Maslak Mahallesi, Maslak Meydan Sk. No:5. Spring Giz Plaza. Maslak/İstanbul
+90 212 282 7700 - info@bilginc.com

Data Pipelines: Workflow and Dataflow for Todays Data Architectures

Learn via: **Classroom**

Duration: **1 Day**

Overview

Data-driven is the modern mantra of business management, but enabling a data-driven organization is complex and challenging. Abundant data sources and multiple use cases result in many data pipelines—maybe as many as one for each use case. Capabilities to find the right data, manage data flow and workflow, and deliver the right data in the right forms for analysis are essential for all organizations that seek to become data-driven.

Multiple and complex data pipelines can quickly become chaotic under pressure from agile development, democratization, self-service, and organizational pockets of analytics. The resulting difficulty in governance and uncertainty of data usage are only the beginning of the troubles. Therefore, data pipeline management must ensure that data analysis results are traceable, reproducible, and of production strength, whether enterprise-level or self-service. Robust pipeline management works across a variety of platforms from relational to Hadoop, and recognizes today's bidirectional data flows where any data store may function in both source and target roles.

Who Should Attend

Analytics architects, BI architects, data warehouse architects, data architects, and anyone in an architect role that intersects with data; data engineers who define, design, and develop data warehouses, data lakes, operational data stores, data sandboxes, master data hubs, or other enterprise data stores; data integration and preparation professionals who define, design, and develop the processes that move data through pathways from sources to consumers.

What You Will Learn

- The challenges and complexities of modern data pipelines
- Why data flow and workflow are critical parts of—and how they fit into—your analytics architecture
- How to define and design data pipelines
- The roles and functions of metadata in pipeline management
- The important relationships between pipeline management and data governance
- The state of tools and technologies to support pipeline management

Outline

Part One: Today's Data Challenges

- Variety and Complexity
  - Sources
  - Ingestion
  - Persistence
  - Management Topology
  - Utility
  - Use Cases
- Time to Value
  - Storing Data
  - From Origin to Destination
  - Finding Data
  - Learning about Data
  - Data Cataloging
  - Data Preparation
  - Analysis and Communication

Part Two: Modern Data Solutions

- Growth and Scalability

    - Data Scalability
    - Process Scalability
    - People Scalability
    - Analytic Scalability
- Rethinking Data Architecture
    - Persistence and Topology
    - Data Flow
    - Services
    - Governance
    - What does this mean for your architecture?
- Building for the Future
    - Future of Databases
    - Future of Data
    - Future of Analytics

Part Three: Data Pipeline Design

- The Big Picture
    - Pipeline Components
- Destination
    - Purpose and End Point
    - Timeliness
- Origin
    - Data Supply and Begin Point
    - Data Type and Velocity
- Data Flow
    - Data in Motion
    - Pipeline Boundaries
    - Blending Batch and Real Time
- Data Storage
    - Data at Rest
    - Choosing Data Storage
- Processing
    - Data Products and Data Value
    - Ingestion
    - Persistence
    - Transformation
    - Delivery
- Workflow
    - Sequence of Activities
- Monitoring
    - Pipeline Health
- Technology
    - Pipeline Tools
    - Abundance of Tools
- Design Summary
    - 7 Steps