Pizzichini, Gisela Isabel

Data Pipeline: una propuesta para la administración y gestión de datos

Licenciatura en Gestión de la Tecnología

Fecha: 20/12/2024

UNIVERSIDAD NACIONAL DE RAFAELA

FACULTAD DE TECNOLOGÍAS E INNOVACIÓN PARA EL DESARROLLO

GESTIÓN DE LA TECNOLOGÍA

"Data Pipeline: Una propuesta para la administración y gestión de datos"

Gisela Pizzichini

Rafaela

Diciembre 2024

Contenido.

1. Introducción.

La creciente disponibilidad de datos en diversas industrias ha propiciado un cambio paradigmático en la manera en que las organizaciones operan y abordan la toma de decisiones. El término "Big Data" se ha convertido en un concepto fundamental en el ámbito de la analítica, haciendo referencia a conjuntos de datos tan voluminosos y complejos que requieren de tecnologías avanzadas para su procesamiento y análisis. Sin embargo, la acumulación de datos no garantiza la obtención de información valiosa, es aquí donde el concepto de "Smart Big Data" cobra relevancia. Este enfoque no solo se centra en la cantidad de datos, sino también en su calidad, relevancia y capacidad para generar datos significativos que faciliten la toma de decisiones (insights).

En un contexto donde la competitividad y la innovación son esenciales para el éxito organizacional, entender cómo convertir datos en decisiones inteligentes se torna imprescindible.

Un estudio realizado por IBM (IBM Institute for Business Value, 2012), ratifica que los enfoques de Big Data en una organización mejoran las predicciones de indicadores económicos, ayudan a detectar tendencias del mercado con la intención de anticipar oportunidades de negocios, visualizar información en tiempo real y a contribuir en la toma de decisiones estratégicas.

Por consiguiente, para que las organizaciones tomen decisiones basadas en datos, previo a la aplicación de analítica de datos, es necesario contar con un conjunto sistemático y automatizado de procesos diseñados para capturar, mover, transformar y gestionar datos de diversos orígenes a un destino específico. Este proceso se denomina “Data Pipeline”, o canalización de datos en español, y se utiliza para procesar grandes volúmenes de datos en sistemas de Big Data, en aplicaciones de inteligencia artificial y aprendizaje automático dado que permite el flujo continuo y eficiente de información a través de diversas etapas de procesamiento. (Datademia, 2023)

El presente trabajo se focaliza en el diseño y modelado de un data pipeline aplicado a una empresa alimenticia de la ciudad de Rafaela. El diseño del pipeline tendrá énfasis en los procesos de extracción, transformación y carga (ETL) que son fundamentales para preparar los datos para su posterior análisis. Este enfoque se justifica con la premisa de que la calidad y la relevancia de los datos son determinantes en la efectividad del análisis; por lo tanto, un pipeline bien estructurado es esencial para garantizar que los datos sean accesibles, consistentes y adecuados para responder a las necesidades que enfrentan las organizaciones.

2. Objetivos.

2.1 Objetivo general.

- Diseñar propuestas de gestión y administración de datos dirigidas a Pequeñas y Medianas Empresas (PYMES) de Rafaela que presentan un mínimo nivel de digitalización en sus procesos, con el fin de mejorar los procesos de toma de decisiones y la competitividad empresarial.

2.3 Objetivo específico.

- Relevar los distintos procesos que intervienen en la organización.
- Seleccionar el proceso de interés donde aplicar solución de gestión de datos.
- Relevar las distintas fuentes de datos involucradas y su estructura en el proceso de interés.
- Establecer el proceso de canalización de datos (Data Pipeline) contemplando el conjunto de procesos automatizados a utilizar en la transferencia de datos de determinadas fuentes a un destino específico.
- Modelar datos y estructura de gestión a utilizar
- Proponer uso de herramientas analíticas para el posterior análisis de datos.

3. Alcance del proyecto.

El presente proyecto se centra en el diseño de un proceso de canalización de datos (data pipeline). Este proceso abarca diversas etapas que van desde la recopilación hasta el almacenamiento de datos, con el objetivo de preparar la información para su posterior análisis de manera que se optimicen los procesos de toma de decisiones en una organización.

En términos de localización, el proyecto está dirigido a pequeñas y medianas empresas (PYMES) situadas en la ciudad de Rafaela, las cuales presentan un nivel significativo de digitalización en sus operaciones. Esta segmentación es esencial, ya que aquellas empresas que carecen de una infraestructura mínima en sus procesos requieren previamente un relevamiento y automatización de diversas tareas organizacionales antes de poder implementar el diseño del proceso de canalización de datos.

