

Base de Datos NoSQL

Universidad de Cuenca

Base de Datos 2

Dr. Villie Morocho

Freddy L. Abad L.

freddy.abadl@ucuenca.edu.ec

Base de Datos NoSQL:

Uso en el Mercado

Facultad De Ingeniería, Universidad De Cuenca

BASE DE DATOS II

Freddy L. Abad L.

freddy.abadl@ucuenca.edu.ec

El siguiente informe busca detallar usos empresariales de las bases de datos NoSQL, y poder responder las siguientes preguntas.

- ¿Que marcas/sistemas/empresas implementan sistemas con base de datos NoSQL?
- ¿Que cantidad de usuarios utilizan estos sistemas?
- ¿Cuántas transacciones por período de tiempo soportan?
- ¿Que arquitectura utilizan?
- ¿Cuál podrá ser el nivel de crecimiento esperado para un futuro?

El presente informe se centra en las base de datos NoSQL de tipo multimodelo, para lo cual, inicialmente se definirá estas bases de datos.

“Una base de datos multimodelo es una plataforma de procesamiento de datos que soporta múltiples modelos de datos, que definen los parámetros de cómo se organiza y acomoda la información en una base de datos. Son capaces de incorporar múltiples modelos en una sola base de datos permite que los equipos de tecnología de la información y otros usuarios cumplan con diversos requisitos de aplicación sin necesidad de implementar diferentes sistemas de bases de datos”[1].

“Los modelos de datos que estas bases de datos que pueden acomodar incluyen relacionales, jerárquicos y objetos, además de documentos, gráficos, columnas anchas y estilos de valor-clave que se agrupan bajo el paraguas de la tecnología NoSQL. Los proveedores de bases de datos individuales ofrecen diferentes combinaciones de modelos de datos, lo que requiere que los compradores potenciales encuentren tecnologías de bases de datos multi-modelos que satisfagan sus necesidades de aplicación particulares. A diferencia de los sistemas que se limitan únicamente al modelo de datos relacional, las bases de datos multimodelo no almacenan de forma uniforme los datos en una estructura de tablas basada en filas. Como resultado, pueden manejar diferentes formas de datos que no se ajustan al esquema rígido del modelo relacional, incluidos los tipos de datos no estructurados y semiestructurados. En algunos casos, sin embargo, el enfoque multimodelo puede limitar la integridad transaccional que utilizan los sistemas de gestión de bases de datos relacionales para mantener la exactitud y consistencia de los datos.” [1]

Este tipo de Base de Datos NoSQL Multimodelo, tienen múltiples soluciones tecnológicas, tales como: MarkLogic, InterSystem, Rocket D3 DBMS, Rocket mvBase DBMS, Rocket U2 Universe, Rocket U2 Unidata, OpenQM, Caché InterSystems.

Siendo una de las más utilizadas MarkLogic, “el cual admite modelos de datos de documentos, gráficos, relacionales y de otro tipo en su base de datos. Este SGBD es utilizado en distintas empresas públicas y privadas, como: BBC, Boeing, Marina de los EE.

UU., Chevron , JPMorgan Chase , Erie Insurance Group , Johnson & Johnson y el Ejército de los EE . UU”. [2]

Los casos de estudio del presente informe son:

- CISCO
- SONY
- HEALTHCARE
- BBC iPlayer
- BBC Sport
- MITCHELL1

NBC



- Tipo de Base de Datos NoSQL

NBCUniversal es un líder de la industria en la digitalización y modernización de la entrega de contenido, como lo demuestra su aplicación SNL extremadamente popular. La aplicación basada en AWS ayuda la búsqueda de video streaming online. Fue construida en solo cuatro meses. Su crecimiento permite soportar agregar las nuevas producciones de la empresa sin problemas mayores. Necesitaban un sistema que pudiera captar la atención de un usuario y predecir un boceto de seguimiento que los obligaría a seguir viendo cuando terminaba su video actual.

