



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE INGENIERÍA

**Transformación de un negocio a
partir de los servicios de nube y
su proceso de venta**

INFORME DE ACTIVIDADES PROFESIONALES

Que para obtener el título de

Ingeniero en Computación

P R E S E N T A

Verónica Ambrosio Velázquez

ASESOR DE INFORME

M.I. Aurelio Sánchez Vaca



Ciudad Universitaria, Cd. Mx., 2017

Agradecimientos

Gracias a mi familia, por haberme acompañado en este largo camino. Por brindarme su amor y apoyo a pesar de todas las adversidades.

Gracias por su ejemplo de superación incansable, por enseñarme el valor de la perseverancia, el trabajo y el esfuerzo. Gracias por haber sembrado en mí desde muy pequeña el gusto por el estudio, legado más grande que pudieron darme para hacer frente a la vida. Los amo con todo el corazón y nunca terminaré de agradecer por todo lo que han hecho por mí.

Gracias Melissa, por ser mi cómplice, mi amiga, por apoyarme en cada etapa de mi vida y regalarme un poquito de tu enorme nobleza.

Gracias a Ricardo Martínez, por estar conmigo en tantos momentos importantes, por darme tu cariño y aliento en los momentos difíciles y en pocas palabras, por compartirme esas parcelas de tu vida.

Gracias a mis amigos de la Facultad, con los que compartí estrés, pero también muchas satisfacciones. Gracias Carlitos por darme tu invaluable amistad.

Gracias a mis colegas de Microsoft México, por compartirme su conocimiento, por darme la oportunidad de trabajar en sus proyectos, por inspirarme y motivarme a ser mejor ingeniera cada día.

Gracias a mi tutor Aurelio Sánchez Vaca, por haberme guiado y apoyado desde que pisé la Facultad de Ingeniería hasta la culminación de mi formación profesional.

Gracias al coro Ars Iovialis de la Facultad de Ingeniería que ha llenado ese lado artístico que me hacía falta y que me ha llevado a cantar a tantos lugares que ni siquiera habría soñado.

Agradezco a la Universidad Nacional Autónoma de México por haberme cambiado la vida desde que me abrió sus puertas al ingresar a la Escuela Nacional Preparatoria y desde entonces me ha colmado de alegrías y satisfacciones.

Verónica Ambrosio Velázquez

Contenido

Agradecimientos.....	1
Contenido.....	3
Introducción	5
I. Descripción de la Empresa.....	7
1.1 Misión	7
1.2 Estrategia	7
1.3 Historia de Microsoft.....	8
1.4 Organigrama.....	10
1.5 Contexto del Puesto	12
1.6 Responsabilidades clave	15
II. Antecedentes	17
III. Definición del Problema	21
IV. Desarrollo.....	23
4.1 Ingeniería de requerimientos.....	23
4.1.1 Clasificación y organización de requerimientos	24
4.1.2 Diagrama de casos de uso	25
4.2 Propuesta de solución	33
4.2.1 Arquitectura	33
4.2.2 Simbología.....	34
4.3 Ambiente creado en Microsoft Azure.....	36
4.3.1 Especificación técnica de Máquina Virtual.....	36
4.3.2 Virtual Private Network (VPN) Forticlient	38
4.3.3 Desarrollo de Scripts.....	39
4.4 SQL Database	43
4.5 Cron de Linux	51
4.6 Análisis con Power BI	52
V. Resultados.....	57

5.1 Visualizaciones.....	57
a. Wordcloud de búsquedas.....	58
b. Productos consultados en catálogo.....	59
c. Visitas por departamento.....	60
d. Compras.....	60
e. Seguimiento por cliente.....	63
f. Top Referer.....	64
5.2 Validación con Matriz de pruebas.....	65
5.3 Nuevas Iniciativas.....	66
5.4 Propuesta Comercial.....	68
VI. Conclusiones.....	71
VII. Referencias.....	73
VIII. Anexos.....	77
A. Glosario.....	77
B. ¿Qué es Microsoft Azure?.....	79

Introducción

El presente reporte ofrece una descripción de las actividades profesionales realizadas en Microsoft México, desde el puesto de *Cloud Technical Specialist* (Especialista Técnico en Nube).

Durante mi estancia en Microsoft, participé activamente en diferentes proyectos relacionados con la transformación digital e implementación de soluciones en la nube pública Microsoft Azure.

Microsoft Azure es una colección de servicios integrados en la nube que permite crear soluciones de análisis, proceso, bases de datos, móviles, redes, almacenamiento y Web.

Mi participación profesional fue enfocada a la preventa técnica, lo cual implica hacer una evaluación técnica de la iniciativa o problema de un cliente en particular, hacer demostraciones y pruebas de concepto que muestren al cliente el valor que Azure puede aportar a su negocio, así como ser un consejero de confianza para el mismo, empoderándolo para tomar decisiones, transformar su negocio y crear nuevos escenarios en la nube.

Tuve la oportunidad de trabajar en proyectos de Estafeta Mexicana, Grupo Modelo, Pemex, Seguros Monterrey, Universidad Anáhuac, Instituto Tecnológico y de Estudios Superiores de Monterrey, por mencionar algunos ejemplos. En estas empresas colaboré principalmente en temas de infraestructura como en el diseño e implementación de arquitecturas en la nube, ambientes híbridos, migración de máquinas virtuales, creación de Virtual Private Networks (VPNs) y análisis de datos.

Este reporte describe el tipo de iniciativas que hoy en día debemos buscar en nuestros clientes para impulsarlos en su transformación digital. Además, se detalla una oportunidad de *Advanced Analytics* (Analítica avanzada) para el departamento de comercio electrónico de uno de nuestros clientes.

Por los términos de privacidad y confidencialidad que acepté al ser contratada por

Microsoft, no me es permitido revelar la identidad del cliente. Para efectos del presente informe, referiré a esta institución con el nombre de Contoso Ltd¹.

Contoso Ltd. es una cadena mexicana de mercados y almacenes con presencia en toda la república. Su departamento de comercio electrónico, nos solicitó diseñar un sistema que le facilitara su tarea de analizar datos generados por su plataforma de comercio en línea, para entender el comportamiento e intereses de los usuarios, con el fin de poder lanzar promociones y ofertas personalizadas, facilitar la toma de decisiones y con ello incrementar las ventas.

Los datos presentados son de referencia y no pertenecen a ningún cliente en particular.

¹ Contoso Ltd. es una empresa ficticia usada por Microsoft para brindar ejemplos.

I. Descripción de la Empresa

1.1 Misión

Empower every person and every organization on the planet to achieve more.

Empoderar a cada persona y organización en el planeta para lograr más.

La misión de Microsoft se conecta a sus primeros días como empresa y está dirigido particularmente a las oportunidades y desafíos que el mundo enfrentará en las próximas décadas.

Cuando hablamos de empoderar a la gente queremos decir que, brindando las herramientas adecuadas, cualquier persona puede lograr cualquier cosa que desee. Nuestra misión se extiende a nivel global, lo cual nos invita a tener a ambición de impulsar la transformación digital en todo el planeta.

1.2 Estrategia

Build best-in-class platforms and productivity services for a mobile-first, cloud-first world.

Construir las mejores plataformas y servicios de productividad para un mundo de movilidad y nube primero.

El término Mobile-first, se refiere a la movilidad de los seres humanos a través de los dispositivos y no a la movilidad de un aparato en específico. Microsoft se refiere a la movilidad como un estilo de vida no a un dispositivo electrónico.

El término Cloud-first, hace referencia a la importancia que la nube está adquiriendo en nuestras vidas. Además de facilitar nuestra visión de movilidad, la nube es donde vive la verdad, ya que es una fuente omnipresente de información actualizada que se mueve a gran velocidad. La nube permite un mejor desempeño, mantiene los datos donde estos pueden fluir más libremente y nos conecta a todos, lo que ofrece un sinfín de oportunidades.

1.3 Historia de Microsoft

Microsoft es una empresa multinacional dedicada al sector del software y equipos electrónicos. Durante años, ha destacado por su constante innovación, productos y soluciones tecnológicas.

Microsoft fue fundado el 4 de abril de 1975 por Paul Allen y Bill Gates en Albuquerque, Nuevo México. Su visión era cambiar la forma en la que trabajábamos y proveer cada oficina y cada hogar de un equipo informático.

En 1981, IBM comienza a comercializar equipos con el sistema operativo de Microsoft MS-DOS 1.0, con este hecho Microsoft se empieza a posicionar en el mercado de las Microcomputadoras.

En 1985, Microsoft lanza al mercado Windows 1.0, una extensión del sistema operativo MS-DOS que proporciona un entorno operativo gráfico, con interacción de “ventanas” y el mouse.

En agosto de 1995, Microsoft lanza Windows 95 y con esto los equipos informáticos llegan a la madurez. Se trata del mayor anuncio de Microsoft hasta esa fecha, con gran éxito de ventas. Es la era del fax/módem, del correo electrónico, del nuevo mundo online y de los deslumbrantes juegos multimedia y el software educativo. Se presenta en 12 idiomas y con conexión telefónica a Internet. En Windows 95 aparecen por primera vez el menú Inicio, la barra de tareas y los botones para minimizar, maximizar y cerrar ventanas.

Durante sus siguientes años, Microsoft desarrolló mejores y novedosos sistemas operativos, así como su famosa suite de Office, promoviendo la productividad y revolucionando la forma de trabajar.

En noviembre 2001 llega la primera versión de la videoconsola con XBOX, que ofrece las más poderosas experiencias de juego.

Abril de 2003 – Microsoft lanza Windows Server 2003, el cual se convierte en el corazón de la estrategia de servidores de Microsoft y se diseña en torno a una arquitectura común que asegure la interoperabilidad. Es el proyecto de desarrollo de software más grande en la historia de Microsoft hasta la fecha.

El 2010 fue año de importantes lanzamientos de Microsoft: aparece Office 2010, el kinect para XBOX 360, se anuncia el lanzamiento del primer Windows Phone, se lanza Lync (herramienta de comunicaciones unificadas para el sector profesional) y se lanza al mercado Windows Azure.

Para el 2011, llega Office 365 como servicio, lo que implica las diferentes herramientas siempre están actualizadas y la información es fácilmente compartida gracias a la nube de Microsoft. Además, adquiere Skype y la red social corporativa Yammer!

En febrero de 2014 Satya Nadella es nombrado nuevo CEO (Chief Executive Officer o director ejecutivo) de Microsoft y con ello empieza una importante transformación de valores, objetivos y estrategias en Microsoft, tales como la colaboración con herramientas de software libre.

En enero de 2015, Microsoft revela su nueva versión de sistema operativo: Windows 10 la cual promete ser el mejor sistema operativo hasta el momento, incluyendo novedades como el nuevo asistente personal Cortana y el sistema unificado con Surface, Windows Phone, XBOX y los futuros HoloLens.

Hoy en día Microsoft se mantiene en el mercado como un de las empresas más influyentes en el ramo de la tecnología.²

² (Prensa, Microsoft, 2017)

1.4 Organigrama

La estructura organizacional de la empresa está distribuida en tres grandes bloques:

- USA
- Canadá
- MSI (Microsoft International)

La subsidiaria de México se ubica en el bloque MSI, en la región de Latinoamérica (LATAM). Dentro de cada subsidiaria se distinguen principalmente seis segmentos que se dedican a cubrir diferentes necesidades del negocio.

En la Figura 1.1 se muestra la estructura organizacional de la empresa, en la parte inferior del diagrama se muestran los segmentos principales de la subsidiaria de México.

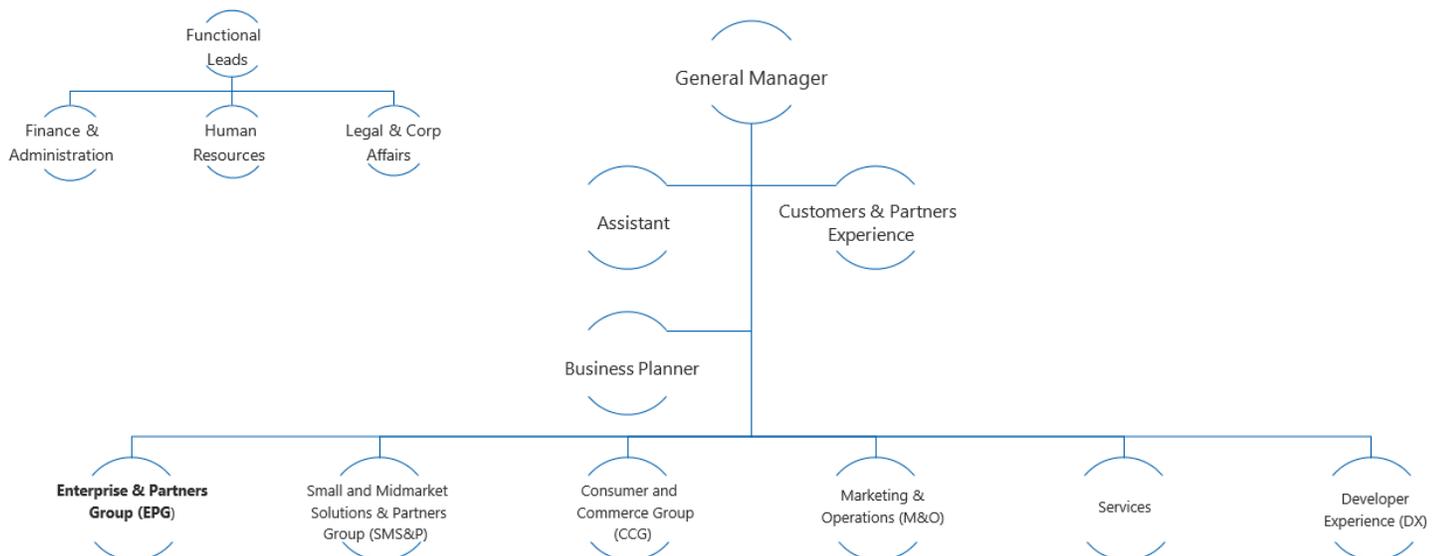


Figura 1.1. Estructura Organizacional de Microsoft.³

El puesto de Cloud Technical Specialist, se ubica en el segmento de Enterprise and Partner Group (EPG). Este segmento se conforma de un amplio equipo de ventas, partners y especialistas técnicos de diferentes tecnologías, cuya misión es dirigir, apoyar e inspirar a los clientes empresariales en su viaje de transformación a través de dispositivos conectados y servicios en la nube.

En la figura 1.2 se muestra el organigrama, considerando los puestos más relevantes para dar un adecuado contexto del segmento donde desempeñé mis actividades.

³ Basado en el diagrama de "Subsidiary structure" (Vaks, 2017)



Figura 1.2. Organigrama Microsoft. ⁴

⁴ Elaboración propia. Información extraída de LinkedIn.

1.5 Contexto del Puesto

El puesto de Cloud Technical Specialist (CTS) surge como respuesta a la transformación digital que está surgiendo a nivel global y a la estrategia de Microsoft: Mobile-First, Cloud-first.

La transformación digital es movimiento tecnológico en el cual las empresas buscan reinventarse utilizando la tecnología como ventaja competitiva. Esto implica repensar cómo opera nuestra organización y como eso genera valor en nuestros clientes.

A manera de resumen, Microsoft identifica cuatro pilares en la transformación digital:

1. Conquistar a los clientes

Buscar nuevas formas de atraer y enganchar a nuestros clientes. Mejorar y ofrecer una experiencia de compra unificada y personalizada.

2. Fortalecer a nuestros empleados

Poner al alcance de los empleados el conocimiento y las herramientas que necesitan, de tal manera que puedan aprender y tomar mejores decisiones con mayor rapidez.

3. Optimizar nuestras operaciones

Reducir riesgos. Acelerar la capacidad de respuesta para descubrir y predecir ineficiencias y hacer los ajustes necesarios.

4. Transformar nuestros productos

Las empresas deben de estar en una búsqueda de formas de entregar valor a sus clientes y con ello mejorar sus ganancias.



Figura 1.3. Pilares de la Transformación Digital⁵

⁵ (Microsoft, Digital Transformation, 2017)

Como CTS operé en el segmento de EPG, enfocándome en las cuentas HiPo. El término HiPo viene de “High Potential” y corresponden a esta clasificación aquellos clientes de Microsoft que tienen potencial y posibilidad económica para hacer realidad su transformación digital mediante el uso de la nube.

Microsoft México tiene más de 30 cuentas HiPo entre las cuales figuran empresas transnacionales de diferentes sectores productivos, por ejemplo: gobierno, educación, servicios financieros, retail, entretenimiento y bienes de consumo.

El equipo HiPo se organiza en células, conformadas por un Principal Solution Specialist (PSS), dos Cloud Solution Architects (CSA) y un Data Solution Architect (DSA).

Cada célula se dedica a atender alrededor de diez cuentas. Por otra parte, los CTS son un recurso dinámico que puede trabajar en cuentas de diferente célula. Puede haber tantos CTS como sea necesario en cada célula.

A continuación, describo brevemente las funciones de los miembros de la célula:

1. Principal Solution Specialist (PSS)⁶

Figura comercial, persona responsable de detectar una oportunidad de venta de Azure. Detecta escenarios de Azure que puedan dar valor a los clientes, así como la disposición del cliente a adoptar estos escenarios. Involucra al resto de los roles para diseñar soluciones, mantener conversaciones técnicas y sustentar la venta.

2. Cloud Solution Architects (CSA)⁷

Persona encargada de diseñar arquitecturas y soluciones en la nube pública Azure. Ayuda a sustentar de manera teórica/técnica la venta. Realiza pruebas de concepto y demostraciones que permitan mostrar al cliente el valor de las soluciones diseñadas. Es responsable de guiar al cliente en su transición hacia la nube y el consumo de sus servicios. Habilita a los clientes con su conocimiento experiencia para poder operar Azure.

3. Data Solution Architects (DSA)⁸

Persona encargada de diseñar soluciones de Big Data y analítica avanzada de datos. Ayuda a sustentar de manera teórica/técnica le venta. Realiza pruebas de concepto y demostraciones que permitan mostrar al cliente el valor de las soluciones diseñadas.

