In Memory Databases
A Real Time Analytics Solution

Bob Wakefield
Principal
bob@MassStreet.net
Twitter:
@BobLovesData

Mass Street Analytics

Who is Mass Street?

- Boutique data consultancy
- Looks to provide organizations with analytics expertise

Mass Street Analytics

What We’re Gonna Talk About

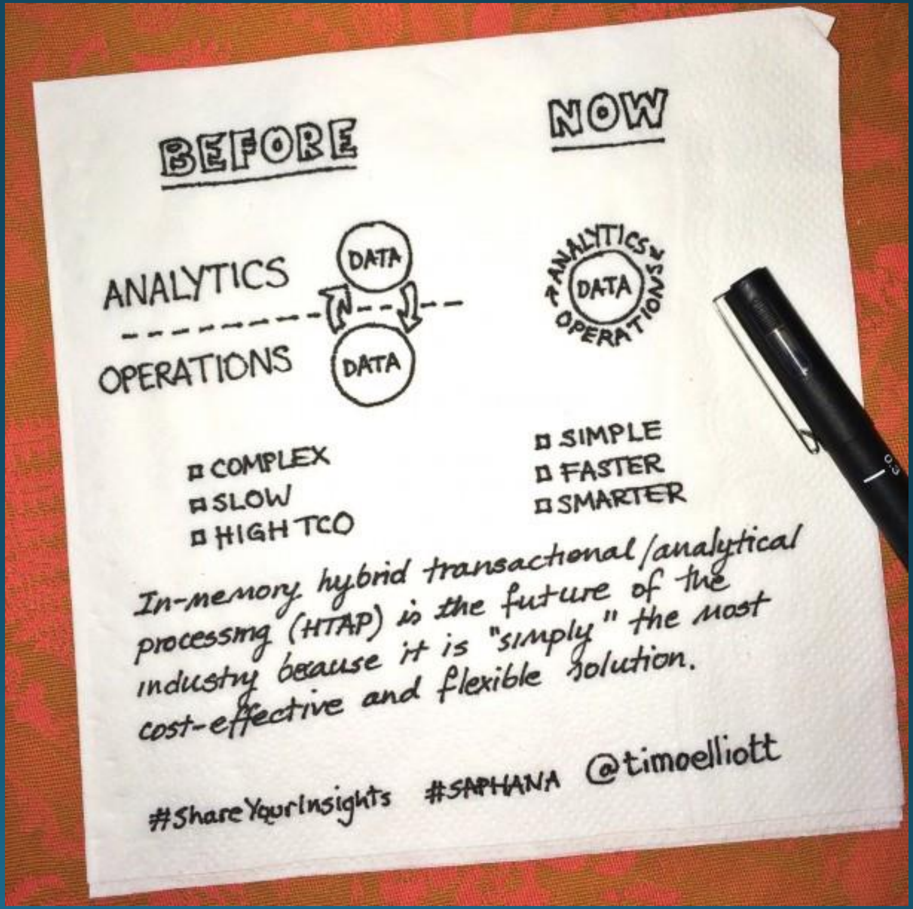
1. HTAP
2. Real Time Analytics
3. In memory databases

What We’re NOT Gonna Talk About

- Deep technical info on how in memory DBs work.
- The advancement of in memory technology.

What Is HTAP?

Hybrid Transactional Analytic Processing



□ COMPLEX
□ SLOW
□ HIGH TCO

□ SIMPLE
□ FASTER
□ SMARTER

In-memory hybrid transactional/analytical processing (HTAP) is the future of the industry because it is "simply" the most cost-effective and flexible solution.

#ShareYourInsights #SAPHANA @timoelliott

Image Source: Timo Elliott

Challenges Tackled By HTAP

- Removes need to ETL data to warehouse
- Transactional data readily available
- Aggregates points to fresh HTAP data
- Cuts the need for copies of data

Source: Gartner, Inc.

Mass Street Analytics

What Is Real Time

- I don’t think we have a standard definition yet
- Real time = instantaneous
- More Practical
  - Arbitrarily close enough to instantaneous to go ahead and call it real time.

