuzzocrea, A., Bellatreche, L. & Song, I (2013). Data warehousing and OLAP over big data. *Proceedings Of The Sixteenth International Workshop On Data Warehousing And OLAP - DOLAP '13*. doi: 10.1145/2513190.2517828

Link: <https://dl.acm.org/citation.cfm?doid=2513190.2517828>

Título: Characterization of hierarchies and some operators in OLAP environment

Autor: Elaheh Pourabbas, Maurizio Rafanelli.

Cita: Pourabbas, E. & Rafanelli, M. (1999). Characterization of hierarchies and some operators in OLAP environment. *Proceedings Of The 2Nd ACM International Workshop On Data Warehousing And OLAP - DOLAP '99*. doi: 10.1145/319757.319790

Link: <https://dl.acm.org/citation.cfm?doid=319757.319790>

Título: An in-depth analysis of data aggregation cost factors in a columnar in-memory database

Autor: Stephan Müller, Hasso Plattner

Cita: Müller, S. & Plattner, H. (2012). An in-depth analysis of data aggregation cost factors in a columnar in-memory database. *Proceedings Of The Fifteenth International Workshop On Data Warehousing And OLAP - DOLAP '12*. doi: 10.1145/2390045.2390057

Link: <https://dl.acm.org/citation.cfm?doid=2390045.2390057>

Título: Aggregation strategies for columnar in-memory databases in a mixed workload

Autor: Stephan Müller, Hasso Plattner

Cita: Müller, S. & Plattner, H. (2011). Aggregation strategies for columnar in-memory databases in a mixed workload. *Proceedings Of The 4Th Workshop On Workshop For Ph. D. Students In Information y Knowledge Management-PIKM '11*. doi: 10.1145/2065003.2065015

Link: <https://dl.acm.org/citation.cfm?doid=2065003.2065015>

Título: A common database approach for OLTP and OLAP using an in-memory column database

Autor: Hasso Plattner

Cita: Plattner, H. (2009). A common database approach for OLTP and OLAP using an in-memory column database. *Proceedings Of The 35Th SIGMOD International Conference On Management*

Of Data-SIGMOD '09. doi: 10.1145/1559845.1559846

Link: <https://dl.acm.org/citation.cfm?doid=1559845.1559846>

Título: Real-time analytical processing with SQL server

Autor: Per-Åke Larson, Adrian Birka, Eric N. Hanson, Weiyun Huang, Michal Nowakiewicz, Vassilis Papadimos

Cita: Larson, P., Birka, A., Hanson, E., Huang, W., Nowakiewicz, M. y Papadimos, V (2015). Real-time analytical processing with SQL server. *Proceedings Of The VLDB Endowment*, 8(12), 1740-1751. doi: 10.14778/2824032.2824071

Link: <https://dl.acm.org/citation.cfm?doid=2824032.2824071>

Fuente: elaboración propia.

Tabla 3 *Artículos relevantes de Science Direct*

Artículo

Título: A performance evaluation of in-memory databases

Autor: Abdullah Talha Kabakus, ResulKarab.

Cita: Kabakus, A. & Kara, R. (2017). A performance evaluation of in-memory databases. *Journal Of King Saud University-Computer And Information Sciences*, 29(4), 520-525. doi: 10.1016/j.jksuci.2016.06.007

Link: <https://www.sciencedirect.com/science/article/pii/S1319157816300453?via%3Dihub>

Fuente: elaboración propia.

Tabla 4 *Artículos relevantes de Sage Journals*

Artículo

No encontrados.

Fuente: elaboración propia.

Tabla 5 *IEEE Xplore Digital Library*

Artículo

Título: A Survey on Trajectory Data Management for Hybrid Transactional and Analytical Workloads

Autor: Keven Richly

Cita: Richly, K. (2018). A Survey on Trajectory Data Management for Hybrid Transactional and Analytical Workloads. *2018 IEEE International Conference On Big Data (Big Data)*. doi: 10.1109/bigdata.2018.8622394

Link: <https://ieeexplore.ieee.org/document/8622394>

Título: HyPer: A hybrid OLTP&OLAP main memory database system based on virtual memory snapshots

Autor: Alfons Kemper & Thomas Neumann

Cita: Kemper, A. & Neumann, T. (2011). HyPer: A hybrid OLTP& OLAP main memory database system based on virtual memory snapshots. *2011 IEEE 27Th International Conference On Data Engineering*. doi: 10.1109/icde.2011.5767867

Link: <https://ieeexplore.ieee.org/document/5767867>

Título: A framework for simulating combined OLTP and OLAP workloads

Autor: Anja Bog, Mathias Domschke, Juergen Mueller, Alexander Zeier

Cita: Bog, A., Domschke, M., Mueller, J. & Zeier, A. (2009). A framework for simulating combined

OLTP and OLAP workloads. *2009 16Th International Conference On Industrial Engineering And Engineering Management*. doi: 10.1109/icieem.2009.5344329

Link: <https://ieeexplore.ieee.org/document/5344329>

Título: RC-NVM: Dual-Addressing Non-Volatile Memory Architecture Supporting Both Row and Column Memory Accesses

Autor: Shuo Li, Nong Xiao, Peng Wang, Guangyu Sun, Xiaoyang Wang, Yiran Chen, Hai Helen Li, Jason Cong, Tao Zhang

Cita: Li, S., Xiao, N., Wang, P., Sun, G., Wang, X., Chen, Y. *et al.* (2019). RC-NVM: Dual-Addressing Non-Volatile Memory Architecture Supporting Both Row and Column Memory Accesses. *IEEE Transactions On Computers*, 68(2), 239-254. doi: 10.1109/tc.2018.2868368

Link: <https://ieeexplore.ieee.org/document/8453833>

Título: Data structures for mixed workloads in in-memory databases

Autor: Jens Krueger, Martin Grund, Martin Boissier, Alexander Zeier, Hasso Plattner.

Cita: Krueger, J., Grund, M., Boissier, M., Zeier, A. y Plattner, H. (2010). Data structures for mixed workloads in in-memory databases. *5Th International Conference On Computer Sciences And Convergence Information Technology*. doi: 10.1109/iccit.2010.5711090

Link: <https://ieeexplore.ieee.org/document/5711090>

Fuente: elaboración propia.

1.8.5. Extracción de Información

1.8.5.1. Definición de Criterios de Inclusión y Exclusión de Información. Con base en la selección de fuentes se escogió la información que se encontraba alineada al propósito de este proyecto, mediante la relación de contenido que cada texto posee. Esto según los distintos conocimientos necesarios para el desarrollo del proyecto, a partir de los títulos y resúmenes de esta relación inicial.

Capítulo II. Marco Conceptual

Al tratarse de una investigación evaluativa, fue necesario el desarrollo de un marco conceptual que permitiera definir ciertos conceptos clave para su desarrollo, así como el proceso de implementación y comparación de los sistemas utilizados para la obtención de medidores y contrastación de características de cada uno.

El procesamiento de datos corresponde al manejo organizado de registros de generar información. Este implica una serie de etapas que van desde la recolección, para después pasar a su procesamiento que incluye estandarización de los datos, corrección de errores, así como su estructuración y almacenamiento. El propósito es presentarlos de forma que no sean solamente una serie de valores presentados, sino que tengan un sentido y que permitan representar aspectos de la vida real (Thomsen, 2002).

Para este procesamiento entran en juego las bases de datos. Una base de datos puede ser estructurada o no. Para efectos de este proyecto, hubo un enfoque en las bases de datos estructuradas, debido a que los diseños HTAP y OLAP se basan en estas. El propósito de las bases es el almacenamiento que forma parte del procesamiento de datos, por lo tanto, son las que poseen la estructura en la que estos se almacenan. Entre los tipos de bases de datos estructuradas se manejan las relacionales comunes que utilizan Structured Query Language (SQL) y las No-SQL.

En primer lugar, las bases de datos relacionales comunes se basan en la relación entre entidades que están conformadas por conjuntos de atributos. Esta relación es la que permite convertir los datos en información y estos se almacenan en tuplas o filas de una tabla que corresponde a las entidades y cuyas columnas representan los atributos de estas entidades. Para el manejo de estas bases de datos, se cuenta con el lenguaje SQL que se basa en consultas que permiten manipular y observar los datos almacenados en las tablas (Beaulieu, 2009). SQL, a la vez, se puede clasificar en los siguientes sublenguajes:

- Data Definition Language (DDL): para manejo de estructuras a nivel de

base de datos.

- Data Query Language (DQL): para consultas que obtienen los datos deseados.
- Data Manipulation Language (DML): para manipular los datos.
- Data Control Language (DCL): específico para permisos que se tienen sobre las estructuras y los datos.

Por otro lado, se pueden encontrar las bases de datos No-SQL, cuyo nombre corresponde a Not-Only-SQL, traducido a *no solamente SQL* (Vaish, 2013). Estas bases son un poco más flexibles para la manipulación de datos y sus estructuras, pero pueden tener limitaciones por esta combinación. Sus principales diferencias corresponden al estilo de almacenamiento y ciertas formas de manipular los datos que se salen de las conocidas en el marco SQL. En el marco No-SQL, existen las bases de datos columnares que son similares a las relacionales tradicionales, pero con la diferencia de que los registros se arreglan por columna y no por filas. Esto permite un manejo computacional más potente para grandes volúmenes de datos.

Para este proyecto, junto a estas bases de datos columnares, se trató el concepto de bases de datos en memoria (In-memory), estas implican el manejo en memoria constante de ciertas estructuras de datos, enfocadas principalmente en mejorar el rendimiento de las transacciones que toman lugar en la base de datos (Müller y Plattner, 2011). Asimismo, las bases de datos pueden clasificarse según el propósito que van a cumplir, delimitado al área de investigación del presente proyecto, se trataron los conceptos de procesamiento transaccional y procesamiento analítico en las bases de datos.

El procesamiento transaccional se enfoca en el manejo de datos en vivo que conforman un sistema informático y son los actuales para cada entidad que se representa, este es el punto en el que se manejan las bases de datos OLTP. Por

otro lado, el procesamiento analítico se enfoca en el análisis de los datos para generar información más detallada que, a la vez, permite soportar procesos de toma de decisiones con base en datos históricos almacenados, entrando en consideración el concepto OLAP (Thomsen, 2002).