⁶ (Microsoft, Pricipal Solution Spec, 2017)

⁷ (Microsoft, Cloud Solution Architect, 2017)

⁸ (Microsoft, Data Cloud Solution Architect, 2017)

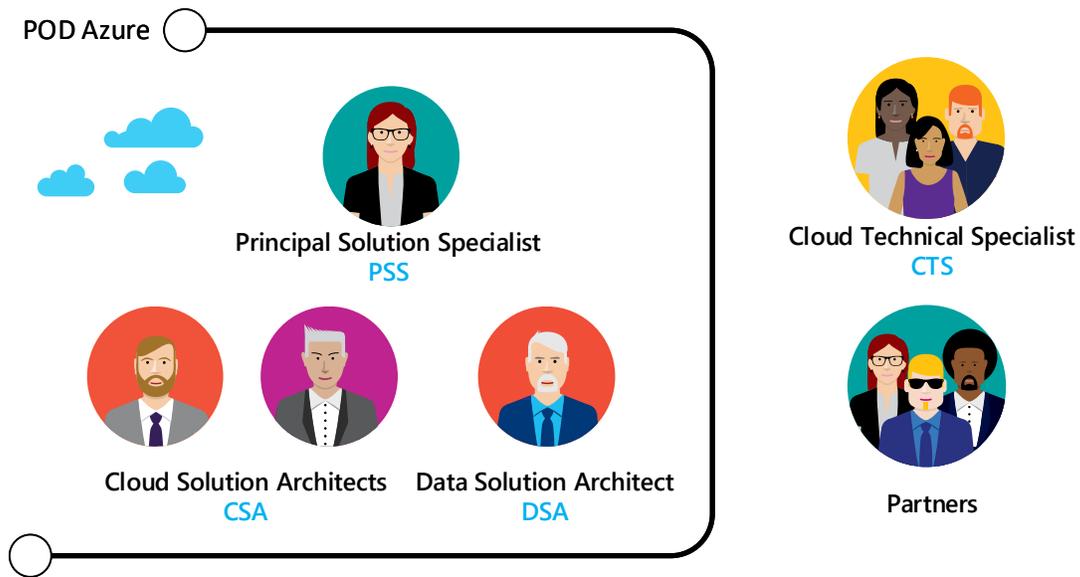
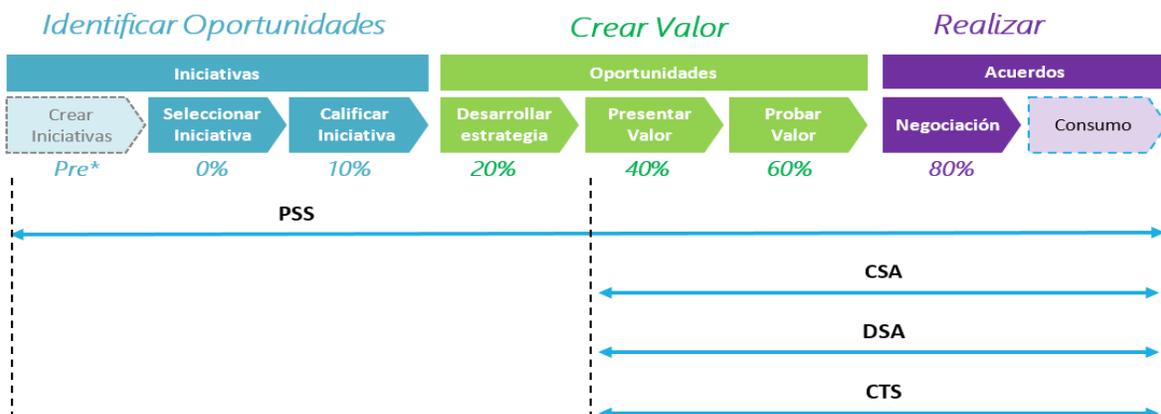


Figura 1.4. Organización de una célula de EPG.

El CTS es un rol técnico, de cara al cliente, que es responsable de la experiencia de despliegue y consumo de nube de clientes. El CTS participa en sesiones de diseño arquitectónico, proyectos específicos de implementación y / o Pruebas de Concepto (PoC).

Colabora con la célula principalmente cuando la oportunidad está al 40% del proceso de venta de Microsoft (Ver Figura 1.5) y hasta la negociación final. Trabaja de la mano con los arquitectos de la célula para acelerar el proceso de venta e impulsar a los clientes a usar Azure. Como parte de un equipo de ventas, mi objetivo es impulsar el crecimiento de los ingresos y la cuota de mercado frente a los competidores a través de la solución y la venta técnica de servicios en la nube.



1.6 Responsabilidades clave

- Colaborar en el despliegue de las cargas de trabajo de los clientes en Azure y aumentar su consumo de la plataforma, proporcionando orientación de implementación, apoyando el desarrollo del modelo de adopción de nube de los clientes y proporcionando recomendaciones apropiadas para superar obstáculos técnicos.
- Construir relaciones sólidas a largo plazo con clientes y socios. Posicionarse como un consejero de confianza, centrándose realmente en cómo las tecnologías y los servicios de Microsoft pueden ayudar a clientes a tener éxito y a alcanzar sus objetivos de negocio.
- Identificar, validar y aumentar las oportunidades de acelerar el consumo en cuentas de clientes de alto potencial (HiPo), en colaboración con el equipo de ventas, mediante la arquitectura de soluciones para Microsoft y soluciones de terceros.
- Acelerar el consumo en las cuentas de clientes HiPo proporcionando experiencia técnica.
- Mantenerse al día con las tendencias de tecnología, el mercado y las ideas competitivas.
- Ningún CTS o miembro del POD tiene permitido realizar implementaciones en ambientes productivos, ya que no hay manera de que el equipo brinde soporte al cliente. Cuando el cliente requiera la implementación de las soluciones en producción, debe contratar a un socio o bien a Microsoft Consulting Services.

⁹ Imagen Adaptada de (Microsoft, Showcase Azure, 2017)

II. Antecedentes

Contoso Ltd. es una cadena mexicana de mercados y almacenes con presencia en 277 municipios a lo largo de los 32 estados de la República. Su superficie de venta está conformada por más de 4.3 millones de metros cuadrados distribuidos en diferentes formatos de tienda.

Su misión se enfoca en satisfacer las necesidades de productos y servicios de las comunidades mexicanas, asegurar una relación permanente y de valor con sus clientes, colaboradores, proveedores, accionistas, comunidad y medio ambiente, obteniendo así rentabilidad, permanencia y crecimiento.



Figura 2.1. Distribución de Tiendas de Contoso Ltd. en México.

El año 2015 fue año trascendental para Contoso, ya que fue un año de cambios importantes en su cultura organizacional, estrategia comercial, imagen y plataforma tecnológica.

Uno de los cambios más relevantes ocurridos en ese año, fue la adquisición de 143 unidades de autoservicio de otra cadena de supermercados ampliamente reconocida en el país, lo que posicionó a Contoso Ltd. como la principal cadena de autoservicio 100% mexicana. Como parte de esta adquisición, Contoso adquirió bienes inmobiliarios,

contratos de arrendamiento con terceros, equipo operativo, inventario de tiendas, locales y espacios comerciales, la plataforma logística, centros de distribución, el licenciamiento sobre la plataforma tecnológica y sistemas de información, ciertos derechos de uso de marcas y campañas promocionales, entre otros activos estratégicos.

Por otra parte, a inicios del año 2016, lanzaron su tienda en línea www.contoso.com¹⁰, una plataforma de comercio electrónico que busca complementar y ampliar la oferta comercial de sus tiendas físicas con miles de productos novedosos, un catálogo diferenciado y exclusivo para ventas por Internet.

Ante esta importante transformación de la empresa, se pueden vislumbrar grandes retos de TI. Probablemente por esta razón, Contoso Ltd. fue nominada como cuenta HiPo al inicio del año fiscal 2017 de Microsoft como un ambicioso proyecto de transformación digital.

Nuestro primer acercamiento con el cliente fue con el departamento de TI (Tecnologías de la Información), donde la principal preocupación era la actualización y unificación de las plataformas tecnológicas actuales y las adquiridas.

Por otra parte, nos pusimos en contacto con el área de Comercio electrónico, donde rápidamente se detectó interés en el campo de analítica avanzada.

Mi colaboración estuvo enfocada a dos oportunidades:

- **Datawarehouse en Azure**
Contoso reportó problemas de performance (tiempo de respuesta y procesamiento) en su datawarehouse debido a las limitaciones de su infraestructura on-premise¹¹. Además, requería escalar su infraestructura a corto plazo como parte de la unificación de la plataforma tecnológica consecuencia de la reciente adquisición de las unidades de autoservicio.
- **Log Analytics en Azure**
Contoso deseaba explotar la información generada por los Logs de su plataforma de comercio Electrónico SAP Hybris, con el propósito de conocer mejor a sus clientes, tomar mejores decisiones, crear nuevas estrategias de venta y crecer su ventaja competitiva en la industria.

¹⁰ URL ficticia para efectos de presente reporte.

³ On-premise, se refiere a la infraestructura que se encuentra físicamente en un centro de datos.

El presente documento se enfoca en el proyecto de *Log Analytics*, donde mi participación consistió en levantar los requerimientos de dicho sistema, diseñar una arquitectura e implementar el sistema bajo una prueba de concepto.

Por temas de confidencialidad, los detalles del proyecto de DWH quedan fuera del alcance de este informe.

III. Definición del Problema

A partir del lanzamiento de su sitio de comercio electrónico, Contoso creó un departamento de análisis de datos, conformado principalmente por actuarios y matemáticos. Una de sus tareas es analizar la información generada por las plataformas que conciernen al comercio electrónico tales como SAP Hybris y CRM (Customer Relationship Management).

SAP Hybris es una familia de soluciones de SAP¹². enfocadas al comercio electrónico basado en una arquitectura modular que permite su integración con diferentes servicios. Proporciona una avanzada gestión y recolección de datos de clientes, herramientas de marketing y procesos comerciales unificados.

La base de desarrollo de Hybris, está realizada sobre estándares abiertos y pensada para ser extensible de manera que los desarrollos puedan añadir y/o modificar la funcionalidad de la herramienta sin necesidad de modificar el núcleo de la misma.

Contoso tenía una máquina virtual (MV) con sistema operativo Red Hat Enterprise 6.7, la cual estaba alojada en un centro de datos de KIO Networks. Esa MV almacenaba los logs de cuatro servidores de la solución SAP Hybris. Los logs de SAP Hybris son generados diariamente y reportan las operaciones realizadas por los usuarios en el portal de Contoso <https://www.contoso.com/>. Por ejemplo, los usuarios pueden ver las promociones, consultar el catálogo de productos, incluirlos al carrito de compra y comprar. Todos estos movimientos en el portal se guardan en tales logs.

El departamento de Análisis de datos, dependía totalmente del departamento de TI para extraer los logs de SAP Hybris, lo cual era muy ineficiente por la carga de trabajo que tiene dicho departamento. Por otra parte, este departamento deseaba sacar ventaja de los datos generados por SAP Hybris, para entender el comportamiento e intereses de sus usuarios, ofrecer una mejor experiencia, promociones personalizadas e incrementar las ventas de comercio electrónico.

³ On-premise, se refiere a la infraestructura que se encuentra físicamente en un centro de datos.

⁴ SAP, empresa multinacional alemana dedicada al diseño de software para la gestión empresarial.

Se convino con el cliente, desarrollar un sistema bajo una prueba de concepto (PoC) en una suscripción de Azure proporcionada por Microsoft¹³, que le permitiera automatizar la extracción diaria de los logs, transformar los datos para ser almacenados en una base de datos relacional y generar un panel para consultar el conocimiento generado.

Una prueba de concepto implica que el sistema debe estar en un ambiente controlado, lo más parecido posible al del cliente y con datos del cliente. Sin embargo, esto no implica que el sistema pueda ser migrado de manera transparente a un ambiente productivo. Por otra parte, una prueba de concepto tiene como objetivo mostrar al cliente, con hechos, la funcionalidad de los servicios y el valor que Azure puede traer a su negocio. También, implica que el cliente está interesado y tiene el presupuesto para consumir Azure e implementar la solución.

¹³ Contoso no contaba con ningún tipo de contrato. A nivel corporativo, Azure se adquiere por un *Enterprise Agreement*. A partir de ese contrato se crean las suscripciones del cliente, donde puede consumir los recursos que requiera de Azure.

IV. Desarrollo

4.1 Ingeniería de requerimientos

Para obtener los requerimientos de una prueba de concepto es necesario tener varias sesiones con el cliente, lo cual nos permite entender sus necesidades y sus expectativas. Esto nos conduce a un modelo iterativo o en espiral.

Citando a Sommerville (2011), "La adquisición y el análisis de requerimientos es un proceso iterativo que se representa como una espiral de actividades: descubrimiento de requerimientos, clasificación y organización de requerimientos, negociación de requerimientos y documentación de requerimientos" (p.115).

Este modelo en espiral además de arrojar los requerimientos, nos permite comprender la necesidad del cliente, sus expectativas, oportunidades de negocio y presupuesto.

Para el descubrimiento de los requerimientos, se elaboró una matriz que reflejara los criterios de éxito de la Prueba de Concepto. Este ejercicio nos ayuda mucho a aterrizar las ideas del cliente, clasificar los requerimientos y empezar la evaluación técnica del escenario. A continuación, se presenta dicha matriz:

#	Criterio de éxito	Cumple	No cumple
1	Está montado en la nube.		
2	Funciona sin intervención humana.		
3	No perjudica el performance de la MV en KIO.		
4	Almacena la información en una base de datos relacional.		
5	Acceso a elementos del sistema mediante contraseñas.		
6	Acceso a máquina virtual vía SSH.		
7	Conexión VPN		
8	Los logs crudos deben estar en un repositorio en la nube organizados por servidor y fecha.		
9	Los logs procesados deben estar en un repositorio en la nube organizados por servidor y fecha.		
10	Funciona fuera de las horas con mayor tráfico.		
11	Es integrable con nuevos sistemas y áreas del negocio.		
12	Genera informes automáticamente cada día.		
13	Arroja información relevante para el negocio.		
14	La información desplegada refleja lo que está ocurriendo en tiempo real.		
15	Solo el cliente tendrá acceso a las visualizaciones realizadas.		
16	Las visualizaciones podrán realizarse desde un navegador web.		
17	Disponibilidad del sistema		
18	Interoperabilidad entre tecnologías OS y Windows.		

Tabla 4.1. Criterios de éxito para prueba de concepto

4.1.1 Clasificación y organización de requerimientos

A partir del ejercicio anterior, se clasificaron y puntualizaron los siguientes requerimientos por prioridad:

- *Requerimientos Funcionales*

1. El sistema debe estar totalmente montado en la nube.
2. La información debe extraerse, limpiarse y cargarse de forma automática.
3. Se debe generar un reporte visual con el resumen de los datos analizados.
4. Se requiere almacenar cada registro de los logs en una base de datos relacional.
5. La conexión a la máquina virtual debe realizarse vía Secure Shell (SSH).
6. Se debe usar Forticlient como asistente para establecer la VPN.
7. Las visualizaciones podrán realizarse desde un navegador web.

- *Requerimientos no funcionales*

1. La solución propuesta debe tener impacto mínimo en el performance de la máquina virtual que actúa como fuente de datos.
2. La extracción se debe realizar fuera de horas pico.
3. Los accesos a los módulos del sistema deben estar restringidos por contraseñas.
4. El sistema debe soportar el manejo de grandes bloques de datos.
5. Los informes pueden ser consultados en Internet solo por las personas interesadas.
6. Se deben almacenar en un directorio los logs en crudo a manera de histórico, y en otro directorio los logs procesados.
7. Facilidad para crear informes.
8. El sistema debe ser fácilmente escalable.
9. El sistema debe tener la flexibilidad de integrarse con nuevos departamentos de Contoso.
10. Interoperabilidad entre tecnologías Open Source y Windows.
11. La prueba de concepto tiene una duración máxima de 3 meses.
12. Microsoft deberá entregar un documento que resuma el funcionamiento del sistema.

4.1.2 Diagrama de casos de uso

Los casos de uso describen en forma de lista de acciones y de interacciones el comportamiento de un sistema, estudiado desde el punto de vista de los actores. Definen los límites del sistema y sus relaciones con el entorno. (Laurent Debraver, 2016).

Dadas las características del sistema, podemos ver que se compone de tres grandes bloques y no tiene gran interacción con el usuario final. Esto se puede visualizar gráficamente en el diagrama de casos de la figura 4.1.

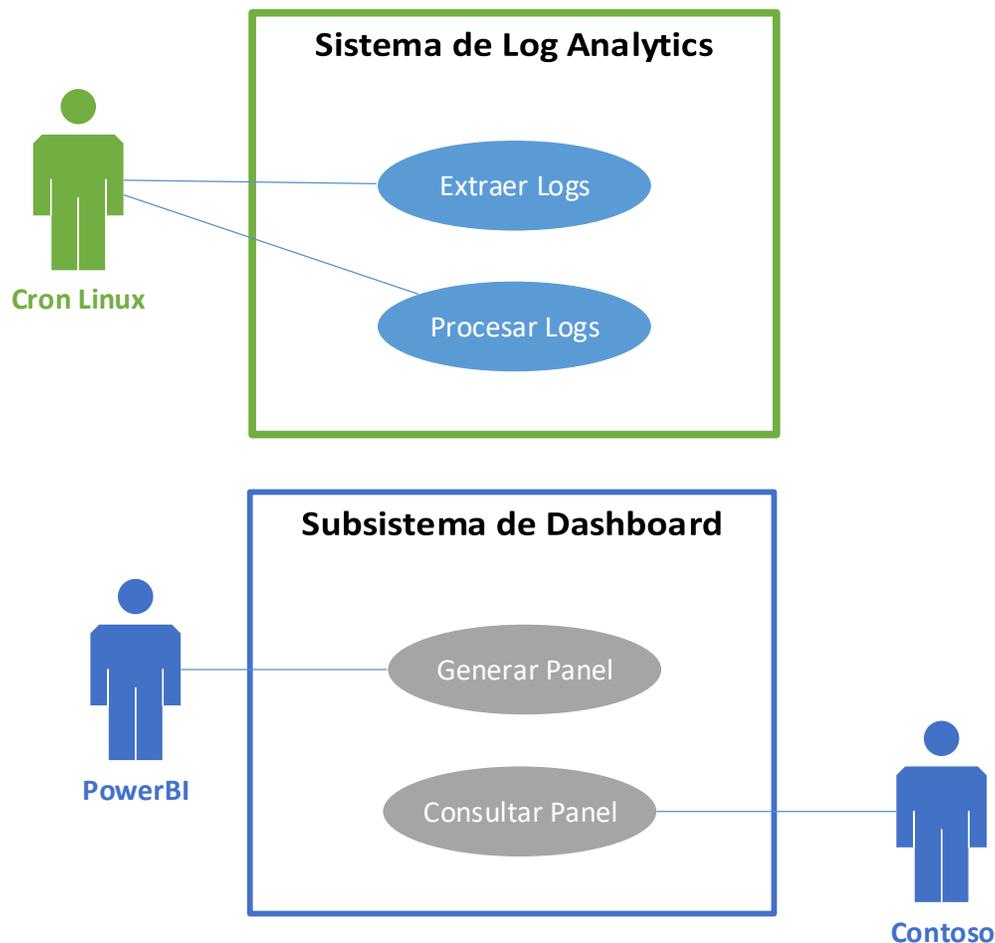


Figura 4.1. Diagrama de Casos de Uso.