Mass Street Analytics

Why Real Time Analytics

- Gaining competitive edge
- Everybody wants their data right no
- Enabling IoT

Mass Street Analytics

Use Cases for Real Time

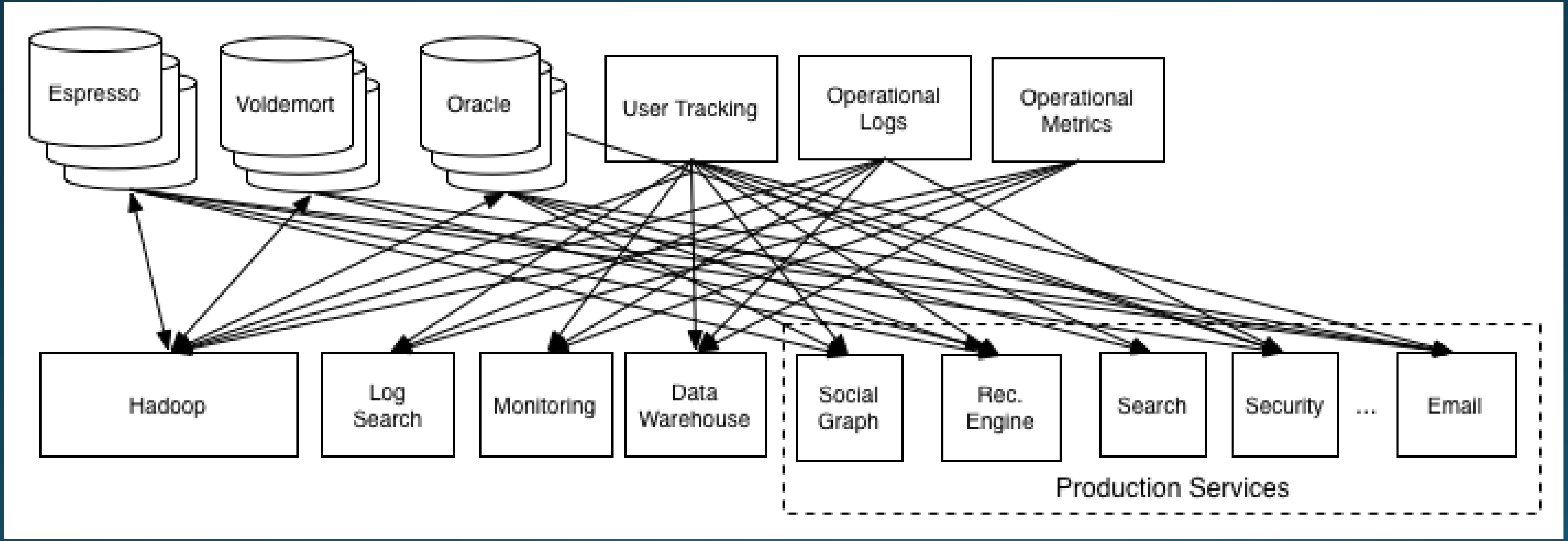
- “Traditional Use Cases”
  - asset trading
  - app performance monitoring
- Um. Everything.

