

Building a Serverless Data Lake

描述

Building a Serverless Data Lake 是为期一天的高级训练营，旨在教您如何使用 AWS 服务设计、构建和运营无服务器数据湖解决方案。本训练营涵盖以下主题：从任何数据源大规模获取数据；安全持久地存储数据；具备使用正确工具处理海量数据的能力；以及了解可用于近乎实时地分析数据的选项。

级别	授课方式	时长
高级	讲师指导的现场课程	1 天

课程目标

通过学习本课程，您将能够：

- 使用 Kinesis Streams 和 Firehose 等服务收集大量数据，并将数据安全持久地存储在 Amazon Simple Storage Service 中。
- 为数据湖创建元数据索引。
- 选择最佳工具来获取、存储、处理和分析数据湖中的数据。
- 将所学知识运用到动手实验中，通过构建完整的解决方案获得实践经验。

目标人群

本课程适用于：

- 解决方案架构师
- 大数据开发人员
- 数据架构师和分析师
- 其他实验数据分析实践人员

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：

- 具备出色的 AWS 核心服务工作经验，包括 Amazon Elastic Compute Cloud (EC2) 和 Amazon Simple Storage Service (S3)

- 具备编程语言或脚本语言工作经验
- 熟悉 Linux 操作系统和命令行界面
- 要求配有笔记本电脑，以完成实验练习 – 不建议使用平板电脑

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

课程大纲

本课程涵盖了以下概念：

- 有助于启用无服务器数据湖架构的关键服务
- 遵循获取、存储、处理和分析工作流程的数据分析解决方案
- 用于实施数据湖解决方案的可重复模板部署
- 构建元数据索引并启用搜索功能
- 利用多个数据源建立大规模数据提取管道
- 使用由事件触发的简单函数转换数据
- 使用最适合该案例的工具和服务进行数据处理
- 可用于更好地分析已处理数据的选项
- 部署与操作最佳实践