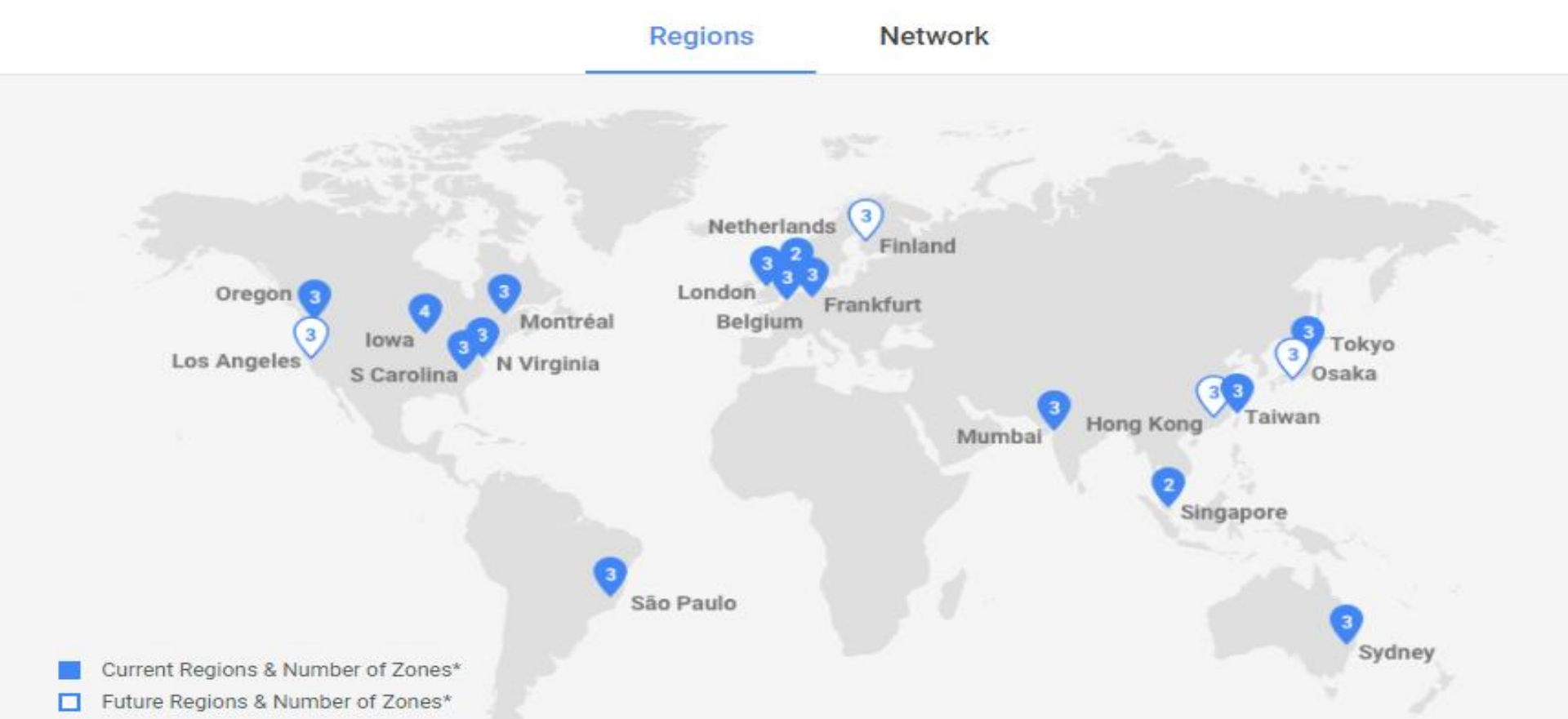
Google Cloud Platform For Data Science

Nazrin Najafzade
Pinar Turkyilmaz

Google Cloud Platform

UPC

Google Cloud Platform Regions and Zones



Payment

Use case

Comparison

Google Cloud

Google Cloud Platform products

COMPUTE

NETWORKING

STORAGE AND DATABASES

BIG DATA

MACHINE LEARNING

IDENTITY & SECURITY

MANAGEMENT AND DEVELOPER TOOLS

Google BigQuery

BigQuery

BigQuery is

scalable

interactive

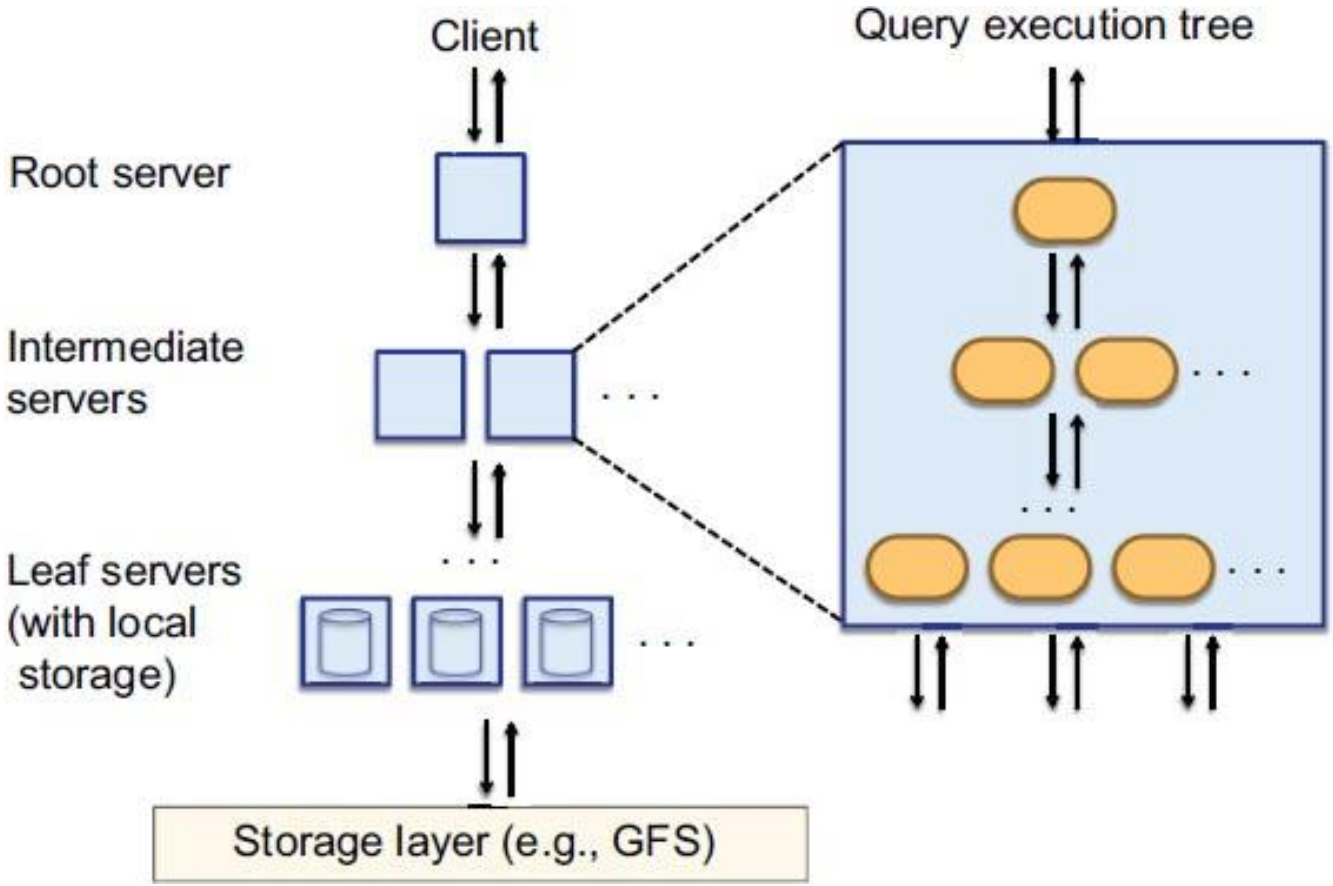
ad hoc query

Implementation details

SHARED MULTI-TENANT ARCHITECTURE