Para el procesamiento analítico en el marco OLAP se manejan los conocidos Data Warehouses que corresponden a un tipo de base de datos relacional que los maneja por fecha y con una categorización más global para enfocarse en el procesamiento más agilizado de los grandes volúmenes de datos (Cuzzocrea, Bellatreche y Song, 2013). Para los almacenes, existen propuestas de diseño que corresponden al diseño estrella y el copo de nieve. En el caso del diseño estrella, se maneja una relación directa entre las dimensiones y la tabla de hechos, mientras que en el copo de nieve existe cierta normalización de dimensiones, por lo que existen ciertas dimensiones que necesitan otra intermediaria para relacionarse con las tablas de hechos (Jensen, Bach y Thomsen, 2010).

Para esto, se manejan dos conceptos clave que son las dimensiones y los hechos. Las dimensiones se encargan de manejar las cualidades de los datos mientras que los hechos se encargan de los aspectos cuantificables de los datos, con su marca de fecha histórica (Jensen, Bach y Thomsen, 2010).

Aunque los sistemas OLAP manejan los datos históricos de un sistema. En este punto los procesos ETL son determinantes debido a que son los encargados de esta carga de datos. El concepto está conformado por tres partes:

- Extracción: se toman los datos del sistema OLTP.
- Transformación: se hacen los cambios necesarios sobre estos datos. para que se acomoden a la estructura creada previamente en el sistema OLAP.
- Carga: se introducen los datos en el sistema OLAP para su uso posterior, se considera una marca de tiempo, ya que estos corresponden a datos históricos.

Finalmente, en el concepto HTAP convergen los demás conceptos detallados, esta es una propuesta que busca combinar en un único sistema, tanto el área transaccional como la analítica. Esto busca la eliminación del proceso ETL intermedio que genera una latencia que impide al lado analítico mantener información lo más en línea posible (Coelho, Paulo, Vilaça, Pereira y Oliveira, 2017). En pocas palabras, es la combinación de un sistema OLTP con uno OLAP, esto aprovecha la tecnología actual y conceptos como las bases de datos en memoria que corresponden a un pilar de HTAP para su implementación.

Capítulo III. Marco Metodológico

3.1. Tipo de Investigación

El tipo de investigación que se desarrolló en el presente proyecto corresponde a una investigación evaluativa, debido a que el propósito de esta fue analizar los aspectos que puede ofrecer el modelo HTAP, específicamente en cuanto a procesamiento analítico de datos. Para esta evaluación se comparó el modelo más tradicional y universal OLAP de forma tal que esta evaluación pudiera partir de un punto familiar y tener una noción más acertada de los aspectos, ya sean positivos o negativos, que el modelo HTAP ofrece.

El propósito de la investigación no fue generar conocimiento nuevo sobre los modelos HTAP, sino que se enfocó en evaluar este modelo. Como se mencionó, se comparó la implementación que simula un escenario real de negocio.

3.2. Alcance Investigativo

El alcance investigativo fue inicialmente exploratorio, debido a que lo que se pretendió fue estudiar el modelo HTAP y sus aspectos positivos o negativos que pudieran surgir al compararse con el modelo OLAP. En la actualidad, existe información y literatura referente al modelo HTAP, pero esta no se enfoca en beneficios o desventajas de utilizar este modelo en aspectos de funcionalidades y rendimiento del sistema. Por eso, esta comparativa entre ambos modelos es necesaria para conocer lo que HTAP puede ofrecer en procesamiento analítico de datos.

De igual forma, este alcance investigativo evolucionó a uno descriptivo que se encargó de caracterizar las cualidades del modelo HTAP. Esto con base en los resultados de la comparación y estudio de ambas implementaciones y sus resultados en la evaluación de rendimiento.

3.3. Enfoque

Para este proyecto se utilizó un enfoque mixto, pero centrado principalmente en el aspecto cualitativo. Por parte del aspecto cuantitativo se analizaron medidas de rendimiento de cada uno de los modelos, de forma que se pudieran evaluar los marcadores de rendimiento que ofrece cada uno de los modelos y su comparación de resultados.

Por otro lado, para el aspecto cualitativo se evaluaron ventajas y desventajas que ofrecían un modelo u otro en cuanto a aspectos de implementación o de funcionalidad. De igual forma, se evaluó la manera en las que estas características podrían significar la selección de alguno de los modelos de acuerdo con las necesidades de procesamiento analítico que pudieran existir.

3.4. Diseño

Para este proyecto, se manejó un diseño experimental en el que se plantearon dos sistemas de gestión de bases de datos con el mismo propósito de procesamiento analítico, pero con un modelo de implementación distinto que representa el objeto de estudio. Los sistemas implementados eran ficticios, pero en aspectos de diseño y cargas de datos, se procuró que simularan con detalle sistemas reales para obtener resultados con la mayor confiabilidad posible. Para esto, se desarrolló un modelo relacional común para el área transaccional y un modelo de Data Warehouse para la alternativa OLAP y se crearon proyectos de inteligencia de negocio para la posterior comparación entre ambos. En el modelo HTAP también se considera cierta carga transaccional que este maneja por su naturaleza en el momento en que se obtuvieron los medidores de cada sistema.

En el aspecto de datos que representan información tradicional descriptiva del negocio en el que se desarrolló, se utilizaron arreglos de datos estándar para la industria del negocio. Por otro lado, para los datos específicos de los sistemas, se usaron datos aleatorios generados con código SQL mediante el manejo de cursores en combinación con funciones de generación aleatoria propios del motor en el que se implementaron las soluciones. Una vez implementados, se plantearon

iteraciones de ejecución de consultas que sirvieran como base para la recolección de medidas de rendimiento deseadas para la comparación.

3.5. Instrumentos de Recolección de Datos

La comparación de medidas de rendimiento entre los sistemas OLAP y HTAP, se basó en una simulación de un sistema transaccional. Además, para el modelo OLAP se consideraron dos propuestas diferentes que el mismo modelo ofrece, es decir, Multidimensional Online Analytical Processing (MOLAP) y Relational Online Analytical Processing (ROLAP). Por otro lado, para la solución HTAP se planteó un modelo analítico tabular que utilizó la opción DirectQuery de SQL Server Analysis Services (SSAS) que implica que los datos se toman directamente del sistema transaccional.

3.5.1. Creación de Consultas para Obtener Medidas de Rendimiento

Para la obtención de las medidas de rendimiento que se usaron para la comparación cuantitativa entre los modelos, se elaboraron tres consultas MultiDimensional eXpression (MDX) diferentes de orientación analítica que se aplicaron con las tres soluciones desarrolladas. Estas consultas se pueden apreciar en la *Tabla 6*.

Tabla 6 Consultas de procesamiento analítico

Descripción de la consulta	Consulta MDX
Cuenta de casos por región, nivel de severidad, estatus de agente encargado y tipo de contrato. Excluidos los casos cerrados.	<p>MOLAP/ROLAP:</p> <pre> SELECT NON EMPTY { [Measures].[Fact Caso Soporte Count] } ON COLUMNS, NON EMPTY { ([Dim Region].[Nombre Region].[Nombre Region].ALLMEMBERS * [Dim SLA].[Nombre Nivel Sev].[Nombre Nivel Sev].ALLMEMBERS * [Dim Agente].[Activo].[Activo].ALLMEMBERS * [Dim Contrato].[Nombre Tipo].[Nombre Tipo].ALLMEMBERS) } DIMENSION PROPERTIES MEMBER_CAPTION, MEMBER_UNIQUE_NAME ON ROWS FROM (SELECT (-{ [Dim Estado Caso].[Nombre Estado].[Cerrado], [Dim Estado Caso].[Nombre Estado].[All].UNKNOWNMEMBER }) ON COLUMNS FROM [Sistema Soporte DWH]) CELL PROPERTIES VALUE, BACK_COLOR, FORE_COLOR, FORMATTED_VALUE, FORMAT_STRING, FONT_NAME, FONT_SIZE, FONT_F </pre> <p>HTAP:</p>

