

BigQuery

...for developers



Big Data

Огромнен брой крайни устройства генерират огромно количество информация

Непосредствена нужда от:

- Съхранение
- Обработка
- Анализ

Косвена нужда от:

- скалиране
- алгоритми
- готови решения

Решения от Google



Cloud Pub/Sub

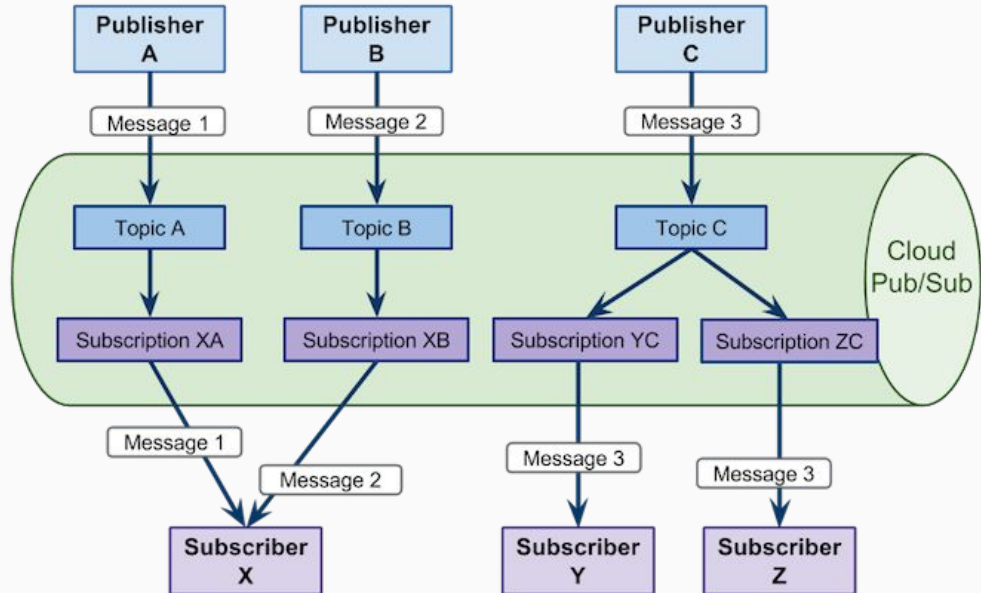


Cloud Dataflow



BigQuery

Cloud Pub/Sub



Cloud Dataflow

BigQuery

- Пълноценна “Data warehouse” инфраструктура
- Базирана на Dremel, GFS, MapReduce и др. научни постижения на Google
- Изцяло разположена върху ресурси на Google
- Съхранение и анализ на петабайти структурирана информация с уникална скорост
- Скалируема според нуждите
- Права и роли за достъп
- Криптирани данни
- Pay-as-you-go model, атрактивна цена с безплатен 1Т всеки месец

Какво ни интересува като програмисти

- Модел
- Въвеждане на данни
- Четене на данни
- SDK
 - Authentication
 - Data model

Как са структурирани данните?

Projects

Datasets

Tables

Как въвеждаме данни?

Таблиците са Append Only!!!

Начини на Append

- HTTP Post / Streaming
 - CSV
 - JSON
- Google Cloud Storage
- Google Cloud Datastore backup

Пример

POST /upload/bigquery/v2/projects/<projectId>/jobs?uploadType=multipart HTTP/1.1

Host: www.googleapis.com

Authorization: Bearer **your_auth_token**

Content-Type: multipart/related; boundary="foo_bar_baz"

Content-Length: **number_of_bytes_in_entire_request_body**

--foo_bar_baz

Content-Type: application/json; charset=UTF-8

```
{
  "configuration": {
    "load": {
      "sourceFormat": "<required for JSON files>",
      "schema": {
        "fields": [
          {"name": "f1", "type": "STRING"},
          {"name": "f2", "type": "INTEGER"}
        ]
      },
      "destinationTable": {
        "projectId": "projectId",
        "datasetId": "datasetId",
        "tableId": "tableId"
      }
    }
  }
}
```

--foo_bar_baz

Content-Type: */*

CSV or JSON data

--foo_bar_baz--

Как четем данни?

Добрият стар SQL

Пример...

Регулярни изрази

Агрегиране

Статистически функции

Прозеречни функции

Специфичен синтаксис за
йерархични данни

UDFs

Table Decorators

Какви клиентски библиотеки се предлагат?

.NET

Objective-C

Java

PHP

Go

Python

JavaScript

Ruby

Node.js

Demo

Изводи

BigQuery е изключително мощен инструмент за анализ на структурирани данни

Предоставя библиотеки на всички основни езици от високо ниво + REST достъп

Дава достатъчно пространство за тестване

Има особености, които трябва да се познават