INF FAKULTÄT FÜR INFORMATIK

Build Your Own OctopusDB. Blinktopus

Ali Hashaam, Ali Memon, Guzel Mussilova, Pavlo Shevchenko

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

May 2, 2017

Table of Contents

Motivation

Modern enterprises need to pick the right DBMSs for their data managing problems.

[1] A. Jindal. The Mimicking Octopus: Towards a one-size-fits-all Database Architecture, 2010

Motivation

Modern enterprises need to pick the right DBMSs for their data managing problems.

1. Use specialized solution for each application.

[1] A. Jindal. The Mimicking Octopus: Towards a one-size-fits-all Database Architecture, 2010

Motivation

Modern enterprises need to pick the right DBMSs for their data managing problems.

1. Use specialized solution for each application.
   $\rightarrow$ costly due to licensing fees, integration overhead and DBA costs

[1] A. Jindal. The Mimicking Octopus: Towards a one-size-fits-all Database Architecture, 2010

Motivation

Modern enterprises need to pick the right DBMSs for their data managing problems.

1. Use specialized solution for each application.
   $\rightarrow$ costly due to licensing fees, integration overhead and DBA costs
2. Use a single specialized DBMS for all applications.

[1] A. Jindal. The Mimicking Octopus: Towards a one-size-fits-all Database Architecture, 2010

Motivation

Modern enterprises need to pick the right DBMSs for their data managing problems.

1. Use specialized solution for each application.
   → costly due to licensing fees, integration overhead and DBA costs
2. Use a single specialized DBMS for all applications.
   → compromise heavily on performance. [1]

[1] A. Jindal. The Mimicking Octopus: Towards a one-size-fits-all Database Architecture, 2010

Idea of OctopusDB

Create a new type of database system without fixed store that will mimic several existing systems.

Idea of OctopusDB

Create a new type of database system without fixed store that will mimic several existing systems.

