**✅ Phase 0: Foundations (Week 0–1)**

**Goal:** Strengthen your SQL + Python basics before jumping into Azure tools.

* 🔹 **SQL**: Focus on:
  + Joins, Subqueries, CTEs, Window Functions
  + Data Modeling basics (Star/Snowflake schema)
* 🔹 **Python for Data Engineering**:
  + Data structures, functions
  + File I/O, CSV, JSON parsing
  + Libraries: pandas, os, datetime

✅ Resources:

* SQLZOO, Mode Analytics SQL Tutorial
* [Python for Data Engineering (YouTube)](https://www.youtube.com/watch?v=OTmQOjsl0eg)

**✅ Phase 1: Azure Data Engineering Essentials (Week 2–4)**

**Goal:** Learn core Azure data services & pipeline building

**🧩 Learn the following Azure tools:**

| **Tool** | **What to Learn** |
| --- | --- |
| **Azure Data Factory (ADF)** | ETL Pipelines, Linked Services, Datasets, Data Flows |
| **Azure Data Lake Gen2** | Folder structure, security, file ingestion |
| **Azure Blob Storage** | Unstructured file storage |
| **Azure SQL Database** | As a sink/source for data |
| **Azure Key Vault** | For storing secrets securely |
| **Azure DevOps (Basics)** | CI/CD for data pipelines |

✅ Resources:

* [Microsoft Learn: Azure Data Engineer Learning Path](https://learn.microsoft.com/en-us/training/paths/data-engineer/)
* ADF Crash Course: [ADF Tutorial - YouTube](https://www.youtube.com/watch?v=l1EssrLxt7E)

**✅ Phase 2: Data Engineering with PySpark & Databricks (Week 5–7)**

**Goal:** Learn processing large datasets using distributed computing.

* 🔹 Concepts:
  + RDDs, DataFrames, Spark SQL
  + Transformations, Actions, Joins, Aggregations
  + Delta Lake and Schema Evolution
* 🔹 Databricks:
  + Create notebooks
  + Mount ADLS to Databricks
  + Read/write files in Parquet, Delta format

✅ Resources:

* Databricks Academy
* YouTube: *"Azure Databricks Tutorial for Beginners"*

**✅ Phase 3: Real-World Projects (Week 8–10)**

**Goal:** Apply skills to practical data workflows

**💡 Sample Projects:**

| **Project Idea** | **Concepts Used** |
| --- | --- |
| Sales Data ETL Pipeline | ADF + ADLS + SQL |
| Log Processing Pipeline | Blob Storage + Databricks + Delta Lake |
| Movie Ratings Aggregator | PySpark + ADF + Azure SQL |
| Incremental Data Load | ADF Mapping Data Flows + Delta |

📝 Document your projects on GitHub with clear **README** files and **data architecture diagrams**.

**✅ Phase 4: CI/CD, Monitoring & Practice (Week 11–12)**

**Goal:** Learn how to productionize pipelines

* 🔹 Azure DevOps:
  + Create pipelines to deploy ADF artifacts
* 🔹 Monitoring:
  + Use ADF Monitoring tab, Alerts, Logs
* 🔹 Interview Prep:
  + Data modeling
  + Data quality and orchestration questions

**📘 Certifications (Optional but Valuable)**

* **DP-203: Azure Data Engineer Associate**
  + Highly recommended to validate your knowledge
  + Covers everything you're learning in this roadmap

**🧠 Additional Tips:**

* Start a **LinkedIn post series** or **blog** explaining what you learn weekly.
* Join communities: [r/dataengineering](https://www.reddit.com/r/dataengineering/), [Microsoft Learn Community](https://learn.microsoft.com/en-us/training/)