Descripción de la consulta	Consulta MDX
	<pre> SELECT NON EMPTY Hierarchize(DrilldownMember(DrilldownMember(DrilldownMember(CrossJoin({[Region].[nombreRegion].[All] ,[Region].[nombreRegion].[nombreRegion].AllMembers },{[NivelSeveridad].[NombreNivel].[All] ,[Agente].[Activo].[All],[TipoContrato].[NombreTipo].[All]})) ,[Region].[nombreRegion].[nombreRegion].AllMembers ,[NivelSeveridad].[NombreNivel] ,[NivelSeveridad].[NombreNivel].[NombreNivel].AllMembers ,[Agente].[Activo] ,[Agente].[Activo].[Activo].AllMembers ,[TipoContrato].[NombreTipo])) DIMENSION PROPERTIES PARENT_UNIQUE_NAME,HIERARCHY_UNIQUE_NAME ON COLUMNS FROM (SELECT ({ [EstadoCaso].[NombreEstado].&[Abierto] ,[EstadoCaso].[NombreEstado].&[Suspendido] ,[EstadoCaso].[NombreEstado].&[Esperando Cliente] ,[EstadoCaso].[NombreEstado].&[Pendiente Interno] ,[EstadoCaso].[NombreEstado].&[Solucion Sugerida] ,[EstadoCaso].[NombreEstado].&[Trabajo en progreso]}) ON COLUMNS FROM [Model]) WHERE ([Measures].[CasoCount]) CELL PROPERTIES VALUE, FORMAT_STRING, BACK_COLOR, FORE_COLOR, FONT_F </pre>
<p>Cuenta de casos por agente y estado del caso, solo para los agentes activos.</p>	<p>MOLAP/ROLAP:</p> <pre> SELECT NON EMPTY { [Measures].[Fact Caso Soporte Hist Count] } ON COLUMNS, NON EMPTY { ([Dim Agente].[Nombre Completo].[Nombre Completo].ALLMEMBERS * [Dim Estado Caso].[Nombre Estado].[Nombre Estado].ALLMEMBERS]) DIMENSION PROPERTIES MEMBER_CAPTION, MEMBER_UNIQUE_NAME ON ROWS FROM (SELECT (-{ [Dim Estado Caso].[Nombre Estado].[All].UNKNOWNMEMBER }) ON COLU FROM (SELECT ({ [Dim Agente].[Activo].&[True] }) ON COLUMNS FROM [Sistema Soporte DWH])) WHERE ([Dim Agente].[Activo].&[True]) CELL PROPERTIES VALUE, BACK_COLOR, FORE_COL FORMATTED_VALUE, FORMAT_STRING, FONT_NAME, FONT_SIZE, FONT_FLAGS </pre> <p>HTAP:</p> <pre> SELECT NON EMPTY Hierarchize(DrilldownMember(CrossJoin({[Agente].[NombreCompleto].[All] ,[Agente].[NombreCompleto].[NombreCompleto].AllMembers} ,{([EstadoCaso].[NombreEstado].[All]}) ,[Agente].[NombreCompleto].[NombreCompleto].AllMembers ,[EstadoCaso].[NombreEstado]}) DIMENSION PROPERTIES PARENT_UNIQUE_NAME,HIERARCHY_UNIQUE_NAME ON COLUMNS FROM [Model] WHERE ([Agente].[Activo].&[True],[Measures].[CasoCountHist]) CELL PROPERTIES VALUE, FORMAT_STRING , LANGUAGE, BACK_COLOR, FORE_COLOR, FONT_F </pre>
<p>Cuenta de casos por tipo de contrato, región, producto y nivel de soporte para los casos de América.</p>	<p>MOLAP/ROLAP:</p> <pre> SELECT NON EMPTY { [Measures].[Fact Caso Soporte Hist Count] } ON COLUMNS, NON EMPTY { ([Dim Contrato].[Nombre Tipo].[Nombre Tipo].ALLMEMBERS * [Dim Region].[Nombre Region].[Nombre Region].ALLMEMBERS * [Dim Producto].[Nombre Producto].[Nombre Producto].ALLMEMBERS * [Dim Nivel Soporte].[Descripcion Nivel].[Descripcion Nivel].ALLMEMBERS)) DIMENSION PROPERTIES MEMBER_CAPTION, MEMBER_UNIQUE_NAME ON ROWS FROM (SELECT ({ [Dim Region].[Nombre Region].&[America] }) ON COLUMNS FROM [Sistema Soporte DWH]) CELL PROPERTIES VALUE, BACK_COLOR, FORE_COLOR, FORMATTED_VALUE, FORMAT_STRING, FONT_NAME, FONT_SIZE, FONT_F </pre> <p>HTAP:</p> <pre> SELECT NON EMPTY Hierarchize(DrilldownMember(DrilldownMember(DrilldownMember(CrossJoin({[TipoContrato].[NombreTipo].[All] ,[TipoContrato].[NombreTipo].[NombreTipo].AllMembers} ,{([Region].[nombreRegion].[All] ,[Producto].[NombreProducto].[All] ,[NivelSoporte].[DescripcionNivel].[All]}) ,[TipoContrato].[NombreTipo].[NombreTipo].AllMembers ,[Region].[nombreRegion] ,[Region].[nombreRegion].[nombreRegion].AllMembers ,[Producto].[NombreProducto] ,[Producto].[NombreProducto].[NombreProducto].AllMembers ,[NivelSoporte].[DescripcionNivel])) DIMENSION PROPERTIES PARENT_UNIQUE_NAME,HIERARCHY_UNIQUE_NAME ON COL FROM (SELECT ({[Region].[nombreRegion].&[America]}) ON COLUMNS FROM [Model]) WHERE ([Measures].[CasoCountHist]) CELL PROPERTIES VALUE, FORMAT_STRING, BACK_COLOR, FORE_COLOR, FONT_FLAGS </pre>

Fuente: elaboración propia.

La primera consulta está orientada principalmente a la creación de varios

agrupamientos al global de casos. La segunda, por otra parte, incluye agrupamientos, sin embargo, tiene un nivel de detalle elevado, ya que agrupa incluso al nivel de agentes y lo aplica sobre los históricos de casos. Estos superan considerablemente el global de casos, razón por la cual esta fue una consulta más pesada para todos los sistemas. Finalmente, la tercera consulta es similar a la primera y se orienta a los agrupamientos, pero con un mayor énfasis en los filtros en cuanto a procesamiento y usa como base los históricos de casos, por lo que igualmente tuvo una mayor carga en su procesamiento.

En el caso de los sistemas OLAP, estos utilizaron la misma consulta MDX. Sin embargo, para el sistema HTAP la consulta MDX fue diferente, aunque su propósito coincidía con las primeras. Además, para la recolección de medidas del sistema HTAP se previó un escenario en el que se simulan las transacciones regulares que forman parte del diariamente de un sistema OLTP para obtener medidas similares a las que se obtendrían en un ambiente HTAP real.

Para esta simulación de transacciones se utilizó el programa SQLTest de la compañía SQL Workshops. Este permite ejecutar consultas y comandos SQL que se aplican repetidamente por periodos constantes a lo largo de un lapso definido y a través de una cantidad de hilos de ejecución definidos por el usuario. Para estas consultas se tomaron números aleatorios que contemplaban los posibles identificadores de los casos y con base en esta selección aleatoria se realizaba la consulta, inserción o actualización correspondiente.

3.5.2. Iteraciones de Ejecuciones de Consultas

Para evaluar mejor las medidas de rendimiento se optó por hacer iteraciones de cada consulta en los distintos sistemas. Para cada consulta se realizaron 500 iteraciones por medio del *script* de PowerShell que se aprecia en la *Figura 1*.

```

For ($i=0; $i -500; $i++) {
Invoke-Sqlcmd -ServerInstance <ServidorTransaccional> -Database '<BaseDeDatos>' -Query 'DBCC DROPCLEANBUFFERS'

Invoke-ASCmd -Server:<ServidorAnalítico> -Query:"<ClearCache xmlns='http://schemas.microsoft.com/analysisservices/2003/engine'>
  <Object>
  |
  | <DatabaseID><BaseDeDatosAnalítica></DatabaseID>
  |
  | </Object>
  |
  | </ClearCache>
  |
  | "
Invoke-ASCmd -Server:<ServidorAnalítico> -Database:"<BaseDeDatosAnalítica>" -Query:"<ConsultaAnalítica>"
}

```

Figura 1 Código Powershell para la iteración de las consultas

Fuente: elaboración propia.

Al inicio de cada iteración se limpia la caché de los sistemas involucrados en el procesamiento de la consulta para manejar valores que representarán el esfuerzo real del sistema para presentar los resultados. Una vez limpiadas las caché, se ejecutaba la consulta. Como se aprecia en la *Figura 1*, para las conexiones a bases de datos se utilizó el comando `Invoke-Sqlcmd` y para las conexiones a SSAS el comando `Invoke-ASCmd`.

3.5.3. Herramientas Utilizadas para la Recolección de Medidas de Rendimiento

La recolección de las medidas de rendimiento se basó en dos herramientas:

- SQL Server Profiler: para la recolección de medidas de tiempo de ejecución de las consultas y de tiempo de CPU.
- Performance Monitor de Windows: por medio de este se obtuvieron las estadísticas del proceso durante la ejecución de las iteraciones. Las medidas obtenidas fueron:
 - Porcentaje del tiempo de procesador.
 - I/O de bytes de datos totales por segundo.
 - I/O de operaciones de datos totales por segundo.
 - I/O de lectura de bytes por segundo.
 - I/O de operaciones de lectura por segundo.

- I/O de escritura de bytes por segundo.
- I/O de operaciones de escritura por segundo.
- Errores de página por segundo.
- Número de conjuntos de páginas de memoria utilizadas.

En el caso de SQL Server Profiler, el proceso para la recolección de datos es sumamente simple. Se seleccionaron solo los eventos correspondientes al inicio y final de las consultas y por cada evento se registró una fila en una tabla de base de datos. Cada consulta de cada sistema se almacenó en diferentes tablas a las que se les aplicó una limpieza para eliminar datos innecesarios y se les efectuó una consulta que permitió obtener los promedios y los límites de duración total y de tiempo de Central Processing Unit (CPU) de todas las iteraciones. Para el caso de los sistemas OLAP, el SQL Server Profiler se utilizó en el servidor SSAS. Por otra parte, el sistema HTAP lo utilizó en el servidor de bases de datos, ya que el procesamiento ocurre en la misma base de datos transaccional.

Por otro lado, en Performance Monitor se creó un Data Collector Set que registraba las métricas indicadas y con el archivo que se obtuvo como resultado se pudo elaborar un reporte, ya fuera con valores promedio o valores límite, lo que indica el lapso entre el comienzo de la primera iteración y el final de la última. En este caso, los sistemas MOLAP y ROLAP se midieron con el proceso de SSAS y el sistema HTAP se midió con el proceso de SQL Server. Una vez recopilados los datos, tanto en SQL Server Profiler como en Performance Monitor, se elaboraron gráficos para la presentación de resultados que facilitaron su análisis posterior y comparación entre los sistemas de acuerdo con las medidas obtenidas.

3.6. Técnicas de Análisis de Información

Para analizar los resultados de las pruebas que se efectuaron en ambas implementaciones, se utilizaron herramientas de graficación que permitieron observar de forma sencilla los resultados, con base en ejecuciones iteradas de transacciones comunes en ambientes de procesamiento analítico de datos. De

igual forma, en los aspectos cualitativos se planteó una síntesis que permitió apreciar las facilidades o dificultades que presentan ambos modelos al compararse. Esto sin dejar de lado los aspectos que pueden depender de las necesidades de analítica que se puedan tener, así como los puntos en los que los dos modelos podrían coincidir.

Capítulo IV. Análisis del Diagnóstico

4.1. Resultados Referentes a Métricas de Tiempo

Para la presentación de los resultados que reflejan los tiempos de duración promedio que tuvieron las distintas implementaciones, se optó por la del gráfico de barras apilado que se puede apreciar en la *Figura 2*. Además, se presentaron los valores máximos y mínimos de la duración total de ejecución para cada conjunto de iteraciones en la *Tabla 7*, también representados en milisegundos. Las medidas por evaluar corresponden a la duración total de las consultas y tiempo de CPU, ambas medidas en milisegundos.