- Storage Views

Like "real" octopuses can mimic other creatures and adjust to the environment

Idea of OctopusDB

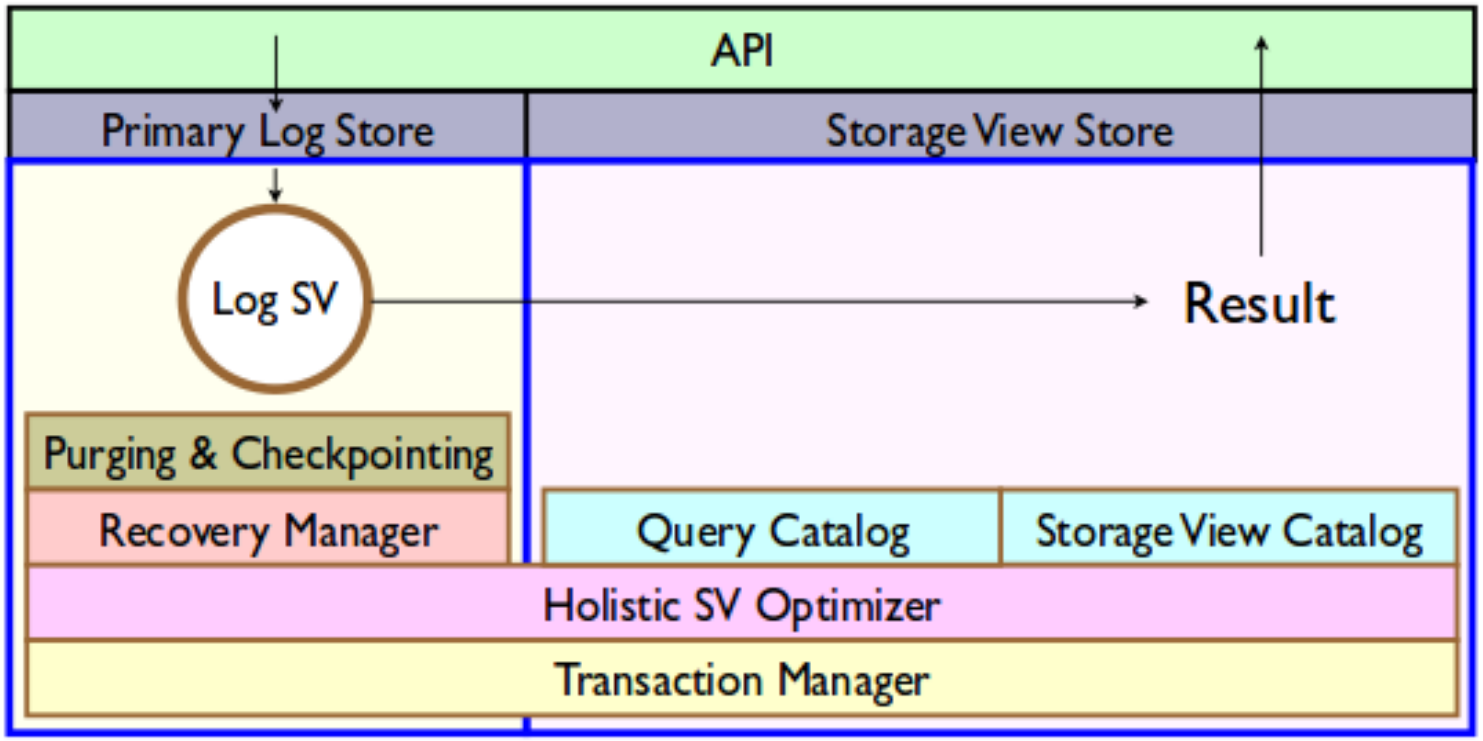


Figure 1: OctopusDB Architecture

2

[2] A. Jindal. The Mimicking Octopus: Towards a one-size-fits-all Database Architecture, 2010

Our Goal

- Not to **clone** OctopusDB

Our Goal

- Not to **clone** OctopusDB
- Provide a **framework** that gives user a chance to act as *Holistic SV Optimizer*

Our Goal

- Not to **clone** OctopusDB
- Provide a **framework** that gives user a chance to act as *Holistic SV Optimizer*
- Add **Approximate Query Processing (AQP)**. BlinkDB

Our Goal

- Not to **clone** OctopusDB
- Provide a **framework** that gives user a chance to act as *Holistic SV Optimizer*
- Add **Approximate Query Processing (AQP)**. BlinkDB
- **Evaluate** performance depending on choice of SV

AQP. Motivation

The goal is to provide approximate answers with acceptable accuracy in orders of magnitude less time than that for the exact query processing.[3]