Use Cases for Real Time

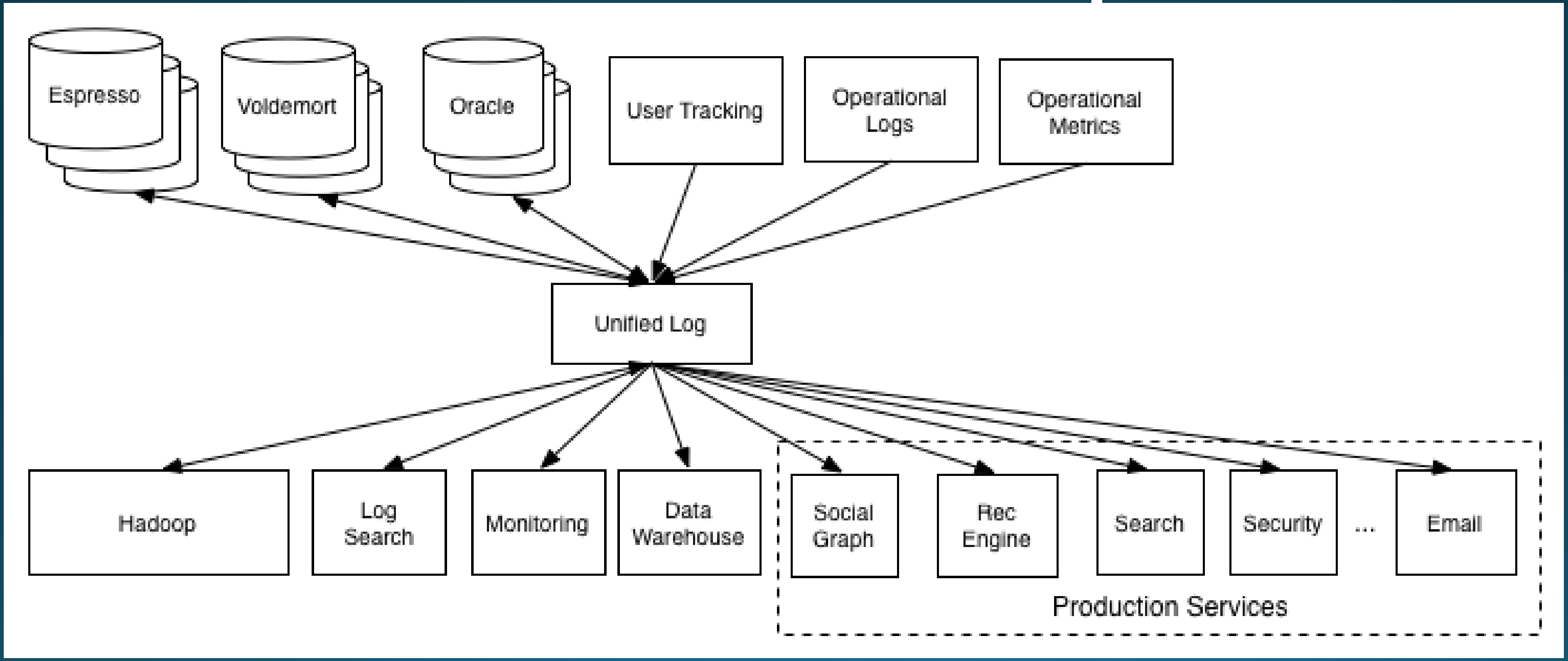
Everything is an event that occurs over time.

Mass Street Analytics

The Log: A Unifying Abstraction For Data Transport



The Log: A Unifying Abstraction For Data Transport



Real Time In The Real World

- Kafka/Storm Hadoop < 1 minute
- SQL Server Replication ~ 5 min
- SSIS Job ~ 15 min

I want to speed up analytics but…

- I’m having performance issues in my DB.
- My architecture is clunky.
- My read/writes are colliding.
- I don’t know how to make it happen.

Enter In Memory Databases

- Not the same as having a caching layer
- Data MUST be stored in main memory
- Data accessed without disk I/O instructions

Comparison Of Various DB Technologies

	Open Source	In Memory	Notes
MemSQL		X	My pick
VoltDB	X	X	In memory only
NuoDB		X	What?
MySQL Cluster	X	X	Since when?
Netezza			$$$
MS PDW			$$$
Cassandra	X		OLTP only

Introduction to MemSQL

- Built for real time
- Horizontal scale out on commodity hardware
- ACID compliant
- SQL complaint
- Mixed OLTP and OLAP workloads

Mass Street Analytics

Introduction to MemSQL

- Uses MySQL wire protocol
- MVCC + lock free data structures
  - Goodbye WITH NOLOCK
- JSON data type
- Row store and a column store

Introduction to MemSQL

- JDBC/ODBC compliant
- Connects to Tableau
  - enables self serve BI
- Spark Connector

Introduction to MemSQL

- Client with 500 nodes.

Introduction to MemSQL

- All queries get turned into compiled code.
- Shared nothing architecture.