Marklogic se utilizó ya que permite:

- Escalabilidad para soportar décadas de contenido y millones de usuarios.
- Elasticidad para flujos de trabajo que cambian rápidamente.
- Entrega de contenido personalizado a través de múltiples dispositivos
- Rápida integración de datos para cumplir con un ambicioso plazo.
- Bajo costo y excelente experiencia del cliente.

Las entregas de Marklogic fue:

- Nube (AWS)
- Centro de datos operacionales

La aplicación SNL se basa en contenido inteligente generado con la base de datos MarkLogic, lo que permite a los usuarios encontrar videos favoritos de forma rápida y fácil de los últimos 40 años. La aplicación incluye un motor de recomendación, impulsado por MarkLogic Semantics, que se adapta a las preferencias de los fanáticos para ayudarles a descubrir nuevos personajes y bocetos de los que no tenían conocimiento previamente.

Plataforma a prueba de futuro

A medida que surgen nuevos contenidos con las nuevas temporadas de SNL, NBCUniversal puede integrar rápidamente sus datos, independientemente del formato, para una distribución perfecta de los medios a su leal base de los usuarios.

- Número de transacciones

El contenido se entrega rápidamente, con **50,000 solicitudes por minuto en las horas punta**. La aplicación también pone en cola tipos similares de contenido que a un usuario le gustaría, según el historial de navegación. El sistema responde a cada transacción en un aproximado de 100-150 milisegundos.

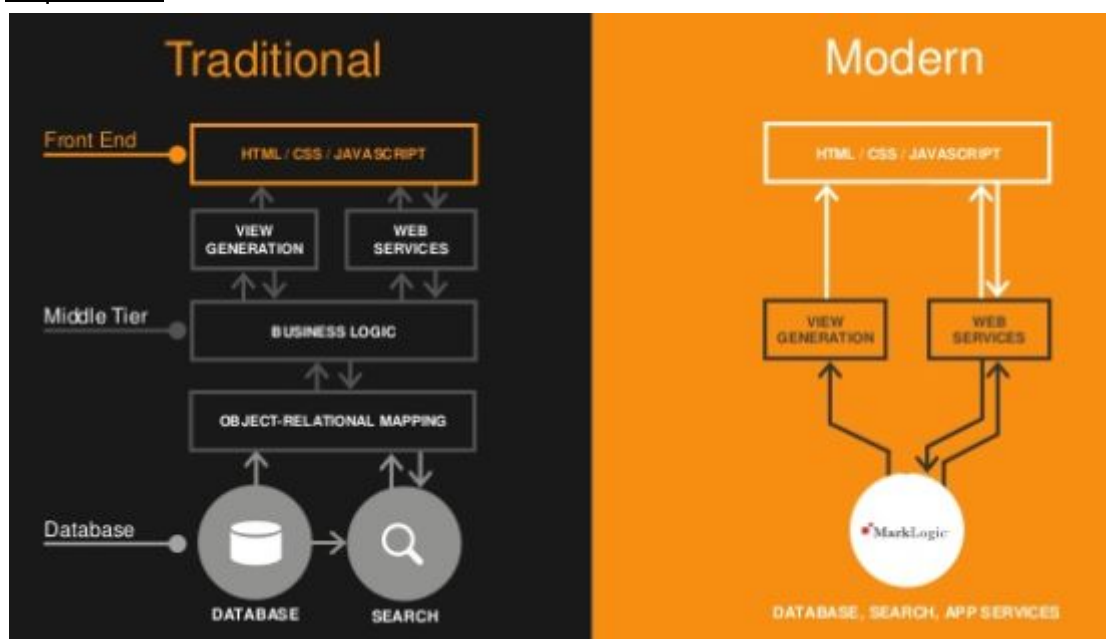
- Usuarios

El uso en la actualidad a soportado la visita de más de 100 millones de visitas en sus primeros tres meses, y tiene la capacidad de respaldar a una alta audiencia durante los estrenos de la temporada y otros eventos especiales. 200 millones de video streaming, 2 millones de descargas.

- Nivel de Crecimiento

Alcanzará a cuadruplicar el número de solicitudes manteniendo los tiempos de respuestas, en los casos de lanzamientos exitosos, sin que el sistema tenga problemas.