Duración (ms)

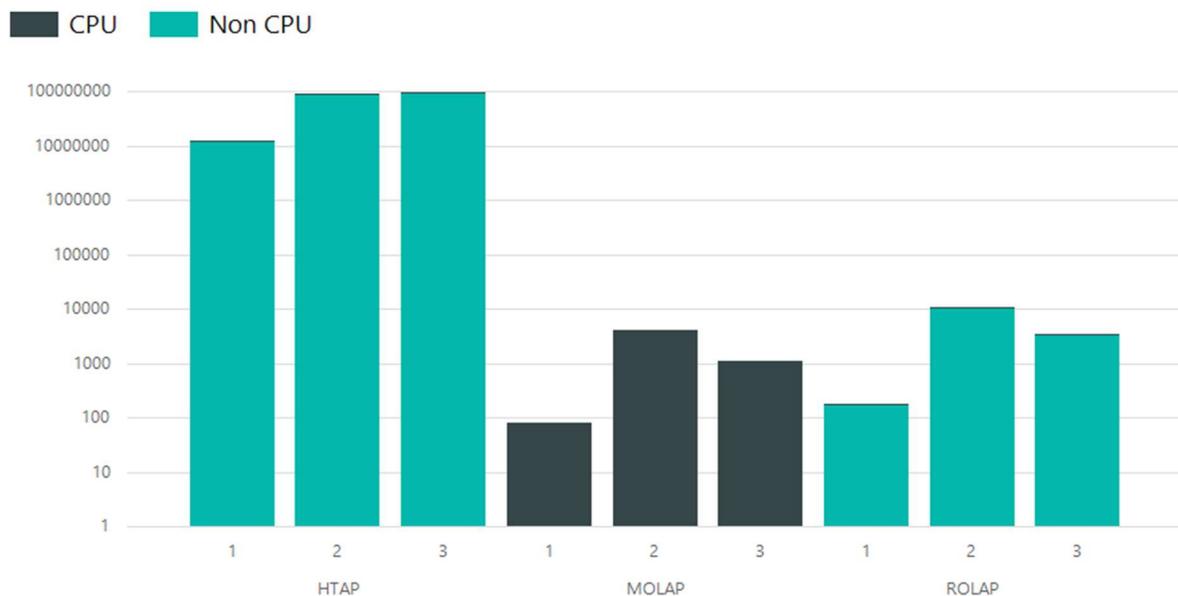


Figura 2 Gráfico para comparación de medidas de duración

Fuente: elaboración propia.

La representación presenta una escala logarítmica para permitir la apreciación de los resultados debido a la diferencia que existió entre el sistema con mayor tiempo de duración (HTAP) y los otros dos. Cada barra completa representa el tiempo total promedio de cada consulta, pero esta está segmentada en tiempo de CPU en negro y el tiempo restante en verde claro. En la *Figura 2* se

observa que el sistema HTAP superó por grandes márgenes el tiempo de duración de las implementaciones OLAP. Asimismo, se evidencia que el tiempo de CPU es mínimo para las implementaciones HTAP y ROLAP que actúan sobre la base de datos directamente, a diferencia de la implementación MOLAP cuyo tiempo correspondió casi en su totalidad a tiempo de CPU. En la *Tabla 7* se puede apreciar cómo el comportamiento que se aprecia en los valores promedio coincide con los valores límite que arrojaron los resultados.

Tabla 7 *Valores límite de duración de las consultas*

		Consulta 1	Consulta 2	Consulta 3
HTAP	MAX	27,497,962.000	142,386,548.000	144,686,323.000
	MIN	10,223,116.000	80,691,030.000	76,775,567.000
MOLAP	MAX	59.000	4,735.000	686.000
	MIN	25.000	1,277.000	253.000
ROLAP	MAX	412.000	14,532.000	4,474.000
	MIN	145.000	10,051.000	3,102.000

Fuente: elaboración propia.

4.2. Resultados Referentes a Operaciones de Entrada y Salida de Datos

Para analizar las métricas referentes a los datos utilizados por cada sistema, se optó por un gráfico de líneas que, de igual forma, permitiera analizar el comportamiento para todos los escenarios evaluados. Se continuó con el uso de una escala logarítmica para apreciar los tres tipos de datos que se representan: totales, lectura y escritura. Además, con el fin de facilitar la apreciación de los datos se elaboraron dos gráficos de líneas por separado: la *Figura 3* para el promedio de bytes por segundo y la *Figura 4* para el promedio de operaciones de datos por segundo.

Datos por segundo (Bytes)

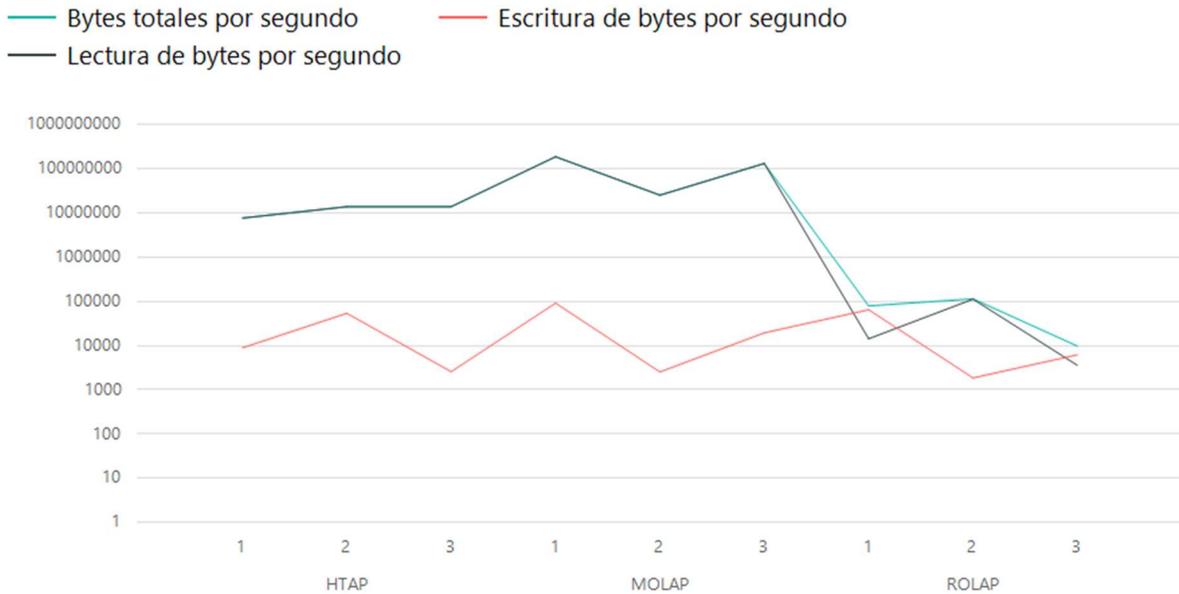


Figura 3 Gráfico para comparación de volúmenes de bytes por segundo

Fuente: elaboración propia.

En ambos gráficos se puede apreciar que la cantidad de operaciones resulta proporcional para todos los escenarios que se contemplan. Además, es claro que sistema MOLAP presenta un número significativamente mayor de operaciones y cantidades de datos manipuladas al compararlo con el modelo HTAP, pero ROLAP es el modelo que menos operaciones y volúmenes de datos maneja.

Datos por segundo (Operaciones)

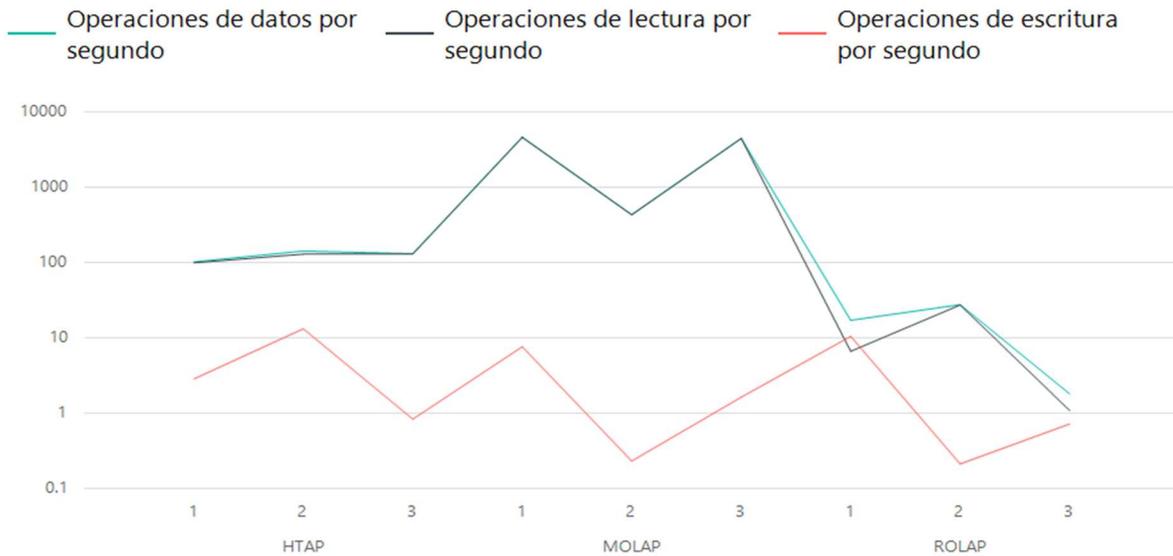


Figura 4 Gráfico para comparación de operaciones de datos por segundo

Fuente: elaboración propia.

Además, para la mayoría de los escenarios se aprecia que las lecturas de datos representan la mayoría de las operaciones de entrada y salida de datos en la mayoría de los escenarios. Sin embargo, en el caso de ROLAP, las diferencias entre operaciones de lectura y las de escritura son muy pocas, incluso tienen más operaciones de escritura para las consultas 1 y 3, correspondientes a carga de agrupamiento y de filtrado.

Por otro lado, también se midieron los fallos de páginas por segundo y su promedio por cada escenario, esto se presentó en un gráfico de áreas que corresponde a la *Figura 5*. En este se puede extraer la gran cantidad de errores que tuvo el modelo HTAP en comparación con los modelos OLAP. Este comportamiento fue consistente para todas las consultas y, como es de esperar, la consulta que maneja menor cantidad de datos (Consulta 1) fue la que tuvo mayores fallos en todos los escenarios.