Tal como se muestra en el diagrama, el sistema se compone de dos casos de uso: Extraer Logs y Generar Logs. Por otra parte, el subsistema de paneles (dashboard), usa los datos

generados por el sistema de Log Analytics y los carga en los paneles de BI. Además, el cliente Contoso funge como actor puesto que es quien hará las consultas de los paneles.

Para ayudarme a especificar los requerimientos funcionales, realicé la representación textual de los casos de uso, ya que de esta forma se puede dar más detalle de los componentes, acciones y reacciones del sistema.

A continuación, se presenta textualmente los casos de uso identificados:

Caso de Uso	Extraer Logs
ID. DEL CU	1
ACTORES PARTICIPANTES	Cron Linux – Demonio del sistema operativo Linux el cual permite ejecutar automáticamente comandos o scripts a una hora y/o fecha específica.
BREVE DESCRIPCIÓN	Permite levantar el túnel VPN hacia KIO Networks y copiar los logs generados por la plataforma Hybris mediante Secure Copy (SCP) a una máquina virtual en Azure. Así mismo, permite subir los logs en crudo a una cuenta de almacenamiento en Azure.
PRE-CONDICIONES	Del proceso <ul style="list-style-type: none"> • Los scripts deben tener permisos de ejecución. • Se deben tener permisos de acceso a la VPN de Contoso a KIO. Del sistema <ul style="list-style-type: none"> • El cron de Linux debe estar configurado para disparar la acción. • Debe de existir una Base de datos.
FLUJO PRINCIPAL	Este Caso de Uso inicia cuando el cron de Linux dispara la ejecución de un script. El script establece un canal VPN de una máquina virtual en Azure a una máquina virtual alojada en KIO networks (E-1) , asegura la conexión y verifica en la base de datos la fecha del último Log almacenado para posteriormente hacer una lista de fechas al día en curso. Mediante SCP (secureCopy) y a través del puerto 22, lleva a cabo la copia de todos los logs que correspondan a la lista generada (E-2) . Si la copia se ha realizado con éxito, estos se suben a un contenedor en una cuenta de almacenamiento en Azure (E-3) .

	<p>Por último, se crea un registro en la base de datos con los datos del Log copiado. En el registro se establece un atributo de "Status" con valor de cero (0=no procesado) (E-4).</p> <p>Termina el Caso de Uso.</p>
FLUJOS ALTERNOS	No Aplican.
FLUJOS DE EXCEPCIÓN	<p>E-1 No se puede establecer conexión VPN En caso no poder establecer comunicación con KIO Networks, reintenta la conexión hasta 10 veces. Si nunca se logra la conexión se reporta en el archivo ErroresVPN.txt</p> <p>E-2 No se pudo copiar Log En caso de no poder copiar el Log se registra el error en un archivo almacenado en /home/vmguest/contosoLogSystem/ErroresGetLog.txt</p> <p>E-3 No se pudo subir Log a cuenta de Almacenamiento en Azure En caso de no poder subir el Log a Azure, se registra el error en un archivo almacenado en /home/vmguest/contosoLogSystem/ErroresBlob.txt</p> <p>E-4 No se pudo registrar en la base de datos En caso de no poder crear el registro en la base de datos por duplicidad u otra excepción, se registra el error en un archivo almacenado en /home/vmguest/contosoLogSystem/ErroresInsertBDLog.txt</p>
POST-CONDICIONES	<ul style="list-style-type: none"> • El sistema queda en espera hasta la ejecución del caso de uso "Procesar Logs"

Tabla 4.2 Caso de Uso "Extraer Logs".

Caso de Uso	Procesar Logs
ID. DEL CU	2
ACTORES PARTICIPANTES	Cron Linux – Demonio del sistema operativo Linux el cual permite ejecutar automáticamente comandos o scripts a una hora y/o fecha específica.
BREVE DESCRIPCIÓN	Orquesta el procesamiento de logs y genera tablas que facilitan la creación de visualizaciones en Power BI. Así mismo, permite subir los logs procesados a una cuenta de almacenamiento y a una base de datos en Azure.
PRE-CONDICIONES	<p>Del proceso</p> <ul style="list-style-type: none"> • Los scripts deben tener permisos de ejecución. <p>Del sistema</p> <ul style="list-style-type: none"> • El cron de Linux debe estar configurado para disparar la acción. • Debe de existir una Base de datos.
FLUJO PRINCIPAL	<p>Este Caso de Uso inicia cuando el cron de Linux dispara la ejecución de un script.</p> <p>El script consulta en la base de datos la fecha del último Log procesado y hace una lista de fechas al día en curso.</p> <p>A continuación, se procesa cada log de la lista mandando a llamar a otro script que se encarga de leer cada línea del log y que a su vez le da un formato predeterminado para su posterior análisis. Cada línea procesada se escribe en un nuevo archivo de extensión csv (Comma Separated Value) y se inserta como un nuevo registro en la base de datos en la tabla de Operaciones (E-1).</p> <p>Una vez terminado el procesamiento de todos los logs de la lista, estos se suben a un contenedor en una cuenta de almacenamiento en Azure (E-2).</p> <p>Posteriormente se actualiza el atributo de "Status" con valor de uno (1= procesado) en la tabla de Logs de la Base de datos (E-1).</p> <p>Por último, se crean varias vistas (Departamento, Compra, Producto, Búsqueda), cuyo objetivo es quitar carga de procesamiento a la herramienta de BI. En estas vistas se hacen correcciones de ortografía y se filtran las operaciones correspondientes a la Tabla. (E-1)</p> <p>Termina el Caso de Uso.</p>
FLUJOS ALTERNOS	No Aplican.

<p>FLUJOS DE EXCEPCIÓN</p>	<p>E-1 No se pudo registrar en la base de datos En caso de no poder crear el registro en la base de datos por duplicidad u otra excepción, se registra el error en un archivo almacenado en /home/vmguest/contosoLogSystem/ErroresInsertBDLog.txt</p> <p>E-2 No se pudo subir Log a cuenta de Almacenamiento en Azure En caso de no poder subir el Log a Azure, se registra el error en un archivo almacenado en /home/vmguest/contosoLogSystem/ErroresBlob.txt</p>
<p>POST-CONDICIONES</p>	<p>No Aplican.</p>

Tabla 4.3 Caso de Uso "Procesar Logs".

Caso de Uso	Generar Panel
ID. DEL CU	3
ACTORES PARTICIPANTES	Power BI – colección de servicios de software, aplicaciones y conectores que funcionan conjuntamente para convertir orígenes de datos sin relación entre sí en información coherente, interactiva y atractiva visualmente.
BREVE DESCRIPCIÓN	Permite conectar a la base de datos generada por el sistema de Log Analytics y generar los paneles de visualizaciones en intervalos de tiempo específicos.
PRE-CONDICIONES	Del proceso <ul style="list-style-type: none"> • Debe existir una base de datos compatible con el conector de PowerBI online. Del sistema <ul style="list-style-type: none"> • Debe estar alojado en Power BI online, las plantillas diseñadas para los paneles de visualizaciones. • Se debe haber configurado el tiempo de refrescamiento de datos en Power BI.
FLUJO PRINCIPAL	Este Caso de Uso inicia cuando el temporizador de Power BI dispara un proceso que permite conectarse a la base de datos alojada en Azure. (E-1) Una vez establecida la conexión, PowerBI extrae la información de los últimos quince días y genera los paneles de acuerdo a las plantillas cargadas en el sistema. Termina el Caso de Uso.
FLUJOS ALTERNOS	No aplican.
FLUJOS DE EXCEPCIÓN	E-1 No se pudo conectar a la base de datos En caso de no poder conectarse a la base de datos, se registra el error en la pestaña de "Historial de ejecución" y no se actualizan los datos del panel.
POST-CONDICIONES	Se generan paneles que solo son visibles para los usuarios especificados en la plataforma. Los permisos son de lectura.

Tabla 4.4 Generar Panel.

Caso de Uso	Consultar Panel
ID. DEL CU	4
ACTORES PARTICIPANTES	Contoso – Es cualquier persona de la empresa Contoso a quien se le haya dado permisos de visualización sobre el link de powerBI.
BREVE DESCRIPCIÓN	Permite al usuario interactuar con los paneles generados por el sistema mediante el portal de PowerBI online.
PRE-CONDICIONES	Del proceso <ul style="list-style-type: none"> • El usuario debe permisos de lectura en la plataforma de Power BI online. Del sistema <ul style="list-style-type: none"> • El usuario debe ingresar sus credenciales en la plataforma de powerBI online.
FLUJO PRINCIPAL	Este Caso de Uso inicia cuando el usuario ingresa al portal de Power BI online. El portal muestra los paneles generados a partir de la información generada por el sistema de Log Analytics. El portal permite al usuario dar clic a las visualizaciones de su interés y filtrar la información que desee consultar, éstas pueden ser consultadas en cualquier orden, también puede hacer “drill down” a algunas visualizaciones para obtener información diferente a la que fue previamente programada en los paneles. A-1 El portal permite al usuario imprimir reportes de las visualizaciones de su interés. Termina el Caso de Uso.
FLUJOS ALTERNOS	A-1 Crear nuevas visualizaciones El portal cuenta con una función que permite realizar preguntas sobre la información expuesta para crear nuevas visualizaciones. Para ello, se deben formular preguntas en inglés, por ejemplo: Which is the most visited department?, What is the most unpopular product? How many purchases are made on September 15 th ?
FLUJOS DE EXCEPCIÓN	No aplican.
POST-CONDICIONES	Después de realizar las consultas, se debe cerrar sesión.

Tabla 4.5 Caso de Uso “Consultar Panel”.

4.2 Propuesta de solución

4.2.1 Arquitectura

El sistema propuesto, se compone de procesos ETL (Extract, Transform, Load).

“Los procesos ETL, permiten a las organizaciones mover datos desde múltiples fuentes, reformatearlos y cargarlos en otra base de datos con el objetivo de analizarlos.” (Espinosa, 2016)

A continuación, se describen los tres procesos en el contexto del problema:

- **Extracción:** Se trae los datos desde la máquina virtual alojada en KIO Networks a una máquina virtual alojada en Azure. Los datos crudos se depositan en un contenedor dentro de una cuenta de almacenamiento de datos en Azure.
- **Transformación:** Mediante scripts de Python se transforman los datos. La transformación consiste en quitar datos irrelevantes, caracteres non-ascii y dar formato csv a la información.
- **Carga:** La información procesada se sube a una cuenta de almacenamiento en Azure y a una base de datos SQL en la nube.

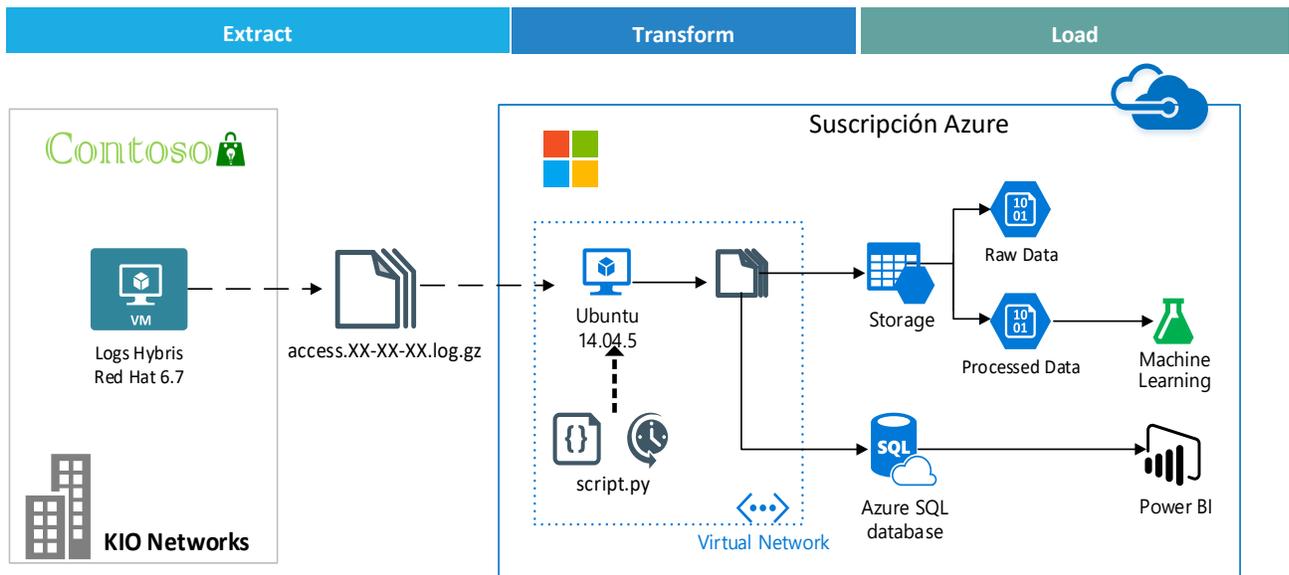


Figura 4.2. Arquitectura de solución propuesta

4.2.2 Simbología

Símbolo	Descripción
	<p>Máquinas Virtuales</p> <p>Compatibles con Linux, Windows Server, SQL Server, Oracle, IBM y SAP, ofrecen una amplia flexibilidad de virtualización para una variada gama de soluciones informáticas: desarrollo y pruebas, ejecución de aplicaciones y ampliación del centro de datos. Azure ofrece flexibilidad para configurar el tamaño y el rendimiento de la máquina virtual para ajustarse incluso a las necesidades empresariales más exigentes. Permite combinar el rendimiento de un súper equipo de clase mundial con la escalabilidad de la nube.</p>
	<p>Logs Contoso</p> <p>Logs generados por Hybris, los cuales son extraídos desde la máquina virtual en KIO Networks para su procesamiento y almacenamiento en la nube.</p>
	<p>Cuenta de Almacenamiento (Azure Storage Account)</p> <p>Almacenamiento de Azure es la solución de almacenamiento en la nube para las aplicaciones modernas que dependen de la durabilidad, la disponibilidad y la escalabilidad para satisfacer las necesidades de sus clientes. Es muy escalable, elástico y accesible globalmente, y equilibra automáticamente la carga de los datos en función del tráfico existente. Azure Storage proporciona cinco servicios que cubren las necesidades de las aplicaciones: Blob Storage, Table Storage, Queue Storage, Disk Storage y File Storage.</p>
	<p>Binary Large Object (BLOB)</p> <p>Objetos escalables de forma masiva para datos no estructurados (imágenes, vídeos, audio, documentos, etc.) con tecnología de eficacia probada a escala de exabytes. Azure Blob Storage administra billones de objetos almacenados con una media de millones de solicitudes por segundo para clientes de todo el mundo.</p>

Símbolo	Descripción
	<p>Base de datos SQL</p> <p>SQL Database es un servicio de bases de datos relacionales de Microsoft Cloud que usa el motor de Microsoft SQL Server, es un servicio Plataforma as a Service (PaaS), diseñada para desarrolladores, con inteligencia integrada que aprende patrones de aplicaciones y se adapta para maximizar el rendimiento, la confiabilidad y la protección de datos.</p>
	<p>Cron Linux</p> <p>Demonio del sistema operativo Linux el cual permite ejecutar automáticamente comandos o scripts a una hora y/o fecha específica.</p>
	<p>Machine Learning</p> <p>Técnica de ciencia de datos que ayuda a los equipos a aprender de los datos existentes para prever tendencias, resultados y comportamientos futuros. En Azure es un servicio en la nube totalmente administrado que le permite crear, implementar y compartir fácilmente soluciones de análisis predictivo.</p>
	<p>Power BI</p> <p>Es una herramienta de Business Intelligence que transforma los datos de una compañía en objetos visuales enriquecidos que puede analizar, recopilar y organizar.</p>

Tabla 4.6. Arquitectura de solución propuesta

A continuación, se explica brevemente cómo interactúan los elementos de la arquitectura propuesta:

1) Máquina virtual

Aloja y orquesta scripts en el lenguaje de Python creados para la extracción de los logs de Hybris y el procesamiento de los mismos. Tales scripts se ejecutan mediante el cron del usuario Root de Linux.

2) *Azure Storage*

Almacena los logs en crudo, así como los logs procesados, ordenados por IP del servidor, año y mes.

3) *SQL Azure DB*

Almacena un catálogo de los logs que se encuentran almacenados (crudos/procesados) hasta el día actual, los logs procesados y tablas que facilitan las visualizaciones en Power BI.

4) *Power BI*

PBI se alimenta con la información contenida en la base de datos para generar visualizaciones que reportan lo que hacen los usuarios en el portal web.

5) *Machine Learning*

La información procesada sirve como fuente de datos al servicio de Machine Learning para generar conocimiento y análisis predictivo.

4.3 Ambiente creado en Microsoft Azure

En las siguientes secciones se detallará cómo se implementó la arquitectura mostrada en la Figura 4.2

4.3.1 Especificación técnica de Máquina Virtual

Azure cuenta con un amplio catálogo de máquinas virtuales. Sin embargo, escogí utilizar una máquina de la serie¹⁴ D3v2 ya que por sus características satisface los requerimientos. Las máquinas virtuales de la serie Dv2 proporcionan procesadores más rápidos, una mayor proporción de memoria a vCPU y una unidad de estado sólido (SSD) para el disco temporal. Las máquinas de la serie Dv2, "se basan en el procesador Intel Xeon E5-2673 v3 (Haswell) de 2,4 GHz de la última generación; y con Intel Turbo Boost Technology 2.0, puede alcanzar los 3,1 GHz.

Están diseñadas para ejecutar aplicaciones que exigen mayor capacidad de proceso y rendimiento de disco temporal.

¹⁴En Azure, las máquinas virtuales se clasifican en Series. Una serie es un conjunto de máquinas que comparten características de hardware y están optimizadas para satisfacer propósitos específicos. Por ejemplo, las MV de la serie N, se especializan representación de gráficos pesados y la edición de vídeo. Están disponibles con uno o varios GPU.

A petición del cliente, se creó una máquina virtual con sistema operativo Ubuntu en el Portal de Azure. El cliente indicó que su personal estaba familiarizado con esta distribución por lo que le era funcional, además, así se cubrió el requerimiento de integrar tecnologías Open Source.

Esta MV, aloja y ejecuta automáticamente los scripts que se encargan del proceso de ETL.