- Arquitectura



HEALTHCARE

HealthCare.gov

- Tipo de Base de Datos NoSQL

Healthcare es un sistema que se presenta como solución a la Ley del Cuidado de Salud a Bajo Precio, los Centros para Servicios de Medicare y Medicaid (CMS) del Departamento de Salud y Servicios Humanos de los EE. UU. Su vínculo con MarkLogic se da ante la necesidad de diseñar, construir e implementar una plataforma tecnológica capaz de inscribir a millones de estadounidenses en nuevos planes de atención médica. Un CMS que acelere el tiempo de producción, teniendo una experiencia de compra en línea sin problemas.

Inicialmente se intentó lograr esto con una base de datos relacional, pero el proyecto fracasó, por lo que CMS seleccionó a MarkLogic para cumplir con el plazo de 18 meses. En

la actualidad, CMS se ejecuta en MarkLogic como el proyecto de integración de datos personales más grande en la historia del gobierno de EEUU.

Los desafíos que solucionar ese sistema fueron la:

- Complejidad de los datos (datos de salud y financieros) y volumen de datos (registros para millones de estadounidenses)
- Cronograma comprimido según lo establecido por el Congreso
- Múltiples fuentes de datos, como compañías de seguros y proveedores de atención médica.
- Consolidación de datos
- Gestión de datos maestros (MDM)
- Gestión de contenido y entrega

Los resultados obtenidos fueron:

- Alto rendimiento y escalabilidad
- Seguridad
- Transacciones ACID
- Alta disponibilidad y recuperación de desastres

Ante la heterogeneidad de los datos, MarkLogic ingiere los datos tal como están y se adapta a las modificaciones a medida que se agregan nuevas fuentes de datos cuando cambian las políticas o regulaciones.

Uno de los logros de CMS con MarkLogic fue el no sacrificar ninguna de las características esperadas de una base de datos tradicional, como seguridad de nivel gubernamental, transacciones ACID y HA/DR. Además, CMS también puede aprovechar un modelo de documento NoSQL que maneja datos heterogéneos y no estructurados, y se escala fácilmente, lo cual es esencial durante la inscripción abierta.

Otro logro fue un tiempo de producción más rápido y soporte para tráfico pesado de sitios CMS lanzó HealthCare.gov dentro de los 18 meses.

La experiencia de compra en línea fue perfecta para HealthCare.gov.

La integración de los datos no requiere mayores cambios, se lo hizo con los datos tal como están, por lo que no fue necesario recodificar los datos de varios estados, agencias gubernamentales y proveedores de planes de salud. Esto facilitó la provisión de opciones de atención médica competitivas para los usuarios del sitio.

- Número de transacciones

En menos de cinco meses después del lanzamiento, MarkLogic realizó más de 5,500 transacciones por segundo para ayudar a más de ocho millones de personas a inscribirse en un seguro de salud.

- Usuarios

Medicare 59.7 millones de usuarios [5]

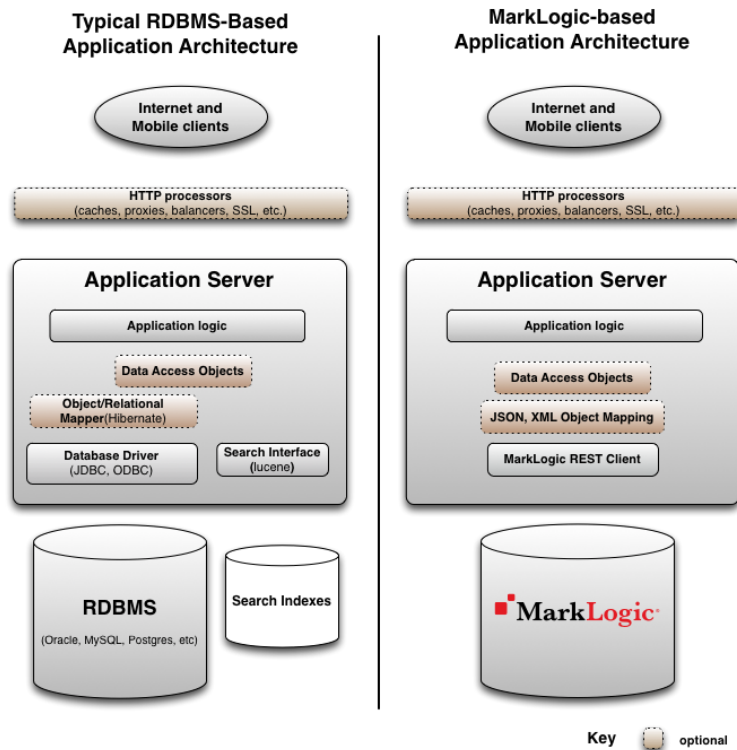
Medicaid 75.1 millones de usuarios [5]