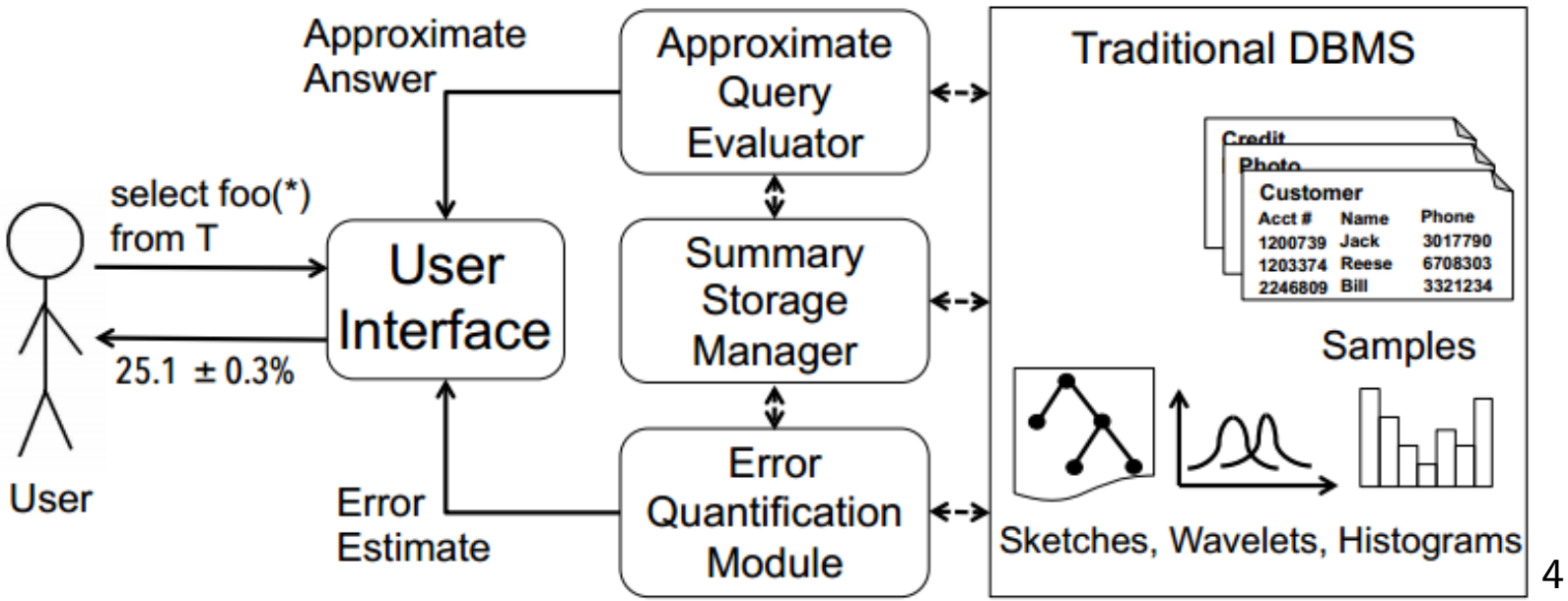
Revised query optimization goals:

- Efficiently Accessing all query-relevant tuples.
- Choose the best query plan among the available equivalent ones.

In many cases, not all query-relevant data needs to be accessed.

[3] Liu, Qing. Approximate Query Processing (Reference work entry) in: Liu, Ling, and M. Tamer zsu. Encyclopedia of database systems. Vol. 6. Berlin, Heidelberg, Germany: Springer, 2009.

AQP. Workflow



[4]

