

Google Cloud Next '20: OnAir | Learn best practices and what's new in Dataflow, now available on demand. [Watch session DA210 today.](https://cloud.withgoogle.com/next/sf/sessions?session=DA210#data-analytics)  
(<https://cloud.withgoogle.com/next/sf/sessions?session=DA210#data-analytics>)

[JUMP TO \(#\)](#)



# Dataflow

Unified stream and batch data processing that's serverless, fast, and cost-effective.

Try Dataflow free (<https://console.cloud.google.com/freetrial>)

- ✓ Fully managed data processing service
- ✓ Automated provisioning and management of processing resources
- ✓ Horizontal autoscaling of worker resources to maximize resource utilization
- ✓ OSS community-driven innovation with Apache Beam SDK
- ✓ Reliable and consistent exactly-once processing

FORRESTER

Google named a Leader in The Forrester Wave™: Streaming Analytics, Q3 2019

Register to download the report (</forrester-streaming-analytics>)

## Streaming data analytics with speed

Dataflow enables fast, simplified streaming data pipeline development with lower data latency.

## Simplify operations and management

Allow teams to focus on programming instead of managing server clusters as Dataflow's serverless approach removes operational overhead from data engineering workloads.

## Reduce total cost of ownership

Resource autoscaling paired with cost-optimized batch processing capabilities means Dataflow offers virtually limitless capacity to manage your seasonal and spiky workloads without overspending.

## Key features

### Autoscaling of resources and dynamic work rebalancing

Minimize pipeline latency, maximize resource utilization, and reduce processing cost per data record with data-aware resource autoscaling. Data inputs are partitioned automatically and constantly rebalanced to even out worker resource utilization and reduce the effect of "hot keys" on pipeline performance.

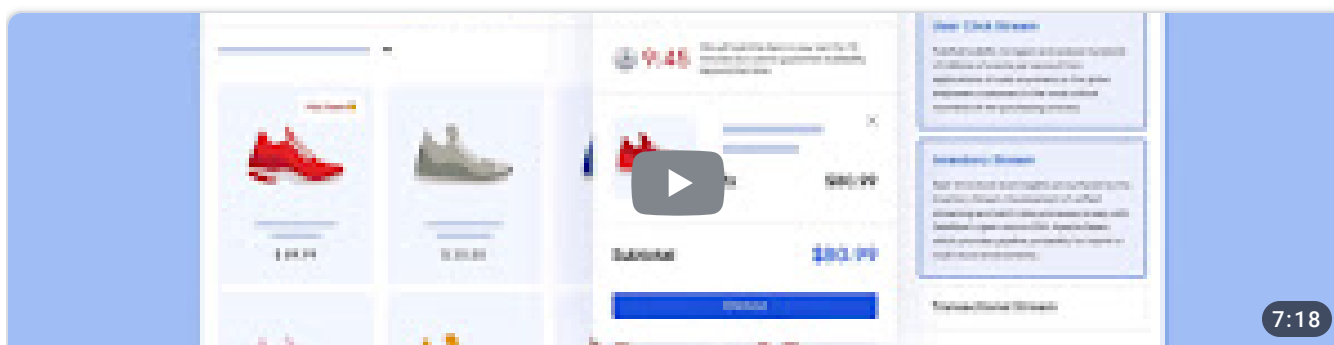
## Flexible scheduling and pricing for batch processing

For processing with flexibility in job scheduling time, such as overnight jobs, flexible resource scheduling (FlexRS) offers a lower price for batch processing. These flexible jobs are placed into a queue with a guarantee that they will be retrieved for execution within a six-hour window.

## Ready-to-use real-time AI patterns

Enabled through ready-to-use patterns, Dataflow's real-time AI capabilities allow for real-time reactions with near-human intelligence to large torrents of events. Customers can build intelligent solutions ranging from predictive analytics and anomaly detection to real-time personalization and other advanced analytics use cases.

