Data Engineering Basics for Everyone

Provider: IBM (Coursera)
Completion Date: May 22, 2025

Overview

This course introduced the fundamentals of data engineering and its role in the modern data ecosystem. It covered how data engineers design and maintain data pipelines, repositories, and architectures that ensure reliable, secure data for analysis and decision-making.

Key Topics Covered

- Roles in the modern data ecosystem
- Data types and file formats (CSV, JSON, XML, XLS, PDF)
- Sources: databases, APIs, IoT, and web scraping
- Repositories: relational databases, NoSQL, data warehouses, data lakes
- Data pipelines and ETL/ELT processes
- Big Data tools: Hadoop, Hive, Spark
- Lifecycle: architecture, wrangling, querying, performance tuning
- Governance, compliance, and privacy considerations

Practical Applications

- Designing pipelines to transform and move data
- Managing repositories for scalable, secure access
- Using SQL and scripting for data preparation
- Applying governance and compliance standards
- Querying cloud-based databases in lab settings

Personal Reflection

This course gave me a strong foundation in data engineering principles. I now understand how pipelines, repositories, and governance deliver quality data for analytics. These skills complement my cybersecurity focus by showing how secure data architectures support compliance and risk management.