Fallos de páginas por segundo

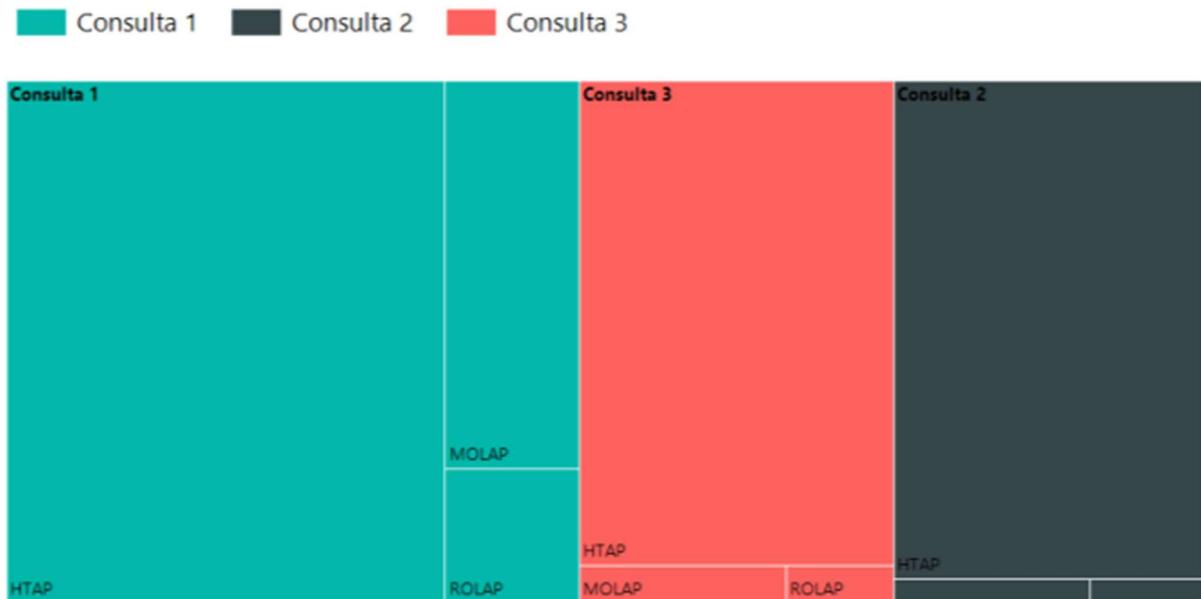


Figura 5 Gráfico para comparación de fallos de página por segundo

Fuente: elaboración propia.

4.3. Resultados Referentes al Porcentaje de Tiempo de Procesador Usado

El promedio del porcentaje de tiempo de procesador se representó con un gráfico de columnas agrupadas que representa el porcentaje promedio para cada consulta en cada sistema implementado, la *Figura 6* presenta este gráfico. En esta se puede apreciar que el sistema HTAP requiere de un procesamiento mucho mayor, incluso supera el 100 % que corresponde a la capacidad total de un único CPU. La consulta 3 que corresponde a la consulta con el gran volumen de datos, agrupamiento y filtros es la única en la que se observa cierta semejanza entre HTAP y MOLAP, debido a que para las otras dos consultas el segundo promedió alrededor del 50 %. Finalmente, se puede apreciar el poco tiempo de procesador que requirió ROLAP en comparación con los otros dos.

Porcentaje de tiempo de procesador

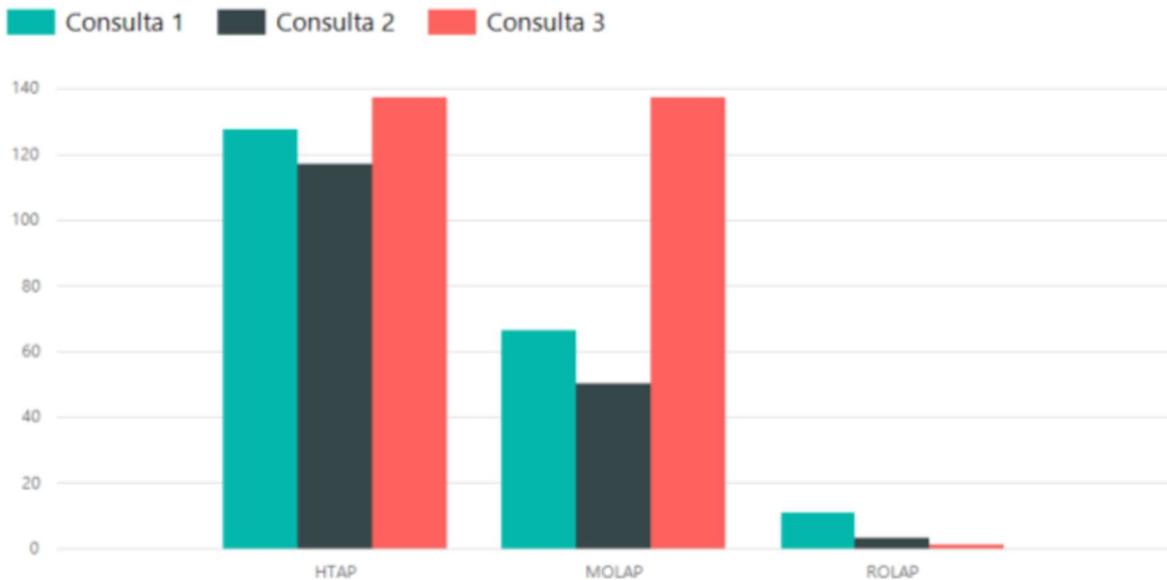


Figura 6 Gráfico para comparación de porcentaje de tiempo de procesador utilizado

Fuente: elaboración propia.

4.4. Resultados Referentes al Uso de Memoria

Por último, se presentaron los resultados referentes al promedio del uso de conjuntos de páginas de memoria en el gráfico que la *Figura 7* representa. En este gráfico se muestran los resultados de cada consulta y se puede apreciar por los tamaños de las barras, la diferencia que hubo entre cada sistema. En todos los casos, el sistema HTAP requirió una cantidad mucho mayor de conjuntos de páginas de memoria para la ejecución de las consultas. Este comportamiento fue el esperado debido al uso de los índices columnares que se basan en el uso de memoria y estos fueron la base para la implementación HTAP.

En la *Tabla 8* se pueden apreciar los valores límites referentes al uso de estos conjuntos de memoria durante las iteraciones de cada consulta en las distintas implementaciones. En esta tabla se puede observar que el comportamiento en los límites es consistente con los promedios, ya que incluso los resultados límite de la implementación HTAP multiplican hasta casi 18 veces

los resultados de las implementaciones OLAP.

Conjuntos de páginas de memoria

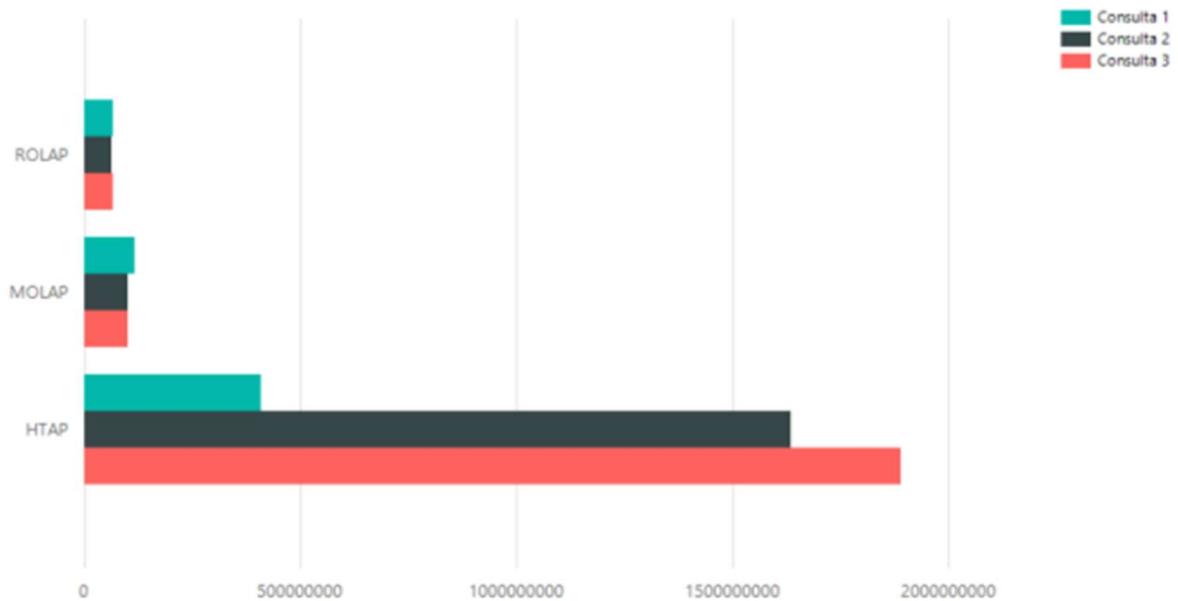


Figura 7 Gráfico para comparación de conjuntos de páginas de memoria utilizados

Fuente: elaboración propia.

Tabla 8 Valores límite de utilización de conjuntos de páginas de memoria

		Consulta 1	Consulta 2	Consulta 3
HTAP	MAX	456,007,680.000	1,836,965,888.000	1,973,542,912.000
	MIN	322,150,400.000	437,628,928.000	1,771,155,456.000
MOLAP	MAX	116,973,568.000	112,246,784.000	109,568,000.000
	MIN	96,641,024.000	97,566,720.000	99,143,680.000
ROLAP	MAX	68,947,968.000	71,950,336.000	69,591,040.000
	MIN	59,973,632.000	60,608,512.000	66,121,728.000

Fuente: elaboración propia.

4.5. Análisis de Implementaciones

En el aspecto cualitativo de este proyecto se evaluó la dificultad de implementación entre los modelos, así como de su utilización general, enfocándose en el procesamiento analítico que estos sistemas ofrecen. Por parte de la implementación HTAP, al ser la más nueva, existió cierta dificultad de

localización de documentación para proceder con la misma. Por otro lado, los modelos OLAP llevan años posicionados en el mercado, por lo que el acceso a información sobre estos es sumamente amplio.

El sistema HTAP, resultó sencillo de implantar en el aspecto de base de datos, debido a que en el ambiente SQL Server se requirió únicamente la creación de los índices de columnas apropiados en las tablas base para los contadores. Estos son los únicos, por lo que su mantenimiento es más sencillo, así como la creación de las dimensiones de tiempo. Por otra parte, las implementaciones OLAP requiere diseñar el Data Warehouse con sus dimensiones de tiempo, así como la elaboración de paquetes ETL incrementales que llevan cierta complejidad en la carga de datos nuevos. Además, para ROLAP es sumamente importante la creación de los índices adecuados.

En cuanto a los proyectos de Análisis desarrollados en SSAS, tanto para HTAP como para las soluciones OLAP su desarrollo es similar. Para ambos casos se necesita un conocimiento base de la herramienta, pero con la capacidad de manejar el modelo multidimensional o el tabular y la curva de aprendizaje cada uno puede ser exponencial.

