

Big Data on AWS

描述

Big Data on AWS 课程将向您介绍基于云的大数据解决方案，例如 Amazon (EMR)、Amazon Redshift、Amazon Kinesis 以及其余的 AWS 大数据平台。在本课程中，我们将向您介绍如何使用 Amazon EMR 来利用诸如 Hive 和 Hue 等广泛 Hadoop 工具体系处理数据。授课内容还包括如何创建大数据环境、如何结合使用 Amazon DynamoDB、Amazon Redshift Amazon Quicksight、Amazon Athena 和 Amazon Kinesis 以及如何利用最佳实践来设计大数据环境，从而实现安全性和经济性。

级别	授课方式	时长
中级	讲师指导、现场授课或通过虚拟课堂授课	3 天

课程目标

通过学习本课程，您将能够：

- 适用大数据体系内的 AWS 解决方案
- 在 Amazon EMR 环境中使用 Apache Hadoop
- 了解 Amazon EMR 群集的组件
- 启动并配置 Amazon EMR 群集
- 利用 Amazon EMR 的常用编程框架，包括 Hive、Pig 和 Streaming
- 利用 Hue 提高 Amazon EMR 的易用性
- 在 Amazon EMR 上借助 Spark 使用内存分析
- 选择适当的 AWS 数据存储选项
- 确定使用 Amazon Kinesis 以近乎实时的速度处理大数据的优势
- 利用 Amazon Redshift 有效地存储和分析数据
- 理解并管理大数据解决方案的费用和安全性
- 保护大数据解决方案
- 确定用于获取、传输和压缩数据的选项
- 利用 Amazon Athena 进行临时查询分析
- 借助 Amazon QuickSight 使用可视化软件描述数据和查询
- 利用 AWS Data Pipeline 编排大数据 workflow

目标人群

本课程适用于：

- 负责设计和实施大数据解决方案的人员，即解决方案架构师
- 数据科学家和数据分析师有兴趣了解 AWS 上大数据解决方案背后的服务和架构模式

先决条件

我们建议参加学习本课程的人员符合以下先决条件：

- 基本熟悉大数据技术，包括 Apache Hadoop、MapReduce、HDFS 和 SQL/NoSQL 查询
- 学员应完成 [Big Data Technology Fundamentals](#) 网络培训或具备同等经验
- 具有核心 AWS 服务和公有云实施工作经验
- 学员应完成 [AWS Technical Essentials](#) 课程或具备同等经验
- 了解数据仓库、关系数据库系统和数据库设计基础

授课方式

本课程将结合以下方式授课：

- 讲师指导培训 (ILT)
- 动手实验室

动手实践活动

本课程允许通过各种实践操作来试验新的技术并将所学知识应用到您的工作环境中

课程大纲

注意：课程大纲可能不尽相同，具体取决于授课的地区位置和/或语言。

第 1 天

- 大数据概述
- 大数据获取和传输
- 大数据流式处理和 Amazon Kinesis
- 实验室 1：使用 Amazon Kinesis 流式处理和分析 Apache 服务器日志数据
- 大数据存储解决方案
- 大数据处理和分析

- 实验室 2：使用 Amazon Athena 查询 Amazon S3 的日志数据

第 2 天

- Apache Hadoop 和 Amazon EMR
- 实验室 3：在 Amazon DynamoDB 上存储并查询数据
- 使用 Amazon EMR
- Hadoop 编程框架
- 实验室 4：在 Amazon EMR 上利用 Hive 处理服务器日志
- Amazon EMR 上的 Web 界面
- 实验室 5：在 Amazon EMR 上的 Hue 中运行 Pig 脚本
- Amazon EMR 上的 Apache Spark
- 实验室 6：在 Amazon EMR 上使用 Spark 处理 NY Taxi 数据

第 3 天

- Amazon Redshift 和大数据
- 大数据的可视化及编制
- 实验室 7：使用 TIBCO Spotfire 可视化数据
- 管理大数据费用
- 保护您的 Amazon 部署
- 大数据设计模式