# Products

[BigQuery](#_65tw6jyo5ai0)

[Cloud Dataflow](#_ei3bnu5npsc)

[Cloud Dataproc](#_diah2oqo916)

[Cloud Datalab](#_i9quag358p4i)

[Data Studio](#_e0zjvsicl2i)

[Cloud Dataprep](#_dohe956obqdy)

[Cloud Pub/Sub](#_z41omueydczu)

[Cloud Functions](#_gt1jke102adc)

# BigQuery

**What is it?**

Analytics data warehouse for finding data insights without managing infrastructure

Underlying server and storage architecture, automatic scaling

Use SQL commands

ODBC and JDBC drivers provided for integration

**How does it work?**

Features:

* Federated query: process data across storage types w/o duplicating data. For example, run a query pulling from multiple data sources such as Cloud Storage (object), Cloud Bigtable (transactional databases), or Google Spreadsheets

Integrations:

* ODBC and JDBC drivers
* ETL tools like Informatica and Talend
* REST API and client libraries in popular programming languages
* AI: CloudML Engine, TensorFlow

Supported BI applications:

* Tableau, Microstrategy, Looker, Google DataStudio

**What are its business benefits?**

Increased productivity for your data analysts by handling the backend work.

Gain insights and share throughout the organization

Integration with popular data sources, visualization and reporting applications, existing apps.

**What is it similar to? How is it different?**

It’s similar to other data warehouses.

**When is it used?**

IoT

Predictive digital marketing

Data distribution

Public datasets

**How much does it cost?**

Free up to 1TB of data analyzed each month and 10GB of data stored

Pay for the amount of data analyzed

More information

[Google Blog: Inside Capacitor, BigQuery’s next-generation colunmar storage format](https://cloud.google.com/blog/big-data/2016/04/inside-capacitor-bigquerys-next-generation-columnar-storage-format)

# [Cloud Dataflow](https://cloud.google.com/dataflow/)

**What is it?**

Fully managed cloud service for transforming and enriching data.

This is the step in between sensors and a data warehouse (for example)

**How does it work?**

Processing frequency

* Stream or batch

Features

1. Create transformation steps using Java and Python APIs in Apache Beam SDK, which provides a rich set of windowing and session analysis as well as an ecosystem of source and sink connectors.
2. Serverless
3. Management: Integration with Stackdriver, GCP’s unified logging and monitoring solution to monitor and troubleshoot your pipelines as they are running!

Integrations:

* Cloud Pub/Sub (streaming events ingestion)
* Data warehousing (BigQuery)
* Machine Learning (Cloud ML)

**What are its business benefits?**

Excellent, automatic scaling to improve the efficiency of your data engineers.

Automated resource management to minimize latency and maximize utilization.

**What is it similar to? How is it different?**

ETL tools

[Alooma](https://www.alooma.com/platform)

Apache Airflow and Apache Nifi

Informatic Powercenter

Oracle Warehouse Builder (OWB)

[See this Quora answer for more examples](https://www.quora.com/Which-is-the-best-ETL-tool-to-learn-and-which-has-good-job-opportunities)

**When is it used?**

Retail: Clickstream, POS, and segmentation analysis

Financial services: Fraud detection

Gaming: personalized UX

Manufacturing: IoT analytics

Healthcare: IoT analytics

Logistics: IoT analytics

**How much does it cost?**

[**NEW PRICING MODEL EFFECTIVE 01/09/18**](https://cloud.google.com/dataflow/pricing)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Dataflow Worker Type | vCPU  $/hr | Memory  $ GB/hr | Local storage - Persistent Disk  $ GB/hr | Local storage - SSD based  $ GB/hr | Dataflow Shuffle [3](https://cloud.google.com/dataflow/#three)  $ GB/hr |
| **Batch** [**1**](https://cloud.google.com/dataflow/#one) | **$0.056** | **$0.003557** | **$0.000054** | **$0.000298** | **$0.0216** |
| **Streaming** [**2**](https://cloud.google.com/dataflow/#two) | **$0.069** | **$0.003557** | **$0.000054** | **$0.000298** | **N/A** |

[**Back to the top**](#_v688muj6eupd)

# Cloud Dataproc

**What is it?**

A managed Spark and Hadoop service to easily process big datasets using the powerful and open tools in the Apache big data ecosystem.

Flexible, in-memory framework.

**How does it work?**

Just like a Compute Engine instance, except fully managed Spark and Hadoop.

Features

* Managing clusters using web UI, Google Cloud SDK, RESTful APIs, and SSH access
* Automatic or manual hardware and software configuration

**What are its business benefits?**

Automated cluster management.

Focus workforce on the problems that matter to you. Let Google handle all the backend stuff.

High availability for zero lost revenue from downtime

**What is it similar to? How is it different?**

Most similar to on-prem Spark or Hadoop.

Amazon EMR

**When is it used?**

**How much does it cost?**

Small incremental fee per vCPU in the Compute Engine instances used in your cluster.

|  |  |
| --- | --- |
| MACHINE TYPE | PRICE |
| Standard Machines  *1-64 Virtual CPUs* | **$0.010 - $0.640** |
| High Memory Machines  *2-64 Virtual CPUs* | **$0.020 - $0.640** |
| High CPU Machines  *2-64 Virtual CPUs* | **$0.020 - $0.640** |
| Custom Machines  *Based on vCPU and memory usage* | **$0.010/ vCPU hour** |

[Back to top](#_v688muj6eupd)

# [Cloud DataLab](https://cloud.google.com/datalab/)

**What is it?**

Explore, analyze, transform, and visualize data and build ML models on GCP.