- Nivel de crecimiento esperado para un futuro

Por causas políticas, se vio truncado este sistema, sin embargo los programas dirigidos a niños se mantienen, por lo cual se espera un Crecimiento 2020: 4.8 millones de personas y Crecimiento Esperado 2021: 6.9 de personas.

La Escalabilidad es casi ilimitada y su rendimiento no tiene precedentes, se atiende a 160,000 usuarios concurrentes y ha entregado más del 99.9% de disponibilidad. Más del 99,99% de las consultas han registrado tiempos de respuesta de menos de 0,1 segundos, sin pérdida de datos ni inconsistencias. [5][6]

Arquitectura [11]



BBC



- Tipo de Base de Datos NoSQL:

Multimodelo, usando MarkLogic. MarkLogic fue implementado por la BBC para ofrece contenido en tiempo real, personalizable e interactivo que satisface las necesidades de los clientes y aumenta los ingresos. Inicialmente en el sitio web de la BBC Sport para los Juegos Olímpicos de verano 2012, después el iPlayer de la BBC para la transmisión de vídeo, finalmente un Centro de datos operacionales. Buscando la maximización de su valor de contenido con una rápida integración de datos y desarrollo de aplicaciones. Este desarrollo contrasto la respuesta de una base de datos relacional (20 segundos) a una no relacional (200 milisegundos).

Los desafíos a solucionar con este sistema de base de datos relacional, incluye la ralentización de la productividad, la innovación y la entrega de contenido. El creciente contenido requería una escalabilidad masiva y tiempos de respuesta más rápidos

Los resultados obtenidos fueron:

- Tiempos de respuesta más rápidos
- Escalabilidad mejorada
- Agilidad
- Eficiencia de costo

BBC iPLAYER

El sistema desarrollado para La BBC iPlayer, el cual era una actualización de su sistema existente, esto para admitir una base de usuarios en crecimiento, además de entregar contenido que pudiera verse en múltiples dispositivos, personalizar este contenido y hacerlo todo más rápido que la implementación original.

Transacciones: BBC iPlayer, manejó tres mil millones de solicitudes dentro del primer año de producción a través de la nube.

La Infraestructura ágil con MarkLogic, el contenido estaba disponible para los espectadores en minutos, mas no en días como el sistema anterior.

El Costo total de desarrollo se redujo notablemente, incluyendo la reducción de la cantidad de sistemas en su pila de tecnología, la BBC también ha reducido drásticamente los costos de TI.

Los desafíos que se solucionaron fue:

- Integración de diversos datos, como estadísticas y videos.
- Desarrollo rápido de aplicaciones para cumplir con los plazos de los juegos.
- Escalabilidad para soportar decenas de miles de transacciones por segundo
- Utilización de estándares abiertos para la entrega móvil.

Los resultados obtenidos:

- Escalabilidad y velocidad mejoradas
- Éxito de la industria
- Synchronicity: interacción de múltiples pantallas con actualizaciones en tiempo real
- Alta disponibilidad

BBC SPORT

La BBC SPORT quería una entrega sin paralelo de contenido en línea, personalizable e interactivo en tiempo real para los Juegos Olímpicos de Verano. Su tecnología relacional comenzaba a retrasarse bajo la carga de tipos de datos variables; y anticipando un crecimiento exponencial del contenido, la BBC recurrió a MarkLogic para la integración de datos, la semántica y la entrega de dispositivos múltiples.

