Data ingestion:

1. **Network traffic Logs** is stored in AWS S3 as a data lake that stores raw web server logs. The logs need to be processed and relevant information extracted from them with the help of AWS Lambda Function. Resultant data will then be written to Amazon Kinesis Data Firehose by Lambda function which will upload it to an Amazon S3 bucket where it will sit for staging.

2. **Transactional data:** use AWS Database Migration Service (DMS) to move customer orders and payment data from relational database into an Amazon S3 bucket that can stage it.

3. **Product Catalog:** Use AWS Glue read CSV files full of product specifics, stage them on an AWS S3 bucket.

4. **Marketing campaign Data:** real-time marketing campaign interactions are captured using Amazon Kinesis Data Streams over messaging queue. For this, we may need an AWS lambda function that shall process and save the output data into another amazon s3 storage area for staging.

Data Transformation:

Implement an Apache Spark pipeline using AWS Glue so as to modify and connect various data sources. This could be done through a sequence of stages:

1.Treat missing values and delete duplicated records

2. Normalize formats by doing necessary calculations or aggregations

3. Merge data with diverse sources

4. Develop a single view of the information

Data Modelling:

Prepare a data model for the transformed data in an Amazon Redshift Data Warehouse based on star schema. The dimensions and facts that will make up the data model are:

Dimension:

1. Date
2. Product

3. Customer
4. Marketing Campaign

Fact:

1.Network Traffic

2. Transactions

3.Product Sales

Data Loading:

Use AWS Glue to load transformed data into Amazon Redshift Data Warehouse. Apply batch loading strategy for scheduled data loads.

Orchestration and Automation:

Utilize Apache Airflow for automating the complete data pipeline, which involves scheduling tasks for data ingestion, transformation, and loading to run at consistent intervals, along with monitoring the pipeline for errors and informing relevant parties, and implementing retry mechanisms and error handling for failed tasks.

Sample Data

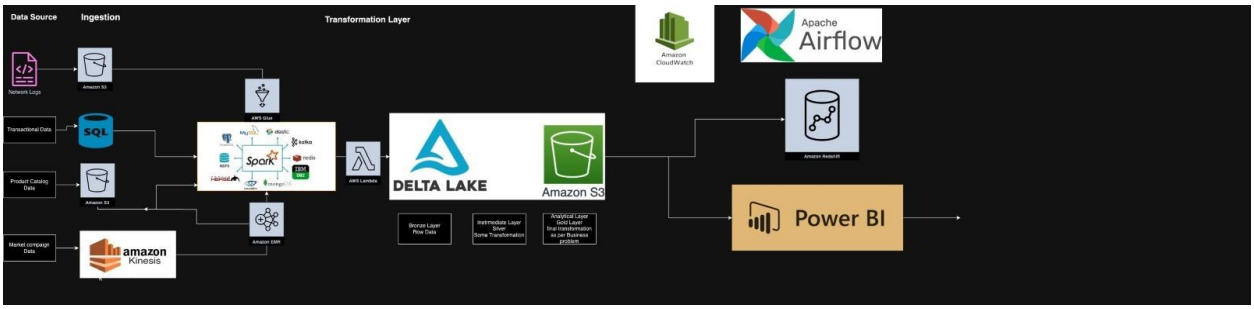
Generate sample data for each of the data sources to demonstrate the data pipeline:

- Network Traffic Logs: 100,000 rows of sample data with columns for user ID, page views, and click events
- Transactional Data: 10,000 rows of sample data with columns for customer ID, order date, and payment amount
- Product Catalog: 1,000 rows of sample data with columns for product ID, description, and price
- Marketing Campaign Data: 10,000 rows of sample data with columns for campaign ID, email opens, and click-throughs

Power BI Integration

Integrate the Amazon Redshift data warehouse with Power BI to enable advanced analytics and business intelligence capabilities. Create dashboards and reports to demonstrate the insights that can be gained from the data, such as:

- Customer behavior and preferences
- Product sales and revenue
- Marketing campaign effectiveness
- Network traffic and usage patterns



Achitecture Diagram