Capítulo V. Propuesta de Solución

Con el propósito de obtener las métricas requeridas para la comparación entre los sistemas para la simulación del sistema transaccional que actúa como base en la comparación entre ambos modelos, se desarrolló una base de datos orientada al manejo de tickets de soporte que una compañía desarrolladora de *software* ofrece a sus clientes. La base de datos se encarga de manejar:

- Agentes: son los encargados de manejar el ticket.
- Clientes: crean los tickets para sus productos.
- Contacto de cliente: persona física que tiene acceso a crear tickets y con quien se mantiene la comunicación.
- Productos: son los paquetes de *software* que tiene la compañía a la venta.
- Módulos de productos: corresponden a componentes que cada producto posee.
- Contratos: son contratos de soporte que existen entre los clientes y la compañía. Cada contrato representa solamente un producto.
- Tipos de contratos: son los tipos de contrato que pueden existir y que ofrecen diferentes tipos de servicio.
- Nivel de soporte: especifica el nivel de soporte de los agentes, de acuerdo con su especialización.
- Regiones: regiones geográficas que identifican la ubicación de los agentes, así como la región en la que se brindará soporte a cada contrato.

- Grupos de soporte: identifican los grupos que existen para cada producto según los niveles de soporte que se le brindan.
- Casos: son los tickets de soporte. Estos indican cuál es el problema que se tiene, así como los detalles del sistema y del manejo de este.
- Niveles de urgencia: identifican los niveles de urgencia que puede tener un caso según la necesidad del cliente.
- Niveles de impacto: identifican el nivel impacto que causa el problema que se tiene en el sistema.
- Nivel de severidad: es un nivel promedio entre el nivel de urgencia y el de impacto del caso que permite tener una noción de la prioridad del caso.
- Service Level Agreements (SLA): indicadores que se basan en los niveles de urgencia y de impacto que definen el tiempo límite esperado para el primer contacto con el cliente.
- Tipos de problemas: identifican los posibles tipos de problemas que pueden justificar la creación del ticket.
- Tipos de soluciones: identifican los posibles tipos de soluciones con las que se puede resolver el problema del ticket.

5.1. Diseño de la Base de Datos Transaccional

En la *Figura 8* se puede apreciar el modelo relacional utilizado para la elaboración de la base de datos transaccional. Esta base de datos actúa como base para la presente comparación.

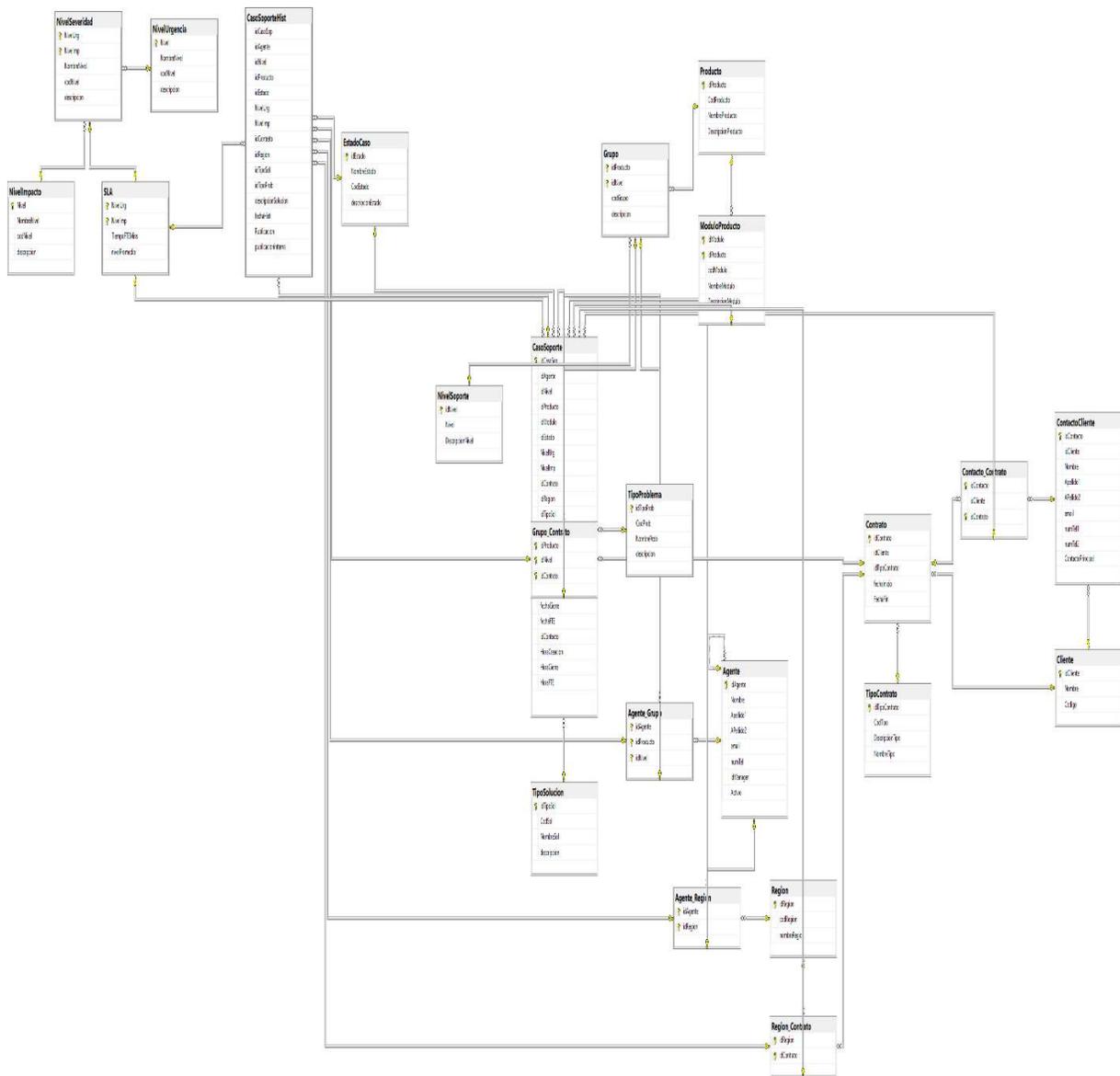


Figura 8 Modelo relacional de la base de datos transaccional

Fuente: elaboración propia.

5.2. Carga de Datos Simulados

Para cargar las tablas de los datos se elaboraron procedimientos almacenados que llevan a cabo esta carga de datos ficticios. Para nombres de personas, tanto de agentes como de clientes, se tomó como base el padrón electoral de Costa Rica, obtenido a través de la página electrónica del Tribunal Supremo de Elecciones. Por otra parte, para nombres de empresas de clientes,

productos, módulos y otros componentes identificadores se utilizaron nombres secuenciales que se basan en el nombre de la entidad acompañados de un número para diferenciar entre las tuplas. Por ejemplo, los productos se nombraron Producto 1, Producto 2, Producto 3 y así consecutivamente.

Por otro lado, para tablas que se utilizan, para configuración sobre cómo actuará el sistema de soporte, se establecieron datos predefinidos que permitirán representar el desarrollo de los tiquetes de soporte adecuadamente. Para las tablas, en las que correspondía, se tuvo la tarea de mantener la integridad referencial en los datos en los procedimientos almacenados.

Además, para la tabla de casos, de igual forma se generaron procedimientos almacenados, pero con una mayor complejidad para mantener la integridad referencial debido a que estas tablas son las que manejan una mayor cantidad de llaves foráneas. Para la tabla de casos, se generó una consulta masiva que reflejara todas las posibles combinaciones de las tablas cargadas previamente y, a partir de esta, se tomaron muestras aleatorias a las que se les asignaron fechas de creación que podían ser de hasta dos años y medio anteriores al 5 de abril del 2020. Para las fechas de cierre, se tomó la fecha de creación y se le añadieron hasta 40 días aleatoriamente. Se incluyó la fecha de cierre para los casos que todavía se estaban trabajando.

Por último, para la tabla histórica de casos, se hizo un procedimiento que iba tiquete por tiquete de la tabla de casos y se generaba aleatoriamente una cantidad de entradas por día, desde la fecha de creación en la actualidad de cierre o hasta el 5 de abril en caso de que todavía estuviera abierto. A cada registro histórico se le dio una posibilidad de que hubiera un cambio de estado del caso para evidenciarlo en la historia de los casos.

En la *Tabla 9* se puede apreciar la cantidad de registros que se añadieron a cada tabla para tener una mejor noción del tamaño significativo del sistema. En la misma se encuentra, además, una breve descripción de lo que cada tabla representa.

Tabla 9 Descripción y número de registros de las tablas transaccionales

Tabla	Descripción	Cantidad de registros generados
Agente	Listado e información de los agentes que dan soporte y sus supervisores.	3798
Producto	Listado e información de productos de <i>software</i> .	29
NivelSoporte	Nivel de soporte con que se manejan los casos.	3
Grupo	Grupo de soporte por nivel y producto.	87
Agente_Grupo	Relación entre los agentes y grupos a los que pertenecen.	3497
Region	Región geográfica a la que se le da soporte.	3
Agente_Region	Relación entre los agentes y la región a la que pertenecen.	3497
Cliente	Listado e información de los clientes a los que se les da soporte.	700
ContactoCliente	Listado de personas de contacto para los casos de soporte por cliente.	2100
TipoContrato	Tipos de contratos de servicio de soporte.	2
Contrato	Contrato que adjudica un servicio de soporte por producto y cliente.	4823
Contacto_Contrato	Relación entre los contratos y los contactos por cliente.	14469
Grupo_Contrato	Relación de grupos de soporte asignados a un contrato.	14469
ModuloProducto	Listado de módulos que conforman un producto.	260
Region_Contrato	Relación entre los contratos y la región a la que pertenecen.	4823
EstadoCaso	Estados que indican en qué situación se encuentra un caso.	7
NivelImpacto	Nivel que define el impacto del problema que originó la creación del caso.	4
NivelUrgencia	Nivel que define la urgencia de resolver el problema que originó la creación del caso.	4

Tabla	Descripción	Cantidad de registros generados
NivelSeveridad	Combinación entre el nivel de impacto y el de urgencia.	16
SLA	Definición de tiempo de respuesta después de abierto el caso, basado en la severidad.	16
TipoProblema	Tipos de problemas raíz que causaron la apertura del caso.	5
TipoSolución	Tipos de soluciones que resolvieron el caso.	5
CasoSoporte	Casos de soporte que definen todas las características de estos.	4149773
CasoSoporteHist	Histórico de todos los estados por los que han pasado los casos, así como comentarios realizados en los mismos.	23819787

Fuente: elaboración propia.

