

# Data Lake Foundation on AWS

Using AWS services, including Amazon S3, Amazon Redshift, Amazon QuickSight, and Amazon SageMaker



## Challenges

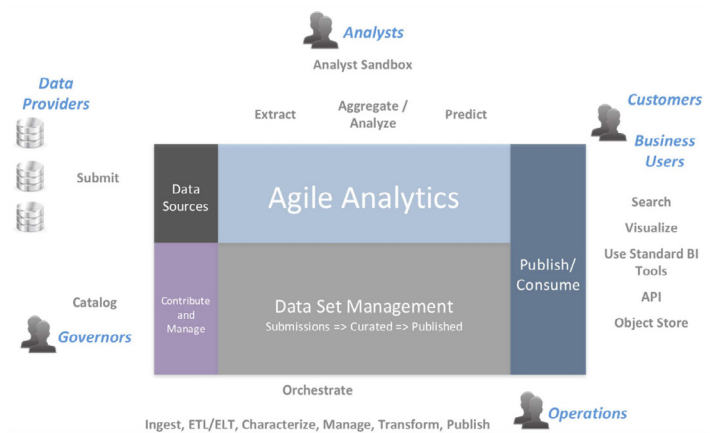
Organizations are often challenged to access and analyze their data when its stored in different formats across various, disparate locations. With data continuously being collected from a variety of sources, if not properly addressed, this challenge only grows bigger as organizations age and grow. A lack of simplified access to data creates workflow bottlenecks as employees often need to request help from IT to access this information.

To address these challenges, organizations are forced to seek out a solution that offers a single source of truth, that is readily accessible when, and to the employees that need it. This is often difficult and costly to execute and maintain as an on-premises solution.

## Data Lake Foundation on AWS

The data lake foundation on Amazon Web Services (AWS) solution enables organizations to store and analyze any type of data (structured or unstructured) in a single, centralized repository. Data can easily be monitored, stored, and analyzed without the need to convert it beforehand.

This data lake solution integrates with a variety of diverse AWS services including Amazon Simple Storage Service (S3), Amazon RedShift, and Amazon Kinesis to provide a fully functional data lake (see figure). The data lake foundation is designed to provide you with a single source of truth, with additional features including data submission, ingest processing, aggregation, analysis, and searching capabilities. A data lake foundation on AWS enables you to easily collect and ingest data, storing it in a cost-effective facility. A portfolio of descriptive, predictive, and teal-time agile analytics built on this data lake foundation can help you answer your most important business questions. Additionally,



you gain quick and easy access to the ecosystem of AWS services and APN partner services that enhance and complement your big data solutions.

Once the foundational layer of your data lake has been deployed, you can leverage various AWS services to complement your data lake with big data solutions.

## Solution Space

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



### Data Submission Cost-Effective Data Storage:

Batch submissions to Amazon S3 and streaming submissions



### Ingest Processing

Data validation, metadata extraction, and indexing via Amazon S3 events



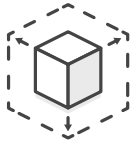
### Search Capabilities

Index metadata in Amazon Elasticsearch Service (Amazon ES) and expose it via Kibana dashboards



### Dataset Management

Performed by Amazon Redshift transformations and Amazon Kinesis Analytics



### Data Transformation

Transformation, aggregation, and analysis via Amazon Athena



### Publish & Visualize

Store data in an Amazon S3 bucket for use by visualization tools

## Data Lake Foundation on AWS Quick Start

47 Lining, an APN partner with Big Data Competency designation, and AWS have collaborated together to create an AWS Quick Start, enabling you to deploy a data lake foundation that integrates with Amazon S3, Amazon Redshift, Amazon Kinesis, Amazon Athena, Amazon ES, and Amazon QuickSight. When the Quick Start launches, it configures and runs the AWS compute, network, and storage necessary to deploy workloads on AWS.

This ready-to-deploy architecture leverages AWS CloudFormation templates to automate the deployment and provide you with the ability to

customize your solution as needed. Once completed, you gain a secure, scalable, and highly available data storage system. This Quick Start also includes an optional, sample dataset that can be loaded into Amazon Redshift, to demonstrate the analytics, querying, and visualization capabilities that come with a data lake foundation on AWS. After you setup this foundational data lake layer, you may choose to augment the data lake with ISV and Software-as-a-Service (SaaS) tools.

Get Started with a Data Lake Foundation on AWS here:  
[Data Lake Foundation on AWS Quick Start](#)



About AWS: For 10 years, Amazon Web Services has been the world's most comprehensive and broadly adopted cloud platform. AWS offers over 100 services for compute, storage, databases, analytics, mobile, Internet of Things (IoT) and enterprise applications from 49 Availability Zones (AZs) across 18 geographic regions in the United States, Canada, Europe, Asia, Australia and South America. AWS services are trusted by more than a million active customers around the world – including the fastest growing startups, largest enterprises, and leading government agencies – to power their infrastructure, make them more agile, and lower costs. To learn more about AWS, visit <http://aws.amazon.com>.

© 2018, Amazon Web Services, Inc. or its affiliates. All rights reserved.