Certification
Google - Cloud Architect

Google Cloud Platform

Google
Big Data Solutions 101

Google Cloud Platform

Big Data Solutions - PAAS

Google offers Big Data Solutions as a platform service.
Different options on BigData Service are

Google BigQuery

Google Cloud Dataflow /Apache Beam

Google Cloud Dataproc

Google Cloud Datalab

Big Data Solutions – Other

Google Cloud Platform

Other Service which can be treated as Bigdata service

Google Cloud Pub/Sub

Google Genomics

Google BigQuery

Fully-managed enterprise data warehouse for large-scale data analytics

BigQuery is Google's fully managed, petabyte scale, low cost enterprise data warehouse for analytics. BigQuery is serverless. There is no infrastructure to manage and you don't need a database administrator, so you can focus on analyzing data to find meaningful insights using familiar SQL.

BigQuery: Product Features

- ✓ Flexible Data Ingestion
- ✓ Global Availability
- ✓ Security & Permissions
- ✓ Cost Controls
- ✓ Highly Available
- ✓ Fully Integrated
- ✓ Connect with Google Products
- ✓ Automatic Data Transfer Service

Google Cloud Platform

Google Cloud Dataflow/Apache Beam

Stream and Batch data processing

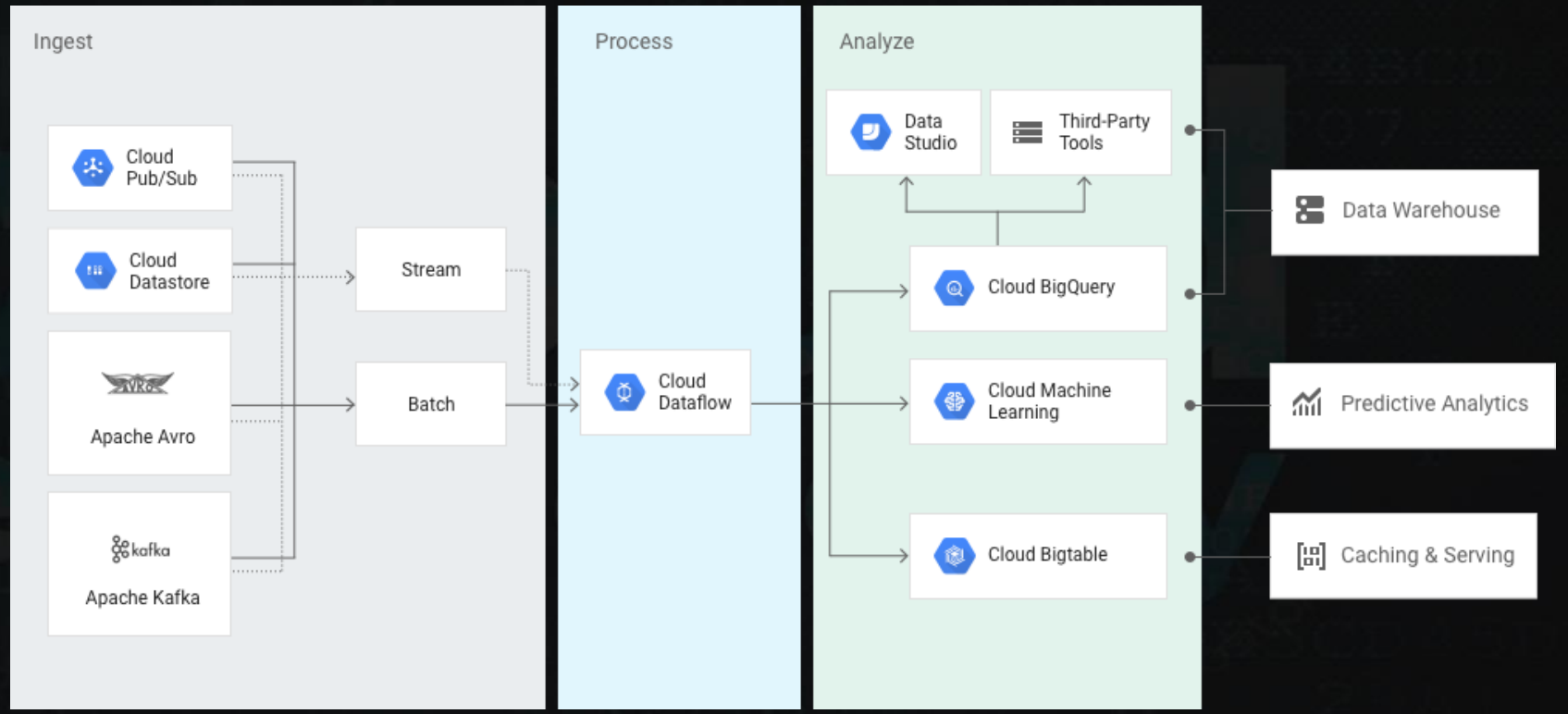
Google Cloud Dataflow offers a unified programming model and a managed service for executing a wide range of data processing patterns including streaming analytics, ETL, and batch computation

