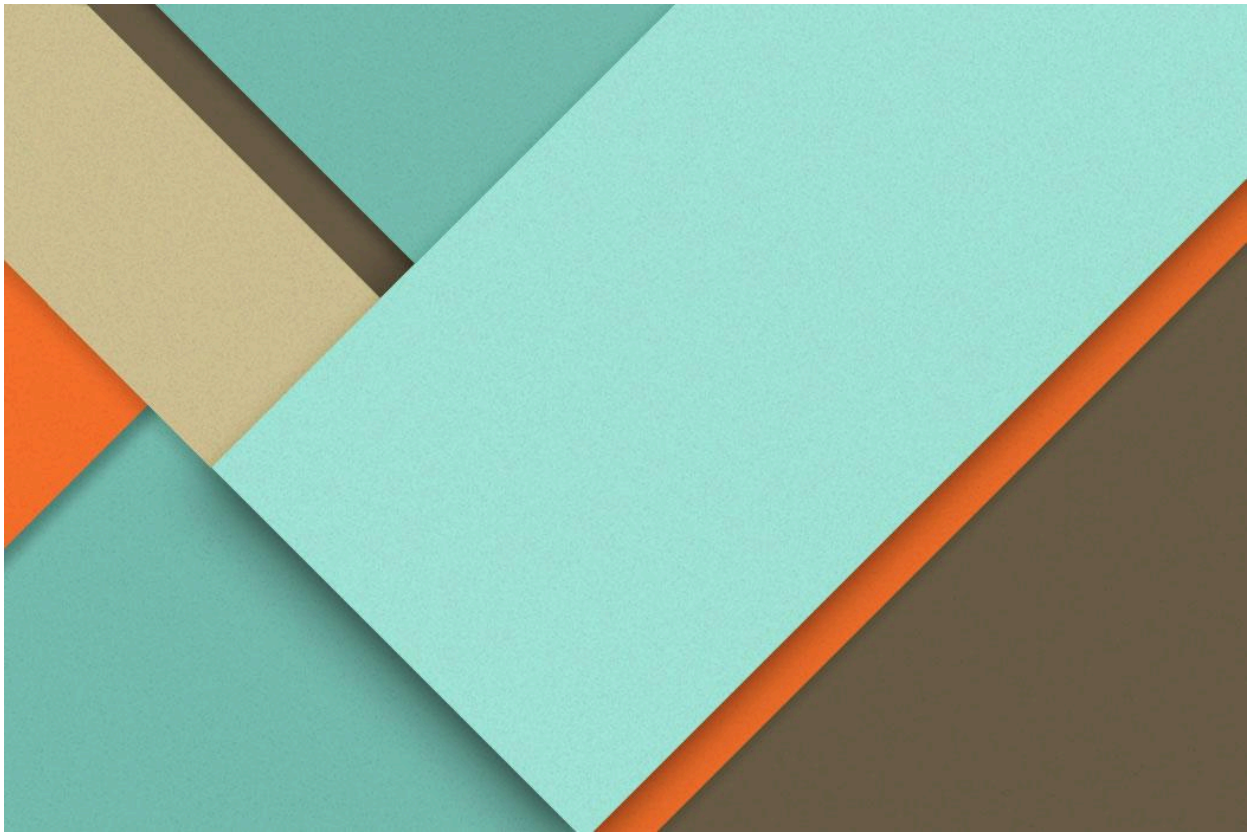

Training Topics for Data Engineering on Microsoft Azure

05.14.2024

—

Himanshu Bhadra
BayAreaLa8s | www.bayareala8s.com | https://github.com/bayareala8s
186 Cameo Drive
Livermore, CA 94550

Breaking down the training topics into granular sub-topics provides a clearer roadmap for learning data engineering on Microsoft Azure.

Let's dive into each topic:

1. **Introduction to Azure (30 Mins)**

- Overview of Azure services
  - Compute (Virtual Machines, Azure App Service)
  - Storage (Azure Blob Storage, Azure Data Lake Storage)
  - Networking (Virtual Network, Azure Load Balancer)
  - Databases (Azure SQL Database, Azure Cosmos DB)
- Azure Management Tools
  - Azure Portal
  - Azure CLI
  - Azure PowerShell
- Azure Resource Management
  - Resource Groups
  - Resource Locks
- Azure Security
  - Identity and Access Management (Azure Active Directory)
  - Network Security Groups
  - Azure Security Center

2. **Azure Data Services Overview (45 Mins)**

- Azure Data Storage Services
  - Azure Blob Storage
  - Azure Data Lake Storage Gen2
- Azure Database Services

