

**Data Analytics on AWS**

**Course Number:** DATA-119
**Duration:** 2 days

**Overview**

This Data Analytics on AWS training course gives attendees a comprehensive overview of AWS' core offerings and teaches students how to use AWS Lake Formation Service, Simple Storage Service (S3), Snowflake Cloud Data Platform, Amazon QuickSight, visualization and reporting tools, AWS Glue  PySpark extensions, and more. Participants learn to choose which big data AWS services best align with their desired outcomes and power their projects with the appropriate AWS data analytics tools.

**Prerequisites**

* Successful completion of class [PYTH-100, Introduction to Python Programming](file:////training/python) (or equivalent Python knowledge)
* Successful completion of class [PYTH-252WA, Advanced Data Analytics with PySpark](file:////training/data-analytics-pyspark-advanced)
* Knowledge of SQL

**Materials**

All AWS for Data training students will receive comprehensive courseware.

**Software Needed on Each Student PC**

A modern web browser and an Internet connection.

**Objectives**

* Use S3 Through Management Console
* Sign Up for the Free Trial of Snowflake
* The Snowflake Web UI
* Create and Work with Databases in Snowflake
* Understand AWS Glue
* Use AWS Glue Crawlers and Classifiers
* Create an S3 Bucket for AWS Glue ETL Script Output
* Create and Work with Glue Scripts
* Use PySpark API Directly
* Understand AWS Glue ETL Jobs

**Outline**

* Introduction
* The AWS Lake Formation Service
	+ First, What is a Data Lake?
	+ Data Lakes vs. Traditional Data Warehouses
	+ Characteristics of Data Warehouses and Data Lakes
	+ Now, What is AWS Lake Formation?
	+ What are the Benefits of Using Lake Formation?
	+ How Lake Formation Works
	+ The Lake Formation Dashboard
	+ AWS Lake Formation Pricing
* AWS Simple Storage Service
	+ What is AWS Simple Storage Service (S3)
	+ AWS S3
	+ Storage
	+ Regions
	+ S3 Regions
	+ Getting started with S3
	+ Using BitTorrent
	+ More on Buckets
	+ Bucket Configurable Properties
	+ Advanced S3 Bucket Properties
	+ The Bucket Creation Dialog in the Management Console
	+ Bucket Permissions
	+ Bucket-level Operations
	+ Authorization of REST Requests
	+ Adding Cross-Origin Resource Sharing Configuration
	+ Event Notifications
	+ The Requester Pays Option
	+ The Object Key
	+ Object Versioning
	+ Example of Object Properties
	+ Object Storage Class Levels
	+ Object-level Operations
	+ Object Lifecycle Configuration
	+ Amazon S3 Data Consistency Model
	+ Observable Data Consistency Behaviors
	+ Eventually Consistent Reads vs Consistent Reads
	+ Amazon S3 Security
	+ S3 Use Case: Backup and Archiving
	+ Another S3 Use Case: Static Web Hosting
	+ More on Static Web Hosting
	+ S3 Static Website Hosting Dialog in Management Console
	+ S3 Use Case: Disaster Recovery
	+ AWS S3 Pricing
	+ Storage Pricing
	+ Request Pricing
	+ Data Transfer Pricing
	+ Amazon S3 Transfer Acceleration
	+ How to Enable Transfer Acceleration
	+ Enabling Transfer Acceleration in the Management Console
	+ Amazon S3 SLA Definitions
	+ Amazon S3 SLA Service Commitment
	+ S3 CLI
* Introduction to the Snowflake Cloud Data Platform
	+ What is Snowflake?
	+ Certifications
	+ Snowflake Conceptual Architecture
	+ Core Underlying Design Considerations
	+ Core Services and Tools
	+ Snowflake Editions
	+ The Standard Edition
	+ The Enterprise Edition
	+ The Business Critical Edition
	+ Virtual Private Snowflake
	+ Billing: The Cost Components
	+ Data Storage Segments
	+ Parts of Snowflake that Incur Compute-related Costs
	+ Snowflake Quickstart
* Snowflake's Web UI
	+ Web UI (Web Portal)
	+ The Landing Page
	+ Snowflake Roles
	+ The Roles UI
	+ Databases
	+ Shares
	+ Data Marketplace
	+ The Warehouses UI
	+ Worksheets
	+ History
	+ A History Sample
	+ Account
	+ Operational Transparency: Controlling the Usage of Your Account
	+ Create Network Policy Dialog (under Account > Policies)
	+ Preview App
* Visualization and Reporting
	+ Amazon QuickSight
	+ SPICE
	+ Data Analyses
	+ Visuals
	+ Sheets
	+ Dashboards
	+ Typical Amazon QuickSight Workflow
	+ Create a Data Set
	+ Create an Analysis
	+ Create a Visual Manually
	+ Amazon Athena
	+ Amazon Athena and AWS Data Catalog
	+ Query Data Using Amazon Athena
	+ What is Tableau?
	+ Create a Report Using Tableau
	+ Tutorial: Get Started with Tableau Desktop
* Introduction to AWS Glue
	+ What is AWS Glue?
	+ AWS Glue Components
	+ Managing Notebooks
	+ Putting it Together: The AWS Glue Environment Architecture
	+ AWS Glue Main Activities
	+ Additional Glue Services
	+ AWS Glue Pricing
	+ When To Use AWS Glue?
	+ Integration with other AWS Services
* AWS Glue PySpark Extensions
	+ AWS Glue and Spark
	+ The DynamicFrame Object
	+ The DynamicFrame API
	+ The GlueContext Object
	+ Glue Transforms
	+ A Sample Glue PySpark Script
	+ Using PySpark
	+ AWS Glue PySpark SDK
* Conclusion