Mass Street Analytics

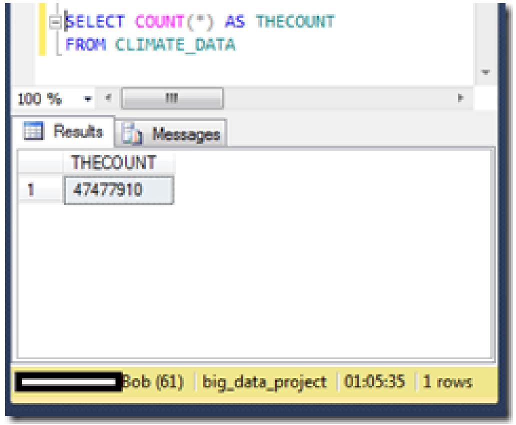
Experimentation with MemSQL

- Assignment: Store some “big data” and analyze it
- Easy installation
- Connect in with SQuirreL

My Personal Computer Lab Setup

- 1 off the shelf windows box
  - Intel 3.3 GHz
  - 4 GB RAM
  - 1TB HDD
- 1 Custom box running Ubuntu 14.4
  - AMD 3 GHz
  - 32 GB RAM
  - 1 TB HDD
- 2 off the shelf servers running Ubuntu 14.4
  - AMD 3GHz
  - 32 GB RAM
  - 3TB HDD

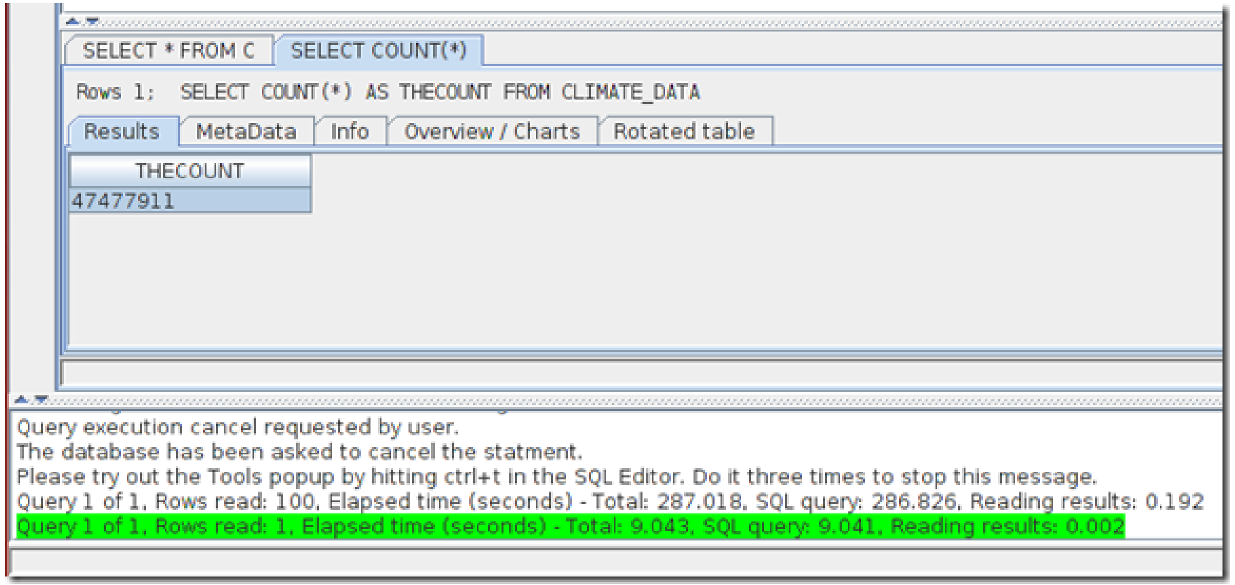
SQL Server vs. MemSQL



```
SELECT COUNT(*) AS THECOUNT
FROM CLIMATE_DATA
```

```
  THECOUNT
1 47477910
```