D3_V2 Standard	
4	vCPUs
14	GB
	8 Data disks
	8x500 Max IOPS
	200 GB Local SSD
	Load balancing

Figura 4.3. Características de una máquina virtual D3_v2

Configuración máquina virtual Ubuntu

1. Se creó el árbol de directorios que existe en Hybris Consoto Ltd.

```
$mkdir -p /home/vmquest/RAW/172.35.45.5  
$mkdir -p /home/vmquest/RAW/172.35.45.6  
$mkdir -p /home/vmquest/RAW/172.35.45.7  
$mkdir -p /home/vmquest/RAW/172.35.45.8
```

2. Para poder habilitar la ejecución de los scripts, se requirió bajar los siguientes paquetes desde bash:

```

#Habilitar herramienta pip y paquetes de python
sudo apt-get install sshpass
sudo apt-get install expect
sudo apt-get install python-pip python-dev build-essential
sudo apt-get install build-essential libssl-dev libffi-dev openssl
sudo apt-get install python-setuptools
sudo pip install --upgrade pip
sudo pip install cryptography
sudo pip install pyopenssl ndg-httpsclient pyasn1

#Biblioteca para subir a Azure Storage desde python
sudo pip install azure-storage

#Biblioteca para habilitar SQL Azure desde Python
sudo apt-get --assume-yes update
sudo apt-get --assume-yes install freetds-dev freetds-bin
sudo python -m pip install --no-binary pymssql pymssql

#Bibliotecas para cliente Fortinet
apt-get update
apt-get install lib32z1 lib32ncurses5 lib32bz2-1.0
apt-get install lib32stdc++6

```

4.3.2 Virtual Private Network (VPN) Forticlient

De acuerdo a los requerimientos del cliente, se utilizó Forticlient para establecer el túnel VPN.

Forticlient es un cliente de VPN para conectarse a dispositivos Fortigate. Fortigate, es un firewall basado en hardware desarrollado por Fortinet.

Se ejecutaron los siguientes comandos para descargar e instalar el cliente de Fortinet:

```

root@vmAzurePoC# wget -q http://kb.arubacloud.com/files/tar-gz/forticlientsslvpn_linux_4-0-2281-tar.aspx -O fortisslvpn.tgz && tar -xzf fortisslvpn.tgz
cd forticlientsslvpn && ./helper/setup.linux.sh

root@vmAzurePoC# cd /etc/apt/sources.list.d

root@vmAzurePoC# echo "deb http://old-releases.ubuntu.com/ raring main restricted universe multiverse" > ia32-libs-raring.list

```

4.3.3 Desarrollo de Scripts

Los logs generados por el servidor Apache Tomcat de Hybris, correspondían al formato NCSA extended/combined, a continuación, se explica el patrón de la información:

LogFormat "%h %l %u %t \"%r\" %>s %b \"%{Referer}i\" \"%{User-agent}i\""
combined

- %h Host remoto (IP del cliente)
- %l Identificador del cliente
- %u Usuario determinado por autenticación HTTP
- %t Tiempo en que se recibió la solicitud
- %r Petición HTTP del cliente ("GET / HTTP/1.0")
- %>s Código HTTP de status enviado del servidor al cliente (200, 404, 302...)
- %b Tamaño de la respuesta al cliente (in bytes)
- Referer URL anterior del cliente
- User-agent Navegador

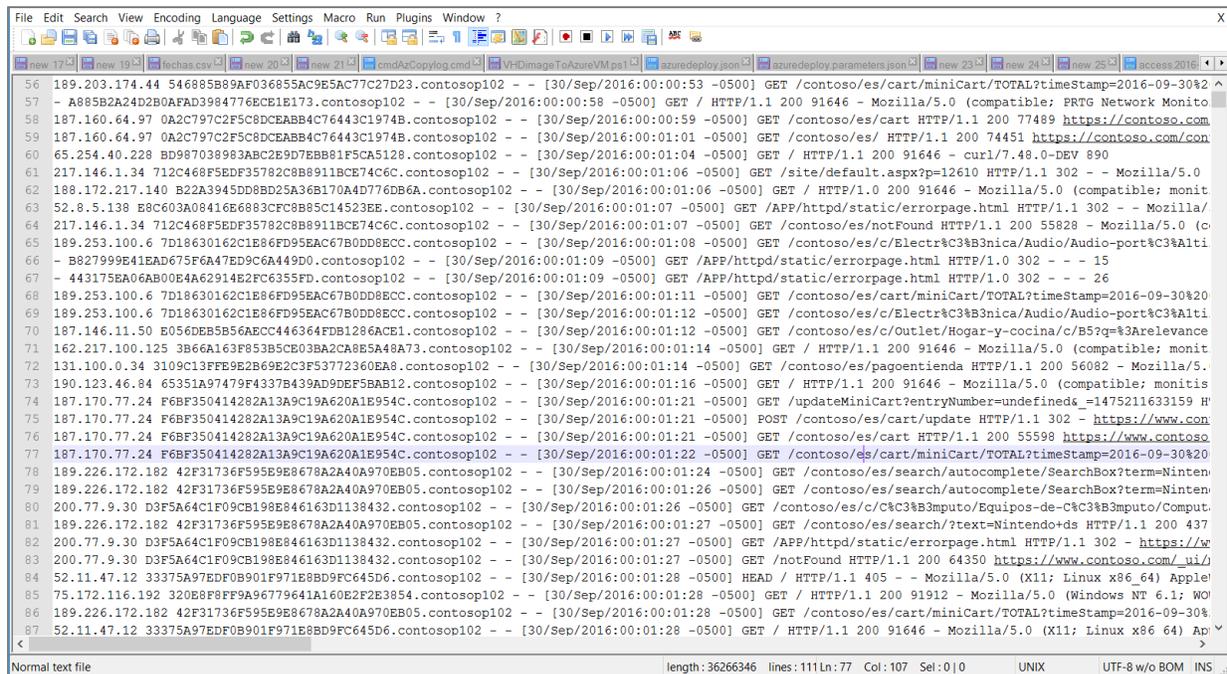


Figura 4.4. Ejemplo de Log generado por Hybris

La solución se compone de scripts en bash y Python. Se eligió Python como lenguaje de programación por ser compatible con los SDK de Azure, así como por la variedad de funciones que facilitan el manejo de cadenas, lo cual era un aspecto importante para el proceso de Transformación de datos definido en la sección 4.2.1.

En esta sección describe la funcionalidad de cada uno de los scripts desarrollados para todo el proceso de ETL¹⁵.

- *rawlogAz.py*

Orquesta el proceso de levantar el túnel VPN hacia KIO Networks y copia los logs generados por Hybris mediante Secure Copy (SCP), que permite extraer la información con la misma autenticación y nivel de seguridad de SSH. Una vez copiados los logs en una unidad de almacenamiento temporal, se suben a una cuenta de almacenamiento en Azure.

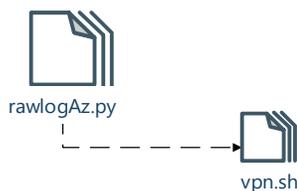


Figura 4.5. El script rawlogAz.py orquesta el proceso de copiar los logs desde KIO Networks a través de un túnel VPN.

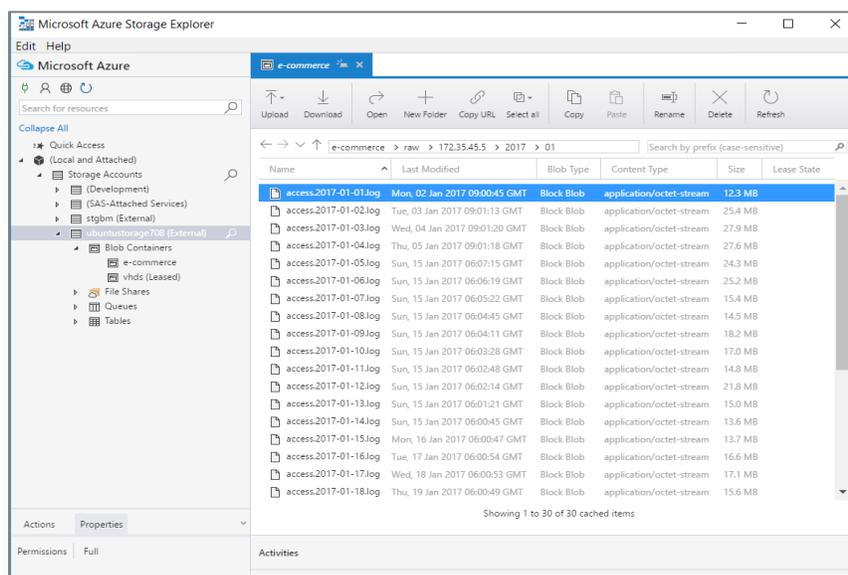


Figura 4.6. Logs en crudo en la cuenta de Almacenamiento en Azure.

¹⁵ Por los términos de privacidad y confidencialidad que firmé con Microsoft, no me es permitido compartir el código de los scripts.

- *vpn.sh*

Script Expect¹⁶. Levanta un túnel VPN de Azure a KIO Networks, es llamado por rawlogAz.py. Este script debe estar en el directorio de forticlientsslvpn que se genera durante la instalación del cliente de fortinet.

```
#!/usr/bin/expect -f
#Establecer conexión VPN

spawn -ignore HUP /home/vmguest/forticlientsslvpn/forticlientsslvpn_cli --server
202.115.17.37:443 --vpnuser msinfo
expect "Password for VPN:"
sleep 1
send "LJGD963$%&fo\r"
expect -exact "N)"
sleep 1
send "Y\r"
expect {
    -exact "STATUS::Tunel running" {exit 0}
    -exact "STATUS::Set up tunnel failed" {exit 1}
}
eof
}

-- INSERT --
```

Figura 4.7. Script para levantar VPN con forticlient.

- *proclgAz.py*

Orquesta el procesamiento de los logs. Sube los logs procesados a Azure Storage y Azure SQL DB. Genera tablas que facilitan la creación de visualizaciones en Power BI.

En la Figura 4.8 se muestra como interactúa con los demás scripts.

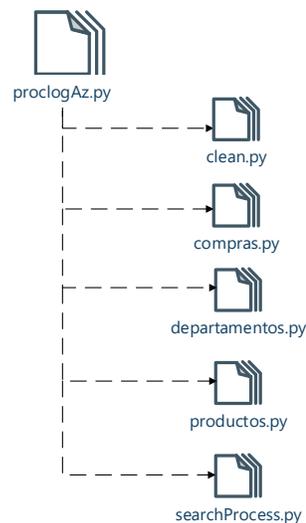


Figura 4.8. proclgAz.py orquesta el procesamiento de logs.

¹⁶ Expect es un lenguaje de scripting de automatización que funciona de la misma manera que los humanos cuando interactúan con un sistema.

- *clean.py*

Da estructura y formato a los logs sin procesar, quita caracteres innecesarios y deja información relevante en la base de datos y en formato CSV (Comma-Separated Values). Es llamado por `proclogAz.py`.

Aunque los logs tenían una estructura predecible, este script se encarga de quitar ciertos elementos para poder analizar la información. Por ejemplo, para el campo `%h` (host remoto) había casos en los que se seguía el patrón `%h%h` y `%h%h%h%h`, esto debido a que se sigue uno o más proxys al acceder al portal de Contoso. Además eliminé líneas incompletas o que no correspondieran al formato NCSA extended/combined.

- *Compras.py*

Hace una subconsulta a la base de datos e inserta información relevante sobre las compras realizadas por los clientes (Tabla Compras). Es llamado por `proclogAz.py`.

- *departamentos.py*

Corrige codificación de las cadenas que se refieren a los Departamentos de la tienda, de manera que se puedan obtener los departamentos de contoso tal y como están expuestos en la tienda virtual.

Por ejemplo, hace que "electronica", "electr%Anica", "Electronica", entre otras variantes, aparezcan como "Electrónica" en la base de datos.

Hace una subconsulta a la base de inserta toda la información relevante sobre los departamentos visitados por cada cliente (Tabla Departamentos). Es llamado por `proclogAz.py`.

- *productos.py*

Corrige codificación de las cadenas que se refieren a los productos, esto para obtener una oración legible por un ser humano. Hace una subconsulta a la base de datos e inserta toda la información relevante sobre los productos consultados en el catálogo de la tienda (Tabla Productos). Es llamado por `proclogAz.py`.

- *searchProcess.py*

Corrige codificación de las cadenas que se refieren a los productos, esto para obtener una oración legible por un ser humano, elimina metadatos de la búsqueda. Inserta en la base de datos toda la información relevante sobre las búsquedas realizadas por los clientes. Es llamado por `proclogAz.py`.

4.4 SQL Database

En esta sección se describe las características de la base de datos que utilicé.

SQL Database permite tener bases de datos completamente aisladas optimizadas para cargas de trabajo cuando las demandas de rendimiento son un tanto predecibles. Permite escalar o reducir verticalmente una base de datos única entre los niveles de servicio Básico, Estándar y Premium para obtener el rendimiento y las características que la aplicación necesita.

Dado que el sistema pretende ser lo más autónomo posible, se eligió SQL Database por sus características de Plataforma como Servicio (PaaS), de tal modo que requiere administración mínima ya que no requiere de una máquina virtual adicional con SQL Server.

SQL Database se clasifica por niveles. Cada nivel se distingue principalmente por el rendimiento, que se mide en unidades de transacción de base de datos (Database Transaction Unit, DTU).

Las bases de datos standard son la mejor opción para las aplicaciones en la nube con pocos o medios requisitos de rendimiento de E/S. En muchos casos, los niveles de rendimiento estándar S2 y S3 son un buen punto de partida, en la siguiente tabla se listan las características de la base de datos elegida:

Nivel de servicio Estándar

Nivel de servicio	S3
DTU máx.	100
Tamaño máximo de base de datos	250 GB
Máximo de inicios de sesión simultáneos	200
N.º máximo de sesiones simultáneas	2400
Point in time restore	Sí
Geo-Replicación	Sí

Tabla 4.7. Características de SQL Database

4.4.2 Diagrama Entidad Relación

A continuación, se presenta el diagrama Entidad Relación de la base de datos generada.

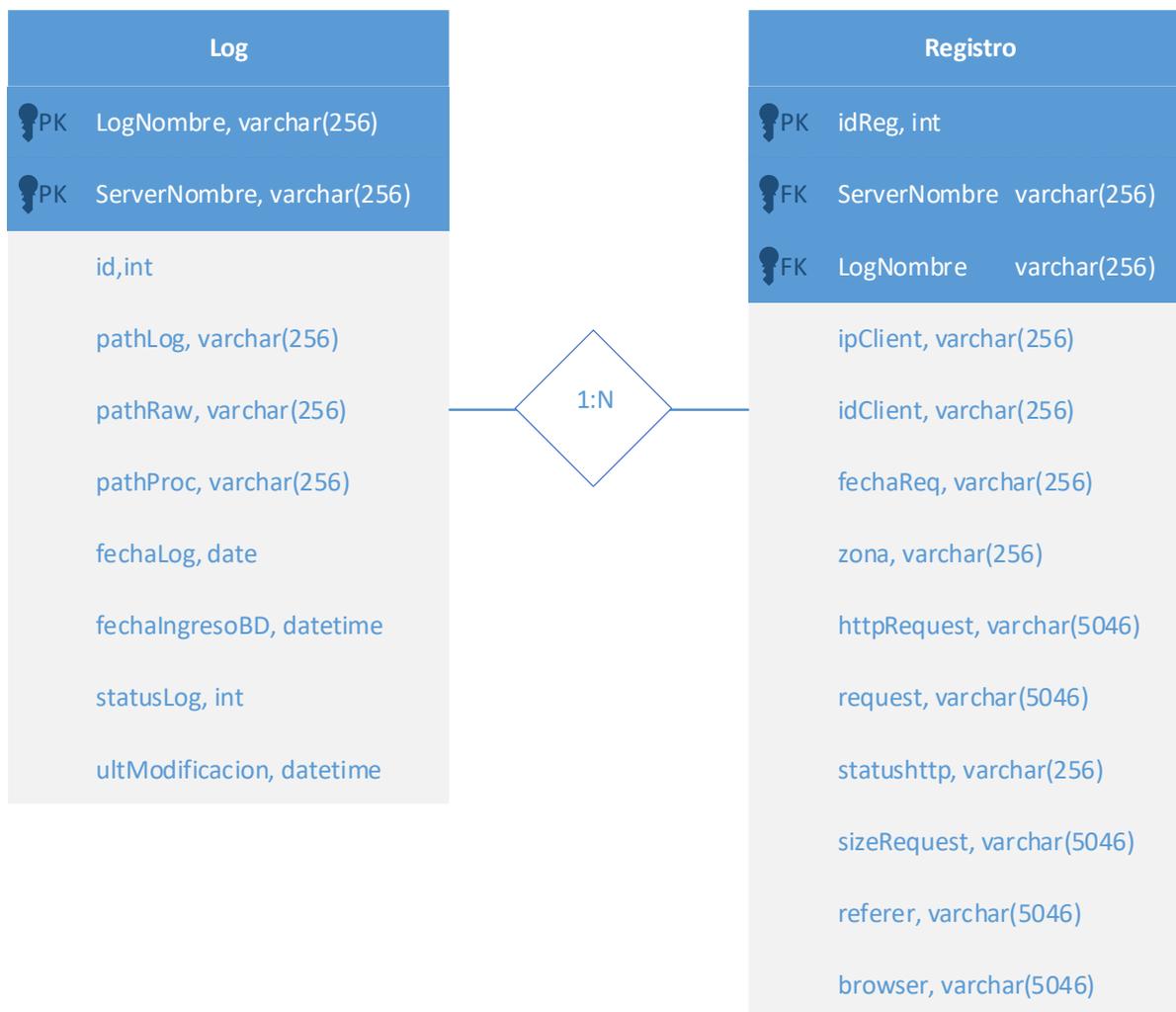


Figura 4.9. Modelo Entidad Relación de la base de datos.

Adicionalmente, para aprovechar el motor de base de datos y mejorar el performance de PowerBI, se crearon cuatro vistas, cuyo Diagrama de Entidad Relación se presenta en la Figura 4.10.

