3-Month Azure Data Engineering Learning Plan

This document provides a structured 3-month learning plan for Azure Data Engineering.

The plan is divided into weekly sections with topics and practical examples to guide you through core Azure services,

data engineering concepts, and practical implementations. The goal is to build a strong foundation in Azure,

data engineering concepts, and hands-on skills.

Month 1: Foundation and Core Azure Services

Week 1: Introduction to Azure and Data Engineering Basics

Topics:

- Introduction to cloud computing and Azure

- Overview of Azure Data Services

- Key concepts in data engineering (ETL, data warehousing, big data)

Examples:

- Create a free Azure account

- Explore the Azure portal and understand key services (e.g., Resource Groups, Virtual Networks)

Week 2: Azure Storage Solutions for Data Engineering

Topics:

- Azure Blob Storage: storing unstructured data

- Azure Data Lake Storage Gen2: for big data storage

- Data security and access control in storage (Shared Access Signature, IAM)

Examples:

- Set up Blob Storage and Data Lake Storage

- Upload and manage files in Azure Blob Storage using Azure CLI and Python SDK

- Grant permissions using IAM roles

Week 3: Azure SQL Database and Azure Synapse Analytics

Topics:

- Introduction to Azure SQL Database

- Basics of Azure Synapse Analytics (formerly Azure SQL Data Warehouse)

- Understanding relational vs. non-relational data storage

Examples:

- Create an Azure SQL Database and import sample data

- Run basic SQL queries in Azure SQL

- Set up an Azure Synapse Analytics workspace and explore dedicated SQL pools

Week 4: Azure Data Factory (ADF) for Data Integration

Topics:

- Introduction to Azure Data Factory

- ETL and ELT pipelines using ADF

- Data ingestion from multiple sources (Azure Blob, SQL, on-premises)

Examples:

- Create a data pipeline in ADF to copy data from Blob Storage to Azure SQL Database

- Use data transformations in ADF (mapping data flow)

- Schedule pipeline runs and monitor them

Month 2: Advanced Data Engineering on Azure

Week 5: Data Transformation with Azure Databricks

Topics:

- Introduction to Azure Databricks (Apache Spark on Azure)

- Data processing and transformation with PySpark

- Integrating Databricks with Azure Data Lake Storage

Examples:

- Set up an Azure Databricks workspace

- Use PySpark to read and process data from Data Lake Storage

- Run transformations on a sample dataset and save the results back to Data Lake

Week 6: Advanced Azure Data Factory Features

Topics:

- Advanced data flows and transformations

- Data integration with REST APIs and web services

- Incremental data loading, parameterization, and reusable pipelines

Examples:

- Set up an incremental data load from an API source to Azure SQL Database

- Build a reusable parameterized pipeline to handle different datasets

- Integrate ADF with REST API endpoints to fetch data

Week 7: Introduction to Stream Processing with Azure Stream Analytics

Topics:

- Stream processing fundamentals

- Azure Stream Analytics for real-time data processing

- Input, output, and query in Azure Stream Analytics

Examples:

- Create a Stream Analytics job that reads from an IoT Hub or Event Hub

- Process the data in real-time and store it in Azure SQL Database

- Visualize the streaming data with Power BI

Week 8: Orchestrating and Managing Data Pipelines

Topics:

- Azure Data Factory triggers, activities, and monitoring

- Integration with Azure Key Vault for secure credentials

- Best practices for managing and optimizing data pipelines

Examples:

- Add triggers to automate ADF pipelines

- Use Key Vault to securely store database and API credentials

- Monitor pipeline performance and troubleshoot common issues

Month 3: Advanced Concepts and Project Work

Week 9: Data Analytics and Machine Learning with Azure Synapse and Databricks

Topics:

- Exploratory data analysis in Azure Synapse

- Integrating Synapse with Azure ML for machine learning

- Using Databricks for ML model development and deployment

Examples:

- Perform data exploration in Synapse using SQL on-demand

- Train a simple ML model in Databricks with sample data

- Deploy the model and score new data using Synapse pipelines

Week 10: Data Security, Governance, and Compliance

Topics:

- Data governance using Azure Purview

- Data security best practices (encryption, access control)

- Compliance and regulatory standards in Azure (GDPR, HIPAA)

Examples:

- Set up Azure Purview to catalog and manage data assets

- Configure role-based access controls on key data assets

- Enable encryption for data in Azure Blob Storage and Data Lake Storage

Week 11: Optimization and Performance Tuning

Topics:

- Optimizing data storage costs (tiered storage, retention policies)

- Performance tuning in Azure SQL Database and Synapse

- Best practices for Azure Databricks performance

Examples:

- Use partitioning and indexing in Azure SQL for optimized querying

- Apply storage optimization settings in Blob Storage and Data Lake

- Tune Databricks cluster settings for efficient processing

Week 12: Final Project and Capstone

Project:

- Build a complete data pipeline from data ingestion to analysis

Example use case: Ingest streaming data, process it in real-time, store it in a data lake, perform batch processing, and visualize results in Power BI.

Tasks:

- Create end-to-end data pipeline for an IoT or sales data use case

- Schedule batch and real-time data ingestion

- Analyze and visualize the processed data in Power BI or Synapse Studio