- Usuarios

10 millones de usuarios activos, 44% utilizan iPlayer únicamente, 56% utilizan BBC SPORT y iPlayer.

- Numero de Transacciones

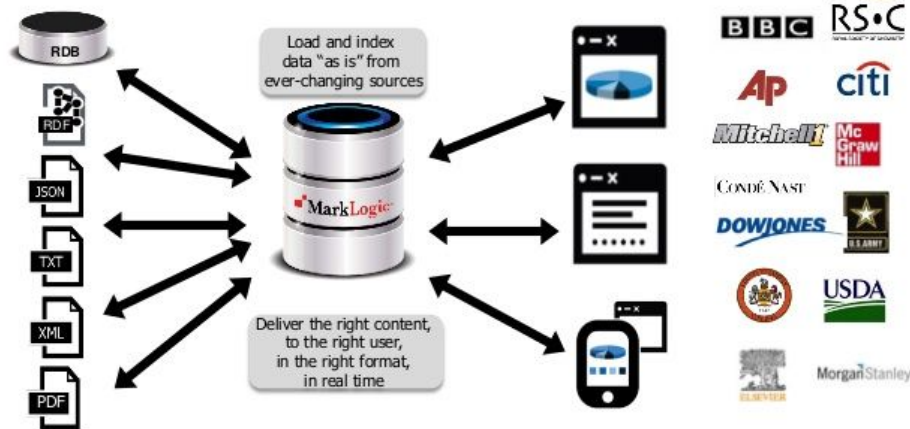
La BBC experimentó 25,000 transacciones por segundo, 45 Mil millones de solicitudes en total, y fue capaz de manejarlas todas. Diariamente se estima llegan 8.5 millones de solicitudes. Este desarrollo contrasto la respuesta de una base de datos relacional (20 segundos) a una no relacional (200 milisegundos)

- Escalabilidad

La Alta escalabilidad de BBC SPORT proporcionó 2.8 petabytes de datos en un solo día pico solo, con cero tiempo de inactividad. Para 2022 se estima crece a cerca de 13 millones de usuarios de iPlayer y BBC Sport, manteniendo sus tiempos de respuesta.

- Arquitectura [12]

MarkLogic



SLIDE: 34

© COPYRIGHT 2015 MARKLOGIC CORPORATION. ALL RIGHTS RESERVED.

MITCHELL1



The First Choice of Automotive Professionals

- Tipo de Base de Datos NoSQL

Multimodelo, Marklogic.

La empresa Mitchell 1 buscar una mejor integra cientos de millones de pedidos de reparación automotriz al tiempo que entrega información de diagnóstico integral en segundos, con un Centro de datos operacionales. Mitchell 1 integro MarkLogic para impulsar sus servicios SureTrack y Pro Demand, y mejora la facilidad de uso de las herramientas de diagnóstico de mano de la empresa matriz Snap-on, que ayudan a los talleres de automóviles a diagnosticar y reparar autos más rápido. Esto se ha traducido en un aumento del negocio, la lealtad del cliente y la cuota de mercado.

Los Desafíos que enfrentó al implementar:

- Gestión compleja de datos e integración de manuales, redes sociales y más.
- Arquitectura flexible para integrar datos más rápidamente
- Base de datos relacional en conflicto que no pudo mantenerse al día con el volumen o la variedad de datos
- Panorama competitivo

Los Resultados obtenidos fueron:

- Tiempo más rápido para valorar
- Aumento de los ingresos
- Experiencia de cliente mejorada
- Costo reducido

El uso de MarkLogic en Mitchell 1 fue una mejor consumo de datos tal como están , reduciendo sustancialmente la necesidad de una transformación manual de los datos y acelerando drásticamente el proceso de integración.

Mitchell 1 utiliza MarkLogic Semantics para enriquecer fuentes de datos dispares en tiempo real, por lo que los talleres de reparación saben que su contenido es preciso, completo y actualizado. Cuando las herramientas están conectadas a un automóvil específico, Mitchell 1 entrega rápidamente toda la información relevante, lo que lleva a diagnósticos automáticos rápidos y precisos, lo que conduce a clientes felices.