4. Justificación del proyecto.

Big data es una de las palabras de moda más importantes en los negocios de hoy. Casi todas las empresas están buscando formas de aprovechar las enormes cantidades de datos que ahora están disponibles. Pero tener acceso a todos estos datos no es suficiente. También necesita una forma de procesarlo de manera rápida y eficiente para que pueda obtener resultados. Ahí es donde entran los pipelines de datos. (Argomedo, 2022)

Los datos han adquirido una importancia significativa como activos para las organizaciones, no obstante, su manejo conlleva diversos desafíos que trascienden al volumen de información. Estos retos incluyen la heterogeneidad de los datos, la arquitectura subyacente, así como consideraciones relacionadas con el uso y la privacidad, la complejidad del procesamiento y la dificultad para distinguir entre datos relevantes e irrelevantes, entre otros factores. Para que una organización pueda fundamentar sus decisiones en datos, reemplazando así la intuición o la especulación, es esencial implementar procesos que gestionen información proveniente de diversas fuentes, permitiendo transformar y preparar los datos en información valiosa y actualizada que posteriormente será objeto de análisis mediante técnicas analíticas adecuadas. Este conjunto de procesos se denomina pipeline de datos o canalizador de datos.

De esta manera, realizar el diseño de un data pipeline en una empresa alimenticia ubicada en la ciudad de Rafaela se justifica por diversas razones que abarcan tanto la optimización de procesos internos como la mejora en la toma de decisiones estratégicas.

La creación de un data pipeline optimiza la recopilación, transformación y carga (ETL) de datos, lo que permite a la empresa manejar información heterogénea y en diversos formatos. En un pipeline bien diseñado se asegura que los datos sean accesibles, limpios y estructurados, lo cual es necesario para realizar análisis precisos y oportunos, mejorando la calidad y fiabilidad de la información evitando errores en las decisiones basadas en datos (DataSpurs, 2023). Además, el diseño del data pipeline, que puede integrar datos provenientes de diferentes departamentos y fuentes externas de una organización, proporciona una visión más holística del negocio.

En conclusión, el proceso de data pipeline representa una inversión estratégica para cualquier empresa de la ciudad de Rafaela que desee mejorar la eficiencia operativa, la calidad en el análisis de datos, y también potenciar la capacidad de innovar y adaptarse a un entorno empresarial dinámico.

5. Marco teórico.

Un pipeline de datos se define como un conjunto de procesos y herramientas diseñados para recopilar información de múltiples fuentes, analizarla y presentar los resultados en un formato accesible. Este sistema opera a través de una serie de etapas interconectadas que facilitan el flujo y el procesamiento de los datos a través de la extracción, transformación, programación, monitoreo, análisis y visualización de los datos.

El Data Pipeline inicia extrayendo datos en su estado natural desde su fuente de origen, permitiendo la integración con diversas fuentes. Una vez que los datos son extraídos, son transformados, es decir, los datos son limpiados, normalizados y puestos en orden. Estos datos transformados se alojan en un sistema de almacenamiento seguro y transitorio por lo general, como puede ser una plataforma de análisis, un data warehouse o data lake.

Otra etapa es la programación, el pipeline de datos puede programarse para ejecutarse de manera automática y periódica, o según la frecuencia de actualización que requiera la

información. En la fase de monitoreo, el data pipeline se encarga de controlar la calidad de los datos, detectar posibles errores y dar alertas para resolver el conflicto de manera rápida. Por último, se genera la visualización, cuando los datos están en su destino final se convierten en objeto de análisis para las decisiones más acertadas en la organización. (Open Sistemas, 2023)