Cloud Dataflow : Product Features

- ✓ Automated Resource Management
- ✓ Dynamic Work Rebalancing
- ✓ Reliable & Consistent Processing
- ✓ Horizontal Auto-scaling
- ✓ Unified Programming Model
- ✓ Community-driven Innovation

Google Cloud Platform

Cloud Dataflow : Product Features



Google Cloud Dataproc

A faster, easier, more cost-effective way to run Spark and Hadoop

Use Google Cloud Dataproc, a managed Spark and Hadoop service, to easily process big datasets using the powerful and open tools in the Apache big data ecosystem.

Cloud Dataproc integrates with storage, compute, and monitoring services across Cloud Platform products, giving you a powerful and complete data processing platform

Cloud Dataproc : Product Features

- ✓ Automatic Cluster Management
- ✓ Resizable Clusters
- ✓ Integrated with other GCP services
- ✓ Image Versioning
- ✓ Highly Available
- ✓ Developer Tools
- ✓ Initializations Actions
- ✓ Automatic or Manual Configuration
- ✓ Flexible Virtual Machines

Google Cloud Platform

Google Cloud Datalab

An easy to use interactive tool for data exploration, analysis, visualization and machine learning

Google Cloud Datalab is an interactive notebook (based on Jupyter) to explore, collaborate, analyze and visualize data. It is integrated with BigQuery and Google Cloud Machine Learning to give you easy access to key data processing services.

Cloud Datalab : Product Features

- ✓ Integrated with other GCP services
- ✓ Multi Language Support
- ✓ Notebook format
- ✓ Pay-per-use Pricing
- ✓ Interactive Data Visualizations
- ✓ Machine Learning
- ✓ iPython Support
- ✓ Open Source

Google Cloud Platform

Google Cloud Pub/Sub

Ingest event streams from anywhere, at any scale, reliable, real-time stream analytics