El tiempo de desarrollo se redujo, Mitchell 1 puede implementar nuevas funciones y mejoras en tan solo dos semanas, en comparación con los 6-12 meses de su anterior sistema relacional.

- Número de transacciones

El número de transacciones por segundo se estima en 150000 transacciones, con picos de hasta 300000 integrando la información de 28 tipos de manufacturas[8]. El Servicio al cliente mejoró notablemente, los clientes pueden encontrar información útil sobre reparaciones en menos de 30 segundos (en comparación con hasta 10 minutos con el sistema anterior) [7].

- Número de Usuarios

Usuarios del sistema : 17'000.000 [9]

- Nivel de Crecimiento

Crecimiento usuarios 2025: 21'500.000 [9]



- Tipo de base de datos Nosql

Utiliza Marklogic el cual es una base de datos NoSQL multimodelo [3], CISCO la implemento para satisfacer creciente demanda de los clientes por una búsqueda similar a Google y la entrega de su documentación técnica, de productos de aproximadamente 30 años, incluidas las adquisiciones, además de una integración rápida y fácil de más de 80 mil URL con información densa como tablas y gráficos. Cisco ahora tiene una base para incorporar fácilmente nuevos productos en el futuro, cumplir con sus objetivos de búsqueda y mejorar el servicio al cliente. Los desafíos que experimentó Marklogic fueron las decenas de miles de casos de soporte por mejorar la capacidad de búsqueda, además de una gran variedad de formatos de datos. Las soluciones implementadas fueron una plataforma segura y escalable. Obteniendo una búsqueda más rápida, además de mejores búsquedas mejoradas y servicios al cliente, además de estar integrados de forma segura y fácil de compartir.

Cisco seleccionó a MarkLogic como la base de su Content Hub para la documentación técnica. Esto favoreció la capacidad de MarkLogic de ingerir datos rápidamente tal como está, lo que lleva a una carga y administración de datos más eficiente. Una de las características importantes fue la entrega acelerada, Cisco pudo lanzar su Content Hub en menos de 6 meses, incluyendo una mayor seguridad y un uso de datos más fácil.

- Número de transacciones

La búsqueda incorporada proporciona una indexación de datos avanzada y rápida y el acceso a más de 80,000 transacciones por segundo de información altamente densa.

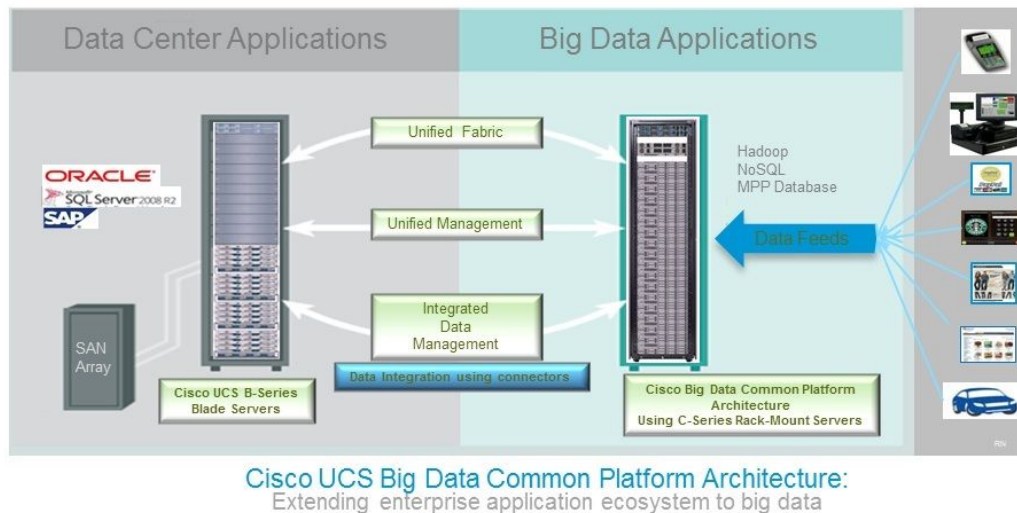
- Usuarios

Como promedio, se tiene 74,200 usuarios activos del sistema concurrentemente, sin embargo, se tiene proyectado un crecimiento hasta 100,000 en 2022 [4].

