INF FAKULTÄT FÜR INFORMATIK

Build Your Own OctopusDB: Blinktopus Edition

Ali Hashaam, Ali Memon, Guzel Mussilova, Pavlo Shevchenko

Scientific Project: Databases for Multi-Dimensional Data, Genomics and Modern Hardware

July 10, 2017

Table of Contents

Motivation

1. Companies need to pick only specialized DBMSs, each tailored to their specific use-case.

Motivation

1. Companies need to pick only specialized DBMSs, each tailored to their specific use-case.
   $\Rightarrow$ Need for *one size fits all system* (e.g. HTAP)

Motivation

1. Companies need to pick only specialized DBMSs, each tailored to their specific use-case.
   $\Rightarrow$ Need for *one size fits all system* (e.g. HTAP)
2. Support OLAP queries for analysis over real-time data (i.e., freshness).

Motivation

1. Companies need to pick only specialized DBMSs, each tailored to their specific use-case.
   ⇒ Need for *one size fits all system* (e.g. HTAP)
2. Support OLAP queries for analysis over real-time data (i.e., freshness).
   ⇒ Explore the techniques related to more interactive queries (e.g. *Approximate Query Processing*)

Background

1. OctopusDB
   - uses logs as a primary storage
   - mimicks several types of systems (OLAP, OLTP, etc.) by representing them as *Storage Views*

Background

1. OctopusDB
   - uses logs as a primary storage
   - mimicks several types of systems (OLAP, OLTP, etc.) by representing them as *Storage Views*
2. BlinkDB
   - Successfully integrates AQP techniques into its architecture

Related Work

1. Apache Samza
   - logs as a primary structure
   - replicates logs on multiple nodes
2. Rodent Store
   - represents data in the various physical layouts
   - provides DBAs a high-level interface to specify the data physical representation by means of storage algebra
3. Snappy Data
   - AQP Support
   - Uses numerous types of synopses
   - User defines the level of accuracy and the number of column sets to approximate the results

Thank you! Any questions?

Literature

1. Jindal, Alekh. "The mimicking octopus: Towards a one-size-fits-all database architecture." VLDB PhD Workshop. 2010.
2. Dittrich, Jens, and Alekh Jindal. "Towards a One Size Fits All Database Architecture." CIDR. 2011.
3. Jindal, Alekh. "OctopusDB: flexible and scalable storage management for arbitrary database engines." (2012).
4. Mozafari, Barzan, and Ning Niu. "A Handbook for Building an Approximate Query Engine." IEEE Data Eng. Bull. 38, no. 3 (2015): 3-29.
5. Cormode, Graham, Minos Garofalakis, Peter J. Haas, and Chris Jermaine. "Synopses for massive data: Samples, histograms, wavelets, sketches." Foundations and Trends in Databases 4, no. 13 (2012): 1-294.