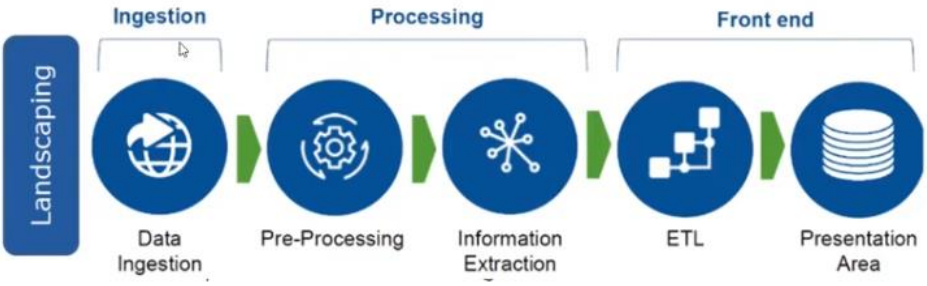


Imagen 1: Modelo Data Pipeline

En el ámbito del manejo de datos, es fundamental entender la diferencia entre dos conceptos: "data pipeline" y "ETL" (extracción, transformación y carga). Un ETL es un tipo específico de data pipeline que se centra en tres etapas cruciales: extracción (obtener datos de diferentes fuentes), transformación (modificar los datos para que sean útiles y estén en el formato adecuado) y carga (almacenarlos en un sistema de destino, como un data warehouse). En este sentido, todos los procesos ETL son data pipelines, pero no todos los data pipelines son ETL, ya que los pipelines pueden operar sin la necesidad de transformar los datos. (Balkenende, 2024) Los pipelines de datos se caracterizan por definir el conjunto de pasos o fases y las tecnologías involucradas en un proceso de movimiento o procesamiento de datos.

Para construir el diseño del data pipeline como indica la imagen*1,* es necesario contar con el modelo entidad relación (ER) de los datos que intervienen a lo largo del proceso. Un modelo ER es una herramienta fundamental para estructurar y organizar los datos, que serán recopilados en el proceso, en un sistema de gestión de bases de datos. Permite representar gráficamente las entidades, sus atributos y las relaciones entre ellas generando un marco visual que sirve de ayuda para entender cómo se relacionan los datos. (School, 2018)

De esta forma, se generarán entidades, que son las tablas donde se guardarán los datos, y los atributos, que son las características que definen cada entidad.

6. Contexto.

6.1 Problemática.

Las organizaciones generan datos continuamente debido a la naturaleza dinámica de sus operaciones y la creciente digitalización de procesos. Cada interacción, transacción y actividad en un entorno empresarial moderno se traduce en la creación de datos. Así mismo, también aumenta la necesidad por parte de las organizaciones de analizar el flujo de información para la toma de decisiones acertadas y para ello es necesario contar con procesos sistematizados que recopilen los datos y los transformen en información valiosa.

La ausencia de un conjunto de procesos y mecanismos para la manipulación de datos en una organización puede generar múltiples problemáticas que afectan tanto la eficiencia operativa como la calidad de la toma de decisiones. Sin un sistema estructurado para mover y procesar datos, las organizaciones enfrentan el riesgo de silos de información, donde los datos se almacenan en diferentes fuentes sin una integración adecuada y que en ocasiones presentan dificultad para poder discernir entre información relevante y no relevante. Esto dificulta el acceso a información coherente y actualizada, lo que puede llevar a decisiones basadas en datos desactualizados o no reales.

De esta manera, un proceso data pipeline garantiza un procesamiento eficiente de los datos, asegurando la calidad y fiabilidad de la información, mitigando posibles errores y decisiones inciertas.

6.2 Perfiles demandantes.

El desarrollo de un pipeline de datos implica la colaboración de profesionales con diversas especializaciones, lo que demanda la conformación de un equipo interdisciplinario. Cada rol del equipo aporta un conjunto específico de competencias esenciales para gestionar las complejidades asociadas al procesamiento y análisis de grandes volúmenes de datos, facilitando así que las organizaciones puedan tomar decisiones fundamentadas en datos precisos y actualizados.

Dentro del equipo interdisciplinario es importante contar con un ingeniero o arquitecto de datos, este rol es el encargado de diseñar, construir y mantener la infraestructura necesaria para el flujo de datos, abarcando desde la ingesta hasta el almacenamiento y procesamiento de la información. Este rol se ocupa de que la infraestructura del sistema de datos sea escalable, segura y eficiente, que los datos puedan fluir sin inconvenientes a lo largo del sistema. En paralelo, el rol de administrador de bases de datos (DBA) gestiona las bases de datos donde se almacenan los datos procesados, garantizando su integridad, seguridad y rendimiento, elementos clave para asegurar que la información esté disponible y operativa cuando se requiera.

Por otra parte, el rol de científico de datos es el encargado del análisis e interpretación de los datos que comúnmente utilizan técnicas estadísticas y algoritmos de machine learning para extraer datos claves y discernir entre datos relevantes y cómo deben ser procesados.

El rol de analista de datos aparece en la última fase del data pipeline al centrarse en la visualización y presentación de los resultados obtenidos, transformando los datos procesados en informes y dashboards accesibles para los stakeholders, de manera tal que facilite la toma de decisiones informadas.