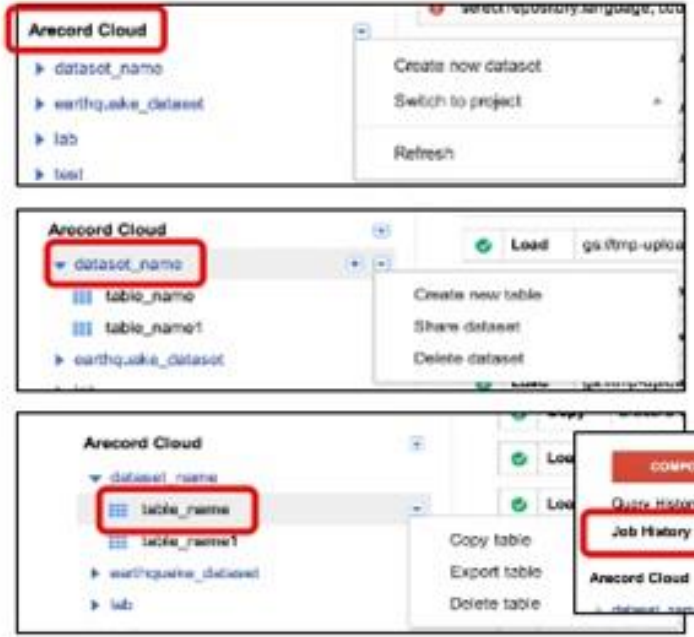
DISTRIBUTED STORAGE AND COLUMNAR DATA LAYOUT

MULTILEVEL EXECUTION TREE



Multi-level execution tree

BigQuery structure



- Project
- Dataset
- Table
- Job

Use case

Cloud Dataflow


```
Pipeline pipeline = Pipeline.create();

PCollection<Tweet> tweets = pipeline.begin()
                                    .apply(new InputFromPubSub())
                                    .apply(new TweetTransfomer());

tweets.apply(new CalculateSentiment());
tweets.apply(new CorrelateKeywords());

pipeline.run();
```

