Get started

Data Lake Foundation on AWS with Apache Zeppelin

Using Apache Zeppelin, Amazon S3, Amazon RDS, and other AWS services




Challenges

Organizations often struggle with the rapid collection of massive volumes of different types of data, from various sources, and storing it efficiently in a way that makes sense for their business. Frequently, this data is stored and distributed across various locations, preventing organizations from having a single source of truth. Additionally, disparate data sources create challenges for IT teams trying to apply multiple analytics and processing frameworks to the same data. In order to overcome these challenges, organizations need a flexible, cost-effective, and scalable storage solution for their data.

Data Lake Foundation on AWS

A data lake on Amazon Web Services (AWS) provides organizations with the ability to store data of any type (structured and unstructured) in a centralized repository. Raw data can be stored, monitored, and analyzed, without having to convert it to a predefined schema beforehand. By storing data in its native format, you are able to accommodate any schema requirements or design changes that may be required in the future. Building a data lake foundation on AWS enables you to separate your storage and compute, which means you can scale out each component individually, as necessary.

A data lake foundation is deployed to integrate with various AWS components to help you migrate your data from your on-premises data center to a data lake on AWS. The data lake foundation relies heavily on Amazon Simple Storage Service (Amazon S3) as the core service in which to store data.

Solutions such as Kibana and Apache Zeppelin can be leveraged on the data lake foundation for visualizing the data stored in Amazon Simple Storage Service (Amazon S3). Kibana serves as a web interface for Amazon Elasticsearch Service (Amazon ES) to provide visualization capabilities for content indexed on an Amazon ES cluster. Apache Zeppelin, when deployed in an AWS data lake provides data ingestion, analysis, and visualization capabilities.

Leverage AWS-hosted data lakes for a variety of use cases, including:

Data Ingestion

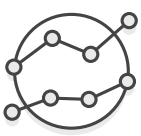
Ingest, store, and analyze original data sets, regardless of native format, without converting to a predefined schema.

Lower TCO

As the amount of data captured grows exponentially, organizations will see a reduction in overall analytics costs.

Data Integration

Integrate and analyze data originating from disparate storage services



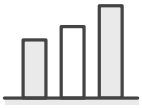
Multiple Analytics Engines

AWS and APN partners offer multiple analytics engines and processing frameworks, leveraging the data stored in Amazon S3

Customer Ready Solutions

Discover scalable solutions that help you achieve your business needs through a combination of AWS services and APN Partners that have attained AWS Competency designations. Based on architectures validated by AWS to accelerate your cloud transformation, you can deploy solutions quickly with AWS Quick Starts and optional Jumpstart consulting offers provided by APN Partners.
Visit here for more information.

Benefits of Data Lake Foundation on AWS with Apache Zeppelin and Kibana



Pre-Build Aggregations and Filters

Easily leverage pre-built aggregations and filters to display data at a higher level, or drill down to specific details.

Interactive Charts

Drag and drop data to create interactive charts and reports, then quickly navigate the vast amounts of data.

Easy Distribution of Dashboards

Collaborate with others by sharing your Kibana dashboard or Zeppelin notebook to easily display your results.

Customize Analysis Results

Analyze both unstructured and structured data generated from a variety of applications to get insights specific to your needs.

Data Lake Foundation on AWS with Apache Zeppelin, Amazon RDS, and other AWS Services Quick Start

Together, Cloudwick Technologies Inc. and AWS, have collaborated to develop a Quick Start reference deployment guide that provides step-by-step instructions for deploying a data lake foundation on AWS. Leverage a Quick Start to help migrate your data from your on-premises environment to an AWS data lake stored on Amazon S3. A data lake portal is also deployed, in which you can upload and download files from your data lake repository, monitor real-time streaming data with Amazon Kinesis Firehose, analyze and explore your data, and check your cloud resources. Additional services such as Amazon Relational Database Service (Amazon RDS), AWS Data Pipeline, Amazon Redshift, AWS CloudTrail, and Amazon Elasticsearch Services (Amazon ES) can also be leveraged during deployment to help you get the most out of your data lake hosted on Amazon S3.

The Quick Start also deploys services like Apache Zeppelin and Kibana to help you analyze and visualize your data, once migrated to Amazon S3. Apache Zeppelin is an open source tool, with an easy to navigate web-based notebook that streamlines the data discovery and collaboration needed to effectively visualize your data.

Kibana is a data visualization tool specializing in analytics and monitoring. With use cases ranging from log and time series analytics to IT operations monitoring, Kibana provides you with a multitude of benefits when leveraged within your data lake foundation.