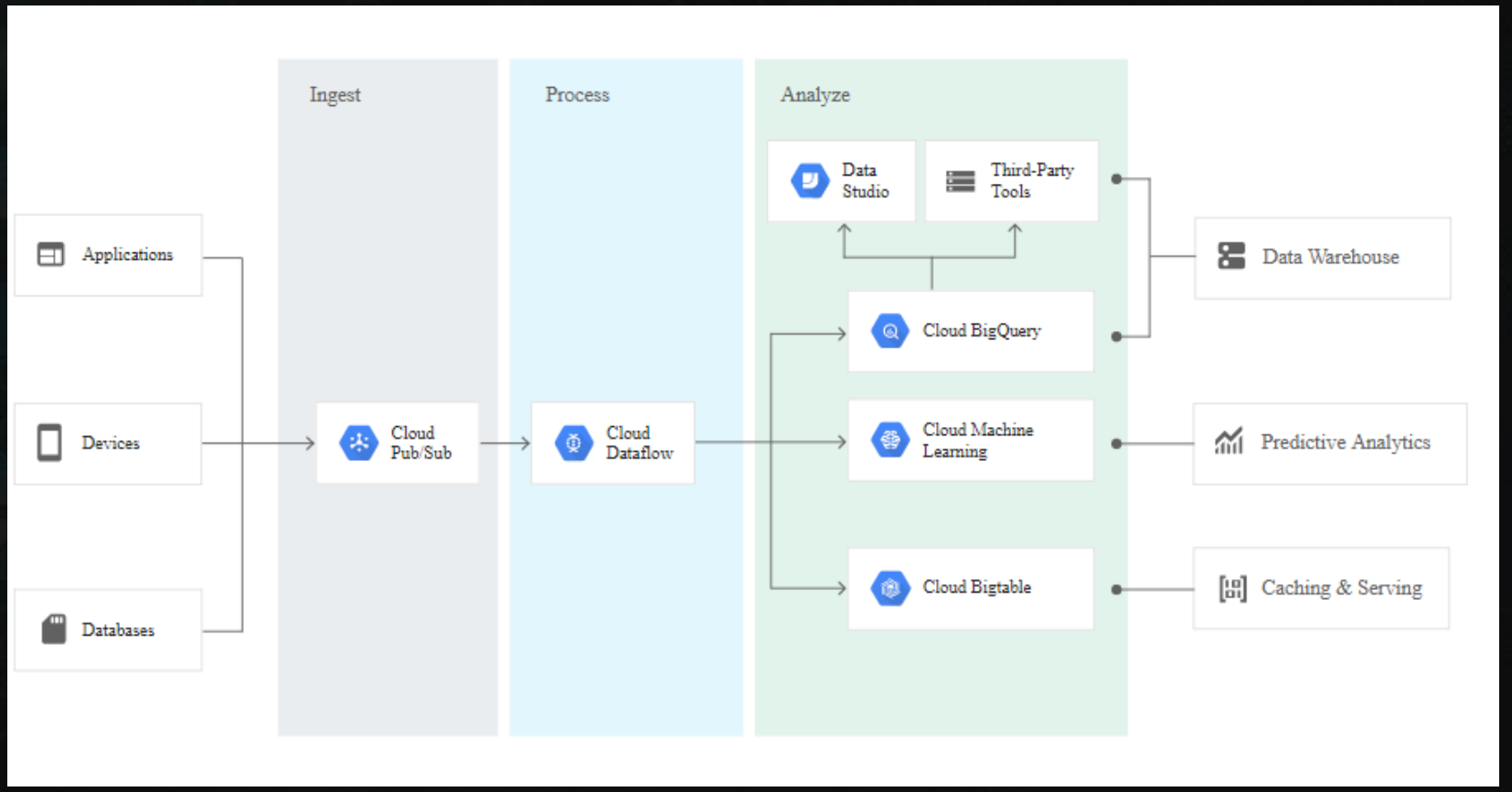
Google Cloud Pub/Sub is a serverless, large scale, reliable, real-time messaging service that allows you to send and receive messages between independent applications. You can leverage Cloud Pub/Sub's flexibility to decouple systems and components hosted on Cloud Platform or elsewhere on the Internet. By building on the same technology Google uses, Cloud Pub/Sub is designed to provide "at least once" delivery at low latency with on-demand scaling to tens of millions of messages per second.

Cloud Pub/Sub : Product Features

- ✓ At-least-once Delivery
- ✓ Exactly-once Processing
- ✓ No Provisioning
- ✓ Integrated with other GCP Services
- ✓ Open API
- ✓ Global by default
- ✓ Compliance and Security

Google Cloud Platform

Cloud Pub/Sub: Product Features



Details of some of these Services coming soon.