IBM Big Replicate

Minimize downtime and maintain business continuity

Solution overview

- Coordinates data operations across heterogeneous platforms achieving consensus and consistency at scale
- Works with Hadoop File Systems, Object Storage, Hive data, IBM Db2 Big SQL metadata and security metadata
- Uses a common administration and management interface putting the *live data* plugins to work on popular data sources ensuring compatibility with multiple systems
- Works across any network topology - LAN, WAN, public and private clouds

Keeping data consistent in a distributed environment is a massive data operations challenge. Big Replicate, an enterprise-class software platform, solves this problem by enabling unstructured data consistency across any environment—on-premises, hybrid-cloud, multi-region cloud, or multi-cloud. Used by enterprises worldwide, our technology is based on a high-performance coordination engine. It uses consensus to keep unstructured data accessible, accurate and consistent in different locations, instead of just moving data from one place to another. With unparalleled throughput, Big Replicate is a foundation for a modern cloud data strategy. It reduces risk in data migration scenarios, prevents data disasters by ensuring zero downtime and simplifies cloud data management in a mixed environment.

Is your data architecture fit for your business needs?

Guaranteeing the safety and accessibility of your business data is vital. If you work in a data-powered business, you need 24/7 access to that data to keep operations moving smoothly and stay competitive. However, as data infrastructures grow more distributed, many companies rely on complex Hadoop and cloud environments, often from a combination of vendors and platforms, in multiple regions. The cost and difficulty of migrating between different platforms—including the risk of outages—means companies often keep the same infrastructure, even if it no longer serves the company's changing business needs. Enterprises need new technology for continuous data consistency in a mixed IT environment.

What is a *live data* platform?

Live data is consistent data everywhere, even when data is changing or moving at petabyte scale. A *live data* platform ensures that your data stays accurate, accessible and consistent across all your business application environments, regardless of geographic location, data platform architecture, or cloud storage provider.

Business benefits

Real-time continuous replication. Provides global data resiliency

Near-zero RPO/RTO. Removes data gaps inherent with point-in-time backups or snapshots

Business continuity during cloud migration. Enables cloud and on-premises environments to operate in parallel even during migrations

Replicates data across a heterogeneous architecture. Reduces latency and points of failure, enabling a transition from separate management of data in private and public clouds to integrated data operations

Self-healing. Automates data movement and lets the system coordinate your data with self-healing non-blocking synchronization

Use cases

Disaster prevention for any environment. Disaster recovery and high availability through near-zero RTO and RPO ensures business continuity and data SLA compliance across on-premises, hybrid and multi-region, multi-cloud environments

Cloud migration without downtime. Zero-disruption data migrations allow users to continue working on data—even migrations at petabyte scale

Data lake consolidation. Synchronize multiple data lakes, unify data silos and maximize line of business data use

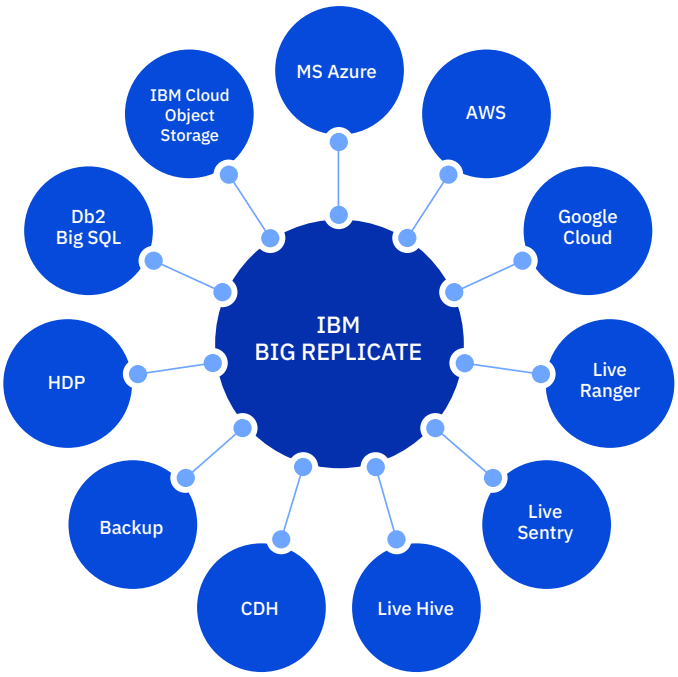
Hybrid-cloud enablement. Keeps on-premises and cloud environments continuously in sync, capturing every change wherever it occurs

IBM

Features

- Flexible selective replication defined at a per-folder level or per data set
- Compatibility with common security protocols—Kerberos, SSL/TLS, LDAP
- Quick comparison and validation of data consistency between zones and replicated environments
- Replication of encrypted at-rest data between different platforms
- Network traffic is between Big Replilcate servers so that application environments can stay isolated
- UI-controlled throttling of network bandwidth use
- API support of server administration and automation

The *live data* platform plugins



Supported environments

Hadoop distributions

- Cloudera CDH
- Hortonworks Data Platform
- MapR
- IBM® BigInsights®
- Microsoft Azure HDInsight

Cloud providers

- Amazon Web Services
- Microsoft Azure
- Google Cloud
- Oracle Cloud

Operating systems

- Centos
- RHEL

File systems

- Amazon S3 compatible
- IBM Cloud Object Storage
- Local and NFS mounted file systems
- OpenStack Swift

Got questions? Ask an IBM expert.

Schedule a free one-on-one consultation with one of the experienced data professionals and distinguished engineers at IBM who have helped thousands of clients build winning data management strategies.

ibm.biz/big-replicate-expert

© Copyright IBM Corporation 2019. IBM, the IBM logo, ibm.com and BigInsights are trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the web at “Copyright and trademark information” at www.ibm.com/legal/copytrade.shtml.

69023369USEN-00