[4]The general anatomy of approximate query processing system

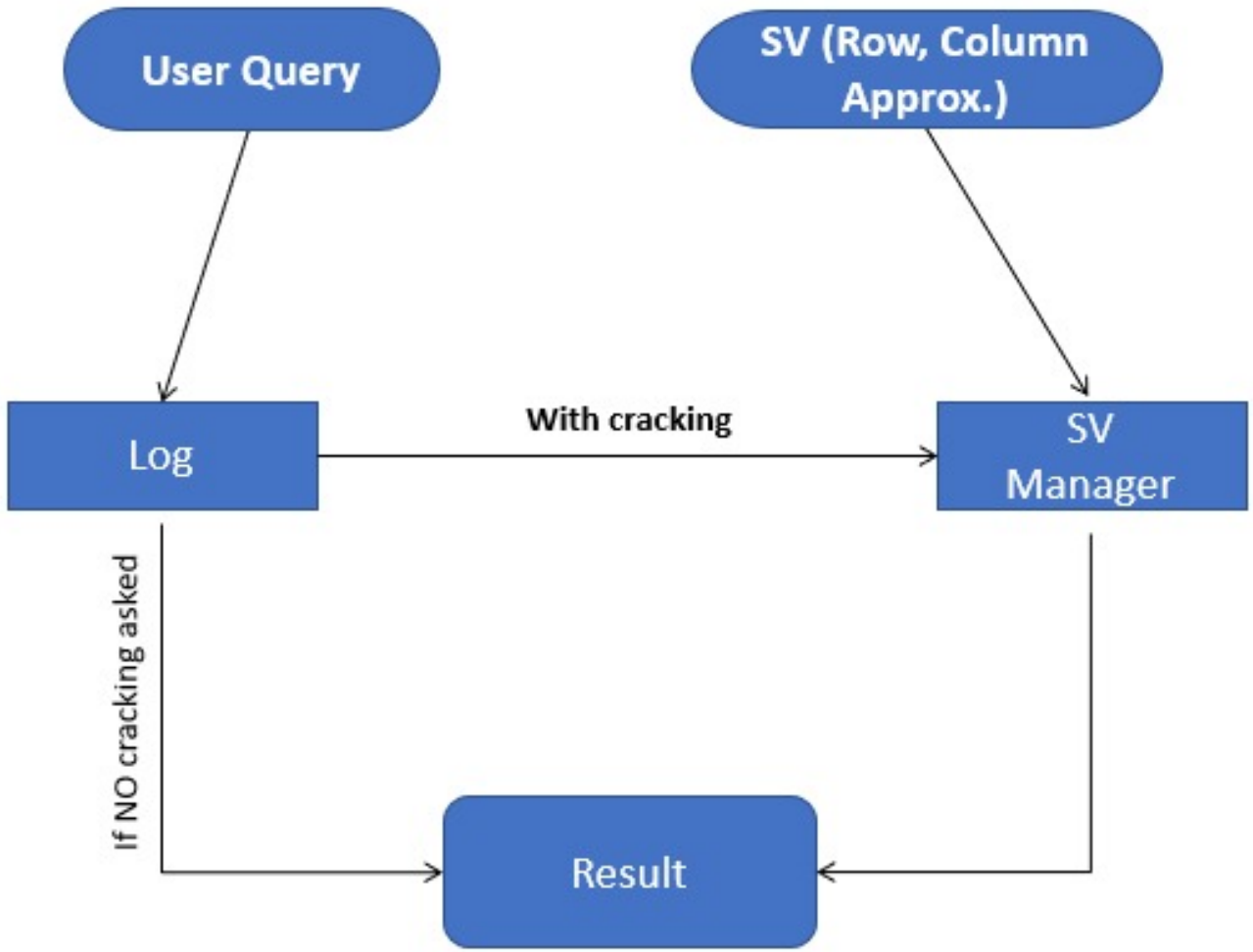
AQP. Synopsis Manager

A synopsis captures essential properties of the real data while taking less space.

The synopses manager is responsible for:

- Type of summary to use(Samples, histograms, sketches, wavelets etc.)
- When to build it (offline vs. online)
- How to store it (to use overlapping samples, how to structure/index/cache the synopses)
- When to update it (batch or online)

Our Vision


User Query
SV (Row, Column Approx.)
Log
With cracking
SV Manager
If NO cracking asked
Result

Building a Blinktopus: What is our challenge?

First, the Octopus:

- Store incoming data in logs.
- Query the logs (just a filter query)
- Allow users to create views (row, column) over certain logs.
- List all views and logs
- Launch the query over views or over logs, see the changes in performance.

Building a Blinktopus: What is our challenge?

Enter AQP:

- What synopsis can we easily support as a view for a specific query? Which will we choose to test? (Samples, histograms?)
- Do Octopuses and AQP match well together?
- How will we allow users to build this view?
- How will we support queries using this view?

Blinktopuses are entirely a novel idea
No one has done this before, ever..!

Project Organisation.Roles

Team:

Guzel - Manager (Team Leader)
Ali H. - Developer
Ali M. - Developer
Pavlo - Researcher

Supervisor:

Gabriel Campero Durand

Changing roles after each milestone.

Project Organisation.Schedule

Milestones

02.05.2017	MS-I (Kick-Off)
23.05.2017	MS-II (Concepts)
13.06.2017	MS-III (Implementation)
04.07.2017	MS-IV (Final)

Meetings

Team Meetings: Mo 14-15
Meetings with supervisor: We 10-11

Thank you for your attention! 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. Idreos, Stratos, Martin L. Kersten, and Stefan Manegold. "Database Cracking." In CIDR, vol. 7, pp. 68-78. 2007.
5. Mozafari, Barzan. "Approximate query engines: Commercial challenges and research opportunities." SIGMOD, 2017.