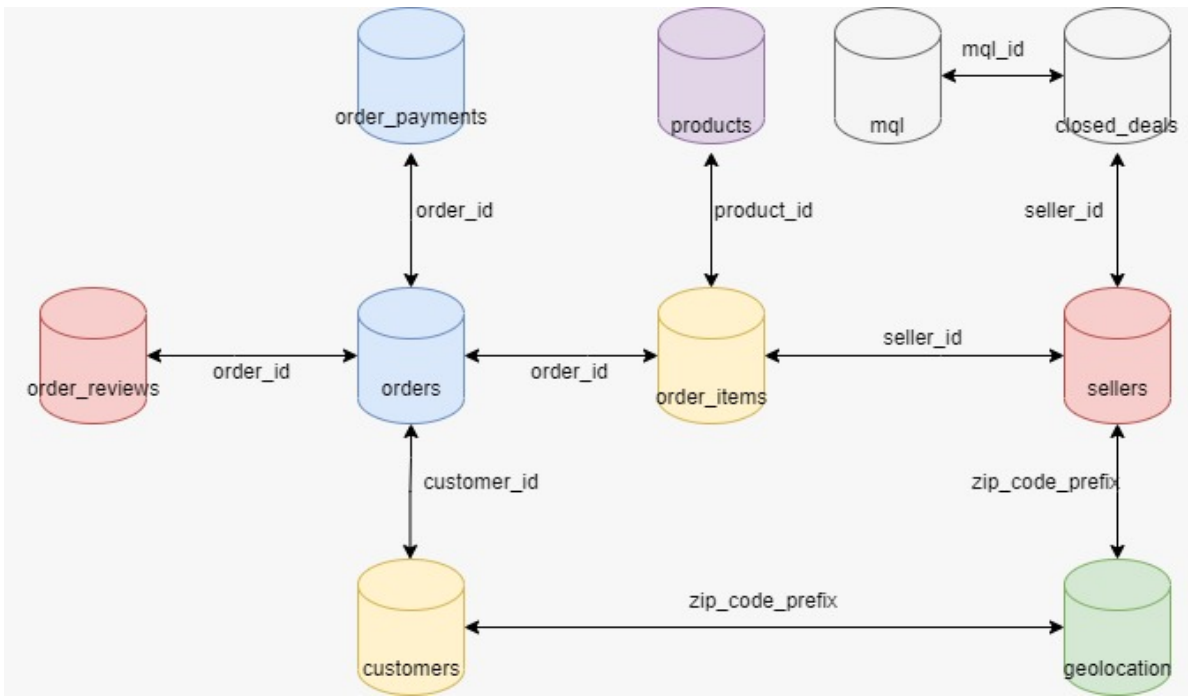


DOCUMENTACIÓN SEMANA 2:

DIAGRAMA DE ENTIDAD-RELACIÓN DETALLADO:



DICCIONARIO DE DATOS:

TABLAS			
closed_deals	Columnas	Tipo de dato	Referencia
	mql_id (FK)	character varying	Potencial Vendedor en Olist, obtiene este id en su primer contacto con alguna de las 'landing page' de Olist.
	seller_id (FK)	character varying(256)	Id del vendedor en Olist.
	sr_id	character varying(256)	Id de 'sales representative' de List.
	sdr_id	character varying(256)	Id de 'sales representative' de Olist.
	won_date	Timestamp without time zone	Fecha de cierre de contrato. Pasa de mql a seller.
	business_segment	character varying(256)	Segmento al cual pertenece o en el cual publicará sus productos.
	lead_type	character varying(256)	Tipo/perfil del vendedor/seller.
	business_type	character varying(256)	Tipo de negocio, del lead (vendedor): fabricante o revendedor (retail)
customer			
	customer_id (PK)	character varying(256)	Identificación de cliente, compra mediante Olist.
	customer_zip_code_prefix	Integer	Prefijo el código postal de Brasil.
	customer_city	character varying(256)	Ciudad del cliente.
	customer_state	character varying(256)	Siglas del estado del cliente.

