

Tarea Juan Pablo González

1. ¿Cuáles son los desafíos si tenemos grandes flujos de datos? ¿Por qué se hace necesaria una plataforma y no se procesa en computadores personales? ¿Qué módulos debería tener esta plataforma?
 - Un retail debe gestionar en tiempo real los datos derivados por su operación, la logística de insumos y la auditoría en las operaciones bancarias, entre otras. Esto termina provocando la necesidad de administrar y gestionar importantes cantidades de datos que oscilan dependiendo del periodo del día o año.
 - El principal objetivo de poder controlar este volumen de datos es lograr conocimiento, lo que permite a las unidades gerenciales y operativas tomar decisiones con base sobre los movimientos de la organización.
 - Al ser una tarea titánica, se requiere levantar data centers los cuales están compuestos por servidores, software especializado, con un diseño de infraestructura dedicado.
 - En términos de diseño de esta infraestructura se debe de tener en cuenta políticas de seguridad que buscan la protección de los datos, la continuidad operacional y escalabilidad. También se requiere tener en cuenta las soluciones de software que serán usadas en esta, entre las cuales encontramos:
 - i. Base de datos: Sistemas que se encargan de gestionar los datos.
 - ii. Soluciones etl: Soluciones que se dedican a recuperar datos, procesarlos y almacenarlos en repositorios especializados.
 - iii. Soluciones de almacenamiento: Soluciones que permiten almacenar y recuperar importantes cantidades de datos en pocos segundos, entre los cuales podemos encontrar datawarehouse o data lake.
 - iv. Soluciones de BI: Esto no es propio del datacenter, pero si queremos visualizar los datos requerimos una herramienta dedicada que nos permita hacer las interpretaciones y consultas que generan valor a partir de los datos disponibles-
 - Un equipo personal queda obsoleto para esta operación ya que estos no poseen la capacidad para administrar hardware lo bastante potente en términos de procesamiento y memoria, los cuales son indispensables.
2. En caso de realizar un desarrollo de plataforma interno: ¿Qué perfiles tecnológicos se recomiendan contratar? ¿Cuáles son las responsabilidades de cada uno?
 - Idealmente, los perfiles que se deberían de contratar son las siguientes:
 - i. Data Analyst: Se encarga de mantener contacto con los equipo gerenciales y generar reportes en orden a las necesidades de estos. Es de un perfil técnico con conocimientos en herramientas de BI y análisis de datos. No está exento de saber programación en lenguajes vinculados a las herramientas que ocupe su organización.
 - ii. Data Scientist: Se encarga de analizar y estudiar los datos de la organización con base en objetivos gerenciales. Disponible de conocimientos estadísticos y técnicos que le permiten implementar soluciones de machine learning, procesamiento del lenguaje natural, etc; permitiendo recuperar información de gran valor que solo es

posible obtener por medio de un proceso investigativo. Son especialistas multidisciplinarios.

- iii. Chief Data Officer: Es gerente interesado en la gobernanza de los datos de su organización, esto significa controlar y auditar los datos existentes además de saber quienes disponen de estos. Al ser un gerente está encargado de apoyar la visión y objetivos de la organización en los desafíos vinculados al big data y estudio de los datos. De este dependen todos los otros perfiles, ya que define los objetivos generales en la administración de los datos.
 - iv. Data Engineer: Es el encargado de disponibilizar la infraestructura tecnológica necesaria para capturar, guardar y disponibilizar datos a los otros roles puedan ejecutar sus funciones. Es un rol de un nivel técnico elevado ya que debe de saber cómo diseñar, implementar y configurar la infraestructura además de dominar el uso de soluciones de software necesarias. En nuestro caso servicios como base de datos, soluciones de ETL, etc.
3. Nombrar un caso de uso que justifique su inversión y detalle cómo debería ser el flujo de datos, actualización y monitoreo de la solución.