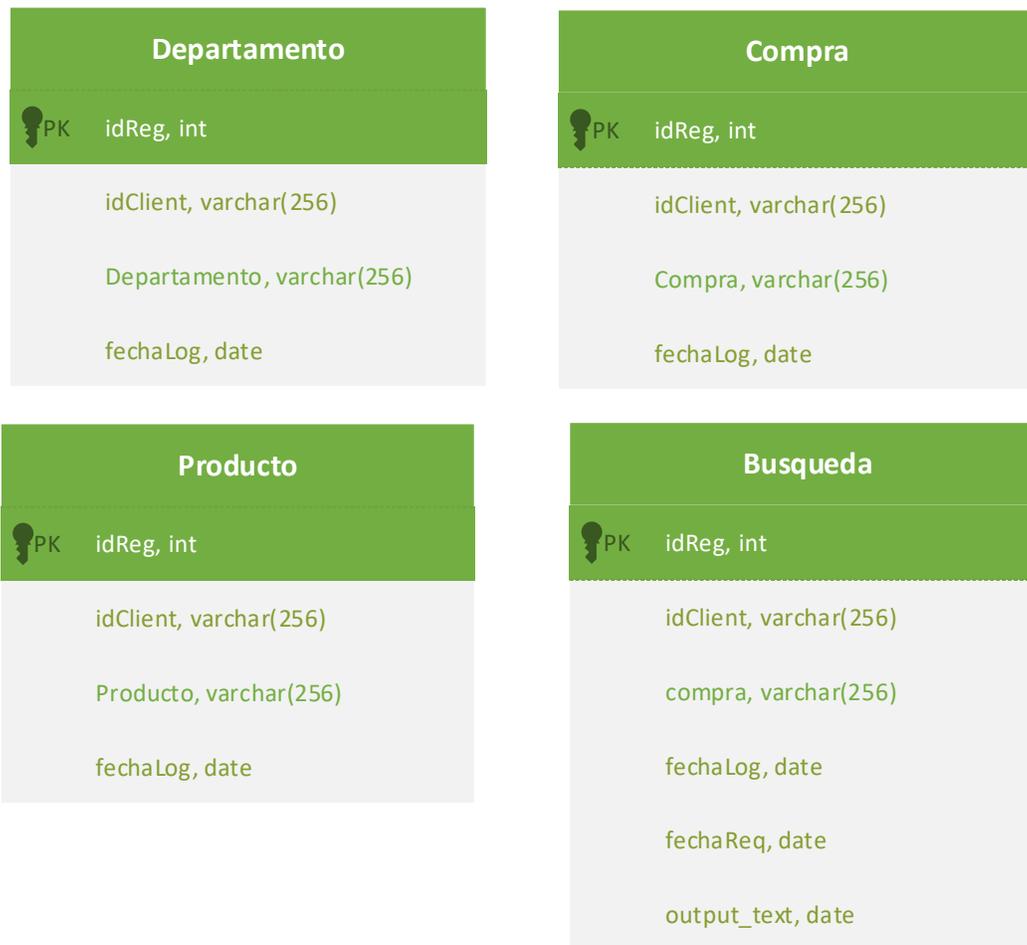


Figura 4.10. Modelo Entidad Relación de las vistas generadas.

4.4.2 Descripción de Tablas

a. Log

Contiene información sobre los logs que han sido cargados a Azure, así como los que han sido procesados.

Campo	Descripción
Id	Identificador único para cada fila en la tabla en la base de datos.
logNombre	Nombre del archivo del que procede el registro.
serverNombre	Servidor del que procede el registro.
pathLog	Directorio local del log en la VM.
pathRaw	URI en la cuenta de almacenamiento para el log en crudo.
pathProc	URI en la cuenta de almacenamiento para el log procesado.
fechaLog	Fecha del log.
fechaIngreso	Fecha en que se registró el log en la Base de Datos.
statusLog	Indica si el log ya fue procesado. 0: Log sin procesar 1: Log procesado - 1: Error, el log no pudo ser procesado.
ultModificacion	Última modificación en el registro actual.

Tabla 4.3. Campos de la Tabla Lista

b. Registro

Contiene cada registro procesado.

Campo	Descripción
IdReg	Identificador único para cada fila en la tabla en la base de datos.
logNombre	Nombre del archivo del que procede el registro.
serverNombre	Servidor del que procede el registro.
fechaLog	Fecha del registro.
ipClient	IP de host remoto, en este caso, del cliente que realiza operaciones en la tienda de Contoso.
idClient	Identidad del cliente, ID del cliente que realiza operaciones en la tienda de Contoso.
FechaReq	Fecha en la que el cliente realiza la petición al servidor.
Zona	Zona horaria, tiempo universal coordinado (UTC).
httpReq	Petición admitida en HTTP / 1.1, la cual se realiza en el recurso identificado por el URI.
Request	El URI de solicitud es un identificador de recurso uniforme e identifica el recurso sobre el cual aplicar la solicitud.
Statushttp	El código de estado es un entero de 3 dígitos en el que el primer dígito del código de estado define la clase de respuesta y los dos últimos dígitos no tienen ninguna función de categorización. El estado permite conocer si la petición se recibió correctamente o con errores.
SizeReq	Es el tamaño de la respuesta al cliente (en bytes).
Referer	Página que enlazó a la URL actual.
Browser	Información relevante del navegador de cada cliente.

Tabla 4.4. Campos de la Tabla Logs

c. Departamentos

Muestra los departamentos visitados por cada cliente.

Campo	Descripción
idReg	Identificador único para cada fila en la tabla en la base de datos.
idClient	Identidad del cliente, ID del cliente que realiza operaciones en la tienda de Contoso.
Departamento	Nombre del departamento visitado
fechaLog	Fecha del registro.

Tabla 4.5. Campos de la Tabla Departamentos

	idReg	idClient	departamento	fechaLog
1	49553441	76277F320256C4BC5627DEC86E0C6C	Hogar	2016-12-25
2	49553443	5DD4ADE361E9B2C86379EFE2C5805E	Videojuegos	2016-12-25
3	49553444	76277F320256C4BC5627DEC86E0C6C	Hogar	2016-12-25
4	49553445	76277F320256C4BC5627DEC86E0C6C	Hogar	2016-12-25
5	49553450	76277F320256C4BC5627DEC86E0C6C	Hogar	2016-12-25
6	49553458	8EEBF023088F0C868DF2AEB12ECA	Computo	2016-12-25
7	49553460	4C204B528D4F68A2A87C5BD915229E	Juguetes	2016-12-25
8	49553461	4C204B528D4F68A2A87C5BD915229E	Juguetes	2016-12-25
9	49553462	646F566059FF546D9D4B6D915E1C67E	Electrodomesticos	2016-12-25

Figura 4.11 Vista de la Tabla Departamentos

d. Compras

Muestra los request relacionados con una compra, así como el ID del cliente que realizó la petición.

Campo	Descripción
idReg	Identificador único para cada fila en la tabla en la base de datos.
idClient	Identidad del cliente, ID del cliente que realiza operaciones en la tienda de Contoso.
Compra	URI relacionada con el proceso de compra, esto puede ser desde que el cliente inicia el proceso de compra, introduce sus datos personales, forma de pago y fin de la compra.
fechaLog	Fecha del registro.

Tabla 4.6. Campos de la Tabla Compras

	idReg	idClient	compra	fechaLog
1	6158012	15CDC0356A23DD96F1B8A40028AC5F4...	/es/cart/checkout	2016-10-01
2	6158013	15CDC0356A23DD96F1B8A40028AC5F4...	/es/login/checkout	2016-10-01
3	6158025	15CDC0356A23DD96F1B8A40028AC5F4...	/es/login/checkout/guest	2016-10-01
4	6158079	91E545415BA904AE0F4A47F6DC3E5315....	/es/cart/checkout	2016-10-01
5	6158080	91E545415BA904AE0F4A47F6DC3E5315....	/es/login/checkout	2016-10-01
6	6158113	91E545415BA904AE0F4A47F6DC3E5315....	/es/login/checkout/guest	2016-10-01
7	6158114	91E545415BA904AE0F4A47F6DC3E5315....	/es/checkout/multi	2016-10-01

Figura 4.12. Vista de la Tabla Compras

e. Productos

Muestra los productos del catálogo de Contoso consultados por los clientes.

Campo	Descripción
idReg	Identificador único para cada fila en la tabla en la base de datos.
idClient	Identidad del cliente, ID del cliente que realiza operaciones en la tienda de Contoso.
Product	Producto consultado en catálogo.
fechaLog	Fecha del registro.

Tabla 4.7. Campos de la Tabla Productos

	idReg	idClient	producto	fechaLog
36	6157820	8A368878B80D91849F55579A7A1DBC20....	Drone DJI Phantom 2 con Zenmuse H4 3D	2016-10-01
37	6157824	9C2A94356F765B736A2D86FBECADA3C...	Centro de Entretenimiento para Pantalla de hasta 55" ...	2016-10-01
38	6157826	EF86E4F3C24DAC886FC3359D54C64415...	Microondas Grill Electrolux EMTK16G4MLM	2016-10-01
39	6157828	3A2634764719EC8F5AC52EA7BE6D7C27...	Lanzador Modulus ECS 10 Blaster Nerf	2016-10-01
40	6157835	EA9D141B82F57F5751A975F9EF7ADED...	Generador de Vapor 1 L Singer Ultimate Finish SHG2...	2016-10-01
41	6157851	26C5C97439D46A697F97EBF5E21EB23E...	Cortina de Bao Poly Boom Dib Multicolor	2016-10-01
42	6157852	0E9A216B1BED9E4ECAC42B48043C3C7...	Gran Barco Colosal de Jake Fisher Price Mattel	2016-10-01
43	6157861	ECE9300E20C107483156F50B2C9C3DB0...	Laptop Asus VivoBook E402SA de 14" Intel Dual Cor...	2016-10-01
44	6157868	07C47220129EAEE4EC5EF9F73A3BF552...	Silln Puff Zen Azul Marino Puff Mx	2016-10-01

Figura 4.13. Vista de la Tabla Productos

f. Búsquedas

Muestra las búsquedas tal y como fueron realizadas por los usuarios.

Campo	Descripción
idReg	Identificador único para cada fila en la tabla en la base de datos.
idClient	Identidad del cliente, ID del cliente que realiza operaciones en la tienda de Contoso.
Compra	URI relacionada con el proceso de compra, esto puede ser desde que el cliente inicia el proceso de compra, introduce sus datos personales, forma de pago y fin de la compra.
FechaLog	Fecha del registro.
FechaReq	Fecha en la que el cliente realiza la petición al servidor.
output_text	Búsqueda realizada por el usuario.

Tabla 4.8. Campos de la Tabla Búsquedas

	idReg	fechaLog	ipClient	idClient	fechaReq	output_text
1	11159230	2016-10-13	177.227.121.22	F956A0E6555CC8B92B977C0069985C7C....	13/Oct/2016:16:53:30	facturacion
2	11159236	2016-10-13	200.79.182.29	327856FF0FE891C88CE2D8F03DBF3872....	13/Oct/2016:16:53:31	Samsung Galaxy S7 Edge 32GB 5.5"
3	11159264	2016-10-13	189.171.43.180	E4685948878F5931E76916A2D70B8D49....	13/Oct/2016:16:53:42	Elptica
4	11159326	2016-10-13	189.171.43.180	E4685948878F5931E76916A2D70B8D49....	13/Oct/2016:16:54:09	Bicicleta elptica
5	11159343	2016-10-13	189.163.145.66	5E872AE459AF7F71266EA7B9110C1B05....	13/Oct/2016:16:54:19	ROPA BEBES
6	11159385	2016-10-13	148.244.68.211	416F3EE5EC61802B1FB68133439A8806.s...	13/Oct/2016:16:54:40	UN75J6300
7	11159439	2016-10-13	189.217.81.26	107C4B2E9E76BE9A9F6ACEE154F13E6D...	13/Oct/2016:16:55:02	gears of war
8	11159446	2016-10-13	187.245.115.71	72E7AD97EEE3BECAACD6F49DC34B3A7...	13/Oct/2016:16:55:06	monster high
9	11159479	2016-10-13	207.248.245.142	10B8A75127EBE2D3E880713D2E898977....	13/Oct/2016:16:55:19	porta celulares
10	11159490	2016-10-13	189.163.145.66	5E872AE459AF7F71266EA7B9110C1B05....	13/Oct/2016:16:55:24	JUGUTES
11	11159519	2016-10-13	187.188.8.203	888081F82B4E3786549FCA7468AC2BB1....	13/Oct/2016:16:55:39	homos de microhondas

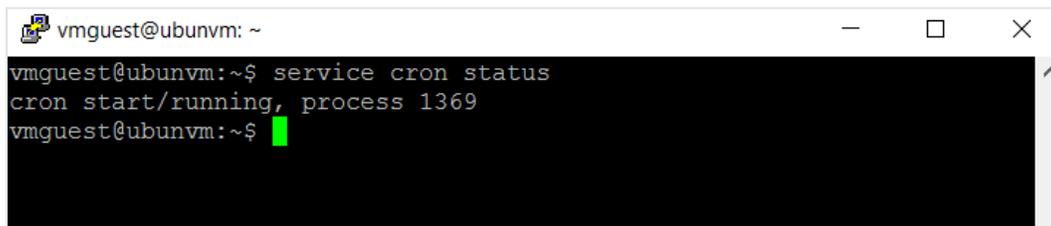
Figura 4.14. Vista de la Tabla Búsquedas

4.5 Cron de Linux

El Cron es un demonio del sistema operativo Linux el cual permite ejecutar automáticamente comandos o scripts a una hora y/o fecha específica.

Para verificar que el Cron del sistema está habilitado en el sistema se utilizó el siguiente comando:

```
#Este comando puede cambiar según la distribución de Linux  
root@vmAzurePoC# service cron status
```



```
vmguest@ubunvm: ~  
vmguest@ubunvm:~$ service cron status  
cron start/running, process 1369  
vmguest@ubunvm:~$
```

Figura 4.15. Verificando status del cron Linux.

a. Modificar permisos del script

Se deben dar los permisos correspondientes a todos los scripts para que puedan ser ejecutados por el cron:

```
root@vmAzurePoC# chmod +x rawlogAz.py  
root@vmAzurePoC# ls -l rawlogAz.py  
-rwx----- 1 root root 0 Oct 20 09:30 log.py
```

b. Modificar archivo crontab

Con permisos de super usuario, introducir el siguiente comando:

```
root@vmAzurePoC# crontab -e
```

Se modificó el archivo crontab para que los scripts se ejecuten con una hora de diferencia cada día de la semana.

```
GNU nano 2.2.6                               File: /tmp/crontab.q00Vn5/crontab
# Edit this file to introduce tasks to be run by cron.
#
# Each task to run has to be defined through a single line
# indicating with different fields when the task will be run
# and what command to run for the task
#
# To define the time you can provide concrete values for
# minute (m), hour (h), day of month (dom), month (mon),
# and day of week (dow) or use '*' in these fields (for 'any').#
# Notice that tasks will be started based on the cron's system
# daemon's notion of time and timezones.
#
# Output of the crontab jobs (including errors) is sent through
# email to the user the crontab file belongs to (unless redirected).
#
# For example, you can run a backup of all your user accounts
# at 5 a.m every week with:
# 0 5 * * 1 tar -zcf /var/backups/home.tgz /home/
#
# For more information see the manual pages of crontab(5) and cron(8)
#
# m h dom mon dow   command
00 06 * * * cd /home/vmquest/contoso_log_analytics && /usr/bin/python /home/vmquest/contoso_log_analytics/rawlogAz.py
00 08 * * * cd /home/vmquest/contoso_log_analytics && /usr/bin/python /home/vmquest/contoso_log_analytics/proclgAz.py
#
```

Figura 4.16. Archivo crontab

4.6 Análisis con Power BI

Power BI es una colección de servicios de software, aplicaciones y conectores que funcionan conjuntamente para convertir orígenes de datos sin relación entre sí en información coherente, interactiva y atractiva visualmente. Tanto si se trata de una sencilla hoja de cálculo de Excel como de una colección de almacenes de datos híbridos locales o basados en la nube, Power BI permite conectar fácilmente los orígenes de datos, visualizar (o descubrir) lo más importante y compartirlo con quien quiera.

Power BI consta de una aplicación de escritorio de Windows denominada Power BI Desktop, un servicio SaaS (Software as a Service) en línea denominado servicio Power BI, y aplicaciones móviles de Power BI disponibles para teléfonos y tabletas Windows, así como para dispositivos iOS y Android.

Un flujo de trabajo habitual de Power BI comienza en Power BI Desktop, donde se crea un informe. Luego, ese informe se publica en el servicio Power BI y después se comparte para que los usuarios de las aplicaciones de Power BI Mobile puedan usar la información.

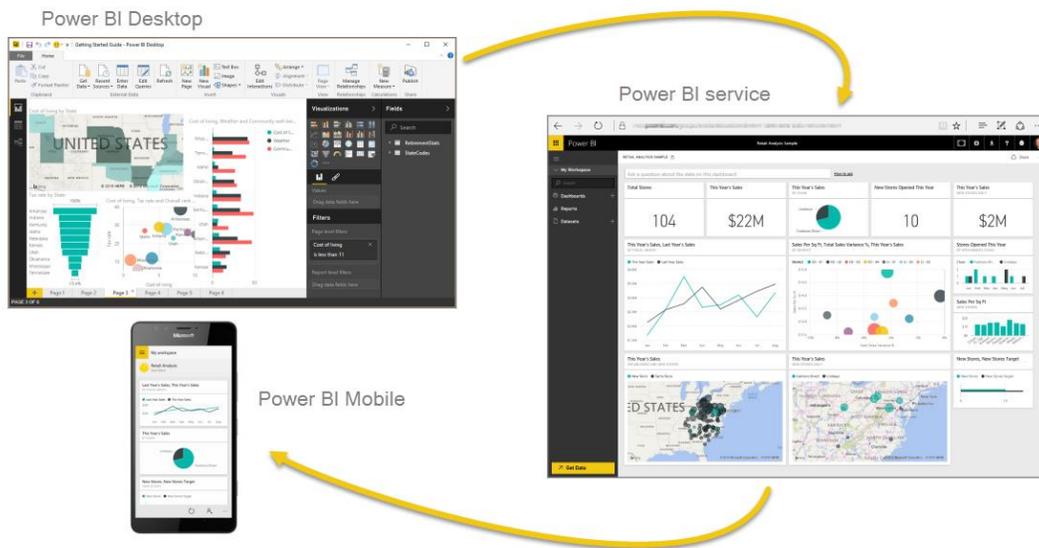


Figura 4.17. Ecosistema de Power BI

Como parte del proceso de venta, se impartió al cliente un taller de Power BI en las instalaciones de Microsoft México, esto con el objetivo de mostrarle el uso de la herramienta y capacitar al cliente.

4.6.1 Fuente de datos

Una vez que los logs fueron procesados y depositados tanto en una cuenta de almacenamiento como en una base de datos, se hicieron visualizaciones con Power BI para comprender el comportamiento de los usuarios que utilizan el e-commerce de Contoso Ltd.

Power BI permite a los usuarios analizar información desde diversas fuentes por ejemplo SQL Server, Sharepoint, Salesforce y SAP Hana.

Para optimizar el funcionamiento de Power BI, se eligió como fuente de datos la SQL Database que vive en Azure, lo que nos permite explotar el poder del motor de bases de datos de SQL.