### View all features (#)



(<https://www.youtube.com/watch?v=JRHy8fUQLYQ>)

#### VIDEO

**[Enhance online retail experiences with real-time, personalized offers: Demo](https://www.youtube.com/watch?v=JRHy8fUQLYQ)**

(<https://www.youtube.com/watch?v=JRHy8fUQLYQ>)

## Customers



**DOW JONES**



**Dow Jones brings key historical events datasets to life with Dataflow.**

Read the story (<https://cloud.google.com/customers/dow-jones>)

### Story highlights

---

- Synthesized 30+ years of news data to assess business impact
- Uncovered hidden data relationships and insights
- Prototype Knowledge Graph delivered with ease in 10 weeks

### Partner

---

[Quantphi](#)

(<https://cloud.withgoogle.com/partners/detail/?id=quantiphi-inc&hl=en-US>)



[Sky updates its big data platform to meet the needs of its next-gen products.](#)

(/customers/sky-uk)



46:29

[Unity uses Dataflow to transform data into insights, decisions, and products.](#)

(<https://www.youtube.com/watch?v=kvOAEb7W3oA&t=1922&autoplay=1>)

**craveable  
brands.**

[Craveable Brands cut data costs 75% developing real-time in-store experiences.](#)

(/customers/craveable-brands)

See all customers (<https://cloud.google.com/customers>)

## What's new

[Sign up](https://cloud.google.com/newsletter) (https://cloud.google.com/newsletter) for Google Cloud newsletters to receive product updates, event information, special offers, and more.

VIDEO

[What is Dataflow? New video explainer from the Google Cloud drawing board series](#)

[Watch video](#)

[. \(https://www.youtube.com/watch?v=KaIJ0VuEM7s\)](https://www.youtube.com/watch?v=KaIJ0VuEM7s)



## Documentation

QUICKSTART

### Dataflow quickstart using Python

Set up your Google Cloud project and Python development environment, get the Apache Beam SDK, and run and modify the WordCount example on the Dataflow service.

[Learn more \(/dataflow/docs/quickstarts/quickstart-python\)](/dataflow/docs/quickstarts/quickstart-python)

TUTORIAL

### Using Dataflow SQL

Create a SQL query and deploy a Dataflow job to run your query from the Dataflow SQL UI.

[Learn more \(/dataflow/docs/guides/sql/dataflow-sql-ui-walkthrough\)](/dataflow/docs/guides/sql/dataflow-sql-ui-walkthrough)

**TUTORIAL****Installing the Apache Beam SDK**

Install the Apache Beam SDK so that you can run your pipelines on the Dataflow service.

[Learn more \(/dataflow/docs/guides/installing-beam-sdk\)](/dataflow/docs/guides/installing-beam-sdk)

**TUTORIAL****Machine learning with Apache Beam and TensorFlow**

Preprocess, train, and make predictions on a molecular energy machine learning model, using Apache Beam, Dataflow, and TensorFlow.

[Learn more \(/dataflow/docs/samples/molecules-walkthrough\)](/dataflow/docs/samples/molecules-walkthrough)

**TUTORIAL****Qwiklab: Google Cloud Big Data and ML Fundamentals**

This one-week on-demand course provides an introduction to the data analytics and ML capabilities of Google Cloud, including creating pipelines with Dataflow.

[Learn more \(https://google.qwiklabs.com/courses/843\)](https://google.qwiklabs.com/courses/843)

**GOOGLE CLOUD BASICS****Dataflow resources**

Find information on pricing, resource quotas, FAQs, and more.

[Learn more \(/dataflow/docs/resources\)](/dataflow/docs/resources)

#### TUTORIAL

### Explore what you can build on Google Cloud

Find Google Cloud technical resource guides pertaining to Dataflow.

[Learn more \(/docs/tutorials#%22dataflow%22\)](/docs/tutorials#%22dataflow%22)

### Not seeing what you're looking for?

[View all product documentation \(/dataflow/docs\)](/dataflow/docs)

### Explore more docs

---

#### Quickstarts

[Get a quick intro to using this product.](#)

[\(/dataflow/docs/quickstarts\)](/dataflow/docs/quickstarts)

#### How-to guides

[Learn to complete specific tasks with this product.](#)

[\(/dataflow/docs/how-to\)](/dataflow/docs/how-to)

#### Tutorials

[Browse walkthroughs of common uses and scenarios for this product.](#)

[\(/dataflow/docs/samples\)](/dataflow/docs/samples)



### APIs & references

[View APIs, references, and other resources for this product.](#)

(/dataflow/docs/apis)

(/dataflow/docs/resources/release-notes-service)

