

Google Cloud BigQuery & Datalab Demo

Ádám Berkecz

AutSoft Ltd.



Google Cloud BigQuery

- Cloud-based analytical Database
- Fully managed
- Run SQL Queries on Google's infrastructure
- Access by using WebUI, CLI, REST API, Client libraries
- Load data via batch loads or streaming
- Infrastructure:
 - 10-100k machines
 - 100k-1M CPU cores
 - 1 Petabit/sec network

Google Datalab

- Interactive tool for exploring, analyzing, and visualizing data on GCP
- Built on top of Jupyter
- Integrated with other Google Cloud Big Data solutions
- Open source -> free
- Usually hosted on Google AppEngine



Citi Bike

- Bicycle sharing system operating in New York City
- Open to the public since May 2013
- ~10 000 bikes
- ~600 stations
- ~40 000 daily rides
- Usage data made public



DEMO