Un perfil clave en el desarrollo de un pipeline de datos es un gestor tecnológico. Este perfil con una formación multidisciplinaria y capacidad de implementar procesos de cambio tecnológico en una organización, se involucra desde la elección de las tecnologías y procesos apropiados a utilizar en el data pipeline, asegurando que todos los componentes del pipeline estén alineados con los objetivos empresariales, hasta cómo llevar adelante su implementación.

De este modo, la colaboración interdisciplinaria entre estos roles permite gestionar de manera efectiva las complejidades del procesamiento y análisis de grandes volúmenes de datos, y asegura que las organizaciones puedan tomar decisiones fundamentadas en datos precisos y actualizados.

7. Factibilidades del proyecto.

La implementación de un proyecto de data pipeline en una PyME de la ciudad de Rafaela, requiere un análisis de viabilidad que comprenda como dimensiones claves el aspecto económico, tecnológico, organizacional y ambiental.

7.1 Factibilidad económica.

En cuanto a la factibilidad económica, el análisis abarca diversos aspectos que son determinantes para la toma de decisiones de inversión. Inicialmente, se requiere una inversión significativa en licencias de software y herramientas, las cuales dependen de las tecnologías seleccionadas en el diseño del data pipeline. Además, es necesario contemplar la inversión en infraestructura tecnológica, ya sea a través de la adquisición de hardware o mediante el uso de servicios en la nube que faciliten el procesamiento y almacenamiento de los datos.

Por otro lado, los costos operativos asociados al proyecto incluyen el mantenimiento continuo del software y hardware, así como el soporte técnico necesario para garantizar su funcionamiento adecuado. A su vez, se deben considerar los gastos vinculados a la contratación de personal capacitado o la capacitación del personal existente involucrados en la implementación del proyecto.

Para hacer frente a la inversión inicial, es relevante explorar líneas de financiamiento que fomenten la adopción de tecnología e innovación en las PyMEs. En este sentido, a nivel local, la ciudad de Rafaela cuenta con la Agencia de Desarrollo e Innovación de Rafaela

(ACDICAR), que no solo ofrece diversas capacitaciones en materia tecnológica, sino que también dispone de líneas de financiamiento específicas para el desarrollo de proyectos tecnológicos en PyMEs.

A pesar de que la inversión inicial puede resultar elevada para las PyMEs locales, el análisis económico sugiere que el proyecto es viable, dado que, a mediano plazo, se anticipan beneficios significativos. Entre estos, se destacan la reducción de costos operativos mediante la automatización del flujo de datos, lo cual disminuye la intervención manual y minimiza los errores humanos. Además, la implementación de un data pipeline incrementará la eficiencia operativa y permitirá identificar nuevas oportunidades de negocio, al proporcionar datos estructurados y depurados que faciliten la toma de decisiones con mayor certeza y la capacidad de adaptarse rápidamente a cambios en el mercado.

7.2 Factibilidad tecnológica.

Desde una perspectiva tecnológica, la implementación de un data pipeline requiere una infraestructura capaz de gestionar la ingesta, procesamiento y almacenamiento de grandes volúmenes de datos. Esta infraestructura puede consistir en una solución local, que implique servidores y almacenamiento físico dentro de la empresa, o en una opción en la nube. La infraestructura local, aunque implica una mayor inversión inicial, ofrece mayor control sobre los datos y elimina la dependencia de terceros. En contraste, la infraestructura en la nube presenta menores costos iniciales, ya que no requiere infraestructura física, pero está sujeta a la calidad de la conectividad y la dependencia de los proveedores de servicios. En este contexto, para una PyME en Rafaela que no cuente con bases de datos transaccionales robustas, la infraestructura en la nube podría ser la opción más conveniente, dado que facilita la flexibilidad y escalabilidad, ajustándose a las necesidades cambiantes del negocio sin grandes inversiones iniciales. Sin embargo, si la empresa ya cuenta con un nivel básico de digitalización y alguna infraestructura de datos, podría ser más beneficioso optimizar y expandir los sistemas existentes en lugar de optar por una nueva infraestructura en la nube.

Además, en las distintas etapas del data pipeline, es necesario utilizar diversas herramientas tecnológicas. Entre ellas se incluyen las herramientas de integración de datos (ETL), las cuales permiten extraer, transformar y cargar los datos; las bases de datos, que pueden ser tanto convencionales como basadas en la nube; y las herramientas de análisis y visualización, que facilitan la creación de reportes y dashboards para la toma de decisiones. La elección de estas herramientas dependerá en gran medida de los requisitos específicos de la empresa, su infraestructura existente y su capacidad para asumir los costos asociados.