geolocation			
geolocation_zip_code_prefix	Integer		Prefijo código postal para geolocalización.
geolocation_lat	Numeric(18,6)		Valor numerico de latitud.
geolocation_lng	Numeric(18,6)		Valor numerico de longitud.
geolocation_city	character varying(256)		Ciudad.
geolocation_state	character varying(256)		Estado.

mql			
mql_id (PK)	character varying(256)		Potencial Vendedor en Olist, obtiene este id en su primer contacto con alguna de las 'landing page' de Olist.
first_contact_date	Timestamp without time zone		Fecha de Primer contacto, con una 'landing page' de Olist.
landing_page_id	character varying(256)		Página en la cual aterriza el 'mql', p/ej al hacer click en un add.
origin	character varying(256)		Tipo de llegada a una landing page. Ej: social, email, paid search.

order_item			
order_id (FK)	character varying(256)		Id de orden compra.
order_item_id	character varying(256)		Cantidad de unidades en la orden.
product_id (FK)	character varying(256)		Id de producto.
seller_id (FK)	character varying(256)		Id del vendedor.
shipping_limit_date			Fecha límite de envío.
price	numeric(18,0)		Precio del producto en Reales(\$RS).
freight_value	numeric(18,0)		Costo del envío en Reales(\$RS)
freight_value_decimal(10,2)	decimal(10,2)		Costo del envío en Reales(\$RS)
price_decimal(10,2)	decimal(10,2)		Precio del producto en Reales(\$RS).

order_payments			
order_id (FK)	character varying(256)		Id de la compra,
payment_sequential	Bigint		Pago secuencial.
payment_type	character varying(256)		Tipo de pago.
payment_installments	Bigint		Pago de cuotas.
payment_value	Numeric(18,0)		Valor del pago, incluye envío.
payment_value_decimal(10,2)	Numeric(10,2)		Valor del pago, incluye envío.

order_reviews			
	review_id (PK)	character varying(256)	Id de las reviews de los clientes con respecto a sus compras.
	order_id (FK)	character varying(256)	Id de las ordenes de compra.
	review_score	Bigint	Puntuación dada por el cliente (del 1 al 5, siendo 5 el mejor puntaje).
	review_comment_title	character varying(512)	Título de la review.
	review_comment_message	character varying(1024)	Mensaje de la review.
	review_creation_date	Timestamp without time zone	Fecha de creación de la review.

orders			
	order_id (PK)	character varying(256)	Id de la orden de compra.
	customer_id (FK)	character varying(256)	Id de cliente.
	order_status	character varying(256)	Estado de la orden de compra.
	order_purchase_timestamp	Timestamp without time zone	Momento de la compra.
	order_approved_at	Timestamp without time zone	Momento de aprobación de la compra.
	order_delivered_carrier_date	Timestamp without time zone	Momento en que comienza el transporte por el 'carrier'.
	order_delivered_customer_date	Timestamp without time zone	Momento en que la compra llega al cliente (fin de recorrido).
	order_estimated_customer_date	Timestamp without time zone	Fecha estimada de llegada del producto al cliente.

products			
	product_id (PK)	character varying(256)	Id de producto.
	product_category_name	character varying(256)	Nombre de la categoría del producto.
	product_weight	Numeric(18,6)	Refiere al peso del producto en gramos.
	product_length_cm	Numeric(18,6)	Refiere al largo del producto en cm.
	product_height_cm	Numeric(18,6)	Refiere al alto del producto en cm.
	product_width_cm	Integer	Refiere al ancho del producto en cm.
	product_category_name_english	character varying(256)	Nombre de la categoría del producto, en Inglés.

sellers			
	seller_id (PK)	character varying(256)	Id del vendedor.
	seller_zip_code_prefix	Integer	Prefijo de código postal del vendedor.
	seller_city	character varying(256)	Ciudad del vendedor.
	seller_state	character varying(256)	Nomenclatura del Estado/Provincia del vendedor.

