Druid pour l'analyse de données en temps réel

Yann Esposito

7 Avril 2016

Intro

Plan

Expérience

- Real Time Social Media Analytics

Demande

- Twitter: 20k msg/s, 1msg = 10ko pendant 24h
- Facebook public: 1000 à 2000 msg/s en continu

En pratique

- Twitter: 400 msg/s en continu, pics à 1500

Origine (PHP)

Introduction

- Traitement de donnée gros volume + faible latence
- Typiquement `pulse`

DEMO

Druid

Who

Metamarkets

Goal

Druid is an open source store designed for real-time exploratory analytics on large data sets.

hosted dashboard that would allow users to arbitrarily explore and visualize event streams.

Concepts

- Column-oriented storage layout
- distributed, shared-nothing architecture
- advanced indexing structure

Features

- fast aggregations
- flexible filters
- low latency data ingestion

arbitrary exploration of billion-row tables tables with sub-second latencies

Proof