#### RELEASE NOTES

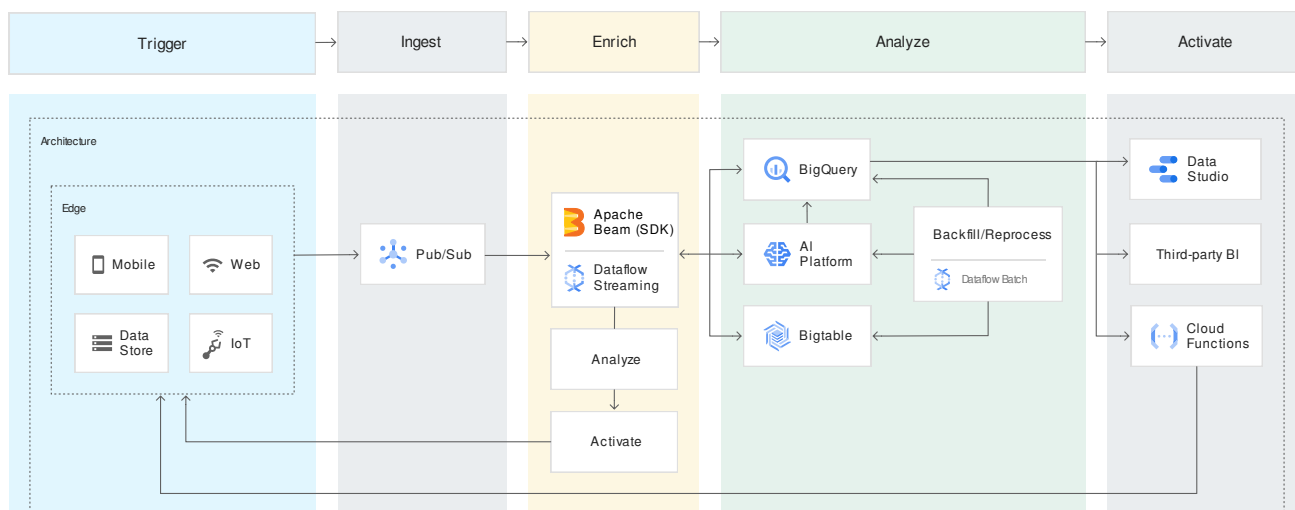
Read about the latest releases for Dataflow

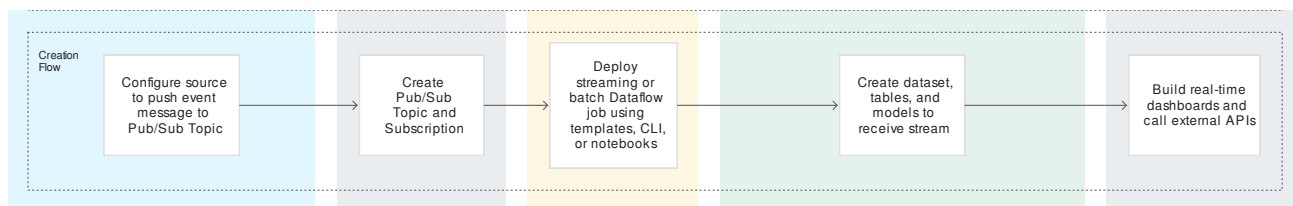
## Use cases

### USE CASE

#### Stream analytics

Google's [stream analytics](#) (/solutions/stream-analytics) makes data more organized, useful, and accessible from the instant it's generated. Built on Dataflow along with Pub/Sub and BigQuery, our streaming solution provisions the resources you need to ingest, process, and analyze fluctuating volumes of real-time data for real-time business insights. This abstracted provisioning reduces complexity and makes stream analytics accessible to both data analysts and data engineers.





## USE CASE

### Real-time AI

Dataflow brings streaming events to Google Cloud's [AI Platform](#) (/ai-platform) and [TensorFlow Extended \(TFX\)](#) (<https://www.tensorflow.org/tfx>) to enable predictive analytics, fraud detection, real-time personalization, and other advanced analytics use cases. TFX uses Dataflow and Apache Beam as the distributed data processing engine to enable several aspects of the ML life cycle, all supported with [CI/CD for ML](#)

(/solutions/machine-learning/architecture-for-mlops-using-tfx-kubeflow-pipelines-and-cloud-build#setting-up-ci-cd-for-ml-on-google-cloud)

through Kubeflow pipelines.

#### PATTERN

### Anomaly detection

Identify and resolve problems in real time with outlier detection for malware, account activity, financial transactions, and more.

**Learn more** ([https://cloud.google.com/solutions/smart-analytics/reference-patterns/overview#anomaly\\_detection](https://cloud.google.com/solutions/smart-analytics/reference-patterns/overview#anomaly_detection))

#### PATTERN

### Pattern recognition

Streamline operations and customer experiences with pattern detection on images, videos, and data.

**Learn more** ([https://cloud.google.com/solutions/smart-analytics/reference-patterns/overview#pattern\\_recognition](https://cloud.google.com/solutions/smart-analytics/reference-patterns/overview#pattern_recognition))

#### PATTERN

### Predictive forecasting

Forecast time series data streams ranging from user activity to equipment health in order to proactively solve problems.