Un factor crítico para el funcionamiento eficaz del data pipeline es la conectividad, especialmente si se opta por una solución en la nube. La calidad de la conexión debe ser estable y de alta velocidad para garantizar que los datos se procesen y almacenen de manera eficiente. Además, la seguridad de la información es esencial. Para proteger los datos y prevenir accesos no autorizados, se deben implementar medidas de seguridad como el cifrado de datos y la

autenticación robusta, lo que garantizará la privacidad y la integridad de la información almacenada.

En consecuencia, la elección de las herramientas y la infraestructura a utilizar en el data pipeline dependerá de varios factores, como el presupuesto disponible, la infraestructura tecnológica preexistente y los requisitos particulares del negocio. Una evaluación cuidadosa de estas variables permitirá a las PyMEs seleccionar las soluciones que más se adecuen al entorno, además de, disminuir la brecha tecnológica que las separa de las grandes empresas que cuentan con una cultura de datos.

7.3 Factibilidad ambiental.

Desde el punto de vista ambiental, la implementación de un data pipeline presenta beneficios significativos, dado que, al automatizar los procesos, contribuye a la reducción del uso innecesario de papel y otros materiales físicos, promoviendo una mayor eficiencia en el manejo de la información. Este tipo de soluciones tecnológicas no solo optimiza la gestión de datos, sino que también puede generar un impacto positivo en la sostenibilidad de las operaciones de la empresa.

No obstante, la adopción de un data pipeline también conlleva desafíos ambientales que deben ser considerados. Uno de los principales retos es la necesidad de alimentar los centros de datos y las infraestructuras tecnológicas con fuentes de energía, lo que podría aumentar la huella de carbono a lo largo del tiempo. Para mitigar este impacto, sería ideal que las empresas implementaran fuentes de energía renovable para abastecer los sistemas involucrados en el procesamiento y almacenamiento de datos. Sin embargo, esta opción puede resultar inviable para la mayoría de las PyMEs del ámbito local, debido a las limitaciones de recursos y la falta de infraestructura.

Por lo tanto, aunque el data pipeline puede contribuir a la sostenibilidad al reducir el consumo de materiales físicos, es importante evaluar las posibilidades de mitigar el impacto ambiental relacionado con el consumo energético. En las primeras fases del proyecto, las PyMEs pueden enfrentar dificultades para acceder a fuentes de energía renovable o no contar con la infraestructura necesaria para implementarlas, lo que podría hacer que la adopción de una solución completamente sostenible sea más difícil.

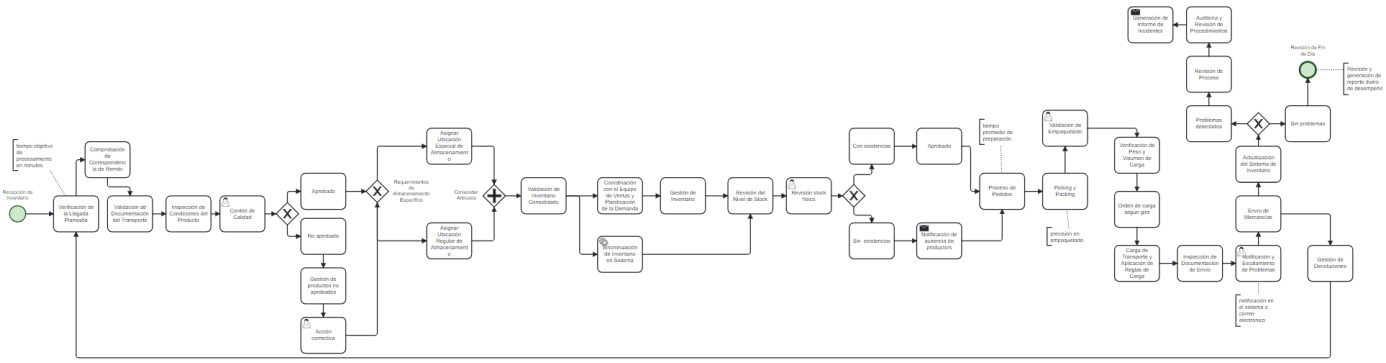
8. Aplicación de Práctica Profesionalizante (PPS)

La práctica profesionalizante se llevará a cabo bajo la supervisión de la profesora Mariel López, correspondiente a la cátedra “Práctica Profesionalizante Supervisada II” y con la colaboración de los estudiantes Guillermo Sastre e Ignacio Joaquín de la Licenciatura en Gestión de la Tecnología, quienes proporcionaron el modelo de negocio de la cadena de suministro de Food Solutions S.A., una empresa ubicada en la ciudad de Rafaela, especializada en el desarrollo de ingredientes personalizados para la industria alimenticia. Para el diseño del

data pipeline, se emplea el modelo de negocio elaborado por los compañeros de la cátedra, omitiendo, no obstante, los datos específicos de la compañía.