5.3. Diseño de Modelos de Analítica

Para la comparación entre las alternativas HTAP y OLAP, se optó por desarrollar 3 modelos que permitieron tener una mejor noción de lo que cada uno puede ofrecer. Para esto, por parte de la alternativa OLAP, se implementó un modelo MOLAP y otro ROLAP en SSAS que comparten el mismo Data Warehouse como fuente de datos. Por otro lado, para la alternativa HTAP, se creó la misma base de datos transaccional que actúa como fuente para el modelo tabular implementado también en SSAS.

5.3.1. Diseño OLAP

A partir de la base de datos transaccional presentada, se elaboró un Data Warehouse copo de nieve que permitiera la creación de los cubos OLAP. El modelo utilizado para este Data Warehouse se ilustra en la *Figura 9*.

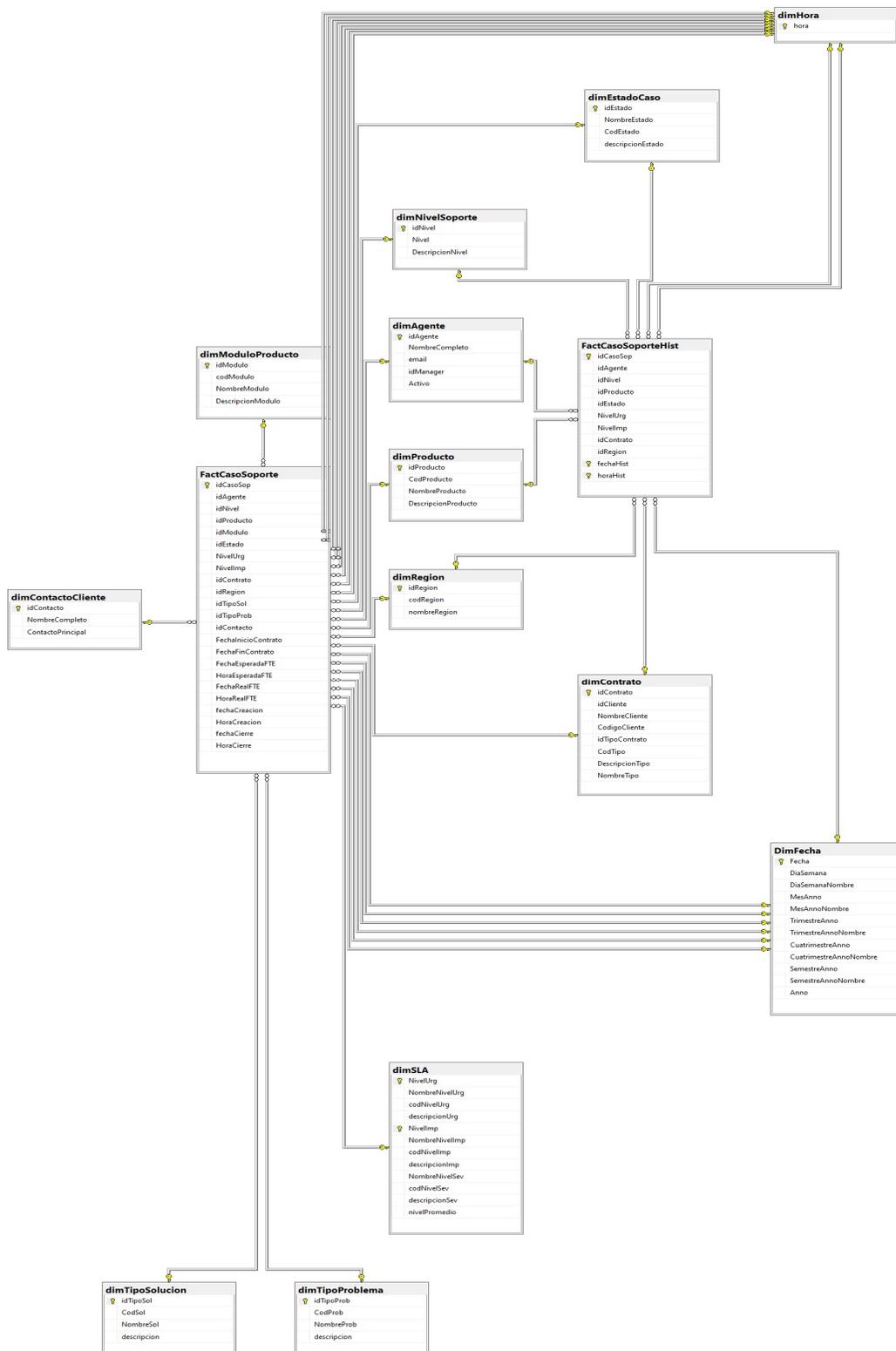


Figura 9 Modelo relacional del Data Warehouse desarrollado

Fuente: elaboración propia.

De igual forma, para la carga de datos en el Data Warehouse fue necesaria la elaboración de paquetes ETL. Se desarrollaron dos paquetes distintos en SQL Server Integration Services (SSIS), uno para la carga inicial que contempla todos los datos existentes en el sistema transaccional y otro para cargas incrementales que se encargan de añadir únicamente los datos que no han sido añadidos previamente.

Además, se procedió con la creación de las dimensiones de tiempo en el *Data Warehouse*. Se creó una dimensión de fecha y otra de hora. Para la dimensión de fecha se utilizó un *script* de creación y de carga recuperado del sitio *web* CodeProject, mientras que para la de hora se elaboró una consulta que pudiera generar un registro de tiempo para cada momento del día con una granularidad de segundos.

Una vez creadas las dimensiones de tiempo, se procedió con la elaboración de los cubos en SSAS. Para esto, como se mencionó, se crearon dos proyectos de SSAS, uno MOLAP y otro ROLAP, el tipo de procesamiento es la única diferencia entre ambos.

5.3.2. Diseño HTAP

Para la implementación utilizando el modelo HTAP, se utilizó la misma base de datos transaccional con la diferencia de que en las tablas que manejan mayor cantidad de datos (la de casos globales y la histórica de los casos), se les removieron los índices tradicionales y se añadieron índices orientados a columnas no agrupadas. Estos últimos, para la implementación de HTAP en SQL Server, representan la base para no afectar el ámbito transaccional del sistema. A cada tabla que requirió de estos índices solo se le creó uno, pero en este se contemplaban todos los campos por los que se esperaba que se hicieran relaciones con las otras tablas en las consultas analíticas.

Además, como única adición al modelo, se incluyeron las dimensiones de tiempo y fecha que se crearon en el *Data Warehouse* de la implementación OLAP. Sin embargo, se tuvo que crear un conjunto de estas dimensiones de tiempo para

cada columna de las tablas de casos que manejan estos tipos de datos, ya que el modelo tabular solo permite una única relación entre dos tablas.

Una vez creados los índices, se procedió con la creación del proyecto tabular en SSAS. Esto se hizo bajo el modelo DirectQuery para obtener esa característica de analítica 100 % en vivo que propone el modelo HTAP.

Capítulo VI. Conclusiones y Recomendaciones

6.1. Conclusiones

Gracias a las implementaciones para la obtención de los datos y la información que permitió la comparación entre los modelos, se presentó un panorama más detallado de lo que cada solución implica:

- Para el modelo OLAP, se pudo apreciar lo consolidado que este sistema se encuentra gracias a los años en el mercado. El proceso de su implementación puede considerarse extenso y complicado por todo lo que implica el diseño de Data Warehouse, elaboración de paquetes ETL y el desarrollo de los cubos requeridos. Sin embargo, gracias a las vastas referencias bibliográficas con las que se cuenta y lo extendido de su implementación de inteligencia de negocios, esta extensión y la complejidad para implementarla es mínima, esto resulta en un proceso de cuidado, pero, a la vez, llevadero.
- Por otro lado, el modelo HTAP contrasta con OLAP en el aspecto de conocimiento que existe. El proceso resulta reducido en secuencia de etapas para lograr su implementación si se compara con los requerimientos de OLAP, pero se necesita un conocimiento especializado sobre administración y diseño de bases de datos, así como de las estructuras de almacenamiento necesarias para su implementación como los índices columnares utilizados en SQL Server y la forma en la que funcionan para lograr un diseño que pueda sacar el mayor provecho.

La elaboración de los sistemas simulados que sirvieron como base para este proyecto, ofrecen una opción sólida para evaluaciones de sistemas, especialmente en el área de bases de datos. Por eso, a pesar de no contar con acceso a un sistema en producción real, este tipo de simulaciones puede contribuir con la elaboración de conocimiento nuevo que pueda ofrecer panoramas para distintos escenarios que se planteen. Tener la capacidad de desarrollar el

sistema propio permite evaluar cualquier circunstancia o área de negocio que se desee, especialmente en un aspecto preliminar.

De esta forma, se puede tener una referencia más clara para tomar la decisión de proceder o no con la evaluación deseada en un ambiente completamente real o clonado de una implementación real. Asimismo, es importante destacar la importancia de llevar a cabo las evaluaciones deseadas en ambientes reales para obtener un mayor grado de confiabilidad en los resultados. Los resultados en las mediciones de métricas de rendimiento produjeron distintas apreciaciones sobre la forma en que cada implementación utiliza los recursos del sistema y la respuesta de la implementación a las necesidades de retribución de datos:

- El tiempo de ejecución de las consultas dejó al modelo HTAP con valores muy altos en comparación con los modelos MOLAP y ROLAP que lograron manejar millones de datos en pocos segundos.
- Por otra parte, las operaciones de lectura y escritura produjeron datos no tan distantes entre las implementaciones. Estos incluso requirieron, en el caso de MOLAP, más operaciones y un mayor volumen de datos que la propuesta HTAP. Sin embargo, los fallos de página fueron considerablemente mayores en el modelo HTAP que en los modelos OLAP, lo que podría impactar el rendimiento de otros procesos que comparta el *hardware* del sistema.
- El uso de CPU, a la vez, tuvo un comportamiento muy elevado para la implementación HTAP, sin embargo, se debe destacar que, para el escenario de consulta número 3 que fue el más complicado, se obtuvieron valores muy similares a los del modelo MOLAP. Por lo tanto, en el aspecto de procesamiento se puede evaluar que, para cargas grandes, este ofrecerá un comportamiento similar a MOLAP, pero ambos se encuentran alejados de los resultados de ROLAP que fueron muy inferiores.
- Por último, el uso de memoria tuvo una proporción superior para el sistema

HTAP en comparación con ambas implementaciones OLAP. Este comportamiento es esperado, debido a los índices basados en memoria que actúan como columna vertebral de esta implementación HTAP.