- Nivel de crecimiento esperado para un futuro

El sistema espera escalar sin ningún problema a las 200,000 transacciones por segundo y un número de usuarios aproximado de 100,000 para el año 2022. [4]

- Arquitectura



SONY

SONY

- Tipo de base de datos NoSQL

Sony implementó Marklogic para su integración de datos de las organizaciones de distribución global que cubrían 175 territorios en todo el mundo, desarrollando un Centro de datos operacionales con este tipo de base de datos. Esta integración presentaba desafíos como la integración de datos de numerosos sistemas dispares, incluido un mainframe obsoleto. Además del sistema de un centro financiero y más con MarkLogic para resolver estos desafíos, agilizando sus procesos globales de ventas a finanzas (S2F) para generar más ingresos.

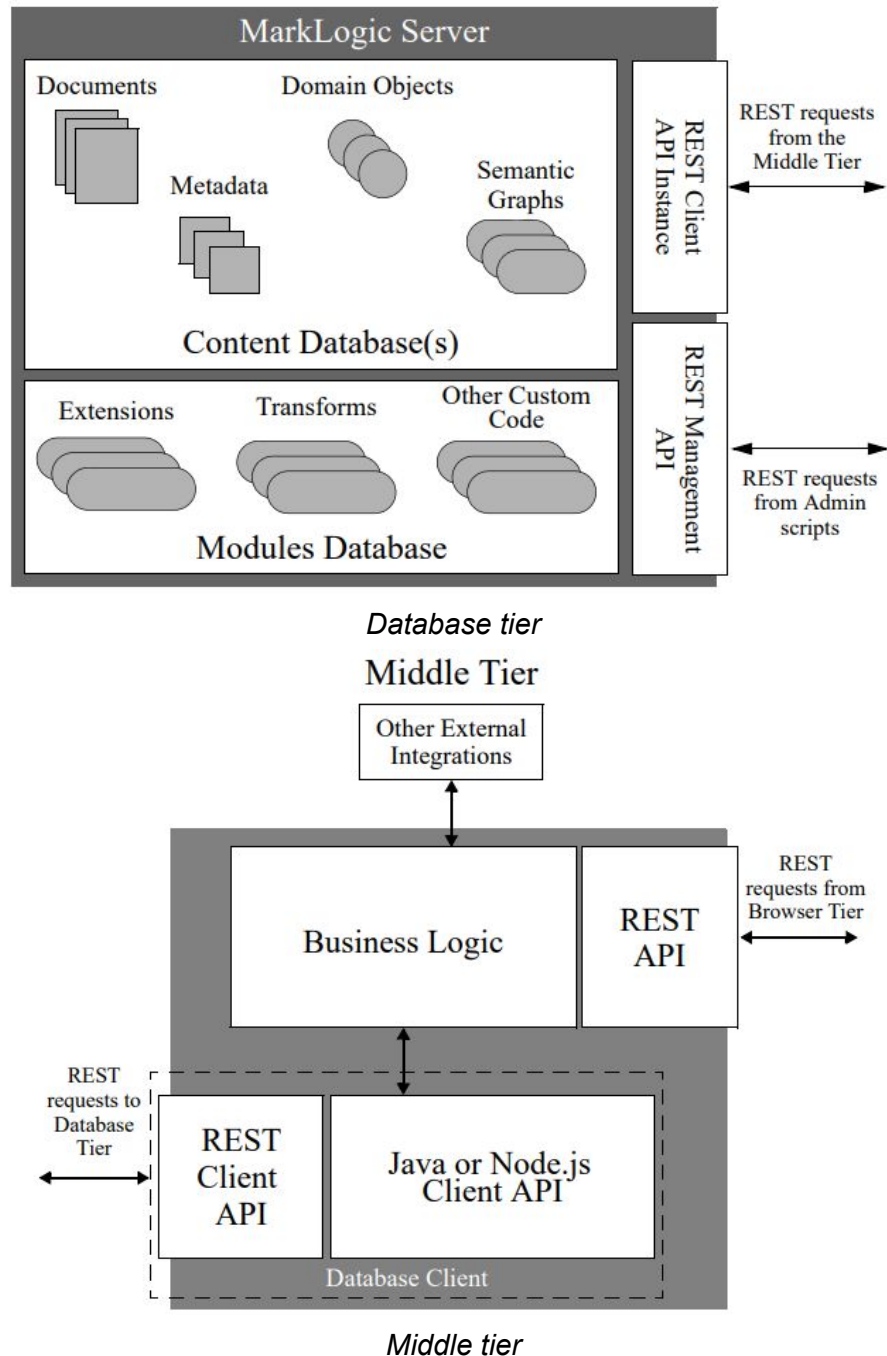
La solución implementada por Marklogic fue un Centro de datos operacionales, el Marco principal de Distribución, la Migración de mainframe y el sistema de búsqueda de productos. Las ventajas del uso de Nosql fue una mejor la agilidad para responder a los cambios contractuales, la Reducción de los requisitos de modelado de datos, Ingesta más rápida de datos y el proyección completa de datos y fácil trazabilidad.