Built on Jupyter (formerly iPython)

**How does it work?**

**What are its business benefits?**

**What is it similar to? How is it different?**

**When is it used?**

**How much does it cost?**

Free. Only pay for the infrastructure you use with it.

[Back to top](#_v688muj6eupd)

# 

# Data Studio

**What is it?**

Turn data into informational dashboards and reports

**How does it work?**

Features

Connect data to reports from DBs (BigQuery, Cloud SQL, MySQL)

Data transformation

Data visualization

Sharing and collaboration

Report templates

User Admin

Report customization

**What are its business benefits?**

Simply generate visual reports from your data for business users.

**What is it similar to? How is it different?**

**When is it used?**

**How much does it cost?**

[Back to top](#_v688muj6eupd)

# [Cloud Dataprep](https://cloud.google.com/dataprep/)

**What is it?**

Service for visually exploring, cleaning, and preparing structured and unstructured data for analysis.

**How does it work?**

Features

Automatically detects schemas, datatypes, possible joins and anomalies such as missing values, outliers, and duplicates

**What are its business benefits?**

Save time

**What is it similar to? How is it different?**

**When is it used?**

**How much does it cost?**

[Back to top](#_v688muj6eupd)

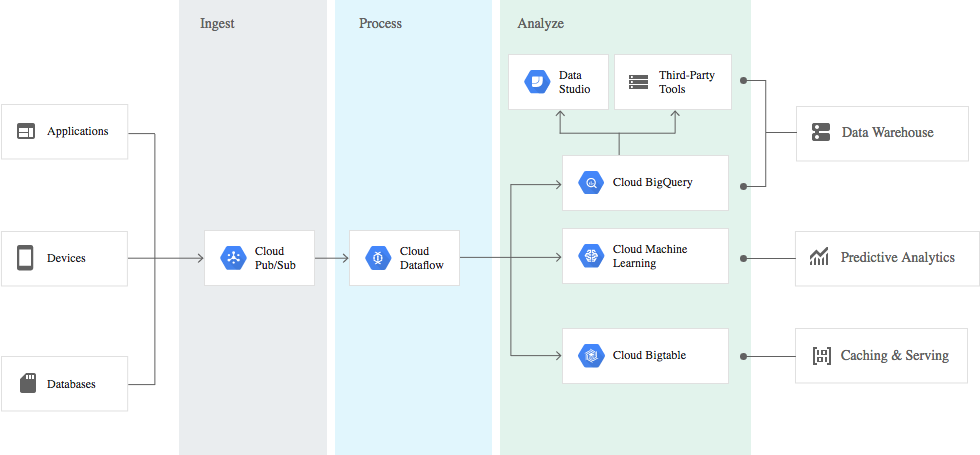
# Cloud Pub/Sub

**What is it?**

Messaging service that allows you to send and receive messages between independent applications.

**How does it work?**

Synchronous

****

**What are its business benefits?**

**What is it similar to? How is it different?**

It’s similar to Apache Kafka.

**When is it used?**

**How much does it cost?**

|  |  |
| --- | --- |
| MONTHLY DATA VOLUME1 | PRICE PER GB |
| First 10GB | **$0.00** |
| Next 50TB | **$0.06** |
| Next 100TB | **$0.05** |
| Beyond 150TB | **$0.04** |

[Back to top](#_v688muj6eupd)

# [Cloud Functions](https://cloud.google.com/functions/)

**What is it?**

Serverless, ephemeral logic (written in JS and executed on standard Node.JS runtime environment)

**How does it work?**

* Triggers invoke a function deployed to Cloud Functions. The following triggers are available: HTTP, Cloud Storage, Cloud Pub/Sub ([source](https://cloud.google.com/functions/docs/concepts/events-triggers)) ([full supported API list](https://cloud.google.com/functions/docs/concepts/services))

**Requirements**

* MUST BE STATELESS: functions are ephemeral

|  |  |
| --- | --- |
| Use Case | Description |
| Data Processing / ETL | Listen and respond to [Cloud Storage](https://cloud.google.com/storage) events such as when a file is created, changed, or removed. Process images, perform video transcoding, validate and transform data, and invoke any service on the Internet from your Cloud Function. |
| Webhooks | Via a simple [HTTP trigger](https://cloud.google.com/functions/docs/calling/http), respond to events originating from 3rd party systems like GitHub, Slack, Stripe, or from anywhere that can send HTTP requests. |
| Lightweight APIs | Compose applications from lightweight, loosely coupled bits of logic that are quick to build and that scale instantly. Your functions can be event-driven or invoked directly over HTTP/S. |
| Mobile Backend | Use Google’s mobile platform for app developers, [Firebase](https://firebase.google.com/docs/functions/), and write your mobile backend in Cloud Functions. Listen and respond to events from Firebase Analytics, Realtime Database, Authentication, and Storage. |
| IoT | Imagine tens or hundreds of thousands of devices streaming data into Cloud Pub/Sub, thereby launching Cloud Functions to process, transform and store data. Cloud Functions lets you do in a way that’s completely serverless. |