

Cloudwick

AWS Data Lake Quickstart

Building a Data Lake on AWS

A data lake is a repository that holds a vast amount of raw data in its native format, including structured, semi-structured, and especially unstructured data, which represents approximately 80% of enterprise data generated. It leverages highly redundant, low-cost storage and provides more flexibility than traditional data management platforms.

Enterprise data lakes enable companies to acquire, store and explore large, disparate data sets, providing users with a broader view of customer, supplier, and employee interactions for a highly optimized business model. But building a data lake that enables users to find what they need quickly and easily without endless data preparation requires proven architectures, solution frameworks and data management practices. With unmatched expertise and extensive experience architecting, scaling and managing enterprise big data solutions, including data lakes, Cloudwick can help.

Approach

Cloudwick engages with clients to understand business and technology requirements for the Data Lake solution and delivers a statement of work along with a high level implementation architecture.

Deliverable

Cloudwick will provision and configure a functional Data Lake solution on your AWS account with the following features:

- Cost optimized collection and storage of data from two (2) sources at scale
- Catalog and search functionality for data in S3
- Data pipeline definition using Elastic MapReduce (EMR) with Spark
- One schema design in AWS Redshift for data warehousing

Effort

This is a four-week engagement using a combination of onsite and remote consultants.

Expertise

- Largest team of certified Spark developers
- Largest team of certified data scientists, administrators and engineers
- Hundreds of thousands of open source and cloud production hours
- Most experience building, monitoring and managing clusters on-premise and in the cloud
- Proven success transforming global enterprises like Target, Walmart, JP Morgan, Nike, 3M, Safeway, Verizon and Nike into leading digital businesses

Certifications

- AWS
- Cloudera
- Confluent
- Databricks
- DataStax
- DataTorrent
- Elastic
- Hortonworks
- MapR
- and more

About Cloudwick

Cloudwick is an AWS Advanced Consulting Partner with the AWS Big Data Competency. It is one of the largest enterprise Big Data-as-a-Service providers with extensive expertise, managing over 50,000 clusters on AWS. Cloudwick has years of experience architecting, scaling, and managing production enterprise big data solutions, including data lakes.

Building a Data Lake on AWS

