

WANDISCO FUSION® AMAZON EMR

Fusion hybrid cloud for Amazon EMR

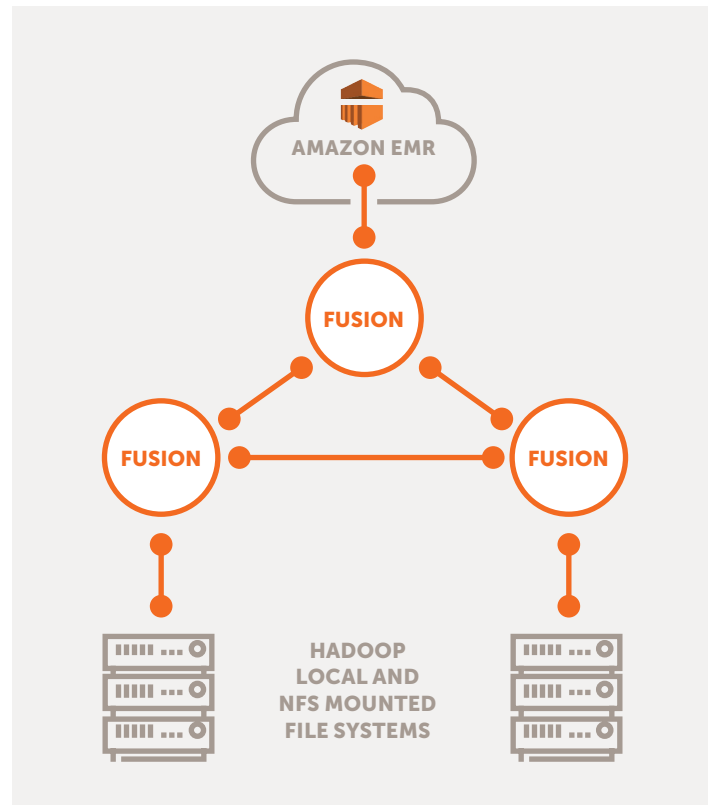
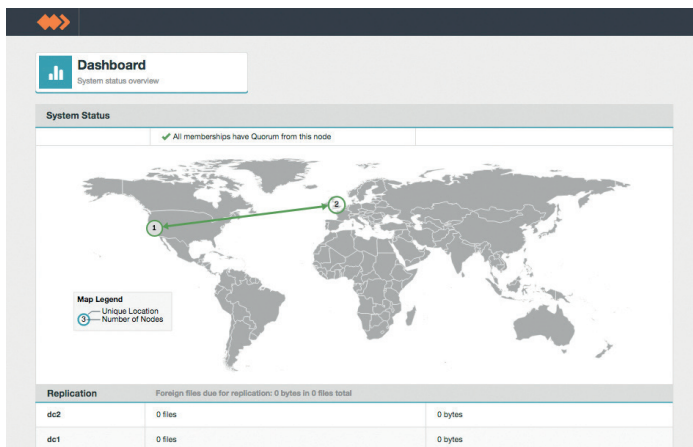
Fusion hybrid cloud for Amazon EMR extends on-premises Hadoop clusters to the cloud for active burst-out processing, offsite DR and data archiving.

Active hybrid

Take total control of your data with WANdisco's Active Data Replication™ from on-premises Hadoop clusters running any distribution to Amazon EMR. Data is replicated between in-house and cloud environments as it changes. Consistency is guaranteed and recovery is automatic after hardware or network outages.

On-demand

Replicate fast streaming data to Amazon EMR as it's ingested on-premises to leverage EMR's scalability and performance for demanding real-time analytics applications. WANdisco Fusion fully supports EMR capabilities that allow clusters to spin up and shutdown on demand, so you only pay for resources when you use them.



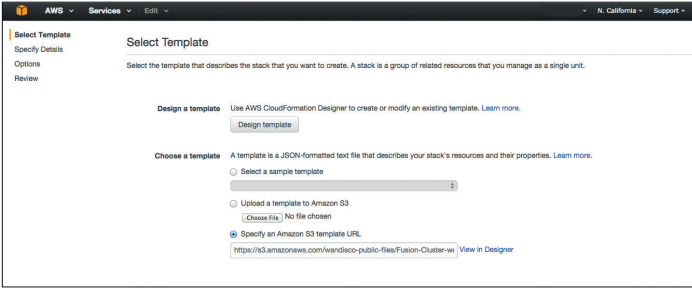
Automatic recovery

Recovery is automatic after planned or unplanned network or hardware outages in both on-premises and cloud environments.

Seamless, flexible and easy to install

Replicates data between on premise clusters deployed on any Hadoop compatible storage and Amazon EMR.

Uses standard AWS utilities for installation and deployment.



Overcomes challenges of other hybrid cloud solutions for Hadoop

Fusion runs as a proxy to on-premises Hadoop clusters and EMR, replicating data as it changes in either environment. DistCp-based solutions offered by Hadoop vendors:

- Run in batch
- Require significant administrator overhead for setup, maintenance and disaster recovery
- Impose significant overhead when moving data that prevents other applications from performing
- Don't guarantee data consistency across on-premises and EMR clusters
- Can't replicate data as it's ingested
- Require manual intervention to handle out-of-sync conditions
- Risk administrator error leading to data loss and extended downtime during recovery.

Supported environments

Hadoop

- EMR
- HDP 2.1.2 – 2.4x
- CDH 5.2 - 5.5
- PHD 3.0
- IBM BI 4.0

Operating Systems

- RHEL 6.1+ (x86-64)
- CentOS 6.1 (x86-64)
- Ubuntu 12.04, 14.04
- SLES 11+ (x86-64)

Cloud

- Amazon S3
- Google Cloud
- MS Azure WASB

About WANDisco

WANDisco (LSE: WAND) is the world leader in Active Data Replication™. Its patented technology provides continuous consistent access to changing data anytime and anywhere with no downtime and no disruption. WANDisco's patented technology serves crucial high availability (HA) requirements, including cloud migration, Hadoop Big Data and Application Lifecycle Management (ALM), including Subversion, Git and Gerrit. WANDisco works directly with Fortune 1000 companies around the world and in all sectors to ensure their data, whether it is on-premises or in the cloud, gives them the real insight they need. WANDisco is committed to the ODPi Interoperable Compliance Program to ensure its products are interoperable across a wider range of commercial Hadoop® platforms.

For more information

About AWS aws.amazon.com

About WANDisco Fusion wandisco.com

Email partnermarketing@wandisco.com

Talk to one of our specialists today
US TOLL FREE +1-877-WANDISCO (926-3472)
EMEA +44 114 3039985
APAC +61 2 8211 0620
ALL OTHER +1925 380 1728

5000 Executive Parkway, Suite 270
San Ramon, California 94583

Copyright © 2016 WANDisco, Inc. All rights reserved.

wandisco.com

Join us online to access our extensive resource library and view our webinars

Follow us to stay in touch

[@WANDisco](https://twitter.com/WANDisco)