PRUEBA TÉCNICA ARQUITECTO DE DATOS JIKKOSOFT S.A.S

Sección 5: Habilidades blandas y liderazgo

1. Describa un momento en el que tuvo que dirigir un equipo multifuncional para

Entregar un proyecto de datos. ¿A qué desafíos se enfrentó y cómo los superó?

Recientemente enfrentamos la tarea de construir un integración entre los sistemas de manejo de proveedores de la compañía y el sistema de contratación del estado SECOP, dicha integración tenia el reto de conectividad de distintos sistemas, desde el ERP para extraer la información de contratación, el lago de datos para la extracción de los archivos soporte de los distintos pasos del proceso de contratación estatal, realizar la extracción y trasformar los datos para ajustarlos a las reglas del SECOP y por ultimo preparar los archivos para que pudiesen ser consumidos por un robot que subiría al sistema destino la información.

Como retos importantes teníamos las restricciones que trabajaríamos con un proveedor externo que se encargaría de la parte final del proceso (Robot de carga a SECOP), por tanto fue necesario mantener coordinado el equipo para acordar las reglas que tendrían que tenerse en cuenta par a poder trabajar en paralelo y que dicho proveedor no dependiera de que finalizáramos los procesos iniciales para avanzar con su trabajo. Acordamos contratos, estructuras finales de los archivos JSON, carpetas y rutas, nombres de archivos etc. Así mismo llegado el proceso de integrar las soluciones finales, coordinar las pruebas integrales y coordinar la orquestación de ejecuciones.

Desde las responsabilidades de nuestro desarrollo, constituimos un equipo en dos frentes, uno el encargado de orquestar el proceso, (extracciones, trasformaciones y generación de archivos) este fue el encargado de construir las ETL (pipelines) encargados de extraer desde la base del ERP la información de la metadata, desde el lago de datos, mediante el llamado a APIS, genera la extracción de los archivos. Luego con la información extraída de ambos sistemas procesar la información y construir los archivos para dejarlos en los repositorios que el otro proveedor subiría al SECOP con el robot construido.

El segundo frente trabajo en la construcción de las Azure Logic APS encargadad e consumir las APIS del ERP y con dicha información extraer y decodificar los archivos del lago de datos.

Adicionalmente se presentaron retos de cambios en el camino ya que el sistema SECOP presentaron cambios que afectaron las implementaciones, con cambios de datos e inclusión de nuevos requerimientos.

Los superamos de una manera no tan traumática ya que desde l diseño logramos decidir la implementación en distintos componentes, muchos de ellos con un gran alto grado de parametrizaciones, de esta manera pudimos trabajar en paralelo en la solución y manejo de los controles de cambios presentados.

2. ¿Cómo se mantiene actualizado con las últimas tendencias y tecnologías en

arquitectura de datos y gestión de datos? Explique cómo aplicó una tecnología o

tendencia reciente en un proyecto real.

Regularmente realizo seguimiento a los perfiles de ingenieros de datos, con eso tengo una guía de lo que el mercado está requiriendo, también sigo en videos a referentes en el área. Con esto investigo y me informo de dichas tecnologías. Igualmente, según las necesidades actuales de la compañía se presentan rutas de estudio para aplicar en proyectos futuros y en curso.

Mi área de enfoque principal durante los últimos años se centrado en la integración masiva de datos, en dicha trabajamos arquitecturas de migración y de integración de datos de aplicaciones onpremises a aplicaciones en nube y viceversa. La tendencia actual de estas integraciones ha sido la comunicación multiplataformas. Basado en eso buscamos una plataforma que nos permitiera la conectividad entre diferentes tipos de orígenes y destinos de datos, escogimos Azure Data Factory ya que cumplía con estás características y además podría adaptarse en el futuro para múltiples tipos de proyectos en manejo de datos, analítica y ciencia de datos.

3. Proporcione un ejemplo de un problema complejo que resolvió en su papel anterior

como arquitecto de datos. Explique su enfoque y el resultado, incluyendo un

problema técnico específico relacionado con arquitectura de datos.

La compañía tenía la necesidad de poder generar rápidamente informes legales e información de requerimientos específicos a órganos de control estatal. Ellos contaban con sistemas que permitían generar el dinamismo para la generación de la información como un reporteador, pero tenían límites a la hora de procesar grandes volúmenes de datos. Con lo que nos planteo el reto de la construcción de una solución que permitiese recibir desde el reporteador las consultas dinámicas y desde la solución poder dejar los archivos en un repositorio, notificando al usuario que su información se encontraba lista.

El reto principal era poder construir una solución trasversal que funcionara a manera de componente para que pudiese integrarse tanto con el sistema reporteador cómo en el futuro poder reusarse por otras aplicaciones que requieran de soluciones similares.

Enfocamos la solución teniendo en cuenta que existe de la incertidumbre del tipo de fuente de origen de datos a procesarse (bases de datos SQL server, Oracle, etc) y el dinamismo exige poder determinar o sólo los nombres de los campos que estamos recibiendo, para ello plantemos una estructura modular del pipilene de generación, el núcleo encargado de la extracción de diferentes fuentes, que contaría con implementación para cada tipo de base de datos (diferentes conectores) de esta forma en dicha funcionalidad se contemplarían los mecanismos para determinar la metadata de los campos de la consulta, poder con esta construir la tabla destino temporal en la base de datos Staging, y de acuerdo al tipo de conector generar la extracción de los datos del origen particular a la tabla temporal. Habría un orquestador para estos procesos que podría paramétricamente invocarse indicando el tipo de conector a ser usado, el indicador del sistema al se conectaría (nos determina la cadena de conexión a ser usada) y la consulta a ejecutar.

De esta forma se tendría un desarrollo modular por cada tipo de conector y sería viable implementaciones futuras para incluir nuevos conectores a bases de datos distintas que puedan requerirse.

Ya una vez la información se encuentre en el área de staging, se tendrían un módulo trasversal para poder generar los planos y dejarlos en la ruta. Esta generación tendría la capacidad de generar n archivos, de acuerdo a los parámetros recibidos en los cuales se indicaría el número de filas por archivo, esto con el fin de flexibilizar el manejo de estos por los usuarios finales.

Logramos en poco tiempo suministrar una herramienta que atacara el problema puntual que requería la compañía y dejamos preparada una arquitectura que podría escalar para abarcar problemas futuros con intervenciones de desarrollo modular para suministrar compatibilidad con nuevos conectores a diferentes bases de datos.