Finalmente, con base en el proceso desarrollado a lo largo de este proyecto, así como de los resultados, se lograron analizar escenarios comparativos entre el modelo HTAP y el modelo OLAP. Estos son escenarios que se orientan a consultas de procesamiento analítico.

En cuanto a acceso a la información deseada, ambos sistemas tienen la capacidad de ofrecer la misma información, sin embargo, el modelo HTAP tiene una amplia desventaja en el aspecto de tiempo para obtener esta información. En consultas analíticas que estén ligadas a aplicaciones de usuario o que sean requeridas con inmediatez para su utilización como herramienta de soporte para procesos de toma de decisiones, el modelo OLAP puede ofrecer resultados más satisfactorios.

Lo anterior no significa que haya que descartar al modelo HTAP como una opción viable, debido a que su característica de procesamiento analítico completamente en vivo permite la presentación de resultados lo más recientes posible. Esto puede ser determinante cuando se necesita información del estado actual de los datos y no se cuenta con el factor tiempo requerido para la carga del Data Warehouse y el despliegue (si es necesario) del cubo OLAP.

Por otra parte, los resultados referentes a la forma en que los recursos del sistema se utilizaron permitieron obtener una noción de lo que cada uno de los modelos puede requerir. En el caso de HTAP es claro que el factor principal por considerar en el sistema es la memoria. Que esta última se extensa puede beneficiar la implementación HTAP e incluso ofrecer mejores tiempos de respuesta para volúmenes de datos masivos, mientras que en el caso de las implementaciones OLAP, este recurso es importante, mas no con la magnitud que HTAP lo requiere.

Por otro lado, los aspectos de lectura y escritura de datos muestran no ser

decisivos en la selección del sistema, sin embargo, siempre es importante lograr la mayor optimización posible en este tipo de operaciones indiferentemente del tipo de implementación. Por último, el uso de CPU muestra tener un rol importante en el procesamiento analítico, pero al comparar las alternativas HTAP y MOLAP se pudo apreciar que sus requerimientos pueden ser similares en este aspecto, especialmente cuando se manejan volúmenes extensos de datos.

6.2. Recomendaciones

- Para lograr implementaciones que permitan obtener resultados confiables y útiles para la comunidad es sumamente importante la búsqueda de información sobre estas implementaciones deseadas. Se debe ahondar incluso en aspectos específicos, sin restringirse a sus relaciones con la implementación.
- Es importante considerar el nivel de realidad que deben ofrecer las simulaciones que se desarrollen para este tipo de evaluaciones. En simulaciones de sistemas de datos es importante lograr que la relación se plasme de la forma más real posible. No es necesario que los datos sean específicos y reflejen completamente la realidad, ya que el uso de datos genéricos resulta igual de válido siempre y cuando se considere este aspecto de relación, así como del espacio de almacenamiento que requieran.
- La obtención de resultados, específicamente de medidas de rendimiento, puede muchas considerarse un inconveniente cuando se buscan medios y herramientas que puedan extraerlos. Sin embargo, no se deben descartar medios de uso común que los distintos sistemas ofrecen, ya que pueden tener la capacidad de obtener los datos deseados. Esto permite evitar el uso de sistemas de terceros que pueden complicar la obtención de esos resultados por la falta de conocimiento sobre la herramienta.
- La comparación entre los modelos HTAP y OLAP pudo producir escenarios no tan favorables para HTAP ya que es una alternativa muy reciente. Al

evaluar la viabilidad de un sistema es importante tomar en cuenta las necesidades de negocio que existen y definir si, a pesar de que en ciertos aspectos un sistema es más recomendado que otro, las características del que posee la desventaja pueden ser útiles para lo que se quiere llegar como resultado de implementación.

- En el caso de modelos de procesamiento analíticos, con base en los resultados de medidas, así como los aspectos cualitativos de ambos. Una opción viable sería contar con ambas implementaciones a la vez y utilizar uno u otro de acuerdo con los requerimientos, las circunstancias y las necesidades de negocio.

6.2.1. Trabajos a Futuro (Líneas Futuras de Investigación)

Al considerar el enfoque de este proyecto de resaltar lo que el modelo HTAP puede ofrecer en el aspecto de procesamiento analítico, un área que podría someterse a investigación es en contraste lo que este modelo puede ofrecer para procesamiento transaccional. De igual forma como se desarrolló en este proyecto, sería de suma importancia evaluar esta implementación en un ambiente que presente las condiciones de consulta por parte del área analítica del sistema. Esto permitiría evaluar adecuadamente su parte transaccional.

Referencias

- Arulraj, J., Pavlo, A., & Menon, P. (2016). Bridging the Archipelago between Row-Stores and Column-Stores for Hybrid Workloads. Proceedings Of The 2016 International Conference On Management Of Data-SIGMOD '16. doi: 10.1145/2882903.2915231
- Beaulieu, A. (2009). Learning SQL. Beijing: O'Reilly.
- Bog, A., Domschke, M., Mueller, J., & Zeier, A. (2009). A framework for simulating combined OLTP and OLAP workloads. 2009 16Th International Conference On Industrial Engineering And Engineering Management. doi: 10.1109/icieem.2009.5344329
- Bog, A., Sachs, K., & Zeier, A. (2011). Benchmarking Database Design for Mixed OLTP and OLAP Workloads. Proceedings Of The 2Nd ACM/SPEC International Conference On Performance Engineering. doi: 10.1145/1958746.1958806
- Coelho, F., Paulo, J., Vilaça, R., Pereira, J., & Oliveira, R. (2017). HTAPBench. Proceedings Of The 8Th ACM/SPEC On International Conference On Performance Engineering-ICPE '17. doi: 10.1145/3030207.3030228
- Cuzzocrea, A., Bellatreche, L., & Song, I. (2013). Data warehousing and OLAP over big data. Proceedings Of The Sixteenth International Workshop On Data Warehousing And OLAP - DOLAP '13. doi: 10.1145/2513190.2517828
- Dziedzic, A., Wang, J., Das, S., Ding, B., Narasayya, V., & Syamala, M. (2018). Columnstore and B+ tree-Are Hybrid Physical Designs Important? Proceedings Of The 2018 International Conference On Management Of Data-SIGMOD '18. doi: 10.1145/3183713.3190660
- Jensen, C., Bach, T., & Thomsen, C. (2010). Multidimensional Databases and Data Warehousing. San Rafael, Calif.: Morgan y Claypool.

- Kabakus, A., & Kara, R. (2017). A performance evaluation of in-memory databases. *Journal Of King Saud University-Computer And Information Sciences*, 29(4), 520-525. doi: 10.1016/j.jksuci.2016.06.007
- Kemper, A. & Neumann, T. (2011). HyPer: A hybrid OLTP& OLAP main memory database system based on virtual memory snapshots. 2011 IEEE 27Th International Conference On Data Engineering. doi: 10.1109/icde.2011.5767867
- Krueger, J., Grund, M., Boissier, M., Zeier, A. & Plattner, H. (2010). Data structures for mixed workloads in in-memory databases. 5Th International Conference On Computer Sciences And Convergence Information Technology. doi: 10.1109/iccit.2010.5711090
- Larson, P., Birka, A., Hanson, E., Huang, W., Nowakiewicz, M. y Papadimos, V. (2015). Real-time analytical processing with SQL server. *Proceedings Of The VLDB Endowment*, 8(12), 1740-1751. doi: 10.14778/2824032.2824071
- Li, S., Xiao, N., Wang, P., Sun, G., Wang, X., Chen, Y. et al. (2019). RC-NVM: Dual-Addressing Non-Volatile Memory Architecture Supporting Both Row and Column Memory Accesses. *IEEE Transactions On Computers*, 68(2), 239-254. doi: 10.1109/tc.2018.2868368
- Makreshanski, D., Giceva, J., Barthels, C. y Alonso, G. (2017). BatchDB. *Proceedings Of The 2017 ACM International Conference On Management Of Data-SIGMOD '17*. doi: 10.1145/3035918.3035959
- Microsoft. (2019). In-Memory Database-SQL Server. Recuperado de: <https://docs.microsoft.com/en-us/sql/relational-databases/in-memory-database?view=sql-server-ver15>
- Müller, S. y Plattner, H. (2012). An in-depth analysis of data aggregation cost factors in a columnar in-memory database. *Proceedings Of The Fifteenth International Workshop On Data Warehousing And OLAP - DOLAP '12*. doi:

10.1145/2390045.2390057

Müller, S., & Plattner, H. (2011). Aggregation strategies for columnar in-memory databases in a mixed workload. Proceedings Of The 4Th Workshop On Workshop For Ph. D. Students In Information & Knowledge Management-PIKM '11. doi: 10.1145/2065003.2065015

Özcan, F., Tian, Y. y Tözün, P. (2017). Hybrid Transactional/Analytical Processing. Proceedings Of The 2017 ACM International Conference On Management Of Data-SIGMOD '17. doi: 10.1145/3035918.3054784

Plattner, H. (2009). A common database approach for OLTP and OLAP using an in-memory column database. Proceedings Of The 35Th SIGMOD International Conference On Management Of Data-SIGMOD '09. doi: 10.1145/1559845.1559846.

Pourabbas, E. & Rafanelli, M. (1999). Characterization of hierarchies and some operators in OLAP environment. Proceedings Of The 2Nd ACM International Workshop On Data Warehousing And OLAP - DOLAP '99. doi: 10.1145/319757.319790

Richly, K. (2018). A Survey on Trajectory Data Management for Hybrid Transactional and Analytical Workloads. 2018 IEEE International Conference On Big Data (Big Data). doi: 10.1109/bigdata.2018.8622394

Thomsen, E. (2002). OLAP Solutions. New York: John Wiley y Sons.

UCLAve. (2019). ¿Qué es el procesamiento de datos? Recuperado de:
<http://www.ucla.edu.ve/dac/Departamentos/coordinaciones/informaticai/documentos/PROCESAMIENTO%20DE%20DATOS.htm>

Vaish, G. (2013). Getting started with NoSQL. Birmingham: Packt Publishing.