

AWS Big Data Engineer

Level 2 course

Office Email: hello@stackintel.io

USA Phone: +1 (917) 920-6267

LHR Phone : +92 42 34003951, +92 302 0410424

Instructor: Bilal Rabbani

Email: bilal.rabbani@stackintel.io

AWS Certified - Trainer

Introduction

Do you know what is Data Engineering or Big Data? This training program is in collaboration with AWS and developed to not only introduce to Big Data but provides hand-on experience in Big Data Engineering. This will cover all the content to know to how to streamline data processing, by leveraging the state of the art technology stack, i.e. AWS, Hadoop, Spark, Pandas, Python, Kafka and use the database management tool DocumentDB to store metadata.

What is Big Data

Big data is a term that describes the large volume of data – both structured and unstructured – that inundates a business on a day-to-day basis. But it's not the amount of data that's important. It's what organizations do with the data that matters. Big data can be analyzed for insights that lead to better decisions and strategic business moves.

Training Progression / Objectives

During this training program, we will do the following modules

- AWS Cloud Practitioner Essentials
- AWS Well-Architecture Best Practices
- AWS Serverless Orchestration
- AWS Certified Big Data – Specialty (Optional)



Schedule

Course Duration	1 month
Days	3 days a week
Time Duration	3 hours per session

Learning Material

AWS Documentation: Find user guides, developer guides, API references, tutorials, and more on the following URL:

<https://docs.aws.amazon.com/>

Apache Documentation: The documentation is available in several formats. Downloadable formats including Windows Help format and offline-browsable HTML are available.

Who Should Attend

- Designed for software engineers (entry-level to professional) to design the Cloud-Native Data Platform.
- A CS/EE graduate or final year student can join this course.
- The course is also valuable for architects, testers, and product managers as they too should understand the platform and how development works with data extensive architectures.

Code of Conduct

Attendance: Students are expected to attend every class to the best of their ability - Emergencies may happen, therefore, we understand. If something comes up we ask that you notify the instructor or the management team ahead of time when possible. This includes calling or emailing if you are going to be late for a class. If you miss three or more classes for any reason, we may ask that you make up that time to ensure maximum learning opportunities.

Conduct: Please remember that this training is a job preparedness program for your future career. Conduct yourself in a professional manner and participate regularly in class. Remember that we operate in a diverse environment. Please be respectful to your fellow students, instructor, and management team members. Avoid distractions such as phone calls and texting.

Homework: Homework opportunities will be given regularly and should not take more than a day. The first few minutes of class will be consumed to go over homework and any questions students have regarding the homework. Although homework is optional, it is encouraged to increase understanding of subject matter.



Course Outline

- Introduction to Big Data, Data Engineering, Data Processes, and Data lifecycle
 - Variety, Velocity, Volume
- Data engineering toolbox
- Introduction to Big Data
- Introduction to Data Management
 - Database
 - Data Lake
 - Data Warehouse
 - Data Mart
- Data Lake
 - Ingestion
 - Transformation
 - Curation
 - Consumption
- Extract, Transform and Load (ETL)
- A crash course on Toolbox
 - AWS EMR
 - AWS Red Shift
 - Apache Hadoop
 - Apache Spark
 - Apache Hive
 - Apache Kafka
 - Pandas
 - AWS DocumentDB
- Build, deploy, and run Spark scripts on Hadoop clusters
- Regular Expressions
- Python Labs
- Apache Spark Labs
 - Run scripts on AWS EMR
 - PySpark
 - SparkSQL
 - Datasets, Data Frames, RDDs
- Optimize Spark jobs
 - Partitioning, caching, and other techniques
- Process continual streams of data
 - Spark Streaming
 - Apache Kafka
 - AWS Kinesis
- Data pipeline and Orchestration
 - AWS Data Pipeline
 - AWS Step Functions
- Maintaining Data and Metadata
 - AWS S3
 - AWS DynamoDB
 - AWS DocumentDB
 - MySQL
- Case Studies of Data Platforms
- White Papers on Big Data
- Introduction to Machine learning
- Introduction to Data Analytics