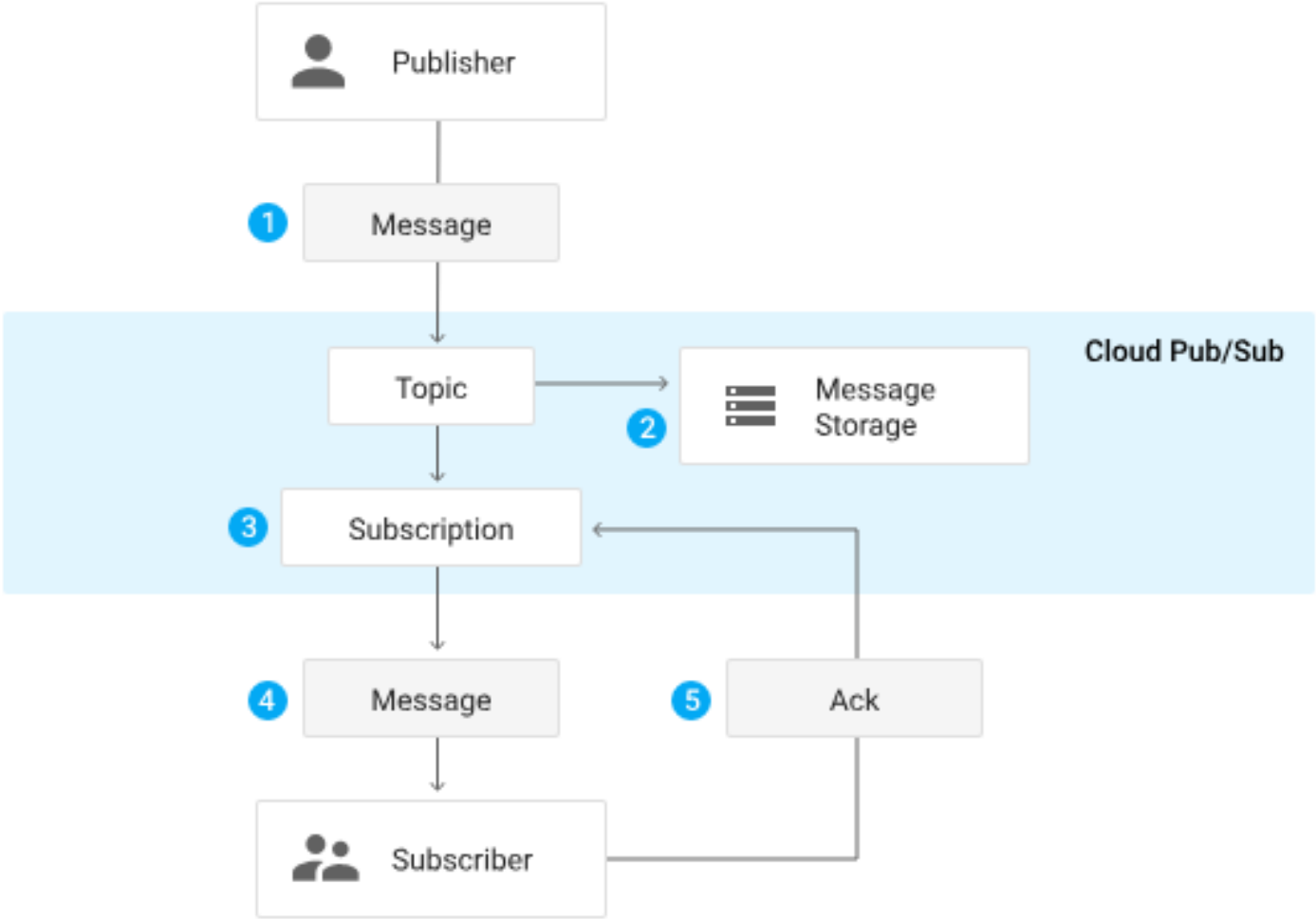
Use case

Fraud detection

IoT analytics in different areas

Logistic

Cloud Pub/Sub


Publisher
1
Message
Cloud Pub/Sub
Topic
2
Message Storage
3
Subscription
4
Message
5
Ack
Subscriber

Use case

Useful sources

- Krishnan, S. P., & L., U. G. (2015). Building your next big thing with google cloud platform: A guide for developers and enterprise architects. New York: Apress.
- https://towardsdatascience.com/want-to-use-bigquery-read-this-fab36822830
- https://dataconomy.com/2014/08/google-cloud-dataflow/
- https://www.enterpriseintegrationpatterns.com/ramblings/82_googlepubsub.html
- https://www.the-swamp.info/blog/simple-serverless-data-pipeline-google-cloud-platform/
- https://cloud.google.com
- https://opencirrus.org/cloud-computing-important/