[redacted] Bob (61) | big_data_project | 01:05:35 | 1 rows



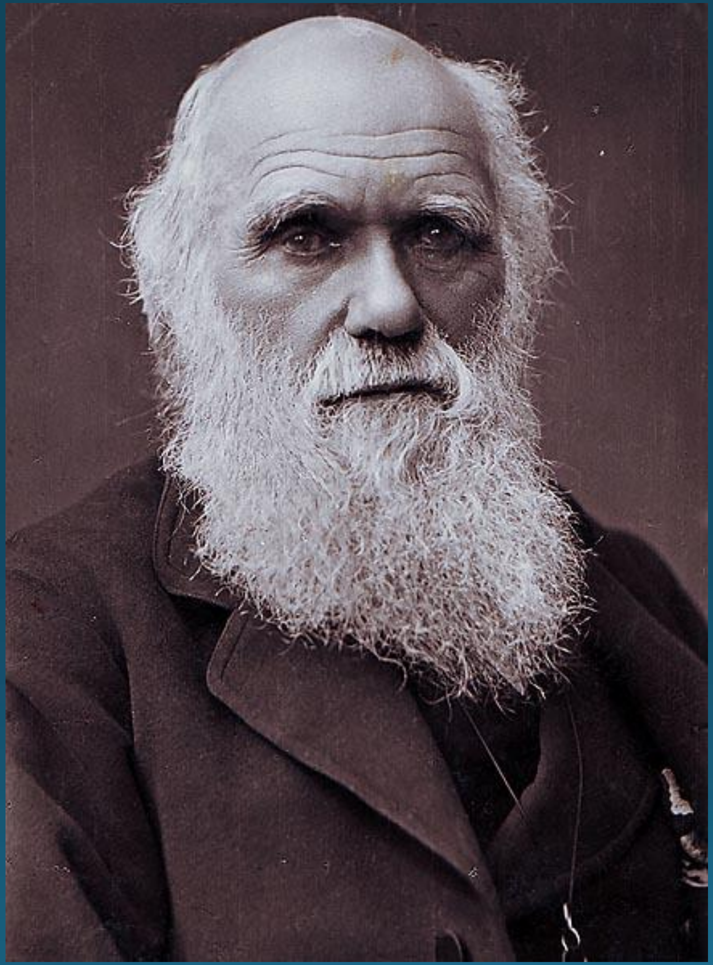
```
Rows 1;  SELECT COUNT(*) AS THECOUNT FROM CLIMATE_DATA
```

```
THECOUNT
47477911
```

```
Query execution cancel requested by user.
The database has been asked to cancel the statment.
Please try out the Tools popup by hitting ctrl+t in the SQL Editor. Do it three times to stop this message.
Query 1 of 1, Rows read: 100, Elapsed time (seconds) - Total: 287.018, SQL query: 286.826, Reading results: 0.192
Query 1 of 1, Rows read: 1, Elapsed time (seconds) - Total: 9.043, SQL query: 9.041, Reading results: 0.002
```

Mass Street Analytics

Evolution Of An Architecture

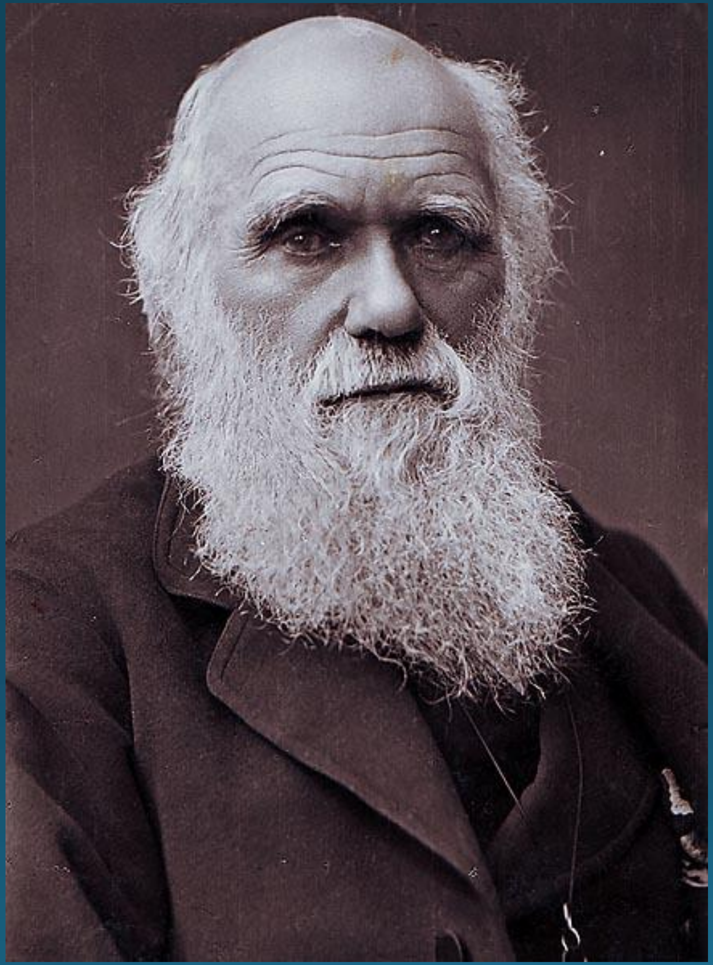


Mass Street Analytics

We are experiencing technical difficulties! For a demonstration of MemSql Ops, please see the link in the video description!

