

Data Warehousing on AWS

Duration: 3 days

Audience: Professionnel TI|Analyste de données

Who is this training for?

- Database architects
- Database administrators
- Database developers
- Data analysts
- Data scientists

Training objectives

- Evaluate the relationship between Amazon Redshift and other big data systems
- Evaluate use cases for data warehousing workloads and review case studies of implementing AWS data and analytics services as part of a data warehousing solution
- Choose an Amazon Redshift node of the right type and size for your data needs
- Understand the right security features for Amazon Redshift, such as encryption, IAM permissions, and database permissions
- Launch an Amazon Redshift cluster and use the components, features, and functionality to create a cloud data warehouse
- Use other AWS data and analytics services, such as Amazon DynamoDB, Amazon EMR, Amazon Kinesis Firehose, and Amazon S3, to support the data warehousing solution
- Evaluate approaches and methodologies for designing data warehouses
- Identify data sources and evaluate requirements for data warehouse design
- Design the data warehouse to make good use of data compression, distribution, and sorting procedures
- Load, unload and perform data maintenance
- Write queries and evaluate query plans to optimize data performance
-

Configure the database to assign resources, such as memory, for query queues and define criteria to route certain types of queries to your configured queues for better processing

- Control, monitor, and receive event notifications about data warehouse activities using features and services such as Amazon Redshift database audit logging, Amazon CloudTrail, Amazon CloudWatch, and Amazon Simple Notification Service (Amazon SNS)
- Prepare operational tasks, such as resizing Amazon Redshift clusters and using snapshots to back up and restore clusters
- Use a business intelligence (BI) application to perform analysis and visualization operations on your data

Summary

In this course, you'll learn new concepts, strategies, and best practices for designing a cloud data warehousing solution with Amazon Redshift, AWS's petabyte-scale data warehouse. We'll look at how to collect, store, and prepare data for a data warehouse using other AWS services, such as Amazon DynamoDB, Amazon EMR, Amazon Kinesis Firehose, and Amazon Simple Storage Service (Amazon S3). We'll also cover using business intelligence (BI) tools to perform analysis on your data.

Course outline

- Evaluate the relationship between Amazon Redshift and other big data systems
- Evaluate use cases for data warehousing workloads and review case studies of implementing AWS data and analytics services as part of a data warehousing solution
- Choose an Amazon Redshift node of the right type and size for your data needs
- Understand the right security features for Amazon Redshift, such as encryption, IAM permissions, and database permissions
- Launch an Amazon Redshift cluster and use the components, features, and functionality to create a cloud data warehouse
- Use other AWS data and analytics services, such as Amazon DynamoDB, Amazon EMR, Amazon Kinesis Firehose, and Amazon S3, to support the data warehousing solution
- Evaluate approaches and methodologies for designing data warehouses

- Identify data sources and evaluate requirements for data warehouse design
- Design the data warehouse to make good use of data compression, distribution, and sorting procedures
- Load, unload and perform data maintenance
- Write queries and evaluate query plans to optimize data performance
- Configure the database to assign resources, such as memory, for query queues and define criteria to route certain types of queries to your configured queues for better processing
- Control, monitor, and receive event notifications about data warehouse activities using features and services such as Amazon Redshift database audit logging, Amazon CloudTrail, Amazon CloudWatch, and Amazon Simple Notification Service (Amazon SNS)
- Prepare operational tasks, such as resizing Amazon Redshift clusters and using snapshots to back up and restore clusters
- Use a business intelligence (BI) application to perform analysis and visualization operations on your data

Approach and methodology

- Hands-on experience in data management and analysis Familiarity with BI or visualization tools (Power BI, Tableau, etc.)
- Knowledge of modern data architectures (lake, warehouse)
- Notions of AWS data services (Redshift, Glue, Athena – asset)
- Understanding of performance, scalability and data quality issues

Prerequisites

- Basics in relational databases (SQL)
- Understanding of data warehousing concepts (modeling, ETL/ELT, schemas)
- Basic knowledge of cloud environments
- Familiarity with AWS services (S3, Redshift, Glue – asset)
- Notions of data analysis and BI (asset)

Recommendations

- Basics in relational databases (SQL)
- Understanding of data warehousing concepts (modeling, ETL/ELT, schemas)
- Basic knowledge of cloud environments
- Familiarity with AWS services (S3, Redshift, Glue – asset)
- Notions of data analysis and BI (asset)