



Google Cloud Datalab

Democratize Big Data for Developers

Dinesh Kulkarni
Cloud Datalab Product Manager





Agenda

01 - Google Cloud Datalab Overview

02 - Features and Benefits

03 - Applications and Use Cases on Cloud Datalab

04 - Conclusion



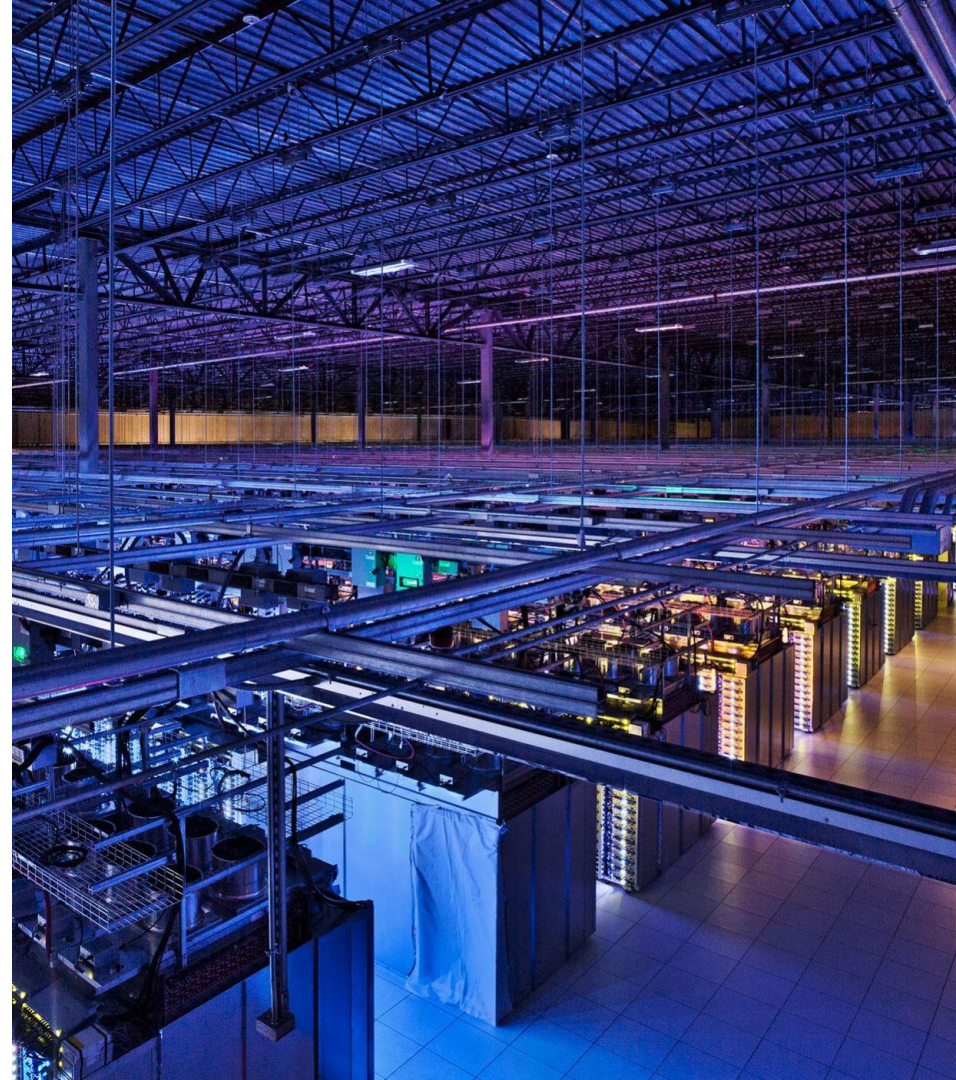
01 Google Cloud Datalab Overview



Cloud Datalab

Google Cloud Datalab is a web-based, interactive tool that makes it easy for you to gain insight into raw data, and present, collaborate, and publish data insights in a fast, simple, and cost effective way.

