Big Data on AWS

Course code: AWSBD

Who is the course for 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 AWSWhat we teach you 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 and Spark SQL on Amazon EMR Choose appropriate AWS data storage options Identify the benefits of using Amazon Kinesis for near real-time big data processing Define data warehousing and columnar database concepts Leverage Amazon Redshift to efficiently store and analyze data Comprehend and manage costs and security for Amazon EMR and Amazon Redshift deployments Identify options for ingesting, transferring, and compressing data Use visualization software to depict data and queries Orchestrate big data workflows using AWS Data PipelineRequired skills Basic familiarity with big data technologies, including Apache Hadoop, MapReduce, HDFS, and SQL/NoSQL querying Students should complete the Big Data Technology Fundamentals web-based training or have equivalent experience Working knowledge of core AWS services and public cloud implementation Students should complete the AWS Technical Essentials course or have equivalent experience Basic understanding of data warehousing, relational database systems, and database designTeaching methodsProfessional explanation with practical samples and examples.
Teaching materialsAmazon Web Services authorized e-book included
Course outlineNote: course outline may vary slightly based on the regional location and/or language in which the class is delivered.
This course will cover the following concepts on each day:
Day 1
Overview of Big Data
Ingestion, Transfer, and Compression
Storage Solutions
Storing and Querying Data on DynamoDB
Big Data Processing and Amazon Kinesis
Introduction to Apache Hadoop and Amazon EMR
Using Amazon Elastic MapReduce
Day 2
Hadoop Programming Frameworks
Processing Server Logs with Hive on Amazon EMR
Processing Chemistry Data Using Hadoop Streaming on Amazon EMR
Streamlining Your Amazon EMR Experience with Hue
Running Pig Scripts in Hue on Amazon EMR
Spark on Amazon EMR
Interactively Creating and Querying Tables with Spark and Spark SQL on Amazon EMR
Managing Amazon EMR Costs
Securing your Amazon EMR Deployments
Day 3
Data Warehouses and Columnar Datastores
Amazon Redshift and Big Data
Optimizing Your Amazon Redshift Environment
Big Data Design Patterns
Visualizing and Orchestrating Big Data
Using Tibco Spotfire to Visualize Big Data

Who is the course for

- 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

GOPAS Praha
Kodaňská 1441/46
101 00 Praha 10
Tel.: +420 234 064 900-3
info@gopas.cz

GOPAS Brno
Nové sady 996/25
602 00 Brno
Tel.: +420 542 422 111
info@gopas.cz

GOPAS Bratislava
Dr. Vladimíra Clementisa 10
Bratislava, 821 02
Tel.: +421 248 282 701-2
info@gopas.sk

Copyright © 2020 GOPAS, a.s.,
All rights reserved

Big Data on AWS

What we teach you

- 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 and Spark SQL on Amazon EMR
- Choose appropriate AWS data storage options
- Identify the benefits of using Amazon Kinesis for near real-time big data processing
- Define data warehousing and columnar database concepts
- Leverage Amazon Redshift to efficiently store and analyze data
- Comprehend and manage costs and security for Amazon EMR and Amazon Redshift deployments
- Identify options for ingesting, transferring, and compressing data
- Use visualization software to depict data and queries
- Orchestrate big data workflows using AWS Data Pipeline

Required skills

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

Course outline

Note: course outline may vary slightly based on the regional location and/or language in which the class is delivered.
This course will cover the following concepts on each day:
Day 1
Overview of Big Data

Ingestion, Transfer, and Compression

Storage Solutions

Storing and Querying Data on DynamoDB

Big Data Processing and Amazon Kinesis

Introduction to Apache Hadoop and Amazon EMR

Using Amazon Elastic MapReduce
Day 2
Hadoop Programming Frameworks

Processing Server Logs with Hive on Amazon EMR

Processing Chemistry Data Using Hadoop Streaming on Amazon EMR

Streamlining Your Amazon EMR Experience with Hue

Running Pig Scripts in Hue on Amazon EMR

Spark on Amazon EMR

Interactively Creating and Querying Tables with Spark and Spark SQL on Amazon EMR

Managing Amazon EMR Costs

GOPAS Praha
Kodaňská 1441/46
101 00 Praha 10
Tel.: +420 234 064 900-3
info@gopas.cz

GOPAS Brno
Nové sady 996/25
602 00 Brno
Tel.: +420 542 422 111
info@gopas.cz

GOPAS Bratislava
Dr. Vladimíra Clementisa 10
Bratislava, 821 02
Tel.: +421 248 282 701-2
info@gopas.sk

Copyright © 2020 GOPAS, a.s.,
All rights reserved

Big Data on AWS

Securing your Amazon EMR Deployments

Day 3

Data Warehouses and Columnar Datastores

Amazon Redshift and Big Data

Optimizing Your Amazon Redshift Environment

Big Data Design Patterns

Visualizing and Orchestrating Big Data

Using Tibco Spotfire to Visualize Big Data

GOPAS Praha
Kodaňská 1441/46
101 00 Praha 10
Tel.: +420 234 064 900-3
info@gopas.cz

GOPAS Brno
Nové sady 996/25
602 00 Brno
Tel.: +420 542 422 111
info@gopas.cz

GOPAS Bratislava
Dr. Vladimíra Clementisa 10
Bratislava, 821 02
Tel.: +421 248 282 701-2
info@gopas.sk

Copyright © 2020 GOPAS, a.s.,
All rights reserved