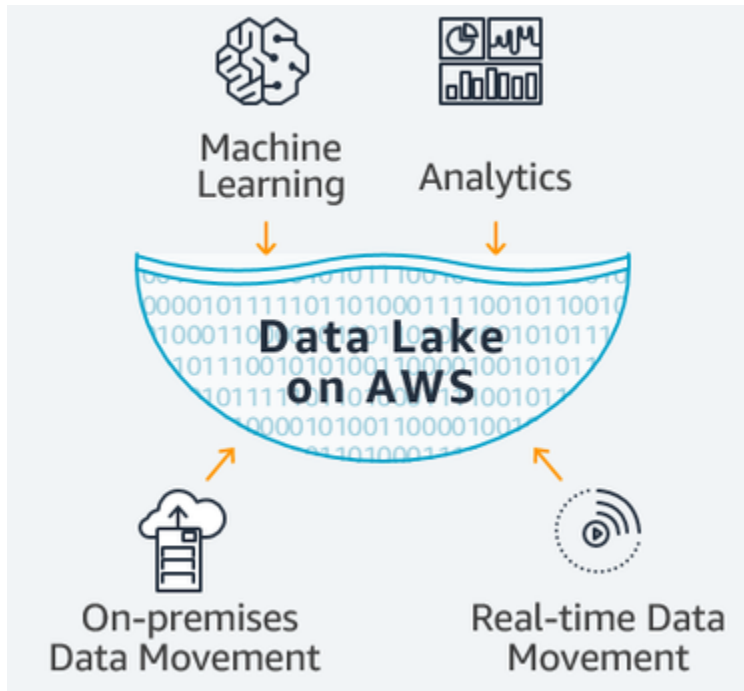


AWS Solutions, Tools & Tutorials for Data Lakes



Data Movement

Import your data from on premises, and in real-time.

Analytics

Analyze your data with the broadest selection of analytics services.

Data Lake

Store any type of data securely, from gigabytes to exabytes.

Machine Learning

Predict future outcomes, and prescribe actions for rapid response.

Step-by-Step Solutions on AWS

[Data Lakes and Analytics on AWS](#) - Amazon Web Services (AWS) delivers an integrated suite of services that provide everything needed to quickly and easily build and manage a data lake for analytics. For various AWS tool usage in the four areas of a data lake click [here](#).

[Data Lake Formation with AWS](#) - AWS Lake Formation is a service that makes it easy to set up a secure data lake in days. A data lake is a centralized, curated, and secured repository that stores all your data, both in its original form and prepared for analysis. A detailed presentation on how the Data Lake Formation solution works can be found [here](#).

[Data Lake Implementation Solution on AWS](#) - AWS offers the data lake solution, which is an automated reference implementation that deploys a highly available, cost-effective data lake architecture on the AWS Cloud along with a user-friendly console for searching and requesting datasets.

Data Ingestion

Ingest - taking in data that goes into the data lake

[AWS Direct Connect](#) - Establish a dedicated network connection to AWS. [more info](#)

[Amazon Virtual Private Cloud \(VPC\)](#)

[How do I configure a VPN over AWS Direct Connect?](#)

[Create a VPC and connect with on premise with Direct Connect](#)

[AWS Snowball](#) - Migrate large datasets using secure devices. [more info](#)

[Using AWS Snowball Edge and AWS DMS for Database Migration](#)

[How to connect to and transfer files to a snowball edge](#)

[AWS Snowball tutorial Import/Export Service](#)

Amazon Kinesis Data Firehose - Easily and reliably stream data into data lakes. [more info](#)

[Batch Data Processing with AWS Kinesis Firehose and S3](#)

[AWS Kinesis Firehose to S3 Tutorial](#)

[AWS Kinesis Firehose tutorial using Python](#)

[AWS - Kinesis Data Stream - Hands On!](#)

AWS Database Migration Service - Migrate databases to AWS quickly and securely. [more info](#)

[AWS Database Migration Service](#)

[AWS DMS - Database Migration Service demo](#)

[AWS Database Migration Service With Replication Tutorial](#)

Catalog & Search

Persist - this is typically where our persisted data is indexed, cataloged and set up to be stored

AWS Glue - Store metadata in a data catalog. [more info](#)

[What is AWS Glue?](#)

