Exemplar: Create an EDA Blueprint

This document presents an exemplar solution to the lab, 'Create an EDA Blueprint'. Please remember that this is just one point of view and not a definitive solution.

Solution:

Task 1: Understand business requirements: Identify the specific data requirements for optimizing supply chain operations

Step 1: Compile a list of all critical data sources

Key data resources
Supplier database
Order management systems
Logistics data
Inventory records
ERP systems

Step 2: Define practical applications for the data

Critical use cases
Inventory forecasting
Supplier performance analysis
Transportation optimization
Demand and supply matching

Step 3: Outline the data compliance standards and regulations that must be followed

Data compliance requirements	Description
GDPR	• Apply if the company handles the personal data of EU citizens • Ensure data minimization, lawful processing, and user consent • Provide mechanisms for data access, rectification, and deletion
Environmental, Social, and Governance (ESG) Regulations	Monitor and report sustainability data in supply chain operations

California Consumer Privacy Act (CCPA)	• Protects data privacy of California residents • Enables users to opt out of data sales and request data access or deletion
Health Insurance Portability and Accountability Act (HIPAA)	• Relevant if healthcare-related goods are shipped • Safeguard sensitive health information
ISO 27001	Implements an Information Security Management System (ISMS) to protect sensitive data

Task 2: Identify and define the core components of data architecture

Step 1: Data sources: Identify and categorize all internal and external data sources

Data source type	**Examples**
Internal	ERP systems, warehouse management systems, production logs
External	Supplier APIs, weather data, market trends

Step 2: Data ingestion: Establish the data ingestion methods

Ingestion method	**Description**	**Examples**	**Tools**
Batch ingestion	Processes data at scheduled intervals	Daily supplier reports	Apache Kafka, AWS Kinesis, Talend
Real-time ingestion	Captures and processes data dynamically as it arrives	IoT devices in transportation	Apache Kafka, AWS Kinesis, Talend

Step 3: Data storage: Define the data storage types

Storage type	**Purpose**	**Examples**
Data Lake	Stores raw, unprocessed data	Hadoop, AWS S3
Data Warehouse	Stores structured, analytical data	Snowflake, Google BigQuery
Archive	Stores historical data for long-term retention	Microsoft Azure Archive Storage, IBM Cloud Object Storage

Step 4: Data processing: Define the data processing methods

Processing method	Description	Examples
ETL/ELT pipelines	Transform raw data for further analysis	Apache Spark, Talend, AWS Glue, Microsoft Azure Data Factory
Stream processing	Process data in real time for immediate use	Apache Flink

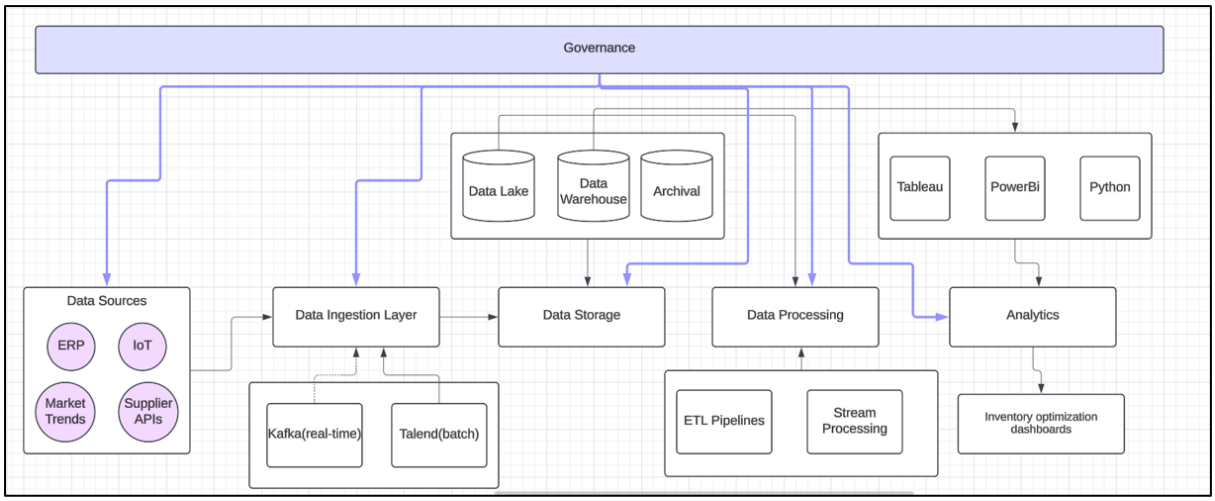
Step 5: Analytics: Identify and report business intelligence capabilities

Analytics type	Description	Examples/Use cases	Tools
Dashboards	Visualize key metrics such as inventory levels, supplier performance, and cost analysis	Inventory levels, supplier performance, cost analysis	Tableau, Power BI
Predictive analytics	Use historical data to forecast future demand	Demand forecasting	Python (for advanced analytics)

Step 6: Data governance: Define the data governance tasks

Governance task	Description	Examples/Tools
Data ownership & stewardship	Define roles and responsibilities for data management	Collibra, Alation
Access policies	Establish guidelines for who can access the data	Okta, Immuta
Compliance & audit	Ensure data usage complies with regulations and maintain audit trails	AWS Audit Manager, IBM OpenPages with Watson
Governance & lineage	Use tools to track data flow and ensure data integrity	Apache Atlas

Task 3: Create the EDA blueprint using Lucidchart



Note: A supply chain company can develop an Enterprise Data Architecture (EDA) blueprint to streamline operations and enhance decision-making by identifying key data sources, defining data flow processes, and integrating real-time tracking and analytics. The company establishes a centralized data lake for seamless storage and retrieval by leveraging cloud-based solutions. The blueprint incorporates data governance policies to ensure security and compliance while utilizing advanced analytics tools to forecast demand, optimize routes, and reduce costs.