



AWS EMR to GCP Serverless Spark

Drive innovation in your AI and Data Analytics, and maintain a competitive edge with Google Cloud's Serverless Spark: The intelligent alternative to AWS EMR. Serverless Spark is a true serverless solution that offers efficient and scalable processing of large data workloads, user-friendly management, and cost-effective pricing models.

With Persistent's deep expertise, numerous organizations have successfully modernized their data platforms, unlocking new opportunities to harness the power of data across all aspects of their business.

Benefits



Cost savings

Google Cloud's Serverless Spark platform is a pay-as-you-go fully managed service, which means that companies only pay for the resources that they use. This can lead to significant cost savings over AWS EMR, which has different complex pricing models.



Increased flexibility

Google Cloud's Serverless Spark platform is more flexible than AWS EMR. We can run jobs on demand, and can scale resources up or down as needed. This gives more control over costs and performance.



Reduced complexity

Google Cloud's Serverless Spark platform is much simpler to use than AWS EMR. There is no need to provision or manage clusters, and one can start and stop jobs with just a few clicks. This can save both time and resources.



Integration with Google Cloud services

The Serverless Spark platform is seamlessly integrated with other Google Cloud services, such as BigQuery, Dataflow, and Pub / Sub. This makes it easier for companies to build end-to-end data pipelines that incorporate multiple services.



Security and compliance

Google Cloud's security and compliance offerings are among the best in the industry, which can be particularly important for companies in highly regulated industries or dealing with sensitive data. The data is encrypted at rest and in transit, and companies can use a variety of security features to protect their workloads.

Program

4 – 10 weeks to work closely with you to provide proof of value (POV) on GCP and ensure a seamless migration plan for migrating AWS EMR workloads. With our proven expertise and commitment to delivering high-quality results, we are confident that we can help you in your data management goals while minimizing disruptions to business operations.

Proof of Value (POV)

- GCP Foundation setup including IaC for infrastructure provisioning
- Advanced permissions and role-based access with GCP IAM
- Setting up Dataproc Metastore (DPMS)
- Data migration of up to 5TB of data
- Batch processing using Dataproc on GCE and Dataproc serverless
- Batch orchestration using Cloud Composer
- Code conversion of current DDLs to BigQuery DDLs
- BI Report using BigQuery as a data store.
- Detailed migration plan, including ROI, that outlines the steps necessary to move from the current databricks-based solution to Google Cloud Platform

Expanded POV to Include

- Convert from Delta Lake to Apache Iceberg
- Creation of BigLake tables
- Persisting data to BigQuery and building a dashboard
- Setup logical data lake organization in Dataplex
- Working with BigQuery connector, BigLake tables and Spark notebooks on Dataproc Serverless
- Simple Spark ML experiment (Optional)
- MLOps using VertexAI (Optional)

Persistent Specialization Badges



For more information on our capabilities.

Visit Us

About Persistent

Persistent Systems (BSE & NSE: PERSISTENT) is a global services and solutions company delivering Digital Engineering and Enterprise Modernization to businesses across industries. With over 23,900 employees located in 19 countries, the Company is committed to innovation and client success. Persistent offers a comprehensive suite of services, including AI-enabled software engineering, product development, data and analytics, CX transformation, cloud computing, and intelligent automation. The Company is part of the MSCI India Index and is included in key indices of the National Stock Exchange of India, including the Nifty Midcap 50, Nifty IT, and Nifty MidCap Liquid 15 as well as several on the BSE such as the S&P BSE 100 and S&P BSE SENSEX Next 50. Persistent is also a constituent of the Dow Jones Sustainability World Index. The Company has achieved carbon neutrality, reinforcing its commitment to sustainability and responsible business practices. As a participant of the United Nations Global Compact, Persistent is committed to aligning strategies and operations with universal principles on human rights, labor, environment, and anti-corruption, as well as take actions that advance societal goals. With 327% growth in brand value since 2020, Persistent is the fastest-growing IT services brand in the 2024 Brand Finance India 100 Report.

USA

Persistent Systems, Inc.
2055 Laurelwood Road, Suite 210
Santa Clara, CA 95054
Tel: +1 (408) 216 7010
Fax: +1 (408) 451 9177
Email: info@persistent.com

India

Persistent Systems Limited
Bhageerath, 402
Senapati Bapat Road
Pune 411016
Tel: +91 (20) 6703 0000
Fax: +91 (20) 6703 0008

