



Migrating Apache Beam Pipelines from GCP to Azure

Apache Beam is a unified batch and stream programming model that provides Cross-language capabilities, portability, extensibility and flexibility in the implementation of large-scale data processing systems. Beam supports Google cloud dataflow runner for GCP and it has multiple features which makes it a viable solution for large scale batch and stream pipeline implementation in Google Cloud. The option to move to other Cloud providers like Azure requires a thorough understanding of Beam, the suitable runners and the time involved in implementing the solution

Migrate your Apache Beam Pipelines from GCP to AZURE with ease

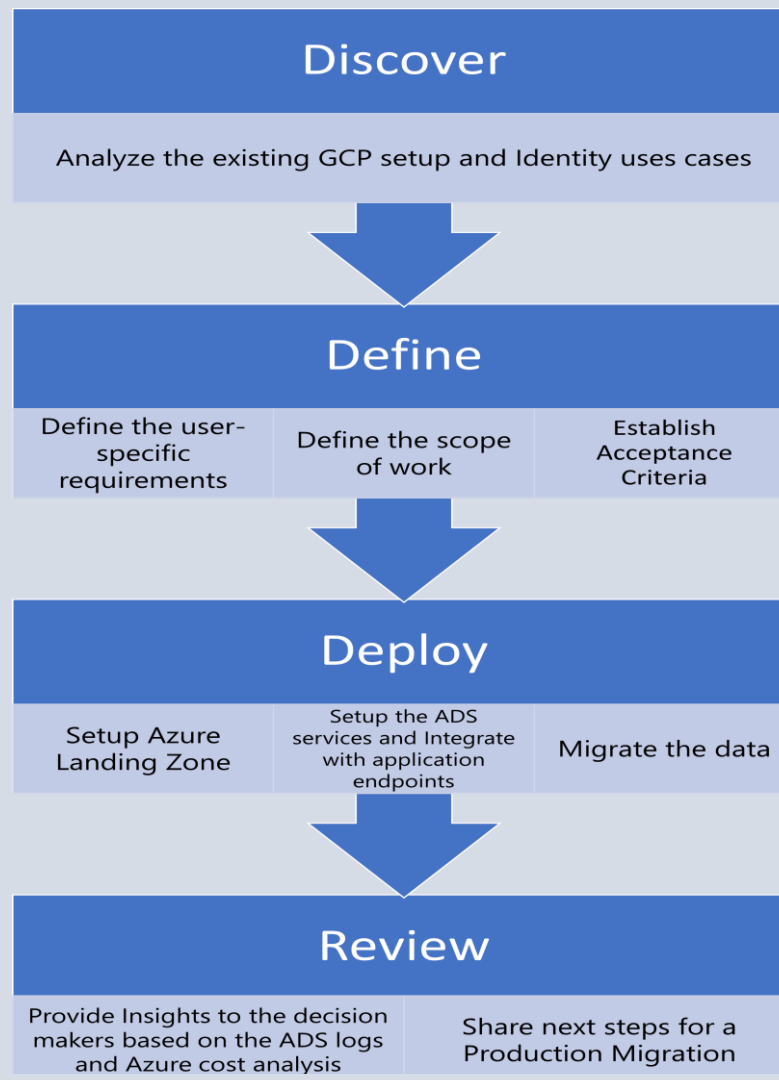
Migrate the existing Stream Analytics ecosystem from GCP (Dataflow / Apache Beam/ Pub sub/ BigQuery) to Azure (AKS/Flink/Apache Beam/Event Hub/Cosmos DB -Synapse Link) ecosystem

Implement robust high performing and scalable services in Azure which offers out-of-the-box equivalent features to replace existing Apache Beam Pipelines deployed in the GCP ecosystem.

Highlights

- Flink jobs running inside Azure AKS containers to host Apache Beam pipelines and to process stream data
- Azure COSMOS DB with Synapse Link no-ETL column data store with near real time data which is made available for downstream users (Data Engineers, Business analysts and data scientists, etc) to use Synapse Spark or SQL to perform their BI/ML/analytics in near real-time
- With the benefits of high performance and optimized for stream data ingestion, cosmos DB is extremely scalable
- These powerful components not only replace the GCP-based Dataflow/BigQuery services to host Apache Beam pipelines but are highly cost-effective and easy to monitor, manage and scale with various tools and APIs that come integrated with Azure Cloud.
- Minimal-Code change to existing GCP Dataflow-based Apache Beam pipeline jobs.
- Azure Event Hubs exposing out-of-the-box Kafka endpoints for seamless integration with Apache Beam pipelines

Our approach to delivery:



Final Deliverables:

Duration: 6 Weeks

Cost: \$35,000

- Azure Cloud Infrastructure/service provisioning using Terraform templates
- Apache Flink cluster setup and configuration in AKS for Data processing
- Implementation of a specific use case (Identified by the customer) to demonstrate end to end data flow

Reach us:

Siva Nagalingam
VP, Microsoft Cloud Services
snagalingam@presidio.com

PRESIDIO®