Caso: Compra de productos en línea

- Al ser un retail un tipo de industria con una alta demanda online, es normal que los clientes deseen pagar de esta misma forma. Esto hace que sus flujos de caja sean literalmente dependientes de tener una infraestructura capaz de soportar tal exigencia. Al mismo tiempo que se genera esta demanda por parte de los clientes, existe otra alta demanda por parte de los partners tecnológicos que entregan los medios para hacer el cobro a los clientes directamente a sus cuentas bancaria o tarjeta de crédito. A estos partner se les menciona como procesadores.
- Si un procesador no me confirma que la transacción fue exitosa o fallida, ¿cómo podría indicar esta información al cliente en tiempo real?. Esto hace que la infraestructura de un retail sea literalmente un embudo donde tanto la información de la compra del cliente como la información de la transacción bancaria desembocan a un mismo punto. Este tipo de infraestructura debe gestionar dicho volumen con una alta disponibilidad, resiliencia y escalabilidad.
- El flujo sería el siguiente:
 - El cliente a través del sitio web/APP selecciona un artículo a comprar, pasa al checkout donde confirma la dirección de entrega y método de pago.
 - La información que entrega el cliente es rescatada y enviada a un servicio web que se encarga de respaldarla en una base de datos dedicada.
 - Esto último puede ser desde la dirección de entrega como la validación de un método de pago recurrente como las tarjetas de crédito.
 - Sobre las tarjetas de crédito suscritas como método de pago se hace una operación donde yo le envío los datos encriptados al procesador y este me responde con una autorización permanente que debe almacenarse en una base de datos con los más altos estándares de

seguridad posibles. La base de datos principal donde se almacena estos datos debería replicarse de forma sincrónica a todas los servicios necesarios para evitar problemas durante el pago del cliente.

- Las direcciones podrían ser utilizadas para investigación de mercado, por tanto podría derivarse desde por medio de un ETL (en este punto se debería cruzar las direcciones de envíos con otros datos del cliente) a un silo dedicado.
- Debería de existir un servicio que permita controlar el inventario dedicado a las ventas por internet, el cual debería estar conectado a una base de datos con un alto capacidad de transacción.
 - Al comprar un producto este servicio debería actualizar en tiempo real su stock.
 - También debería ser un servicio web que se replique en varias bases de datos dedicadas, ya que el stock de productos debe ser escalado a varios interesados. Entre ellos la operación que controla la distribución de insumos, las páginas web/aplicaciones de compras las cuales deben bloquear la opción de compra al acabarse el stock, etc.
 - Entre todos estos servicios la replicación puede ser sincrónica o asincrónica dependiendo de la demanda de estos, por ejemplo los sitios web/APP requieren tener esta información de inmediato.
- Al confirmar la compra me contacto nuevamente con el procesador para indicarles que cobraré X monto al método de pago del cliente.
 - Por nuestra parte registramos el intento de pago a través de un servicio dedicado. La base de datos de este servicio sería solo de escritura. El procesador nos debería de entregar un identificador que más tarde podemos utilizar para confirmar el cobro del cliente, este dato también debería registrarse en el intento de pago.
 - Normalmente el procesador entrega un servicio web al cual se llama para generar el cobro. También dependiendo del método de pago esta operación es directa o indirecta.
 - En Chile estamos acostumbrados a la operación indirecta al pagar por medio de “Transbank” + “confirmación en el sitio del banco”.
 - Nosotros debemos crear un servicio que recepcione notificaciones por parte del procesador, estas notificaciones incorporan la confirmación o rechazo en el cobro al cliente.
 - Este registro debe contener el identificador que originalmente el procesador al intentar transaccionar.
 - Dependiendo del método de pago esta confirmación puede demorar segundos o días. Esto está directamente relacionado al nivel de integración digital de la región donde se compra, en particular en Chile es inmediata.
 - La interfaz de la APP debe mostrar el resultado de la transacción, esto se puede lograr por que el sitio web/app consulta constantemente a un servicio que está sincronizado con el resultado de la transacción.
 - En términos de monitoreo el resultado de una compra puede ser procesado y monitoreado para obtener antecedentes para:

- Cuadrar las ganancias.
- Prevención de fraude, aunque generalmente los procesadores llevan control por este punto en particular.
- Tendencias de compra y generación de perfiles de clientes,
- Control de stock y automatización en la compra de insumos.