8.1 Modelo y Notación de Procesos de Negocio (BPMN)

A continuación, se describe el modelo entidad relación de los datos que participan en cada uno de los procesos del BPMN de la empresa Food Solutions S.A., con el fin de poder desarrollar posteriormente el diseño del data pipeline.



Pedido de productos: La empresa gestiona los pedidos de productos con cada proveedor, y genera una nota de pedido en una planilla excel donde carga la siguiente información: proveedor, productos solicitados, cantidades solicitadas, fecha estimada de entrega, presupuesto.

Recepción de productos: La empresa recibe productos en forma de aditivos alimenticios en polvo, de tipo perecedero, provenientes de diversos proveedores. La mercadería puede llegar con un remito físico en algunos casos, o bien con un remito digital enviado por correo electrónico en otros. La carga de los productos se realiza en una planilla de cálculo Excel, donde se registra la siguiente información: fecha, código del producto, descripción, tipo de producto, cantidad, lote, fecha de vencimiento, proveedor, medio de transporte e información relacionada con el transporte. Actualmente, la compañía no dispone de un módulo de gestión de stock implementado, no obstante, los remitos recibidos son ingresados de igual modo en un sistema ERP (Enterprise Resource Planning), en tanto el inventario se maneja todo por medio de planillas de cálculo.

Inspecciones y control de calidad de productos: La verificación de los productos actualmente se hace manual, un colaborador se encarga de controlar que la cantidad enviada de productos coincida con la solicitada cotejando la recepción de productos, el remito enviado y el pedido realizado por la empresa.

En caso de encontrar fallas en el producto o alguna inconsistencia en cuanto al pedido solicitado, se realiza una acción correctiva como devolver solo la mercadería afectada o dar de baja todo el lote, acción seguida se descuenta del stock registrado en excel.

Almacenamiento: La ubicación del almacenamiento físico de los productos está determinado por parámetros preestablecidos como el tipo de producto, método FIFO o LIFO según lo determine la categorización del producto y planificación de la producción.

Sincronización del stock: Una vez realizado el control de calidad, se consolida el stock en el sistema ERP que utiliza la empresa para luego ser consultado por el módulo de gestión de ventas, actualizando la existencia del stock al concretarse una venta así como también, la ubicación del producto en el almacén.

8.2 Modelo Entidad - Relación (ER)

El siguiente ER involucra los procesos desde que se genera el pedido de productos a los respectivos proveedores, control de calidad, carga del remito y actualización del stock disponible de la compañía.