Cloud Datalab combines the power of Google's Big Data, Machine Learning, and Cloud Storage products in a single tool, eliminating the need for complex product integration.



Why Choose Cloud Datalab?

Simple & Easy to Use

Rich and simple toolkit at your fingertips for:

- Data exploration
- Data transformation
- Data analysis

Complete & Integrated

Integrated set of products:

- Cloud Storage
- BigQuery
- Cloud DataStore
- Cloud SQL

to bring data into a notebook

Productivity for All

Get started easily and scale up to complex analyses:

- Analysts (SQL)
- Developers (Python)
- Data scientists (ML models)



02 Features and Benefits

Features



Choose your language

Write code in multiple languages: Python, SQL and JavaScript.



Notebooks

Launch notebooks to explore, transform and process data on Google Cloud Platform or locally.



Fully Integrated

It leverages the power of Cloud Storage, BigQuery, Cloud DataStore and Cloud SQL for analyses.



Built on Jupyter

Built on IPython/Jupyter which already has a thriving ecosystem of modules and a huge knowledge base.

Benefits



Increase Productivity

Increased productivity through interactive tools and availability of third party libraries.



Simplicity

Explore and analyze data with ad hoc queries and visualizations.



Collaboration

Explore, transform and process data collaboratively or publish data as reports, dashboards or APIs.



Reach

Makes Google's Big Data capabilities easier to use and therefore more accessible across the company.

Pricing

Google Cloud Datalab is a development tool that runs on top of several Google Cloud Platform resources.

Using Datalab will not incur additional charges however you will still be charged for the Google Cloud Platform resources used whilst running Datalab. These typically involve: Google Cloud Compute, Google Cloud Storage and BigQuery.





03 Applications and Use Cases

Use Case #1: Get started with Big Data

“As a developer working on a small team I’m often entrusted with the task of data analysis. I want a simple way to understand what I can do with the data and a roadmap for implementing the analyses. If tools and services connect the dots for me and show some interesting results, I can iteratively build a more useful data pipeline. I can be a part-time data analyst or data scientist and help my company while building new skills.”

Enabling technology

- Cloud project integration: easy way to launch notebooks and authentication
- Connectors for GCS, BigQuery, DataStore and Cloud SQL to bring data into a notebook
- An end-to-end, tool and service-enabled story for going from logs and other semi-structured data to structured analyses and visualizations
- A catalog of common plug-and-play analyses and visualizations referenced below under “Data Analysis & Visualization” scenario

Use Case #2: Analyze & visualize data

“As a data analyst, I want to understand how data varies as a function of time, geolocation and other common variables. I want to be able to chart it to understand it and communicate the meaning to others on the team.”

Enabling technology

- Rich visualization library support
- Extensible visualization support for third party and custom visualization
- Built-in support for time-series and geospatial analysis and rendering
- User segmentation and conversion funnel support

Use Case #3: Build Machine Learning Models

“As a developer or data analyst, I want access to capabilities that have traditionally required Machine Learning experts. I want to get a jump on building basic prediction models and use text analysis on user-generated content such as reviews and comments without starting from scratch or sitting down with a text book.”

Enabling technology

- A library of learning algorithms
- An accessible toolkit or service for building models based on training data and evaluating models based on new data
- Interactive experience for model evaluation and iterative refinement
- Deployment of the model as a service



04 Conclusion

Conclusion

Google Cloud DataLab allows developers to quickly and easily explore, transform and analyze data using familiar languages like Python and SQL. The beta version is fully integrated with the Google Cloud Platform and BigQuery in particular so you can go from data to insights much faster.

- Simpler than developer's traditional Integrated Development Environments (IDEs) for data exploration, transformation and analysis and more powerful than other query/visualization tools. Right tool for the job!
- Fully integrated with the differentiated Big Data offerings in Google Cloud Platform: BigQuery and Google Cloud Storage with more to follow.
- Built on IPython/Jupyter which already has a thriving ecosystem of modules and a huge knowledge base



cloud.google.com