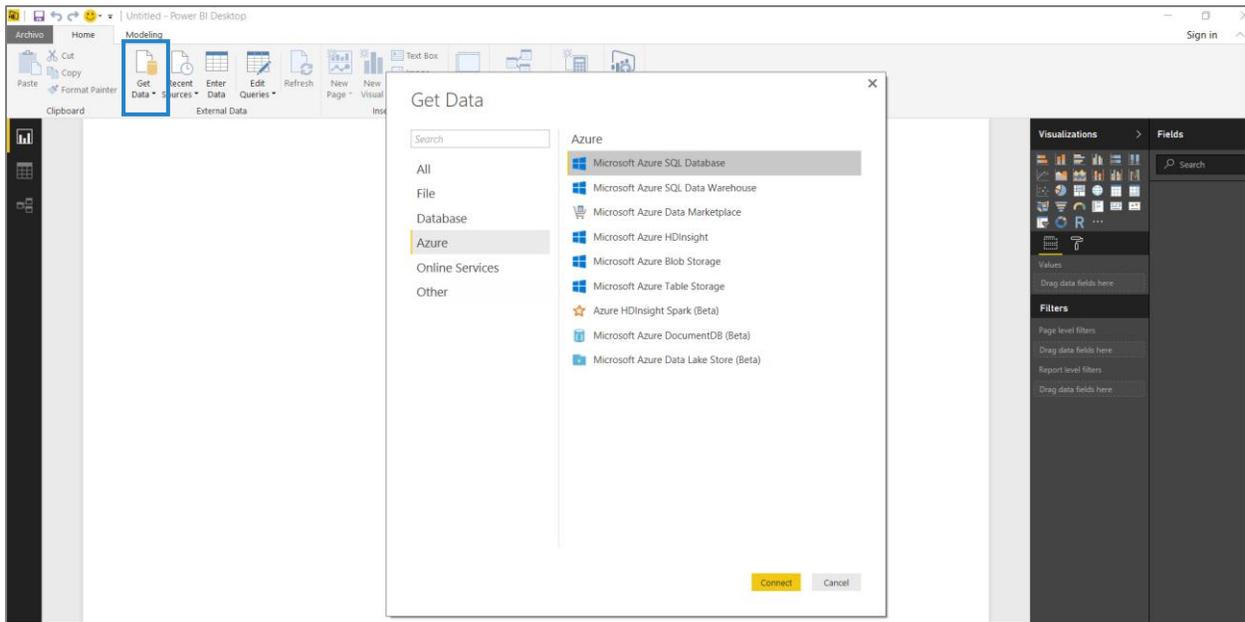


Figura 4.18. Power BI desktop toma SQL Database como fuente de datos.

Luego de introducir las credenciales de SQL Database, Power BI guía al usuario paso a paso para seleccionar las tablas que desea analizar.

Una vez cargadas las tablas de nuestro interés, se utilizó Power Query (conocido informalmente como M Language) para hacer consultas a la base de datos para trabajar las visualizaciones.

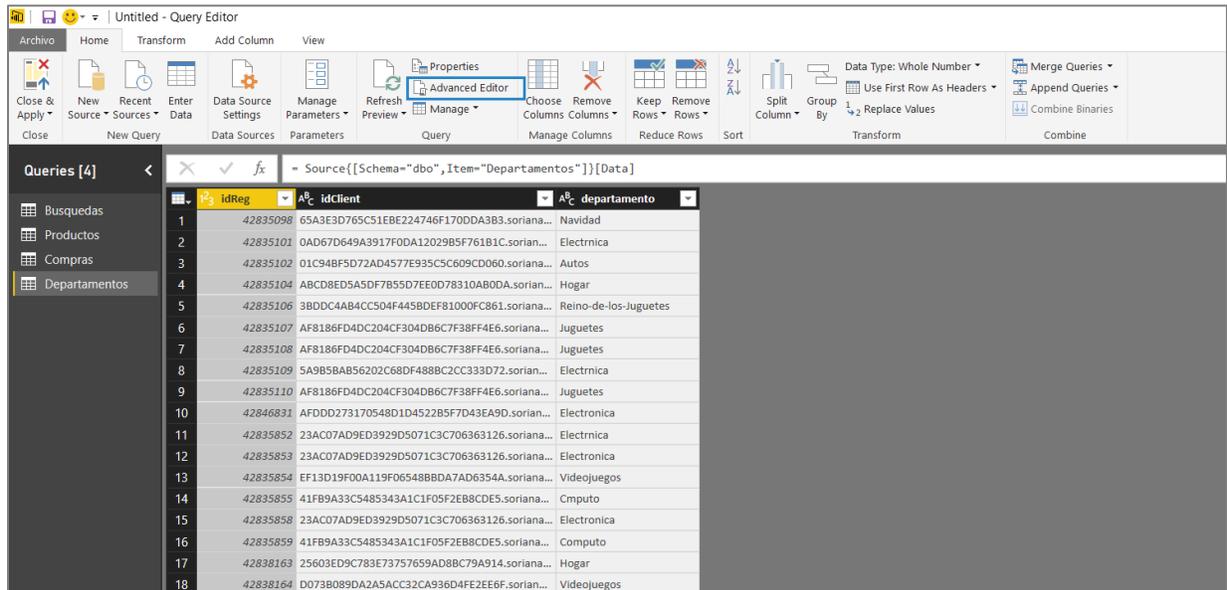


Figura 4.19. Tablas cargadas en Power BI Desktop.

Ejemplos de query:

```
let
    Source = Sql.Database("servertestoct.database.windows.net", "sqlTest",
[Query="SELECT DISTINCT idClient FROM Logs WHERE
DATEDIFF(day,fechaLog,getdate()) between 0 and 15"])
in
    Source
```

```
let
    Source = Sql.Database("servertestoct.database.windows.net", "sqlTest",
[Query="SELECT a.fechaLog, a.idReg, a.idClient, b.palabra_correcta FROM
Departamentos a JOIN Diccionario b ON a.departamento=b.palabra_erronea
WHERE DATEDIFF(day,a.fechaLog,getdate()) between 0 and 15
UNION ALL
SELECT a.fechaLog, a.idReg, a.idClient, a.departamento
FROM Departamentos a LEFT JOIN Diccionario b ON
a.departamento=b.palabra_erronea
WHERE b.palabra_erronea IS NULL AND DATEDIFF(day,a.fechaLog,getdate())
between 0 and 15"])
in
    Source
```



Figura 4.20 La opción Advanced editor permite al usuario realizar queries con Power Query.

V. Resultados

5.1 Visualizaciones

Como parte de los requerimientos, Contoso solicitó crear un Panel que aportara información de valor para el negocio. Utilicé Power BI para satisfacer este requerimiento, ya que posee herramientas para crear visualizaciones dinámicas y estéticas.

En este capítulo se mostrarán algunas de las visualizaciones entregadas a Contoso Ltd. a partir del análisis de sus datos. Además, se detallará la propuesta comercial que derivó de la presente prueba de concepto.



Figura 5.1 Panel publicado en Power BI Service.

b. Productos consultados en catálogo

Muestra en orden descendente los productos existentes en el catálogo de productos más consultados por los usuarios.

Este tipo de visualización permite entender de forma sencilla cuáles son los productos más populares en el catálogo, y que por lo tanto hay que procurar tener en inventario. También podría ser una forma de monitorear la respuesta del público a una promoción.

Por ejemplo, en la Figura 5.4, se observa que el producto más consultado fue una Cámara Reflex Fujifilm.

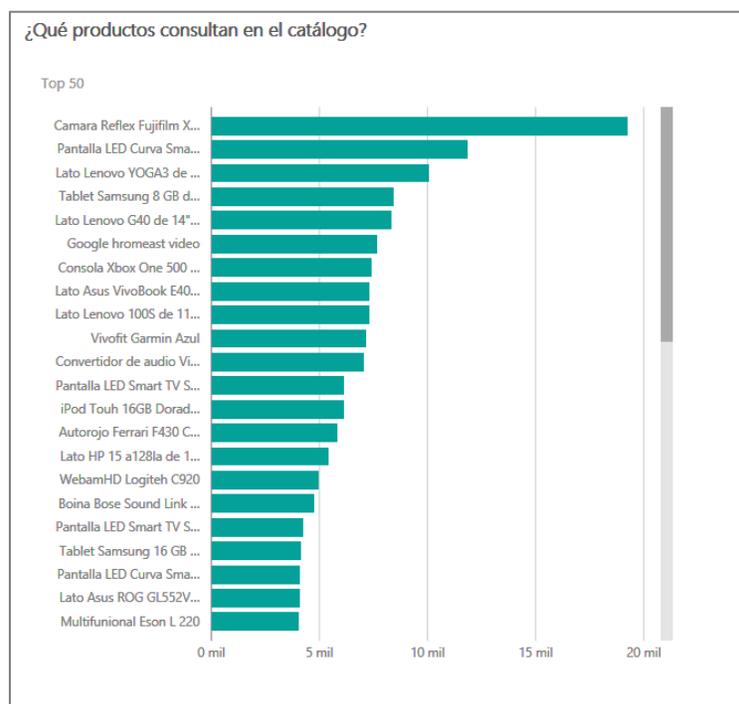


Figura 5.4 Visualización de productos más consultados en Catálogo.

c. Visitas por departamento

Muestra la cantidad de usuarios que visitaron cada departamento en orden descendente.

Este tipo de visualización nos permite saber cuáles son los departamentos más fuertes y que por lo tanto hay que trabajar en mejores estrategias de marketing, promociones o calidad.

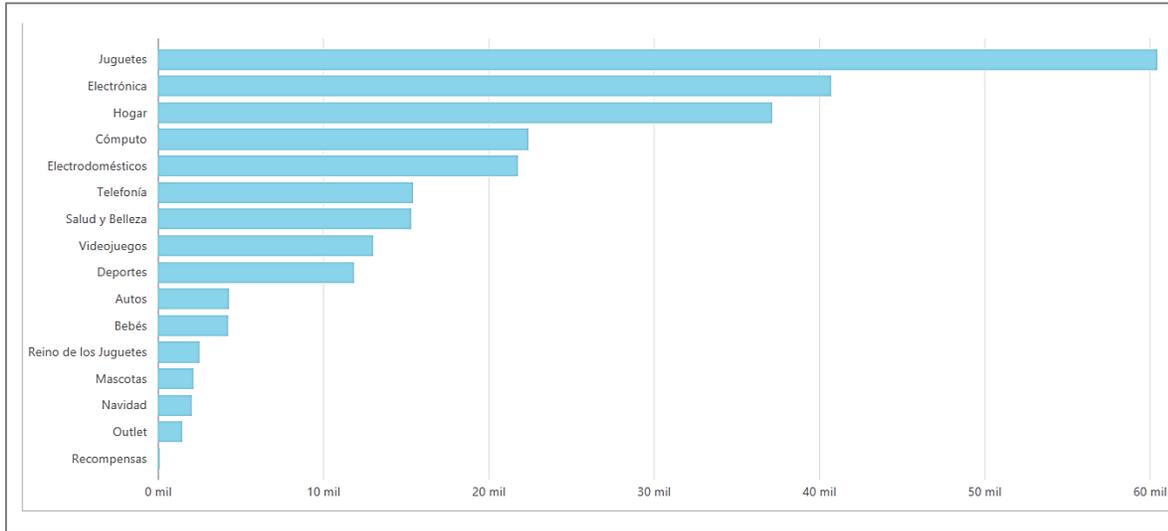


Figura 5.5 Visualización de visitas por departamento

d. Compras

Muestra cuantos clientes entraron al sitio web <https://www.contoso.com/> (basado en el idClient), la cantidad de usuarios que iniciaron el proceso de compra y cuantos compraron. Además, interactúa con la visualización c (Figura 5.5), para mostrar las compras por departamento.



Figura 5.6. Visualización de Compras

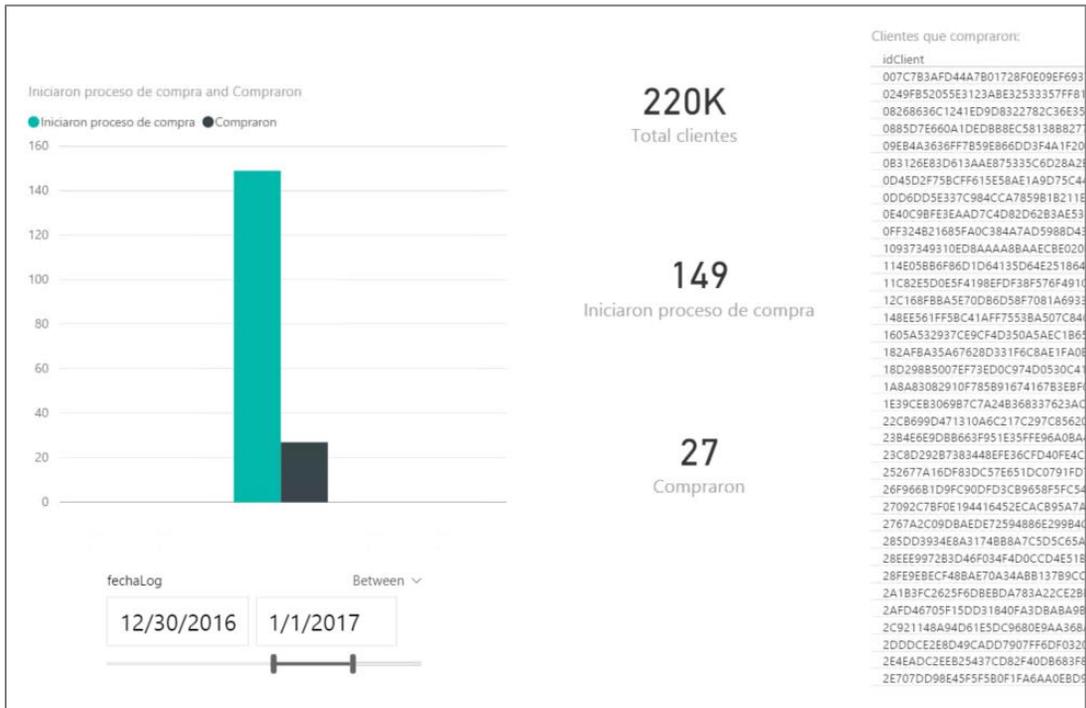


Figura 5.7 Informe de compras por rango de fecha.

Power BI tiene la capacidad crear relaciones automáticamente o de forma manual para ofrecer visualizaciones que interactúan entre sí. Por ejemplo, en la figura 5.9, se puede ver que, al hacer clic en algún departamento de la visualización de "Departamentos", la visualización de "Compras" refleja la cantidad de usuarios que compró en dicho departamento.

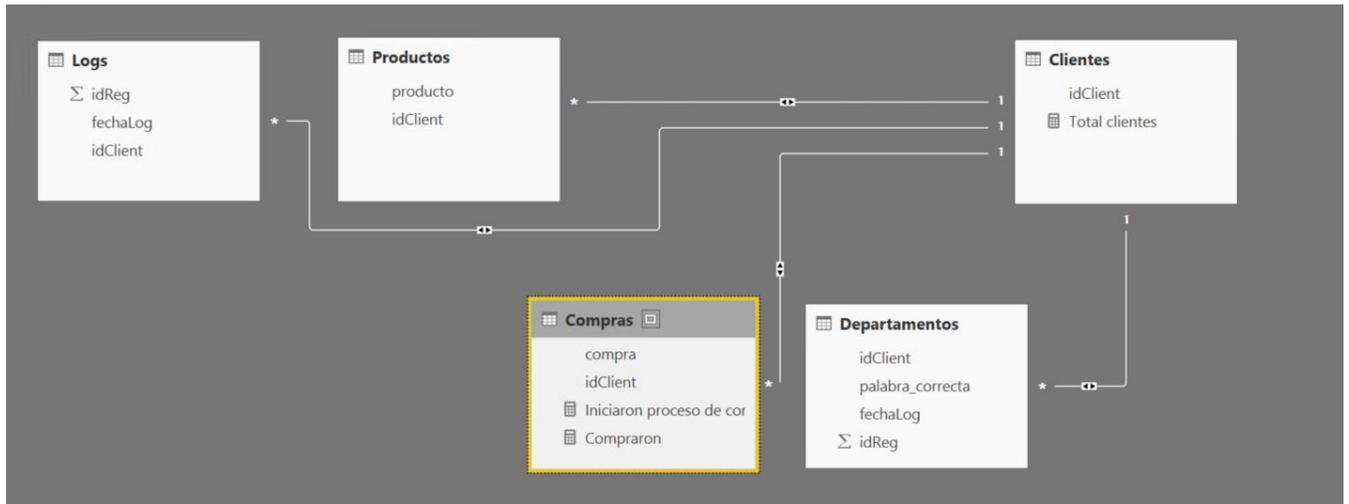


Figura 5.8. Relaciones creadas en Power BI Desktop

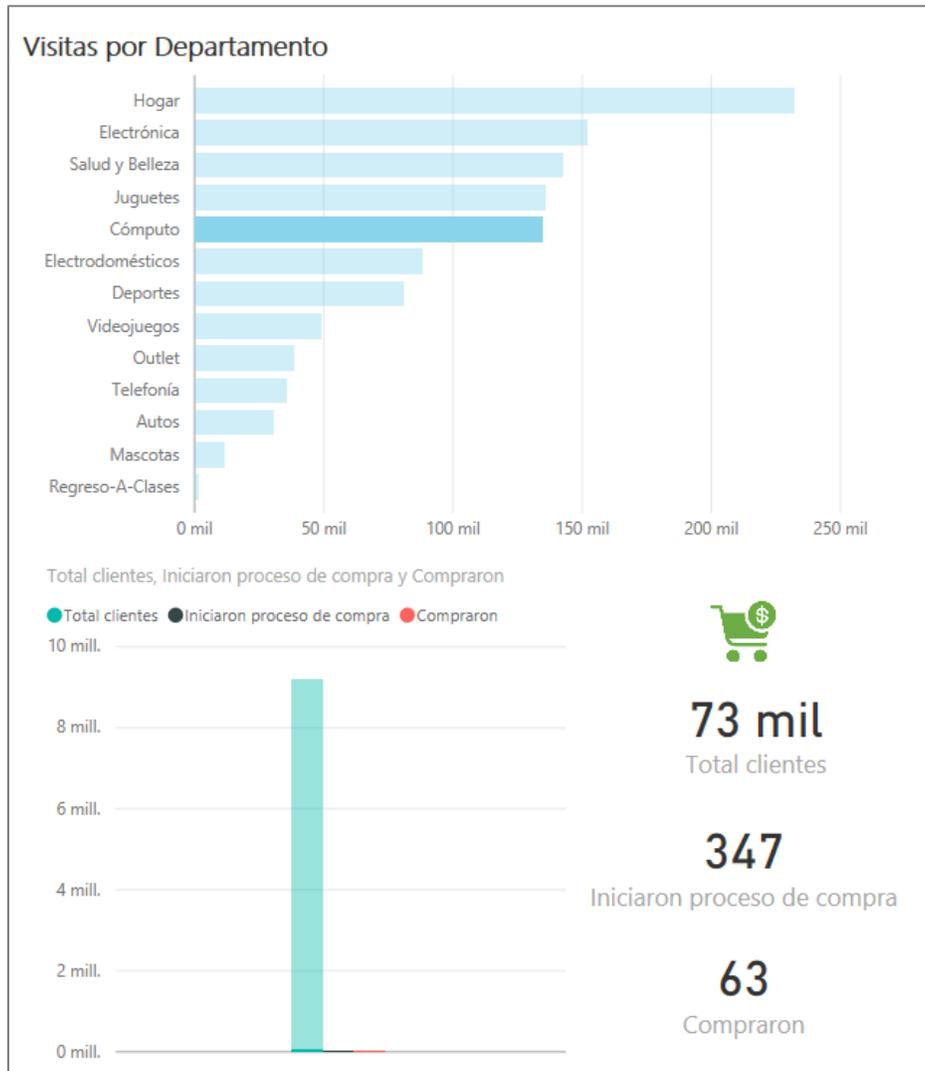


Figura 5.9 Interacción entre visualizaciones.

