

# Data Warehousing 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、ペタバイト規模の AWS のデータウェアハウスである Amazon Redshift を使用して、クラウドベースのデータウェアハウスソリューションを設計するための概念、戦略、およびベストプラクティスを紹介します。このコースでは、Amazon DynamoDB、Amazon EMR、Amazon Kinesis、Amazon S3 といった AWS のサービスを用いてデータウェアハウスに関連するデータの収集、保存、作成を行う方法について取り上げます。また、データ分析と可視化を行うため、Amazon QuickSight も紹介します。

レベル	実施形式	所要時間
中級	クラスルームトレーニング、ハンズオンラボ、グループ演習	3 日間

### コースの目標

このコースの学習内容は、以下のとおりです。

- データウェアハウスの主要概念、およびデータウェアハウスとビッグデータソリューションの共通点について説明する
- クラウドでデータウェアハウスを実装するために、Amazon Redshift クラスターを起動して、コンポーネント、特性、および機能を使用する
- データウェアハウスソリューションに役立つ Amazon DynamoDB、Amazon EMR、Amazon Kinesis、Amazon S3 など、AWS の他のデータサービスおよび分析サービスを使用する
- データウェアハウスを設計する
- パフォーマンスを向上させるために、パフォーマンスの問題を特定し、クエリを最適化し、データベースを調整する
- Amazon Redshift Spectrum を使用して、Amazon S3 バケットのデータを直接分析する
- Amazon QuickSight を使用して、データウェアハウスに対するデータ分析および可視化タスクを実行する

### 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。

# Data Warehousing on AWS

## AWS クラスルームトレーニング

- データベースアーキテクト、データベース管理者、データベース開発者、データアナリスト、データサイエンティスト

### 前提条件

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。

- リレーショナルデータベースおよびデータベース設計の基本概念についての知識
- EC2 や VPC、S3 などの基本的な AWS サービスに関する知識

### 登録

<https://www.aws.training/training/schedule?courseId=10025&countryName=JP&trainingProviderId=1>

### コースの概要

#### 1 日目

##### モジュール 1: データウェアハウジングの概論

- リレーショナルデータベース
- データウェアハウスの概念
- データウェアハウスとビッグデータの交点
- AWS におけるデータ管理の概要

##### モジュール 2: Amazon Redshift の概論

- コンセプトの概要
- ユースケース
- ハンズオンラボ 1: Amazon Redshift の概論

##### モジュール 3: クラスターの起動

- クラスターの構築
- クラスターへの接続
- ハンズオンラボ 2: Amazon Redshift クラスターを起動する

# Data Warehousing on AWS

## AWS クラスルームトレーニング

### 2 日目

#### モジュール 4: データベーススキーマの設計

- スキーマとデータ型
- 列指向の圧縮
- データ分散スタイル
- ハンズオンラボ 3: データベーススキーマの最適化

#### モジュール 5: データソースの特定

- データソースの概要
- Amazon S3
- 大量のデータの移動
- AWS Glue
- Amazon EMR
- Amazon DynamoDB
- Amazon Kinesis Data Firehose
- リモートホストから SSH 接続
- レガシーデータウェアハウスと Schema Conversion Tool
- ハンズオンラボ 4: リアルタイムデータを Amazon Redshift データベースにロードする

#### モジュール 6: データのロード

- データを準備する
- COPY コマンドを使用してデータをロードする
- テーブルを管理する
- 同時書き込みオペレーション
- ロードに関する問題のトラブルシューティング
- ハンズオンラボ 5: COPY コマンドを使ってデータをロードする

### 3 日目

#### モジュール 7: クエリの記述とパフォーマンスのチューニング

- Amazon Redshift SQL
- ユーザー定義関数 (UDF)
- クエリのパフォーマンスに影響する要素
- EXPLAIN コマンドとクエリプラン
- ワークロード管理 (WLM)

# Data Warehousing on AWS

## AWS クラスルームトレーニング

- ハンズオンラボ 6: ワークロード管理を設定する

### モジュール 8: Amazon Redshift Spectrum

- データの「ダーク」サイド
- Amazon Redshift Spectrum
- Amazon Redshift Spectrum のデータ設定
- Amazon Redshift Spectrum クエリ
- ハンズオンラボ 7: Amazon Redshift Spectrum を使用

### モジュール 9: クラスターの保守

- 監査ログ記録
- パフォーマンスモニタリング
- イベントと通知
- ハンズオンラボ 8: クラスターの監査およびモニタリング
- クラスターのサイズ変更
- クラスターのバックアップと復元
- リソースのタグ付け
- 上限と制約
- 高可用性と災害対策
- ハンズオンラボ 9: クラスターのバックアップ、復元、サイズ変更

### モジュール 10: データの分析と可視化

- 可視化がもたらす効果
- ダッシュボードの構築
- Amazon QuickSight のエディションと機能