MarkLogic sirve como el centro para mejorar el sistema de ventas a finanzas (S2F) de Sony para aumentar las ventas globales.

Incrementar la Agilidad, mejoró la capacidad de responder a los cambios contractuales, lo que le permitió a la empresa ganar más dinero.

Su Productividad incremento, los desarrolladores, especialmente los desarrolladores de Java, se hicieron rápidamente más productivos en el nuevo entorno.

- Número de transacciones
700.000 usuarios [100]
24'000.000 transacciones por día [10]
Permite picos 750.000 transacciones por segundo [10]
- Usuarios
700000 usuarios activos [10]
- Arquitectura



REFERENCIAS BIBLIOGRÁFICAS

[1] ACENS - Telefónica, 2016, "Bases de datos NoSQL. Qué son y tipos que nos podemos encontrar", Online available:

<https://www.acens.com/wp-content/images/2014/02/bbdd-nosql-wp-acens.pdf>

[2] Marklogic Editores, 2019, “Cisco Improves Customer Service with World Class Content Delivery Based on the MarkLogic Database”, Online available:

<https://www.marklogic.com/customers/cisco/>

[3] Editores Wikipedia, 2019, “NoSQL”, Online available:

https://en.wikipedia.org/wiki/NoSQL#Multimodel_database

[4] Editores CISCO, 2019, “Cisco Overview”, Online available:

<https://newsroom.cisco.com/overview>

[5] Editores CMS Healthcare, 2019, “CMS Fast Facts”, Online available:

<https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/CMS-Fast-Facts/index.html>

[6] MarkLogic Corporation, “MarkLogic Provides Access to Healthcare

Insurance for Millions of Americans”, 2014, Online available:

<https://www.marklogic.com/wp-content/uploads/2014/10/MarkLogic-CMS-Case-Study-10-2014.pdf>

[7] BBC Editors, 2019, “BBC passes 10 million registered users” , Online available:

<http://informitv.com/2017/09/06/bbc-passes-10-million-registered-users/>

[8] MarkLogic Editors, 2019, “Mitchell 1 Selects MarkLogic® to Power Automotive Software”, Online available:

<https://www.marklogic.com/news/mitchell-1-selects-marklogic-to-power-automotive-software/>

[9] Zaino Jennifer, 2015, “MarkLogic Enterprise NoSQL Database Smooths Car Talk for Mitchell1”, Online available:

<https://www.dataversity.net/marklogic-enterprise-nosql-database-smooths-car-talk-for-mitchell1/#>

[10] MarkLogic Editors, 2017, “Why Sony Is Using a MarkLogic Data Hub for Sales and Financial Processing”, Online available:

<https://www.marklogic.com/blog/why-sony-is-using-a-marklogic-data-hub-for-sales-and-financial-processing/>

[11] MarkLogic Editors, 2017, “MarkLogics Solutions: Architecture Case Studies”, Online available:

<https://developer.marklogic.com/learn/arch/diagram-101>