**FashionMart EDA Blueprint (Task 1.1)**

**1. Data Sources**

Comprehensive coverage of FashionMart’s operations through:

* **POS Systems**: Capturing sales transactions, returns, discounts, and cashier logs.
* **Customer Transactions**: Includes online/offline purchase history, shopping carts, and loyalty program activities.
* **Supplier Databases**: Supplier product catalogs, shipment schedules, invoices, and returns.
* **Third-Party Data Feeds**:
  + Market trend analysis
  + Social sentiment data
  + Weather forecasts affecting sales
  + Competitor pricing (web scraping or APIs)

**2. Data Ingestion**

Efficient ETL/ELT mechanisms:

* **ETL Tools**: Talend, dbt (for transformation), Apache NiFi.
* **Streaming Ingestion**: Apache Kafka, AWS Kinesis for real-time feeds (e.g., live POS data, website activity).
* **Batch Ingestion**: Scheduled jobs to pull supplier and CRM data.
* **API Gateways**: AWS API Gateway, Apigee for standardized communication between internal and external systems.

**3. Data Storage**

Layered to balance real-time access and analytical depth:

* **Cloud Data Warehouse**: Snowflake or Google BigQuery for analytical workloads.
* **Operational Databases**: PostgreSQL/MySQL for transactional data.
* **Real-Time Store**: Amazon DynamoDB, Redis for session data, inventory states.
* **Object Storage**: AWS S3 for unstructured assets (images, logs, customer uploads).
* **CRM & PIM**:
  + CRM: Salesforce
  + PIM: Akeneo

**4. Data Processing**

Ensuring quality and readiness for analytics:

* **ETL/ELT Pipelines**: Built using dbt and Apache Spark.
* **Data Wrangling & Cleansing**: Python (pandas), Spark jobs.
* **Data Orchestration**: Apache Airflow for scheduling, monitoring, and retry logic.

**5. Analytics & BI**

Driving decision-making through:

* **BI Tools**: Power BI, Looker, Tableau for dashboards and reports.
* **Real-Time Analytics**: Flink, AWS Kinesis Analytics for live traffic insights.
* **Predictive Modeling**: Using AWS SageMaker for demand forecasting and customer behavior prediction.
* **Customer Segmentation**: Google Analytics, Segment, or custom ML pipelines.

**6. Security**

Robust protocols across data lifecycle:

* **Encryption**: AES-256 at rest and TLS 1.2+ in transit.
* **Access Control**:
  + Role-Based Access Control (RBAC)
  + OAuth2 and SSO (Okta, Azure AD)
* **Monitoring & Compliance**:
  + Tools: Splunk, AWS Security Hub
  + GDPR & CCPA compliance with platforms like OneTrust