

BIG DATA ON AWS

3-Days Training Duration / Instructor-Led, Classroom-Based Training / Intermediate Level

DELIVERY METHODS



ILT – Instructor-Led Classroom Training

ILT sessions are conducted in a physical classroom environment.



ILO – Instructor-Led Online Training

ILO sessions are conducted via WebEx in a VoIP environment



FLEX Classroom™ – Combined ILT & ILO

FLEX Classroom sessions are delivered via ILT and ILO giving you the ultimate flexibility.

Singapore Office

111 Somerset Road #07-07
Triple One Somerset
Singapore 238164

Malaysia Office

L28-03-03A, Level 28
PJ Exchange No 16A Persiaran Barat
46050 Petaling Jaya

Philippines Office

Unit 522 5th Floor,
Gateway Tower,
Araneta Center Cubao,
Quezon City Philippines 1109

Indonesia Office

Jl. Prof. Dr. Latumenten Raya # 7E,
Special Capital Region of Jakarta,
Indonesia

Taiwan Office

2F-15, No.12, Fuxing 4th Road
Qianzhen Dist
Kaohsiung City 806
Taiwan R.O.C.

Hong Kong Office

Suite 501, 5 Floor
The Hennessy, 256 Hennessy Road
Wan Chai
Hong Kong

COURSE DESCRIPTION

Big Data on AWS introduces you to cloud-based big data solutions such as Amazon EMR, Amazon Redshift, Amazon Kinesis and the rest of the AWS big data platform. In this course, we show you how to use Amazon EMR to process data using the broad ecosystem of Hadoop tools like Hive and Hue. We also teach you how to create big data environments, work with Amazon DynamoDB, Amazon Redshift, Amazon Quicksight, Amazon Athena and Amazon Kinesis, and leverage best practices to design big data environments for security and cost-effectiveness.

BRING YOUR OWN DEVICE

Please note that this class requires you to bring your own device. Plan to come to class with the following:

- A notebook computer with Wi-Fi (labs cannot be completed on an iPad or tablet device)
- Administrator access to the computer, if Windows
- Chrome, Firefox, or Safari (IE not supported)

PREREQUISITES

- Basic familiarity with big data technologies, including Apache Hadoop, MapReduce, HDFS, and SQL/NoSQL querying
- Students should complete the AWS Technical Essentials course or have equivalent experience
- Students should complete the free Big Data Technology Fundamentals web-based training or have equivalent experience
- Working knowledge of core AWS services and public cloud implementation
- Basic understanding of data warehousing, relational database systems, and database design.

INTENDED AUDIENCE

This course is intended for:

- Individuals responsible for designing and implementing big data solutions, namely Solutions Architects
- Data Scientists and Data Analysts interested in learning about the services and architecture patterns behind big data solutions on AWS

HANDS-ON ACTIVITY

This course allows you to test new skills and apply knowledge to your working environment through a variety of practical exercises

BIG DATA ON AWS

3-Days Training Duration / Instructor-Led, Classroom-Based Training / Intermediate Level

DELIVERY METHODS



ILT – Instructor-Led Classroom Training

ILT sessions are conducted in a physical classroom environment.



ILO – Instructor-Led Online Training

ILO sessions are conducted via WebEx in a VoIP environment



FLEX Classroom™ – Combined ILT & ILO

FLEX Classroom sessions are delivered via ILT and ILO giving you the ultimate flexibility.

COURSE OBJECTIVES

This course is designed to teach you how to:

- Fit AWS solutions inside of a big data ecosystem
- Leverage Apache Hadoop in the context of Amazon EMR
- Identify the components of an Amazon EMR cluster
- Launch and configure an Amazon EMR cluster
- Leverage common programming frameworks available for Amazon EMR including Hive, Pig, and Streaming
- Leverage Hue to improve the ease-of-use of Amazon EMR
- Use in-memory analytics with Spark on Amazon EMR
- Choose appropriate AWS data storage options
- Identify the benefits of using Amazon Kinesis for near real-time big data processing
- Leverage Amazon Redshift to efficiently store and analyze data
- Comprehend and manage costs and security for a big data solution
- Secure a Big Data solution
- Identify options for ingesting, transferring, and compressing data
- Leverage Amazon Athena for ad-hoc query analytics
- Use visualization software to depict data and queries using Amazon QuickSight
- Orchestrate big data workflows using AWS Data Pipeline.

Singapore Office

111 Somerset Road #07-07
Triple One Somerset
Singapore 238164

Malaysia Office

L28-03-03A, Level 28
PJ Exchange No 16A Persiaran Barat
46050 Petaling Jaya

Philippines Office

Unit 522 5th Floor,
Gateway Tower,
Araneta Center Cubao,
Quezon City Philippines 1109

Indonesia Office

Jl. Prof. Dr. Latumenten Raya # 7E,
Special Capital Region of Jakarta,
Indonesia

Taiwan Office

2F-15, No.12, Fuxing 4th Road
Qianzhen Dist
Kaohsiung City 806
Taiwan R.O.C.

Hong Kong Office

Suite 501, 5 Floor
The Hennessy, 256 Hennessy Road
Wan Chai
Hong Kong

BIG DATA ON AWS

3-Days Training Duration / Instructor-Led, Classroom-Based Training / Intermediate Level

DELIVERY METHODS



ILT – Instructor-Led Classroom Training

ILT sessions are conducted in a physical classroom environment.



ILO – Instructor-Led Online Training

ILO sessions are conducted via WebEx in a VoIP environment



FLEX Classroom™ – Combined ILT & ILO

FLEX Classroom sessions are delivered via ILT and ILO giving you the ultimate flexibility.

Singapore Office

111 Somerset Road #07-07
Triple One Somerset
Singapore 238164

Malaysia Office

L28-03-03A, Level 28
PJ Exchange No 16A Persiaran Barat
46050 Petaling Jaya

Philippines Office

Unit 522 5th Floor,
Gateway Tower,
Araneta Center Cubao,
Quezon City Philippines 1109

Indonesia Office

Jl. Prof. Dr. Latumenten Raya # 7E,
Special Capital Region of Jakarta,
Indonesia

Taiwan Office

2F-15, No.12, Fuxing 4th Road
Qianzhen Dist
Kaohsiung City 806
Taiwan R.O.C.

Hong Kong Office

Suite 501, 5 Floor
The Hennessy, 256 Hennessy Road
Wan Chai
Hong Kong

COURSE OUTLINE

Note: Course outline may vary slightly based on the regional location and/or language in which the class is delivered.

This course covers the following concepts on each day:

Day 1

- Overview of Big Data
- Big Data Ingestion and Transfer
- Big Data Streaming and Amazon Kinesis
- Lab 1: Using Amazon Kinesis to Stream and Analyze Apache Server Log Data
- Big Data Storage Solutions
- Big Data Processing and Analytics
- Lab 2: Using Amazon Athena to Query Log Data From Amazon S3

Day 2

- Apache Hadoop and Amazon EMR
- Lab 3: Storing and Querying Data on Amazon DynamoDB
- Using Amazon EMR
- Hadoop Programming Frameworks
- Lab 4: Processing Server Logs With Hive on Amazon EMR
- Web Interfaces on Amazon EMR
- Lab 5: Running Pig Scripts in Hue on Amazon EMR
- Apache Spark on Amazon EMR
- Lab 6: Processing NY Taxi data using Spark on Amazon EMR

Day 3

- Amazon Redshift and Big Data
- Visualizing and Orchestrating Big Data
- Lab 7: Using TIBCO Spotfire to Visualize Data
- Managing Big Data Costs
- Securing Your Amazon Deployments
- Big Data Design Patterns