



DATAFLOW FOR

Real-time Log Replication & Analytics

Our website's performance has decreased and I have to wait hours to collect logs... (sigh)

With Dataflow we can analyze logs in real-time and work proactively!

... and by the time we get the logs, the issue will snowball into a major disruption

Here's how Dataflow can help your business...



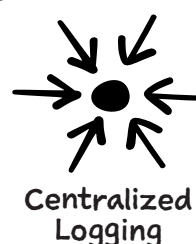
Incident Prevention



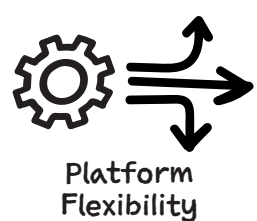
Reduced Downtime



Real-time Troubleshooting

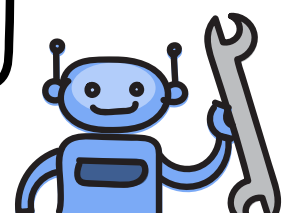


Centralized Logging

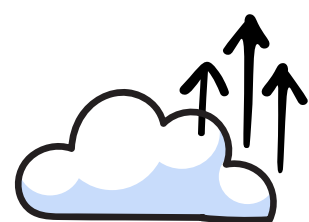


Platform Flexibility

These technical benefits aren't too shabby, either!



Serverless experience eliminates need to manage infrastructure



Limitless scalability with high-performance backend service



High resiliency with snapshots & built-in checkpointing



Achieve sub-second latencies with Streaming Modes



Monitoring tools optimize performance & proactively detect issues

Here's a sample architecture for Log Replication and Analytics...

Logging

Collect and forward logs into Cloud Logging



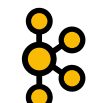
Cloud Logging

Logging Sink

Ingest logs from applications with event streaming platforms



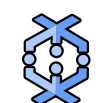
Pub/Sub



Kafka

Replication Pipeline

Enrichment and log routing pipeline for replication



Dataflow

Log Destination

Write replicated logs into Splunk or other logging platforms



Splunk



Log Management Tools

Proactive and reliable. That's the Dataflow difference!

What is Dataflow?

It's a serverless data processing service for both streaming & batch data.

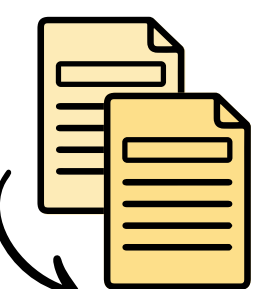


Batch Data

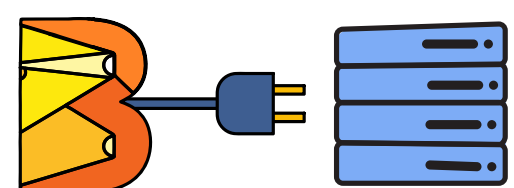


Streaming Data

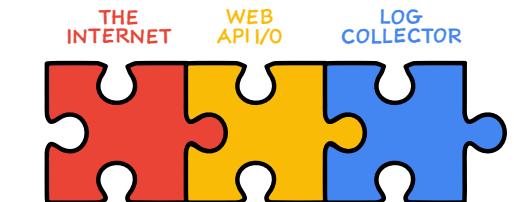
Keep in mind...



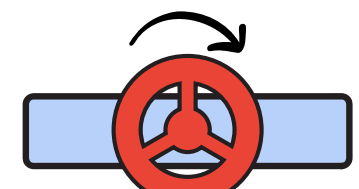
Use Dataflow templates if you need to write to Splunk, DataDog, or Elasticsearch



Leverage off-the-shelf Apache Beam connectors like SplunkIO to write easily into log collectors



Use the Beam WebAPIs I/O connector if your log collector isn't natively supported



Enable Streaming Engine to handle logs at any scale