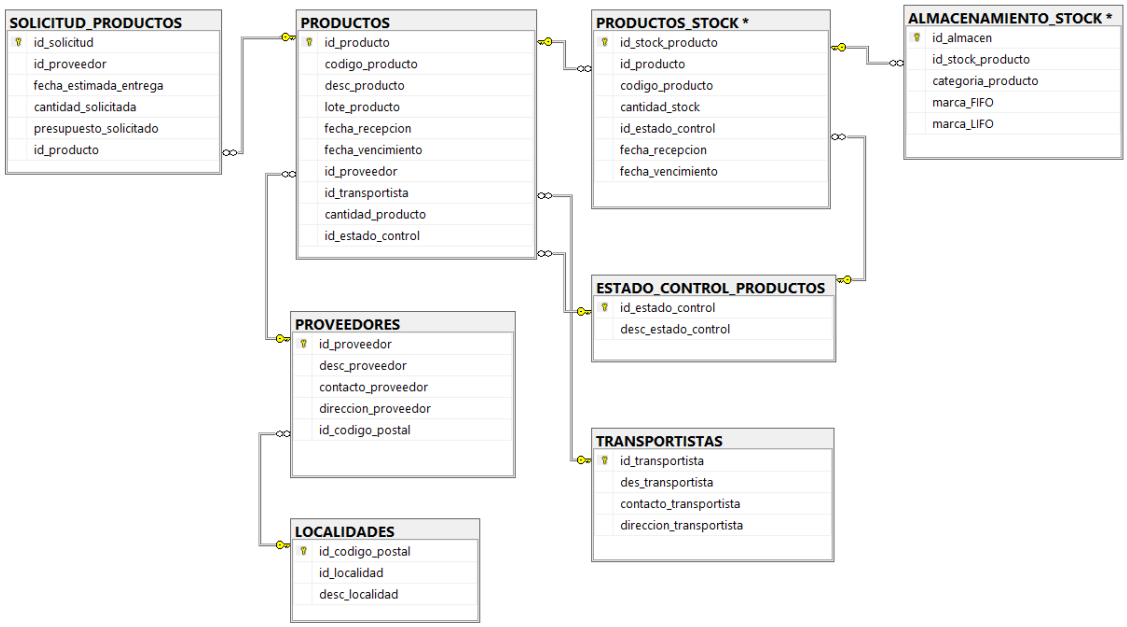


Imagen 2: Modelo Entidad Relación.

8.3 Diseño Data Pipeline.

Un pipeline de datos se conforma de una serie de etapas interconectadas que facilitan el flujo y el procesamiento de los datos a través de la extracción, transformación, programación, monitoreo, análisis y visualización de los datos. Para diseñar el pipeline de datos, se clasifican las etapas como se menciona en la imagen 1.

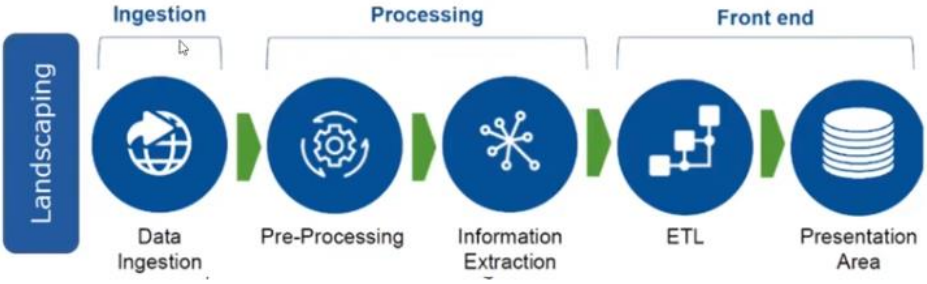


Imagen 1: Modelo Data Pipeline

En la fase de ingestión de datos se recolectan los datos de las distintas fuentes involucradas en el proceso, en este caso los inputs serán archivos de hojas de cálculos y bases de datos del sistema de gestión ERP que utiliza la compañía. Situándonos en el proceso de compra de la cadena de suministro, los pedidos realizados a los diferentes proveedores se cargan en una hoja de cálculo, de igual manera, en una segunda hoja de cálculo se realiza la carga de la mercadería que se recepciona previamente de ser controlada por un colaborador. Por otro lado, de la base de datos del sistema ERP con el que opera la empresa se extraen datos provenientes del maestro de proveedores, productos, localidades, módulo de gestión de ventas, ubicación del almacén y del stock consolidado.

Al tratarse de datos estructurados, es decir, que se disponen en forma de filas y columnas en el origen de la información, en este caso, archivos de excel y tablas alojadas en una base de una base de datos, los datos extraídos en la primera fase se almacenan en un data warehouse dando lugar a la siguiente fase, el procesamiento y transformación de los datos. En el data warehouse se almacenarán las entidades indicadas en el modelo ER (Imagen 2), en cada entidad se almacenarán los datos que fueron extraídos, previamente saneados y normalizados por un proceso ETL.

De esta manera, la entidad denominada SOLICITUD_PRODUCTOS contendrá datos provenientes del pedido de mercadería que se acordó con los respectivos proveedores. La entidad PRODUCTOS almacenará el detalle de los productos recepcionados y la validación de su control de calidad.

La herramienta ETL, componente del data pipeline, por medio de algoritmos se encargará de la lectura del atributo que identifica la validación del control de stock, si esta es positiva, el stock consolidado y disponible a utilizar se registrará en la entidad PRODUCTOS_STOCK quedando listo para operar con los restantes módulos del sistema ERP. En cambio, aquellos productos en la cual su validación fue negativa y requieren tomar acciones por fallas se almacenarán en REVISION_PRODUCTOS.

La asignación de cada producto en el almacén se determinará por medio de algoritmos, identificando los parámetros preestablecidos en el maestro de productos. La fecha de recepción

y vencimiento de cada producto, teniendo en cuenta su categorización, determinará el método bajo el cual se almacenará.

