

Hybrid Data Lake with WANdisco Fusion




Challenges

Maintaining synchronized data between on-premises and cloud environments is crucial to keeping business processes up to date. Yet traditional, on-premises data storage tools cannot support bulk transfers of data to the cloud without disrupting business processes. Legacy tools are often incapable of supporting data centers in a hybrid cloud environment, as they cannot move actively changing production data between on-premises and cloud environments.

Data Lake Foundation on AWS

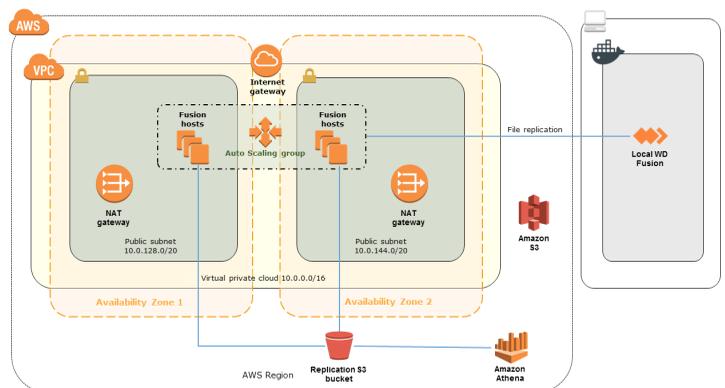
A data lake on AWS is designed to store any type of data (structured and unstructured) in a centralized repository. Data can be stored, monitored, and analyzed without the need to convert it beforehand.

A hybrid AWS data lake offers the features of an AWS data lake, while incorporating on-premises Hadoop clusters with WANdisco Fusion. This allows Apache Hadoop deployments to replicate Hadoop Distributed File System (HDFS) data between clusters that are running various versions of Hadoop.

WANdisco offers fast, active data transfers between any Fusion supported on-premises environment and Amazon Simple Storage Service (Amazon S3) or Amazon Elastic MapReduce (Amazon EMR) to help maintain data consistency. These continuous data transfers enable your environment to keep up with on-premises ingest rates, without causing disruptions.

When leveraged with an AWS hybrid data lake, WANdisco Fusion provides:

- Virtual file system for Hadoop that is compatible with all Hadoop applications
- Single, virtual namespace to integrate storage from different types of Hadoop deployments
- Globally distributed storage
- WAN replication using WANdisco Fusion Active Data Replication



Customer Ready Solutions

For additional information including consulting offers and technical details, visit the [Solutions Page](#).

Benefits of Data Lake Foundation



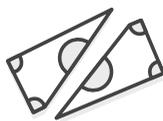
Data Consistency

AWS and on-premises environments stay continuously in sync with automated recovery



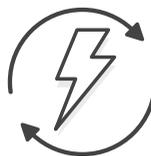
Minimal Downtime and Disruption

Both AWS and on-premises environments operate in parallel during migration allowing workloads to move in phases without disruptions



Low Cost

Analyze data with Amazon Athena and access a range of applications that would be costly and difficult to deploy in house



Disaster Recovery

WANdisco Fusion enables Amazon S3 to be easily configured for on-premises disaster recovery

Data Lake Foundation on AWS Quick Start

WANdisco, Sturdy Networks and AWS have collaborated to develop a Quick Start that deploys a hybrid cloud architecture to integrate on-premises Hadoop clusters with an AWS data lake. The deployment leverages WANdisco Fusion, Amazon S3, and Amazon Athena, and can be set up and running with AWS CloudFormation templates in a matter of minutes.

The hybrid data lake architecture combines on-premises components with AWS components to support burst-out processing in the cloud and cloud migration. An optional Docker container can also be deployed to represent your on-premises Hadoop cluster for demonstration purposes, helping you gain hands-on experience with your new architecture.

Once deployed, WANdisco fusion replicates data from Docker to Amazon S3 continuously, to help maintain strong consistency between on-premises and cloud data. Amazon Athena can then be leveraged to view and analyze the replicated data.

The Quick Start can also be customized to enable a disaster recovery scenario for your on-premises Hadoop cluster by provisioning an Amazon EMR cluster that references data replicated into Amazon S3.

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 70 fully featured services for compute, storage, databases, analytics, mobile, Internet of Things (IoT) and enterprise applications from 33 Availability Zones (AZs) across 12 geographic regions in the U.S., Australia, Brazil, China, Germany, Ireland, Japan, Korea, and Singapore. 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>.

© 2017, Amazon Web Services, Inc. or its affiliates. All rights reserved.