[Tutorial AWS Glue read dataset from S3, etl jobs, workflow, athena SQL](#)

[- Build a data lake foundation with AWS Glue and Amazon S3](#)

[- How to ETL for analytic processing using AWS glue part 1](#)

[- How to ETL for analytic processing using AWS glue part 2](#)

[- Harmonize, Query, and Visualize Data from Various Providers using AWS Glue, Amazon Athena, and Amazon QuickSight](#)

[AWS Glue Tutorial | Getting Started with AWS Glue ETL](#)

[How to run python scripts for ETL in AWS glue?](#)

Move & Transform

Process - where our data is processed or transformed and moved to into storage or analytics

AWS Lambda - Event-driven, serverless computing. [more info](#)

[Introduction to AWS Lambda - Serverless Compute on Amazon Web Services](#)

[AWS Lambda Tutorial For Beginners | What is AWS Lambda?](#)

[Learning AWS Lambda & Serverless Applications](#)

[- AWS Lambda & DynamoDB - AWS Serverless Part I](#)

Amazon Kinesis Data Firehose - Easily and reliably stream data into data lakes. [more info](#)

[Batch Data Processing with AWS Kinesis Firehose and S3](#)

[AWS Kinesis Firehose to S3 Tutorial](#)

[AWS Kinesis Firehose tutorial using Python](#)

AWS Storage Gateway - Gain on-premises access to AWS cloud storage. [more info](#)

[AWS Storage Gateway](#)

[Cloud Storage in Minutes with AWS Storage Gateway](#)

[How to setup AWS Storage Gateway on VMware for file share and S3 backup](#)

Amazon S3 - Secure, highly scalable, durable object storage with millisecond latency for data access. [more info](#)

[Introduction to Amazon Simple Storage Service \(S3\)](#)

[Setting up Amazon Web Services \(AWS\) S3 Bucket and IAM User](#)

[AWS S3 CLI Tutorial: AWS S3 CLI and S3API Basic Commands](#)

Amazon DynamoDB - Fully managed NoSQL database service. [more info](#)

[AWS DynamoDB Tutorial for Beginners](#)

[AWS DynamoDB Tutorial - Beginner to Expert](#)

[AWS DynamoDB Schema Design | How to choose the right key](#)

[- AWS Lambda & DynamoDB - AWS Serverless Part I](#)

Amazon Elasticsearch Service - Fully managed Elasticsearch service. [more info](#)

[Amazon AWS \(ES\) Elastic Search Tutorial](#)

[Elasticsearch - Logstash - Kibana 4 \(ELK Stack\) Setup Tutorial](#)

[Elasticsearch Tutorial | Getting Started with Elasticsearch | ELK Stack Training](#)

AWS Glue - Fully managed ETL service. [more info](#)

[What is AWS Glue?](#)

[Tutorial AWS Glue read dataset from S3, etl jobs, workflow, athena SQL](#)

[- Build a data lake foundation with AWS Glue and Amazon S3](#)

[- How to ETL for analytic processing using AWS glue part 1](#)

[- How to ETL for analytic processing using AWS glue part 2](#)

[- Harmonize, Query, and Visualize Data from Various Providers using AWS Glue, Amazon Athena, and Amazon QuickSight](#)

[AWS Glue Tutorial | Getting Started with AWS Glue ETL](#)

[How to run python scripts for ETL in AWS glue?](#)

[AWS - Kinesis Data Stream - Hands On!](#)

Analytics & Serving

Get insights - reports, dashboards and visualizations can be taken from here

Amazon Redshift - Fast, simple, cost- effective data warehousing service. [more info](#)

[AWS Redshift Concept](#)

[Hands-On Amazon Redshift for Data Warehousing - 6 short tutorials](#)

[Amazon Redshift Tutorial | AWS Tutorial for Beginners](#)

Amazon QuickSight - Fast, cloud-powered business intelligence service. [more info](#)

[Amazon QuickSight - Overview](#)

[Amazon \(AWS\) QuickSight Basic Tutorial](#)

[Tutorial AWS: QuickSight BI Tool with Athena and S3](#)

[QuickSight - Calculated Field, Parameters, Actions & Filter](#)

[QuickSight Dashboard Tutorial & Overview](#)

Amazon DynamoDB - Fully managed NoSQL database service. [more info](#)

