

# Building Batch Data Analytics Solutions on AWS

AWS クラスルームトレーニング



## コースの説明

このコースでは、Apache Spark および Apache Hadoop マネージドサービスである Amazon EMR を使用して、バッチデータ分析ソリューションを構築する方法を学習します。Amazon EMR が Apache Hive、Hue、HBase などのオープンソースや、AWS Glue、AWS Lake Formation などの AWS のサービスと、どのように連携するかを学びます。このコースでは、Spark および Hadoop 環境でのデータ収集、取り込み、カタログ化、ストレージ、処理コンポーネントを取り上げます。EMR Notebooks を使用して、分析と Machine Learning の両方のワークロードをサポートする方法を学びます。また、Amazon EMR のオペレーションに関わるセキュリティ、パフォーマンス、コスト管理のベストプラクティスも学びます。

| レベル | 実施形式                        | 所要時間 |
|-----|-----------------------------|------|
| 中級  | インタラクティブデモ、ハンズオンラボ、ディスカッション | 1 日間 |

## コースの目標

このコースで学習する内容:

- データウェアハウス、データレイク、モダンデータアーキテクチャの機能と利点を比較する
- Amazon EMR を用いたバッチデータ分析ソリューションを設計して実装する
- 圧縮などのデータストレージの最適化に適した手法を見極めて適用する
- データを取り込み、変換、保存するための適切なオプションを選択し、デプロイする
- 特定のビジネスユースケースに適したインスタンスおよびノードのタイプ、クラスター、オートスケーリング、ネットワークトポロジを選択する
- データの保存と処理が、実用的なビジネスインサイトを抽出するために必要な分析と可視化のメカニズムにどのように影響するかを理解する
- 保管中および転送中のデータを保護する
- 分析ワークロードをモニタリングし、問題を特定して修復する
- コスト管理のベストプラクティスを適用する

## 対象者

このコースは以下のような方を対象としています。

- データプラットフォームエンジニア
- データ分析パイプラインを構築し、管理するアーキテクトおよびオペレーター

# Building Batch Data Analytics Solutions on AWS

AWS クラスルームトレーニング



## 前提条件

このコースは、Apache Spark、Apache Hadoop などのオープンソースデータ分析フレームワークの管理経験が 1 年以上ある受講者向けです。このコースを受講するにあたっては、次のことを身につけておくことをお勧めします。

- AWS Technical Essentials または Architecting on AWS の受講修了
- Building Data Lakes on AWS の受講修了

## 登録

<https://www.aws.training/training/schedule?courseId=91270&countryName=JP&trainingProviderId=1>

## コースの概要

### モジュール 1: Amazon EMR の概論

- 分析ソリューションでの Amazon EMR の使用
- Amazon EMR クラスターのアーキテクチャ
- インタラクティブデモ 1: Amazon EMR クラスターの起動
- コスト管理戦略

### モジュール 2: Amazon EMR を使用したデータ分析パイプライン: 取り込みとストレージ

- Amazon EMR のストレージの選択肢
- データ取り込み手法

### モジュール 3: Amazon EMR の Apache Spark を使用した高パフォーマンスのバッチデータ分析

- Amazon EMR の Apache Spark のユースケース
- Amazon EMR の Apache Spark のメリット
- Spark のコンセプト
- インタラクティブデモ 2: Spark シェルを使用して EMR クラスターに接続し、Scala コマンドを実行する
- 変換、処理、分析
- Amazon EMR でのノートブックの使用
- ラボ 1: Amazon EMR の Apache Spark を使用した低レイテンシーのデータ分析

### モジュール 4: Amazon EMR と Apache Hive を使用したバッチデータの処理と分析

- Amazon EMR と Hive を使用したバッチデータ処理
- 変換、処理、分析
- ラボ 2: Amazon EMR と Hive を使用したバッチデータ処理
- Amazon EMR での Apache HBase の概論

# Building Batch Data Analytics Solutions on AWS

**AWS クラstrームトレーニング**



## **モジュール 5: サーバーレスデータ処理**

- サーバーレスデータの処理、変換、分析
- Amazon EMR ワークロードでの AWS Glue の使用
- ラボ 3: AWS Step Functions を使用した Spark でのデータ処理のオーケストレーション

## **モジュール 6: Amazon EMR クラスターのセキュリティおよびモニタリング**

- EMR クラスターの保護
- インタラクティブデモ 3: EMRFS を使用したクライアント側の暗号化
- Amazon EMR クラスターのモニタリングとトラブルシューティング

## **モジュール 7: バッチデータ分析ソリューションの設計**

- バッチデータ分析のユースケース
- アクティビティ: バッチデータ分析ワークフローの設計