- Azure SQL Database
- Azure Cosmos DB
- Azure Analytics Services
- Azure Data Factory
- Azure Databricks
- Azure Synapse Analytics
- Azure Stream Analytics

3. **Azure Data Lake Storage (60 Mins)**

- Introduction to Azure Data Lake Storage Gen2
- Data Ingestion
- Azure Data Factory
- Azure CLI
- Data Management
- Storage Tiers
- Data Lifecycles
- Security and Governance
- Access Control Lists (ACLs)
- Azure AD Integration
- Data Classification
- Auditing and Monitoring

4. **Azure Data Factory (180 Mins)**

- Introduction to Azure Data Factory
- Pipelines and Activities
- Copy Data Activity
- Data Flow Activity
- Execute Pipeline Activity
- Data Movement and Transformation

- Mapping Data Flows
  - Wrangling Data Flows
- Integration with other Azure Services
  - Azure SQL Database
  - Azure Blob Storage
  - Azure Synapse Analytics

5. **Azure SQL Database (120 Mins)**

- Introduction to Azure SQL Database
- Provisioning and Configuration
  - Serverless vs. Provisioned
  - Performance Levels
- Data Management
  - Data Import/Export
  - Data Migration
- Monitoring and Optimization
  - Query Performance Tuning
  - Indexing Strategies
- Security and Compliance
  - Firewall Rules
  - Transparent Data Encryption

6. **Azure Cosmos DB (120 Mins)**

- Introduction to Azure Cosmos DB
- Data Modeling
  - Partitioning
  - Consistency Levels
- Global Distribution
  - Multi-region Writes

- Replication Policies
- Querying and Indexing
  - SQL API
  - MongoDB API
- Monitoring and Scaling
  - Metrics and Alerts
  - Autoscale Provisioning

7. **Azure Stream Analytics (180 Mins)**

- Introduction to Azure Stream Analytics
- Real-time Data Ingestion
  - Event Hubs
  - IoT Hubs
- Query Language and Windowing
  - SQL Queries
  - Tumbling Windows
  - Sliding Windows
- Integration with Output Sinks
  - Azure Blob Storage
  - Azure SQL Database
  - Power BI

8. **Azure Databricks (240 Mins)**

- Introduction to Azure Databricks
- Cluster Configuration and Management
  - Autoscaling
  - Node Types
- Data Processing with Spark
  - RDDs and DataFrames

- Spark SQL
- Machine Learning with MLlib
  - Classification
  - Regression
- Integration with Azure Services
  - Azure Blob Storage
  - Azure SQL Database
  - Azure Data Lake Storage

9. **Azure Synapse Analytics (240 Mins)**

- Introduction to Azure Synapse Analytics
- Provisioning and Configuration
  - SQL Pools
  - Spark Pools
- Data Integration and Orchestration
  - Data Movement
  - PolyBase Integration
- Analytics and Reporting
  - SQL Queries
  - Power BI Integration
- Security and Compliance
  - Role-Based Access Control
  - Auditing and Compliance

10. **Data Governance and Compliance (120 Mins)**

- Data Governance Principles
  - Data Ownership
  - Data Stewardship
- Compliance Standards

  - GDPR
  - HIPAA
- Access Control and Authorization
  - Role-Based Access Control (RBAC)
  - Azure Active Directory Integration
- Data Encryption and Masking
  - Transparent Data Encryption (TDE)
  - Dynamic Data Masking (DDM)

11. **Monitoring and Optimization (120 Mins)**

- Monitoring Azure Data Services
  - Azure Monitor
  - Metrics and Alerts
- Performance Tuning
  - Query Optimization
  - Indexing Strategies
- Cost Optimization
  - Resource Utilization
  - Reserved Instances
- Scaling Strategies
  - Vertical Scaling
  - Horizontal Scaling

12. **Best Practices and Design Patterns (120 Mins)**

- Scalability and Performance
  - Horizontal Partitioning
  - Vertical Partitioning
- Data Modeling Best Practices
  - Star Schema

  - Snowflake Schema
- Data Lifecycle Management
  - Data Archiving
  - Data Retention Policies
- Fault Tolerance and Disaster Recovery
  - Redundancy and Replication
  - Backup and Restore Strategies

These detailed sub-topics provide a structured and comprehensive curriculum for mastering data engineering on Microsoft Azure.

Each subtopic can be further explored through hands-on exercises, tutorials, and real-world projects to deepen understanding and proficiency.