**Learn more** ([https://cloud.google.com/solutions/smart-analytics/reference-patterns/overview#predictive\\_forecasting](https://cloud.google.com/solutions/smart-analytics/reference-patterns/overview#predictive_forecasting))

#### USE CASE

### Sensor and log data processing

Unlock business insights from your global device network with an intelligent [IoT platform](#) (/solutions/iot).

[View all technical guides \(/docs/tutorials#%22dataflow%22\)](/docs/tutorials#%22dataflow%22)

## All features

### Streaming Engine

Streaming Engine separates compute from state storage and moves parts of pipeline execution out of the worker VMs and into the Dataflow service back end, significantly improving autocooldown and data latency.

autoscaling and data latency.

---

## Autoscaling

Autoscaling lets the Dataflow service automatically choose the appropriate number of worker instances required to run your job. The Dataflow service may also dynamically reallocate more workers or fewer workers during runtime to account for the characteristics of your job.

---

## Dataflow Shuffle

Service-based Dataflow Shuffle moves the shuffle operation, used for grouping and joining data, out of the worker VMs and into the Dataflow service back end for batch pipelines. Batch pipelines scale seamlessly, without any tuning required, into hundreds of terabytes.

---

## Dataflow SQL

Dataflow SQL lets you use your SQL skills to develop streaming Dataflow pipelines right from the BigQuery web UI. You can join streaming data from Pub/Sub with files in Cloud Storage or tables in BigQuery, write results into BigQuery, and build real-time dashboards using Google Sheets or other BI tools.

---

## Flexible Resource Scheduling (FlexRS)

Dataflow FlexRS reduces batch processing costs by using advanced scheduling techniques, the

Dataflow Shuffle service, and a combination of preemptible virtual machine (VM) instances and regular VMs.

---

### Dataflow templates

Dataflow templates allow you to easily share your pipelines with team members and across your organization or take advantage of many Google-provided templates to implement simple but useful data processing tasks. With Flex Templates, you can create a template out of any Dataflow pipeline.

---

### Notebooks integration

Iteratively build pipelines from the ground up with AI Platform Notebooks and deploy with the Dataflow runner. Author Apache Beam pipelines step by step by inspecting pipeline graphs in a read-eval-print-loop (REPL) workflow. Available through Google's AI Platform, Notebooks allows you to write pipelines in an intuitive environment with the latest data science and machine learning frameworks.

---

### Inline monitoring

Dataflow inline monitoring lets you directly access job metrics to help with troubleshooting batch and streaming pipelines. You can access monitoring charts at both the step and worker level visibility and set alerts for conditions such as stale data and high system latency.

---

**Customer-managed encryption keys**

You can create a batch or streaming pipeline that is protected with a customer-managed encryption key (CMEK) or access CMEK-protected data in sources and sinks.

---

**Dataflow VPC Service Controls**

Dataflow's integration with VPC Service Controls provides additional security for your data processing environment by improving your ability to mitigate the risk of data exfiltration.

---

**Private IPs**

Turning off public IPs allows you to better secure your data processing infrastructure. By not using public IP addresses for your Dataflow workers, you also lower the number of public IP addresses you consume against your Google Cloud project quota.

## Pricing

Dataflow jobs are billed per second, based on the actual use of Dataflow batch or streaming workers. Additional resources, such as Cloud Storage or Pub/Sub, are each billed per that service's pricing.

**[View pricing details \(/dataflow/pricing\)](/dataflow/pricing)**

## Partners

Google Cloud partners have developed integrations with Dataflow to quickly and easily enable powerful data processing tasks of any size.

(<https://cloud.withgoogle.com/partners/detail/?id=confluent-inc&hl=en-US>)



(<https://cloud.withgoogle.com/partners/detail/?id=snowplow-analytics&hl=en-US>)

**SNOWFLOW**

(<https://cloud.withgoogle.com/partners/detail/?id=talend-inc&hl=en-US>)



---

Cloud AI products comply with the SLA policies listed [here](#)

(<https://cloud.google.com/products#product-launch-stages>). They may offer different latency or availability guarantees from other Google Cloud services.

## Take the next step

Start building on Google Cloud with \$300 in free credits and 20+ always free products.

**Try Dataflow free** (<https://console.cloud.google.com/freetrial>)

### **Need help getting started?**

Contact sales (<https://cloud.google.com/contact/>)

### **Work with a trusted partner**

Find a partner (<https://cloud.withgoogle.com/partners/>)

### **Continue browsing**

See all products (<https://cloud.google.com/products/>)