Esta visualización nos entrega información muy importante sobre el comportamiento del cliente, pues nos permite saber que, aunque miles de personas entran al portal, solo unas cuentas efectúan compras en línea. Este es un comportamiento esperado, ya que las personas suelen comprar en la tienda física y usan el portal en línea para consultar los productos más que para efectuar compras (este dato fue proporcionado por encuestas de Contoso).

Por otra parte, basándonos en los datos de la figura 5.8, vemos que de todos los usuarios que comienzan el proceso de compra, es decir, que ponen productos en el carrito, ingresan su correo y forma de pago, casi la quinta parte no concluye el proceso de venta. Esto podría ser consecuencia de que el portal no es muy intuitivo y la gente no entiende

cómo comprar o bien que hay una parte del proceso de venta que no está siendo funcional e impide a los usuarios avanzar y comprar.

e. Seguimiento por cliente

Muestra los departamentos que visitados por cada cliente y el número de operaciones realizadas en ese departamento.

Este tipo de visualización podría servirnos para hacer el rastreo del cliente por la tienda virtual. De manera similar a las tiendas físicas donde a través del mapa de calor se observan las zonas que son de mayor afluencia. Si entendemos cuales son los departamentos de interés de cada cliente, podemos aprovechar esta información para hacer ofertas personalizadas, tal como sugieren los modelos modernos de comercio.

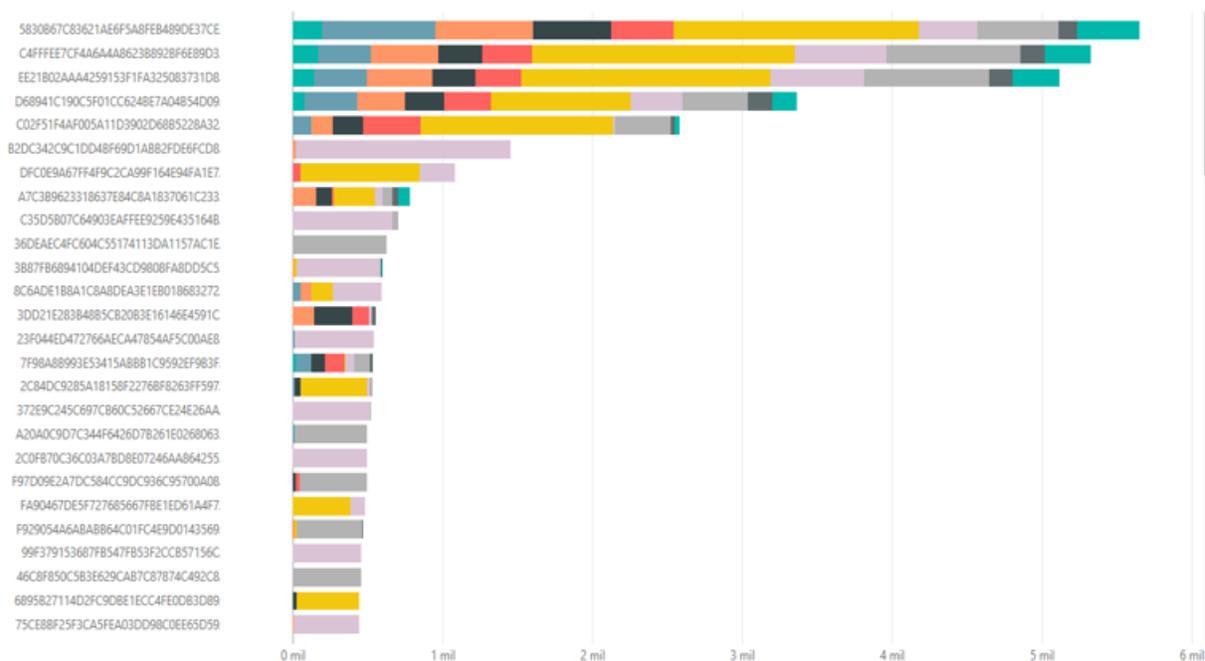


Figura 5.10 Visualización de Seguimiento por cliente

f. Top Referer

Muestra desde que sitio están siendo redireccionados los clientes al portal de Contoso, está gráfica se basa en el campo Referer de la base de datos.

Este tipo de visualización nos da una idea de los medios que podemos explotar para llegar a más clientes, por ejemplo, en la figura 5.11 se observa que la mayoría de clientes accedió por el buscador de Google y Facebook. Esto nos puede sugerir que es conveniente usar los anuncios de ambas plataformas para atraer más clientes.

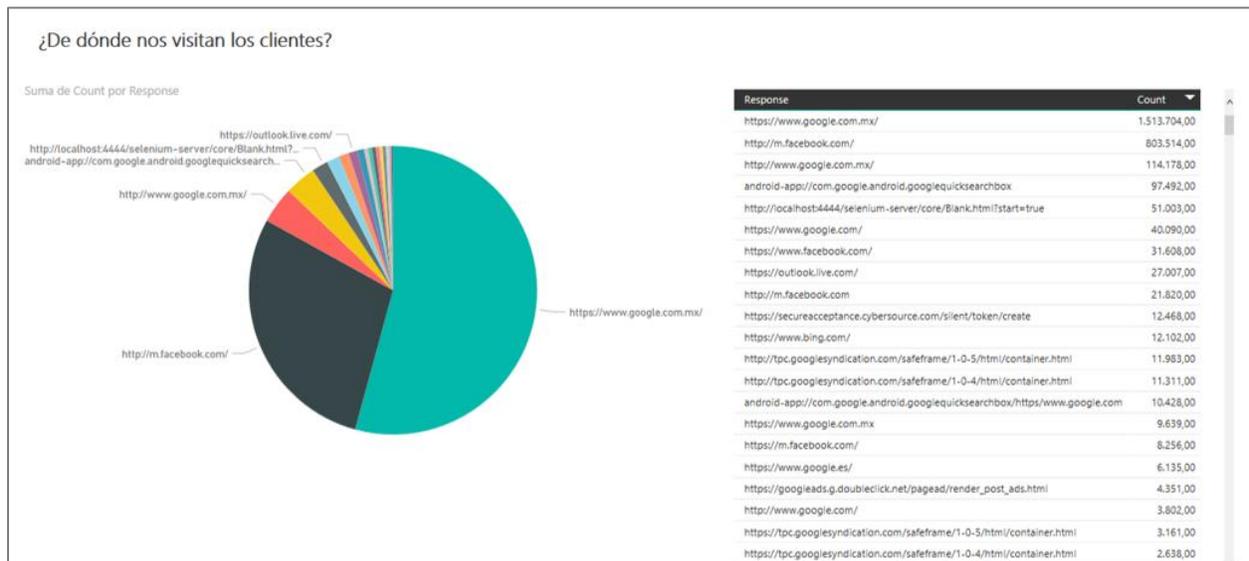


Figura 5.11 Visualización top Referer

5.2 Validación con Matriz de pruebas

Al concluir la prueba de concepto, se validó con el cliente la matriz de pruebas que se generó durante la etapa de requerimientos.

#	Criterio de éxito	Cumple	No cumple
1	Está montado en la nube.	X	
2	Funciona sin intervención humana.	X	
3	No perjudica el performance de la fuente de datos.	X	
4	Funciona fuera de las horas con mayor tráfico.	X	
5	Accesos al sistema restringidos por contraseña.	X	
6	El sistema es capaz de manejar grandes volúmenes de datos.	X	
7	Generación de reportes visuales.	X	
8	Los informes pueden consultarse en Internet bajo acceso restringido.	X	
9	Almacena datos en una base relacional.	X	
10	Acceso a máquina virtual vía SSH.	X	
11	Conexión VPN con Forticlient.	X	
12	Los logs crudos deben estar en un repositorio en la nube organizados por servidor y fecha.	X	
13	Los logs limpios deben estar en un repositorio en la nube organizados por servidor y fecha.	X	
14	Es integrable con nuevos sistemas y áreas del negocio.	*1	
15	Genera informes automáticamente cada día.	X	
16	Arroja información relevante para el negocio.	X	
17	La información desplegada refleja lo que está ocurriendo en tiempo real.		X ⁴
18	Disponibilidad del sistema.	*2	
19	Interoperabilidad entre tecnologías Open Source y Windows.	X	
20	Tiempo de implementación no mayor a tres meses.	X	
21	El sistema debe ser fácilmente escalable.	*3	

¹ Azure ofrece diferentes servicios para integrar los ambientes on-premise con los ambientes de nube, sin embargo, se debe realizar una evaluación técnica exhaustiva para asegurar dicha integración.

^{2,3} Azure proporciona las herramientas necesarias para asegurar la disponibilidad y escalabilidad del sistema, queda a decisión del cliente consumir o no estos servicios.

⁴ Debido a la manera en que opera Hybris, mostrar la información en tiempo real implica un nivel de complejidad que queda fuera de los servicios que el POD puede ofrecer. Sin embargo, es posible cubrir este requerimiento con el apoyo de los socios o de Microsoft Consulting Services.

5.3 Nuevas Iniciativas

Como consecuencia directa de la presente prueba de concepto, logramos detectar nuevas oportunidades de negocio en Contoso, lo que nos permitió posicionar más servicios de analítica avanzada en Azure, tales como Data Lake Analytics, Spark, R Server, Machine Learning y Data Factory.

De manera inmediata, nos abrió puerta para iniciar una nueva prueba de concepto que permitiría explotar información de su programa de Lealtad, el cual le permite al cliente obtener puntos, dinero electrónico u otros beneficios cada que realiza una compra en Contoso Ltd. La iniciativa consiste en realizar extracciones diarias del CRM Lealtad, analizar los datos y utilizar esta información para mejorar la experiencia de compra, completar la canasta, realizar email marketing personalizados, cupones personalizados, así como segmentación de clientes leales.

A manera de resumen las iniciativas detectadas fueron:

1. VPN Site-Multisite
2. Programa de Lealtad CRM
 - a. Market basket
 - b. Email marketing y cupones personalizados
 - c. Segmentación de clientes leales.
3. Migración de sistemas a Azure Data Warehouse.
4. Migración de Aplicaciones de BI (Arthus y .NET) a Power BI.
5. Mejora del proyecto de "Log Analytics" utilizando Azure Functions.

Como se puede ver, la cuenta se empezó a atender apenas con ciertos elementos de analítica avanzada basado en nube, sin embargo, después de varias sesiones con cliente, pudimos posicionar muchos servicios de Azure que impactan de manera directa en el negocio del cliente. En la Figura 5.12 se muestra la arquitectura proyectada para el resto de iniciativas encontradas.

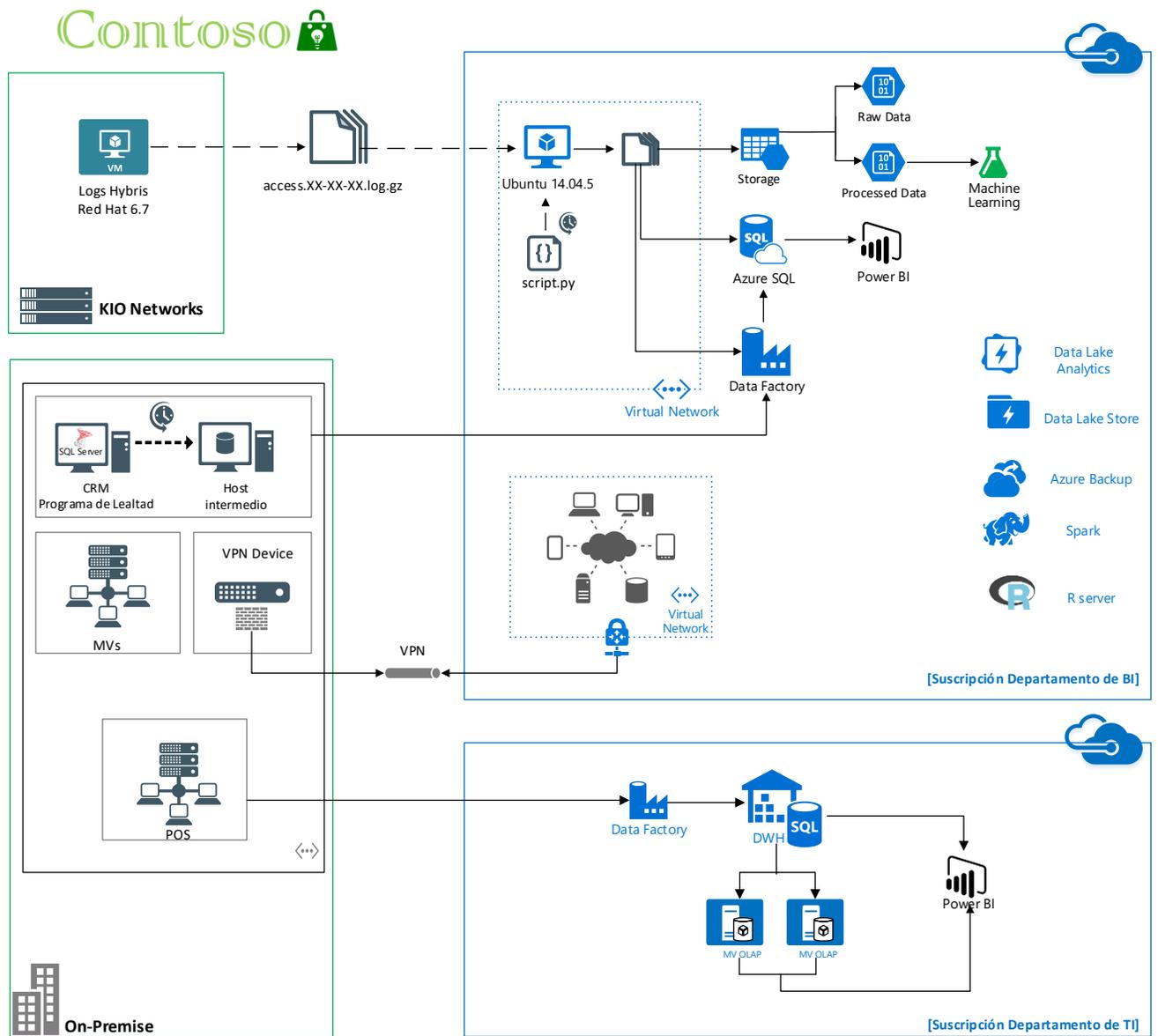


Figura 5.12 Arquitectura proyectada a partir de las nuevas iniciativas.

5.4 Propuesta Comercial

Existen varias formas de adquirir una suscripción de Azure, por ejemplo, a través de un contrato Enterprise, mediante asociados de Microsoft o bajo el modelo Pay-as-you-go.

Las grandes empresas suelen optar el modelo Enterprise, de este modo aceptan un compromiso monetario inicial (anual) para Azure del cual obtienen grandes ventajas, como opciones de facturación flexibles, promociones y mejores precios.

El PSS es la figura encargada de elaborar una propuesta comercial que haga sentido al cliente. Como resultado de la prueba de concepto de Log Analytics, encontramos nuevos workloads que se consideraron para cotizar el tamaño del *Monetary Commitment*. En la tabla 5.2 se hace el desglose de los servicios que Contoso estaría consumiendo de manera mensual y anual.

Una vez que el cliente estuvo conforme con la propuesta comercial, se firmó un contrato por el monto del acordado para Azure, uso licencias de Power BI Pro y un paquete de Soporte (Tabla 5.1).

Producto	Precio
Azure Monetary Commitment (<i>Escenario Operación Inicial</i>)	\$40,920.00
Subscription / Azure Support Commit Standard	\$3,000.00
Power BI Pro per User (30 seats)	\$3162.04
TOTAL ANUAL: \$47,382.04 *	

Tabla 5.1. Cotización final de proyectos.

*Precios Expresados en Dólares Americanos antes de IVA.

Una vez cerrado el contrato, el CSA y el DSA deben elaborar un plan de consumo, que tal como su nombre lo indica, es un documento que define el consumo de servicios de Azure estimado por mes. El objetivo de este plan, es que el cliente consuma la totalidad de su Monetary Commitment en el tiempo estimado en el contrato (3 años para contrato Enterprise).

Cuando se identifican nuevas oportunidades que no pueden cubrirse con el Monetary Commitment actual, se hace un nuevo contrato.

Ambos arquitectos son responsables de dar seguimiento al consumo y ser un facilitador del cliente en su transición a la nube.

En la Tabla 5.2 se detalla la cotización de los servicios de Azure que el cliente espera consumir durante su primer año.

Custom name	Description	Service type	Region	Estimated Cost
Logs ETL VM	Servidor para las tareas de transferencia y transformación de la información de Ecommerce	Virtual Machines	South Central US	\$109.82
R server	Servidor dedicado al análisis estadístico	Virtual Machines	South Central US	\$1,291.93
Spark + R	Cluster de computo de alto desempeño para análisis de bases de datos históricas	HDInsight	South Central US	\$186.51
SQL Database	Base de Datos como Servicio	SQL Database	South Central US	\$512.23
Data Lake HDFS	Sistema de Archivos Distribuido de Hadoop	Data Lake Store	Central US	\$41.82
Data Lake HDFS Analytics	Análisis de los datos almacenados en el HDFS	Data Lake Analytics	Central US	\$597.33
Machine Learning Studio	Plataforma de análisis Machine Learning	Machine Learning	South Central US	\$297.39
Storage	Almacenamiento general	Storage	South Central US	\$27.13
CRM ETL	Extrae, transforma y carga la información del CRM hacia la plataforma de análisis y la Base de Datos	Data Factory	West US	\$30.15
JDBC Bridge Hybris	Servidor para conectar Hybris a bases de datos estándar	Virtual Machines	South Central US	\$109.82
Discos	disco extra para almacenamiento	Storage	South Central US	\$0.00
Backup R Server	Servicio de Backup del Servidor R Server	Backup	South Central US	\$135.36
Backup Logs ETL y JDBC Bridge Hybris	Servicio de Backup de los servidores de Logs ETL y el Bridge de Hybris	Backup	South Central US	\$43.18
Support				\$27.32
				\$3,410.00
				\$40,920.00

Tabla 5.2. Desglose de costos para el proyecto de Log Analytics y análisis de CRM

VI. Conclusiones

Mi experiencia en Microsoft México, me permitió adquirir y fortalecer mis habilidades técnicas como ingeniera en computación, así como desarrollar nuevas habilidades en el ámbito comercial y corporativo.

