Become an Azure Data Engineer for Absolute Beginners

Azure DE Roadmap

1

"Cloud" means using online services like Amazon Web Services, Google Cloud, or Microsoft Azure to build and manage systems for storing, processing, and analyzing data, instead of using their own physical computers. This

2

Python, in simple terms, is a versatile programming language that data engineers use to build and manage systems for handling data.

3

Azure Storage offers highly available, massively scalable, durable, and secure storage for a variety of data objects in the cloud.

4

SQL, or Structured Query Language, in simple words for a data engineer, is the language used to talk to and manage data in relational databases.

5

ADF is a cloud-based service for data integration that allows data engineers to build, schedule, and orchestrate data workflows for tasks like ETL and data movement at scale

6

It's the way to use the popular Python to work with Apache Spark, a powerful engine analyzing massive datasets.

7

a low-code, serverless workflow and integration service used for automating tasks like data orchestration, process automation, and connecting disparate systems

8

"Big data" is about handling extremely large, complex, and fast-moving datasets that traditional systems can't manage

Continues

9

ADB is aa unified data and AI platform that helps engineers build and maintain data solutions, including ETL pipelines and machine learning workflows

10

Snowflake is a cloud-based data warehouse that is used to store, process, and analyze large amounts of data efficiently

11

Airflow is a workflow management platform that acts as a smart manager to programmatically schedule, monitor, and orchestrate complex data pipelines and tasks

12

Kafka is an event streaming platform, essentially a message bus for handling large volumes of real-time data

13

Azure Functions, in simple terms for a data engineer, is a serverless compute service that allows you to run small pieces of code

14

It is a cloud-based data platform large volumes of log and performance data generated by Azure and non-Azure resources.

15

Is most commonly refers to Microsoft Fabric, a unified analytics platform that simplifies the entire data workflow from ingestion and transformation to analysis and reporting

16

Key Vault is a secure digital "safe box" in the cloud used by data engineers to store and manage sensitive information like passwords, API keys, and encryption keys