[AWS DynamoDB Tutorial for Beginners](#)

[AWS DynamoDB Tutorial - Beginner to Expert](#)

[AWS DynamoDB Schema Design | How to choose the right key](#)

[- AWS Lambda & DynamoDB - AWS Serverless Part I](#)

Amazon EMR - Run & Scale Spark, Hadoop, and other Big Data Frameworks

[An introduction to Amazon EMR - Amazon Web Services](#)

[EMR Training - 11 short tutorials](#)

[Amazon EMR Masterclass](#)

Amazon Elasticsearch Service - Fully managed Elasticsearch service. [more info](#)

[Amazon AWS \(ES\) Elastic Search Tutorial](#)

[Elasticsearch - Logstash - Kibana 4 \(ELK Stack\) Setup Tutorial](#)

[Elasticsearch Tutorial | Getting Started with Elasticsearch | ELK Stack Training](#)

Amazon RDS - Distributed relational database service. [more info](#)

[Getting Started with Amazon RDS - Relational Database Service on AWS](#)

[AWS RDS MySQL Database Setup | Step by Step Tutorial](#)

[AWS RDS Tutorial for Beginners 2020: Hands-on Tutorial](#)

Access & User Interfaces

Amazon Athena - Serverless, interactive query service. [more info](#)

[AWS Athena Tutorial | Athena Hands On LAB](#)

[AWS Athena Tutorial - How to Use Amazon Athena in AWS?](#)

[AWS Athena Data Lake Tutorial: Create AWS S3 using AWS Glue](#)

[How to use SQL to Query S3 JSON files with AWS Athena](#)

AWS Glue - Store metadata in a data catalog. [more info](#)

[What is AWS Glue?](#)

[Tutorial AWS Glue read dataset from S3, etl jobs, workflow, athena SQL](#)

[- Build a data lake foundation with AWS Glue and Amazon S3](#)

[- How to ETL for analytic processing using AWS glue part 1](#)

[- How to ETL for analytic processing using AWS glue part 2](#)

[- Harmonize, Query, and Visualize Data from Various Providers using AWS Glue, Amazon Athena, and Amazon QuickSight](#)

[AWS Glue Tutorial | Getting Started with AWS Glue ETL](#)

[How to run python scripts for ETL in AWS glue?](#)

AWS Direct Connect - Establish a dedicated network connection to AWS. [more info](#)

[Amazon Virtual Private Cloud \(VPC\)](#)

[How do I configure a VPN over AWS Direct Connect?](#)

[Create a VPC and connect with on premise with Direct Connect](#)

Amazon Neptune - Fully managed Graph database service. [more info](#)

[Amazon Neptune: Build Applications for Highly Connected Datasets](#)

[Amazon AWS Neptune Graph Database \(AWS Tutorial\)](#)

[Nike: A Social Graph at Scale with Amazon Neptune](#)

Data governance - enforce policies

AWS AppSync - Manage and synchronize mobile app data in real time across devices and users. [more info](#)

[AWS AppSync Introduction](#)

[Creating Your First AWS AppSync GraphQL API](#)

[Creating & Interacting with an AWS AppSync GraphQL API with React & AWS Amplify](#)

Amazon Cognito - Add user sign-up, sign-in, and access control to your web and mobile apps quickly and easily. [more info](#)

[Introduction to Amazon Cognito - User Authentication and Mobile Data Service on AWS](#)

[AWS Cognito Tutorial Part I | Cognito User Pool & AWS Amplify setup](#)

[AWS Cognito Tutorial Part II | Sign in & registration](#)

[AWS Cognito Tutorial Part III | Session persistence & change password](#)

Manage & Secure

Data governance - enforce policies

AWS KMS - Manage cryptographic keys and control their use across services. [more info](#)

AWS CloudTrail - Enable governance, compliance, operational auditing, and risk auditing. [more info](#)

Amazon API Gateway - Fully managed service for creating, publishing, maintaining, and monitoring secure APIs at scale. [more info](#)

[API Gateway With Lambda Function Example](#)

[How to build an API with Lambdas and API Gateway](#)

[AWS API Gateway Tutorial - Input Output Mapping](#)

AWS IAM - Securely manage access to AWS services and resources. [more info](#)

Amazon CloudWatch - Monitor your AWS resources and the applications you run on AWS in real time. [more info](#)