La herramienta ETL permite la integración de tecnologías como inteligencia artificial (IA) y aprendizaje automático (ML), que, sumado a las operaciones matemáticas y estadísticas, es posible programar tareas y generar indicadores para su seguimiento. En la aplicación de este trabajo se crearán algunos indicadores estratégicos como ser; estado del almacén, rotación de productos, desviación respecto al presupuesto, tiempo de permanencia en almacén, tiempo de espera por el proveedor, entre otros. En la fase de visualización, los datos se representan por medio de indicadores y atributos que son disponibilizados en una herramienta analítica para la construcción de reportes y dashboard para su posterior análisis y tomas de decisiones. La elección de la herramienta analítica se determina de acuerdo a la infraestructura que posee la organización, los recursos físicos y financieros.

Los indicadores que se establezcan para su construcción se transforman en KPI estratégicos que facilitan el proceso de toma de decisiones. Por ejemplo, el conocimiento sobre la rotación del inventario permite a las organizaciones modificar sus estrategias de adquisición y producción, lo que conduce a una optimización de sus recursos. En este sentido, si un producto presenta una baja rotación, podría ser necesario reevaluar su precio o la estrategia de marketing asociada. Contar con un KPI que indique el tiempo total necesario para completar un proceso de producción desde la recepción del pedido hasta la entrega del producto, permite organizar eficientemente la planificación de compras de materia prima y ventas de productos.

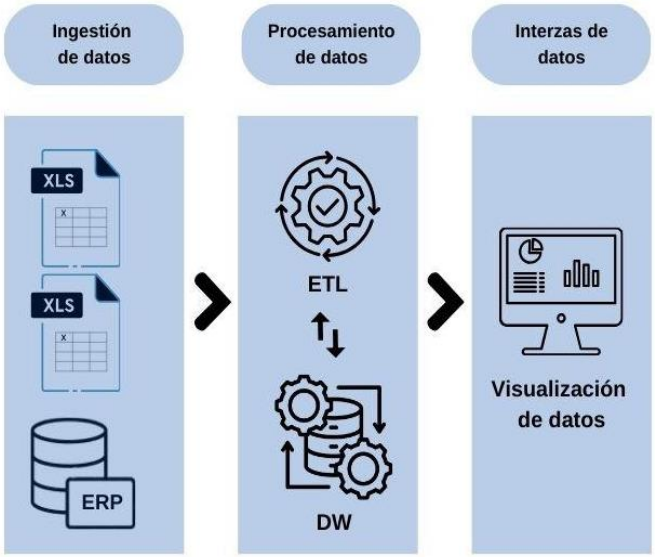


Imagen 3: Diseño data pipeline aplicado en la cadena de suministro de Food Solutions S.A

9 Conclusiones.

El desarrollo del presente trabajo subraya la relevancia del diseño e implementación de pipelines de datos como solución estratégica para optimizar la gestión y procesamiento de información en organizaciones, particularmente en PyMEs situadas en la localidad de Rafaela. A través del empleo de herramientas ETL (extracción, transformación y carga), el diseño del data pipeline propone una arquitectura que facilita la integración de múltiples fuentes de datos, asegurando calidad, consistencia y disponibilidad de la información para análisis avanzados.

El desarrollo del pipeline de datos no solo responde a la necesidad de centralizar y normalizar datos heterogéneos, sino que también incorpora herramientas analíticas que permiten la generación de indicadores claves para las organizaciones (KPIs). Estos indicadores, visualizados mediante dashboards y o reportes, contribuyen a una toma de decisiones más acertadas, basadas en datos y oportuna, mejorando la eficiencia operativa y la capacidad de respuesta de la organización ante los cambios del mercado.

Finalmente, la propuesta destaca la importancia de considerar factores económicos, tecnológicos y ambientales para garantizar la viabilidad del proyecto. La implementación de esta infraestructura tecnológica representa una oportunidad significativa para que las PyMEs superen limitaciones tradicionales de gestión de datos logrando una ventaja competitiva sostenible en un entorno empresarial dinámico, además de reducir la brecha que las separa de las grandes empresas.

10 Bibliografía.

Argomedo, J. G. (2022). RSM. Obtenido de https://www.rsm.global/chile/es

Balkenende, M. (7 de Octubre de 2024). Matillion. Obtenido de https://www.matillion.com/blog/etl-vs-data-pipeline

Datademia. (2023). Datademia.

DataSpurs. (8 de Febrero de 2023). Obtenido de https://dataspurs.com/blog/beneficios-de-implantar-un-pipeline-de-datos-en-tu-empresa/

Open Sistemas. (Agosto de 2023).

School, E. B. (2018). Modelo entidad relación: descripción y aplicaciones.