Case Study – **Rough Outline**

Data Engineering on Supply Chain Management

Week 1: Understanding Requirements and Data Exploration

Kickoff Meeting: Meet with stakeholders to understand project goals, scope, and requirements.

Data Discovery: Identify relevant data sources such as ERP systems, inventory databases, sales records, etc.

Data Profiling: Explore the structure, quality, and volume of available data. Document findings and potential data quality issues.

Requirement Gathering: Work closely with business analysts to define specific data needs for supply chain management analytics.

Week 2: Data Acquisition and Integration

Data Extraction: Develop scripts or workflows to extract data from various sources.

Data Transformation: Cleanse and transform raw data into a format suitable for analysis.

Data Loading: Load transformed data into a centralized data warehouse or data lake.

Integration Testing: Validate data integration processes and ensure data accuracy and completeness.

Week 3: Building Data Models and Pipelines

Data Modeling: Design dimensional models for supply chain data, considering various dimensions such as time, product, location, etc.

ETL Development: Develop Extract-Transform-Load (ETL) pipelines to populate data models.

Workflow Orchestration: Implement workflow management tools (e.g., Apache Airflow) to schedule and monitor ETL processes.

Pipeline Optimization: Optimize data pipelines for performance and scalability.

Week 4: Implementing Analytics and Reporting

Analytics Development: Develop SQL queries, data processing scripts, or analytical models to derive insights from supply chain data.

Dashboard Development: Design and build interactive dashboards to visualize key supply chain metrics and KPIs.

Report Automation: Automate report generation processes to deliver timely insights to stakeholders.

User Acceptance Testing: Engage stakeholders in testing analytics and reporting functionalities to ensure alignment with business needs.

Week 5: Performance Tuning and Optimization

Query Optimization: Identify and optimize slow-performing queries or data processing steps.

Infrastructure Scaling: Evaluate infrastructure requirements and scale resources as needed to accommodate growing data volumes.

Data Governance: Implement data governance policies and procedures to ensure data security, privacy, and compliance.

Documentation: Document data pipelines, data models, and analytics processes for future reference and knowledge sharing.

Week 6: Deployment and Handover

Deployment Planning: Plan for the deployment of data engineering solutions into production environments.

User Training: Conduct training sessions for end-users on accessing and utilizing data analytics tools and reports.

Knowledge Transfer: Transfer knowledge to the internal team or relevant stakeholders for ongoing support and maintenance.

Post-Deployment Support: Provide post-deployment support to address any issues or concerns that arise during initial usage.