Enterprise Computing Solutions - Education Services

TRAINING OFFERING

You can reach us at:

Arrow Enterprise Computing Solutions Ltd, Part 1st Floor, Suite 1D/1, Central House, Otley Road, Harrogate, HG3 1UG

Email: educationteam.ecs.uk@arrow.com
Phone: 0870 251 1000

Big Data on AWS

CODE:	LENGTH:	PRICE:
AMA_AWS-BD	24 Hours (3 days)	£2,245.00

Description

Big Data on AWS (Amazon Web Services) introduces you to cloud-based big data solutions and Amazon Elastic MapReduce (EMR), 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 Pig and Hive. We also teach you how to create big data environments, work with Amazon DynamoDB, Amazon Redshift, and Amazon Kinesis, and leverage best practices to design big data environments for security and cost-effectiveness.

Objectives

- Understand Apache Hadoop in the context of Amazon EMR
- Understand the architecture of an Amazon EMR cluster
- Launch an Amazon EMR cluster using an appropriate Amazon Machine Image and Amazon EC2 instance types
- Choose appropriate AWS data storage options for use with Amazon EMR
- Know your options for ingesting, transferring, and compressing data for use with Amazon EMR
- Use common programming frameworks available for Amazon EMR including Hive, Pig, and Streaming
- Work with Amazon Redshift to implement a big data solution
- Leverage big data visualization software
- Choose appropriate security options for Amazon EMR and your data
- Perform in-memory data analysis with Spark and Shark on Amazon EMR
- Choose appropriate options to manage your Amazon EMR environment cost-effectively
- Understand the benefits of using Amazon Kinesis for big data

Audience

- Individuals responsible for designing and implementing big data solutions, namely Solutions Architects and SysOps Administrators
- Data Scientists and Data Analysts interested in learning about big data solutions on AWS

Prerequisites

- Basic familiarity with big data technologies, including Apache Hadoop and HDFS
- Knowledge of big data technologies such as Pig, Hive, and MapReduce is helpful but not required
- Working knowledge of core AWS services and public cloud implementation
- Students should complete the AWS Essentials course or have equivalent experience
- Basic understanding of data warehousing, relational database systems, and database design

Programme

Day 1

- Overview of Big Data, Apache Hadoop, and the Benefits of Amazon EMR
- Amazon EMR Architecture
- Using Amazon EMR

- Launching and Using an Amazon EMR Cluster
- Hadoop Programming Frameworks

Day 2

- Using Hive for Advertising Analytics
- Using Streaming for Life Sciences Analytics
- Overview: Spark and Shark for In-Memory Analytics
- Using Spark and Shark for In-Memory Analytics
- Managing Amazon EMR Costs
- Overview of Amazon EMR Security
- Data Ingestion, Transfer, and Compression
- Using Amazon Kinesis for Real-Time Big Data Processing

Day 3

- Using Amazon Kinesis and Amazon EMR to Stream and Process Big Data
- AWS Data Storage Options
- Using DynamoDB with Amazon EMR
- Overview: Amazon Redshift and Big Data
- Using Amazon Redshift for Big Data
- Visualizing and Orchestrating Big Data
- Using Tableau Desktop or Jaspersoft BI to Visualize Big Data

Session Dates

On request. Please <u>Contact Us</u>

Additional Information

This training is also available as onsite training. Please contact us to find out more.