## STACK TECNOLÓGICO Y FLUJO DE TRABAJO:

**Amazon S3:** Se encarga del almacenamiento general, genera buckets que son contenedores para objetos almacenados en esta herramienta. Aquí en uno de esos buckets se colocan los archivos raw que son entregados al grupo de trabajo, en otro bucket se colocan los datos ya procesados con Glue y finalmente se crea el bucket donde los datos son almacenados para luego ser tomados por Redshift.

Elegimos esta herramienta por contar con una gran variedad de opciones para la transferencia de datos desde y hacia sus servidores, además de ser sencilla la misma. Te permite configurar todos los archivos, datos y documentos almacenados, y controlar los accesos a ellos. Controlando las versiones, te permite preservar, restaurar y recuperar fácilmente todas las versiones ante alguna acción involuntaria por parte de los usuarios que haya generado problemas. Además, sus principales características son la gran escalabilidad, seguridad, alto rendimiento y disponibilidad de datos. Por lo tanto nos pareció una muy buena opción.

**Amazon Glue con Python:** Su principal uso fue para el ETL (Extraction, Transformation, and Load). Su flexibilidad nos sirvió para traer las tablas raw que teníamos en un bucket de S3, realizar el proceso de ETL a las distintas tablas y transferirlas luego a otro bucket para que Redshift pueda acceder a los datos. Proporciona múltiples conectores, tanto para servicios propios(S3 o Redshift) como para sistemas externos como MongoDB o Apache Kafka. Brinda un IDE de Spark, que nos permite ver una interfaz gráfica y el script de nuestras tareas de desarrollo. Cuenta con Crawlers que sirven para escanear e identificar la información con la que contamos.

**Amazon Redshift:** Bases de datos. Utilidad para crear y administrar Data Warehouse y Data Lakes. Tiene la ventaja de estar concebida para análisis de datos y big data, con un modelo OLAP (Online Analytical Processing) que está menos enfocado en la transacción y más enfocado en las consultas para análisis.

**Amazon EC2 con python:** Crea una instancia de cómputo y su uso es por hora. Se utilizará para computó especialmente en Machine Learning.

**Amazon QuickSight:** Herramienta de visualización y Business Intelligence.

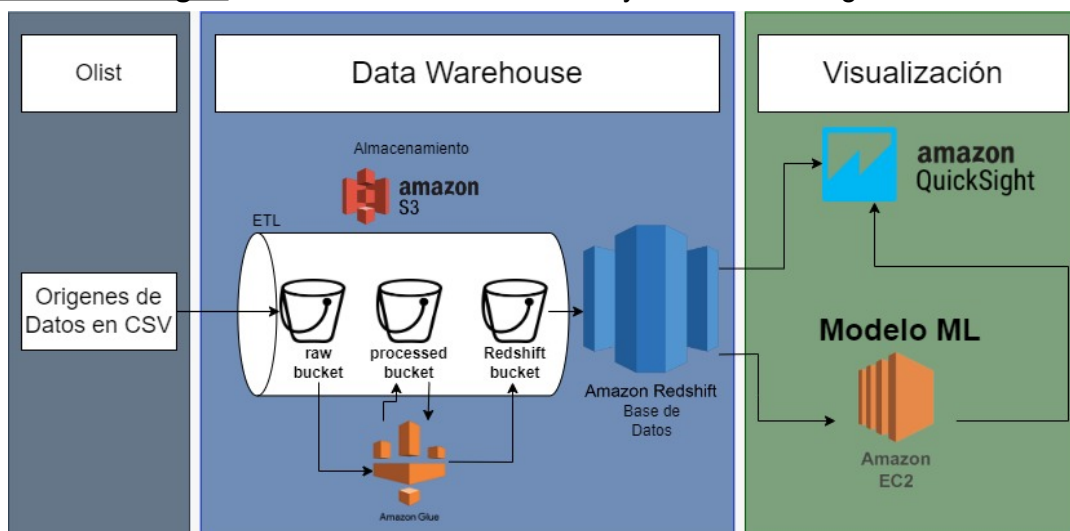


DIAGRAMA DE GANTT:

