Cirata and Google Cloud Platform

Enabling a modern data architecture

The modern business landscape is ruled by data. Analytics and AI are now essential for driving key business transformation. In order to make effective use of their valuable data assets companies are modernizing their data architecture, which means bringing their business-critical data to work in the cloud.

Google Cloud Platform (GCP), provides a suite of cloud computing services that runs on the same infrastructure that Google uses internally for its end-user products, such as Google Search, Gmail, file storage, etc. Two key services for enabling a modern data architecture include Google Cloud Storage and Google Dataproc.

Data migration business risks

While customers have benefited from the performance, flexibility, and cost savings offered by GCP, many enterprises have struggled with migration of their on-premises data to the cloud. Cloud data migrations can be fraught with business risks including disruption of critical business operations, risk of data loss, and overall project complexities that often result in cost overruns or failed initiatives.

According to Bloor Research, more than 80% of data migration projects run over time and over budget. This is despite the fact that nearly 70% of enterprises have started moving data to the cloud as part of their digital transformation. Migrating large data volumes with traditional approaches, such as transfer devices or DistCp (distributed copy), requires disrupting the operation of on-premises applications and doesn't cater to data that is modified or created during migration. Reconciliation at scale is costly and does not guarantee a completely consistent data outcome. The overhead required to achieve non-disruptive, no-downtime big data migration is significant due to repeated scans, systems out of sync and manual intervention for anticipated failures and interruptions.

Organizations need data migration and replication solutions that reduce and eliminate these business risks. They need solutions that let them maintain business operations, are easy to use, ensure a complete and continuous migration with zero data loss, and which maintain consistency across distributed environments.

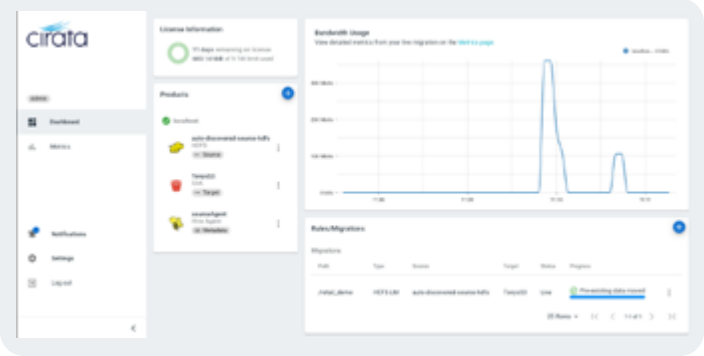
Google Cloud Storage

Google Cloud Storage is a RESTful online file storage web service for storing and accessing data on Google Cloud Platform infrastructure. The service combines the performance and scalability of Google's cloud with advanced security and sharing capabilities making it an ideal cloud storage solution for Google's enterprise customers.

Google Dataproc

Google Cloud Dataproc is Google's version of the Hadoop ecosystem. It includes the Hadoop Distributed File System (HDFS) and Map/Reduce processing framework. The Google Cloud Dataproc system also includes a number of applications such as Hive, Mahout, Pig, Spark, and Hue built on top of Hadoop.

Dataproc actually uses Compute Engine instances under the hood, and it takes care of the management details for you. It's a layer on top that makes it easy to spin up and down clusters as you need them.



Cirata Data Migrator

Data Migrator is an automated, scalable, and high-performance data integration solution that simplifies making your data available in and immediately usable across on-premises environments and the Google cloud.

Data Migrator is non-intrusive and requires no custom code development or changes to existing applications, cluster, node configuration or operation. Data transfer of any scale can begin immediately and be performed while the source data is under active change without requiring any production system downtime or business disruption, and with zero risk of data loss.

Key benefits

Automated data transfer at any scale with zero business disruption, minimized risk, and best time-to-value.

Business continuity

- Automated data transfer of changing data
- No downtime or business disruption
- Immediate data availability

Scalability

- Supports data transfer at any scale
- Horizontal scaling with multiple transfer agents
- Maximizes use of available network bandwidth

Cost and risk avoidance

- Fully automated data transfer
- No custom coding nor application changes
- Minimizes need for IT resource involvement

Use cases

Data modernization
Shift away from legacy data platforms and siloed or underutilized datasets to modern data technologies in the Google cloud that enable advanced analytics, AI, faster decision making, and more flexible and elastic storage and compute to unlock the full value from the data.

Disaster recovery
Maintain a current replica of actively used data in another location (either cloud or on-premises) for failover purposes in case the primary production environment becomes unavailable. Providing the ability to replicate the data in near-real-time is critical to meet any near-zero RTO (recovery time objective) and RPO (recovery point objective) requirements.

Hybrid and multi-region
Implement flexible architectures that maintain data in hybrid environments, which can include on-premises, cloud, and multi-region deployments. As a result, organizations can leverage best in-class capabilities, improve availability and increase regional coverage.

cirata

5000 Executive Parkway, Suite 270, San Ramon, CA 94583
US +1 877 (926-3472) EMEA +44 (0) 114 3039985
APAC +61 2 8211 0620 All other +1 925 380 1728

www.cirata.com

ds-dmgcp-230913