La transformación digital está cambiando la forma en la que se hacen los negocios y en general el estilo de vida de las personas. La generación masiva de información y la constante innovación tecnológica representan retos importantes para los profesionales de TI y a través de la nube podemos resolver de una forma más rentable algunos de estos retos.

Desde mi puesto de Cloud Technical Specialist, tuve la oportunidad de colaborar con este movimiento e inspirar a mis clientes con soluciones basadas en nube. Un claro ejemplo de esto, fue el proyecto de Log Analytics.

La prueba de concepto “Log Analytics” fue exitosa y permitió al cliente comprobar el valor que Azure puede traer a su negocio.

Habilitamos al departamento de Análisis de datos para hacer su trabajo de una manera más sencilla y eficiente, obtener información valiosa que les permita conquistar nuevos clientes, reinventar la estrategia del negocio, predecir lo que desean sus clientes y con ello obtener más ganancias.

La información arrojada por las visualizaciones resultó de gran valor para Contoso, ya que sugiere varias oportunidades de mejora, por ejemplo:

- Analizar lo que buscan los clientes y crecer el catálogo de productos para estar al día de sus expectativas.
- Atender los productos del catálogo que son más consultados nos puede ayudar a procurar que siempre haya en el inventario de la tienda. También puede darnos ideas de nuevas promociones o estrategias de venta basadas en la oferta y demanda de los productos para aprovechar la popularidad de los mismos.
- Comprender hasta que parte del proceso de venta llegan los usuarios, nos puede ayudar a detectar fallas en el canal, como podría ser un proceso erróneo o una interfaz poco amigable.

- Saber desde donde visitan los clientes el portal, es una pista importante para saber en dónde hay que invertir en más publicidad para atraer a los clientes.

Estos resultados, nos entregan tan solo una parte de la gama de posibilidades que se pueden explotar al analizar los datos que se generan día con día.

Por otra parte, los logs procesados y la base de datos generada a partir del proyecto Log Analytics, servirá como origen de datos para proyectos de Azure Machine Learning, el cual puede entregar todavía más conocimiento sobre el comportamiento y/o preferencias de los usuarios. Un ejemplo de aplicación, es la predicción de la canasta de los usuarios para realizar sugerencias (market basket analysis) o promociones oportunas.

Hoy en día, la generación masiva de datos e inteligencia artificial, juegan un rol muy importante para reconocer patrones de comportamiento y análisis de sentimiento de nuestros clientes.

El proyecto realizado fue solo el primer paso para impulsar la transformación digital de Contoso Ltd. ya que la gama de servicios ofrecidos por Microsoft Azure, permitió a Contoso tener nuevas iniciativas alrededor de los modelos de comercio electrónico modernos, donde la experiencia omnicanal, el marketing personalizado y la popularidad en redes sociales, pueden representar una ventaja competitiva significativa en la industria.

Por otra parte, se logró firmar un contrato con Contoso Ltd. lo cual nos habilitó para diseñar y proyectar nuevas soluciones en la nube que identificamos durante las entrevistas con Contoso como la generación de un algoritmo de market basket analysis, unificación de la infraestructura actual y la adquirida en el 2016, así como la migración de sistemas a Azure Data Warehouse.

VII. Referencias

- Colaboradores. (10 de Febrero de 2017). *Documentación SQL Database*. Obtenido de Microsoft Azure: <https://docs.microsoft.com/es-es/azure/sql-database/sql-database-what-is-a-dtu>
- Colaboradores. (26 de Enero de 2017). *Tamaños de MV*. Obtenido de Microsoft Azure: <https://docs.microsoft.com/en-us/azure/virtual-machines/virtual-machines-windows-sizes?toc=%2fazure%2fvirtual-machines%2fwindows%2ftoc.json>
- Espinosa, R. (Diciembre de 2016). *Dataprix*. Obtenido de <http://www.dataprix.com/blogs/respinosamilla/herramientas-etl-que-son-para-que-valen-productos-mas-conocidos-etl-s-open-sour>
- Haishi Bai, S. M. (2015). *Exam Ref 70-534 Architecting Microsoft Azure Solutions* (Primera Edición, Microsoft).
- Hess, K. (4 de Febrero de 2017). *Using Expect Scripts to Automate Tasks*. Obtenido de Admin Magazin: <http://www.admin-magazine.com/Articles/Automating-with-Expect-Scripts>
- Hipertextual. (11 de Diciembre de 2016). *Software*. Obtenido de Hipertextual: <https://hipertextual.com/presentado-por/microsoft/mobile-first-cloud-first>
- Laurent Debrawer, F. V. (2016). *UML 2.5. Iniciación, ejemplos y ejercicios corregidos*. España: Eni Ediciones. Obtenido de https://books.google.com.mx/books?id=sCU_bpelECAC&pg=PA52&dq=uml+casos+de+uso&hl=es&sa=X&redir_esc=y#v=onepage&q=uml%20casos%20de%20uso&f=false
- Microsoft. (2016). *¿Qué es una nube pública?* Obtenido de Microsoft Azure: <https://azure.microsoft.com/es-mx/overview/what-is-a-public-cloud/>
- Microsoft. (10 de Diciembre de 2016). *Company overview & strategy*. Obtenido de Microsoft: <https://www.microsoft.com/en-us/Investor/company-overview.aspx>
- Microsoft. (2 de Enero de 2016). *Facts about Microsoft*. Obtenido de Microsoft: <https://news.microsoft.com/facts-about-microsoft/#sm.000efk9o5174tf0ywby1sbr4sbrfw>

- Microsoft. (18 de Enero de 2017). *What is IaaS?* Obtenido de Microsoft Azure: <https://azure.microsoft.com/en-us/overview/what-is-iaas/>
- Microsoft. (Marzo de 2017). *¿Qué es la transformación digital?* Obtenido de Microsoft Blog para PyMEs: <https://blogs.business.microsoft.com/es-mx/2016/10/07/la-transformacion-digital/>
- Microsoft. (5 de Enero de 2017). *About Microsoft*. Obtenido de Microsoft: <http://www.microsoft.com/en-us/about/default.aspx>
- Microsoft. (6 de Enero de 2017). *Blob Storage*. Obtenido de Microsoft Azure: <https://azure.microsoft.com/en-us/services/storage/blobs/>
- Microsoft. (20 de Agosto de 2017). *Cloud Solution Architect*. Obtenido de Careers: <https://careers.microsoft.com/jobdetails.aspx?ss=&pg=0&so=&rw=13&jid=321820&jlang=EN&pp=SS>
- Microsoft. (21 de Agosto de 2017). *Data Cloud Solution Architect*. Obtenido de Careers: <https://careers.microsoft.com/jobdetails.aspx?ss=&pg=0&so=&rw=2&jid=287164&jlang=EN&pp=SS>
- Microsoft. (23 de Abril de 2017). *Digital Transformation*. Obtenido de Microsoft Enterprise: <https://enterprise.microsoft.com/en-us/digital-transformation/>
- Microsoft. (7 de Enero de 2017). *Documentación de Storage*. Obtenido de Microsoft Azure: <https://docs.microsoft.com/es-es/azure/storage/>
- Microsoft. (12 de Enero de 2017). *Machine Learning*. Obtenido de Microsoft Azure: <https://azure.microsoft.com/es-es/services/machine-learning/>
- Microsoft. (7 de Enero de 2017). *Maquinas Virtuales*. Obtenido de Microsoft Azure: <https://azure.microsoft.com/es-es/services/virtual-machines/>
- Microsoft. (8 de Febrero de 2017). *Power BI learning*. Obtenido de Power BI: <https://powerbi.microsoft.com/es-es/guided-learning/>
- Microsoft. (11 de Febrero de 2017). *Precios Azure*. Obtenido de Microsoft Azure: <https://azure.microsoft.com/es-es/pricing/>
- Microsoft. (22 de Febrero de 2017). *Principal Solution Spec*. Obtenido de Careers: <https://jobsbeta.microsoft.com/us/en/job/318866/Principal%20Solution%20Spec>

- Microsoft. (14 de Marzo de 2017). *Showcase Azure*. Obtenido de Microsoft Technet:
https://blogs.technet.microsoft.com/uspartner_ts2team/2015/08/28/showcase-azure-solutions-with-the-cloud-platform-immersion-sales-program/
- Microsoft. (10 de Enero de 2017). *SQL Database*. Obtenido de Microsoft Azure:
<https://azure.microsoft.com/en-us/services/sql-database/?b=16.50>
- Microsoft. (8 de Febrero de 2017). *What is PaaS?* Obtenido de Microsoft Azure:
<https://azure.microsoft.com/en-us/overview/what-is-paas/>
- Prensa, Microsoft. (5 de Enero de 2017). *Noticias*. Obtenido de Microsoft:
<https://news.microsoft.com/es-es/2015/04/06/historia-microsoft-40-aniversario/#sm.00000vafwst4uecqgyqmy66t414kw#7EolQDouk5Brgkoi.97>
- Sinneux. (10 de Enero de 2017). *BI*. Obtenido de Sinneux:
http://www.sinnexus.com/business_intelligence/
- Sintec. (Mayo de 2017). *De datos a dinero en retail: análisis de la canasta de compra*. Obtenido de Forbes México: <https://www.forbes.com.mx/brand-voice/de-datos-dinero-en-retail-analisis-de-la-canasta-de-compra/>
- Sommerville, I. (2011). *Ingeniería de software* (9na ed.). México: Pearson Educación.
- Vaks, I. (19 de Febrero de 2017). *Finance Today*. Obtenido de SlideShare:
<https://www.slideshare.net/jcibewise/bewisei-loeng-ivo-vaks-performance-management-ja-laienemine-microsofti-nitel-tt-18092014>



VIII. Anexos

A. Glosario

- **Monetary Commitment:** Forma de prepago de Azure. El cliente paga una cantidad equivalente al consumo de Azure que espera tener durante un año. Si al final del año le queda saldo sin usar, los fondos se perderán.
- **Partner:** Socio de Microsoft que se involucra en el proceso de venta, con el fin de participar en las pruebas de concepto y en las implementaciones en producción. Cada proyecto representa una oportunidad de negocio para el socio.
- **Enterprise and Partner Group (EPG):** División de negocios dentro de Microsoft que se enfoca en clientes empresariales con gran potencial y relevancia estratégica para Microsoft. Dentro de esta división, los equipos de ventas se encuentran dentro del Account Team Unit (ATU), especialistas técnicos de preventa están en el Sales Team Unit (STU) y los socios dentro del Partner Team Unit (PTU).
- **Business Intelligence (BI):** Business Intelligence es la habilidad para transformar los datos en información, y la información en conocimiento, de forma que se pueda optimizar el proceso de toma de decisiones en los negocios.
- **Infrastructure as Service (IaaS):** infraestructura informática inmediata que se aprovisiona y administra a través de Internet. Permite reducir o escalar verticalmente los recursos con rapidez para ajustarlos a la demanda y se paga por uso. IaaS evita el gasto y la complejidad que suponen la compra y administración de sus propios servidores físicos y otra infraestructura de centro de datos.
- **Platform as Service PaaS (PaaS):** Es un entorno de desarrollo e implementación completo en la nube, con recursos que permiten entregar todo, desde aplicaciones sencillas basadas en la nube hasta aplicaciones empresariales sofisticadas habilitadas para la nube.
- **Data Transaction Unit (DTU):** Una DTU es una medida que combina CPU, memoria, datos de E/S y E/S de registro de transacciones en una relación determinada por una carga de trabajo de pruebas comparativas de OLTP diseñada

para ser la normal entre las cargas de trabajo OLTP reales. Duplicar el número de DTU aumentando el nivel de rendimiento de una base de datos equivale a duplicar el conjunto de recursos disponibles para esa base de datos. El número de DTUs es una unidad de medida de los recursos que se garantiza que estarán disponibles para una instancia única de Azure SQL Database en un nivel de rendimiento específico dentro de un nivel de servicio de base de datos única.

- **Market Basket:** También conocido como análisis de asociación. Es un análisis matemático que sirve para identificar asociaciones o patrones típicos entre productos comprados. En el área de retail, se suele usar para completar la canasta, acomodo de productos o evaluar efectividad del portafolio y promociones.
- **Análisis de sentimiento:** proceso de determinar el tono emocional que hay detrás de una serie de palabras, y se utiliza para intentar entender las actitudes, opiniones y emociones expresadas en una mención online. El análisis de sentimiento es extremadamente útil en la monitorización de las redes sociales ya que permite hacernos una idea de la opinión pública general sobre ciertos temas.
- **CRM (Customer Relationship Management):** Es una estrategia de negocio orientada a la fidelización de los clientes. Software que rastrea y gestiona de forma eficaz la información de los clientes, permitiendo a los empleados de una empresa, disponer de información sobre clientes actuales y potenciales (nombres, direcciones, números telefónicos, etc.), y sus actividades y puntos de contacto con la empresa, que incluyen visitas a sitios web, llamadas telefónicas, correos electrónicos y más.

B. ¿Qué es Microsoft Azure?

Microsoft Azure es una colección de servicios integrados en la nube pública que permite crear soluciones de análisis, proceso, bases de datos, móviles, redes, almacenamiento y Web.

“La nube pública se define como servicios informáticos que ofrecen proveedores externos a través de la Internet pública y que están disponibles para todo aquel que desee utilizarlos o comprarlos. Pueden ser gratuitos o venderse a petición, lo que permite a los clientes pagar solo por el uso que hacen de ciclos de CPU, el almacenamiento o el ancho de banda que consumen.” (Microsoft, ¿Qué es una nube pública?, 2016)

Las nubes públicas pueden ahorrar a las compañías los enormes gastos que supone tener que comprar, administrar y mantener hardware e infraestructura de aplicaciones locales, ya que el proveedor del servicio en la nube es el responsable de todo el trabajo de administración y mantenimiento del sistema. Las nubes públicas también se pueden implementar con más rapidez que las infraestructuras locales y con una plataforma que permite una escalabilidad casi ilimitada.

Al momento de escribir el presente informe, Azure está disponible en 140 países, y tiene disponibilidad general en 36 regiones de todo el mundo. Microsoft ha dado alta prioridad a la expansión geográfica, a fin de ofrecer mayor rendimiento y satisfacer las necesidades y preferencias en cuanto a la ubicación de los datos.

En la Figura 8.1, se muestra de manera gráfica los servicios más representativos de Microsoft Azure. Como se muestra en la imagen, Azure ofrece servicios para montar aplicaciones móviles y web, servicios de analítica avanzada, servicios de seguridad, cómputo de alto rendimiento, almacenamiento de datos e infraestructura como servicio. Además cuenta con servicios de integración, los cuales permiten conectar los sitios on-premise a la nube.

Actualmente Microsoft Azure continua expandiendo y mejorando sus servicios de nube, compitiendo con grandes empresas como Amazon Web Services y Google Cloud.

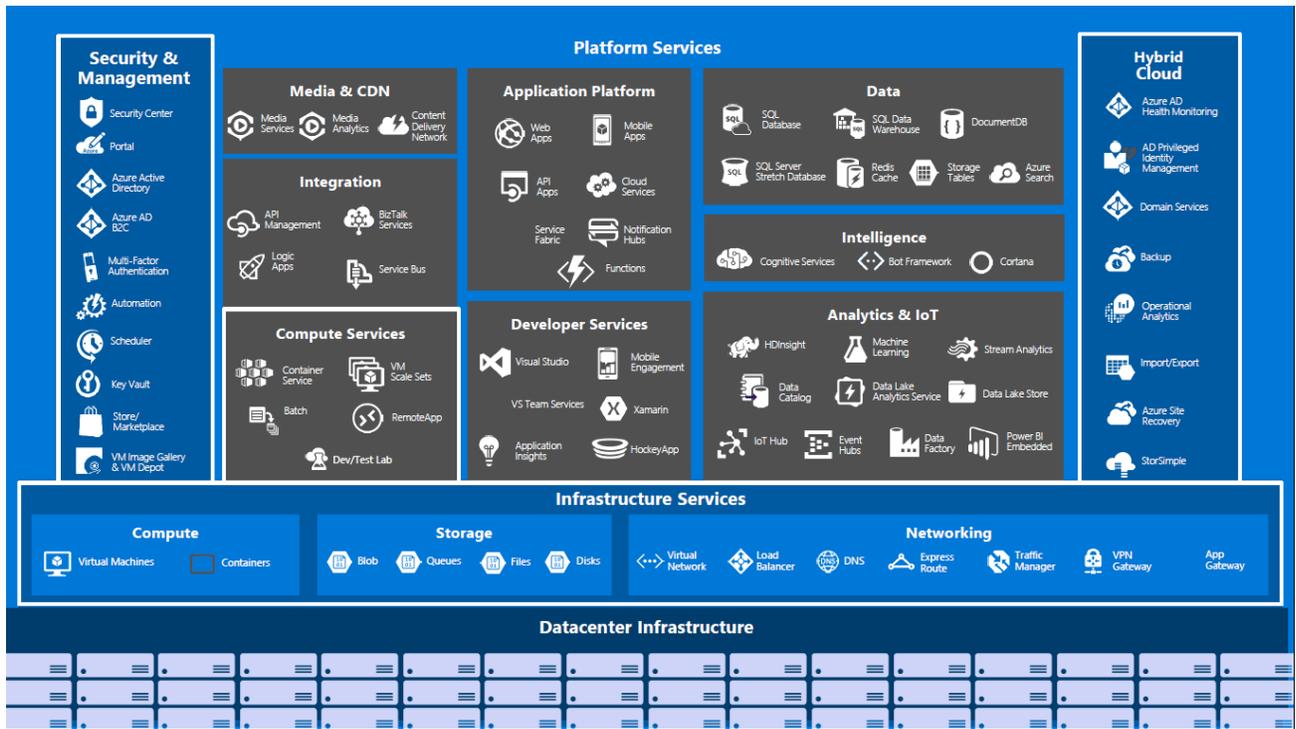


Figura 8.1 Catálogo representativo de servicios de Azure.