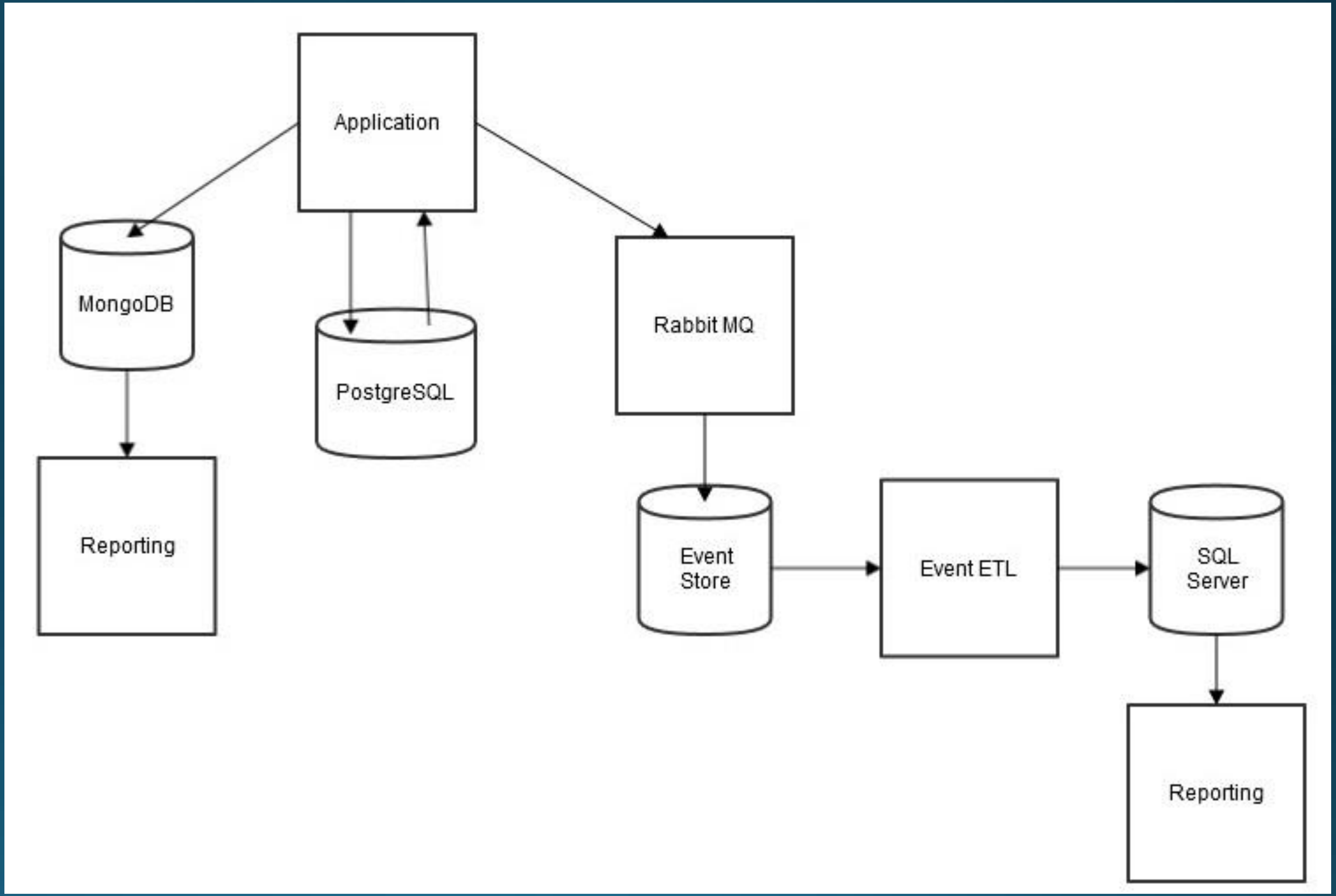
Mass Street Analytics

Evolution Of An Architecture

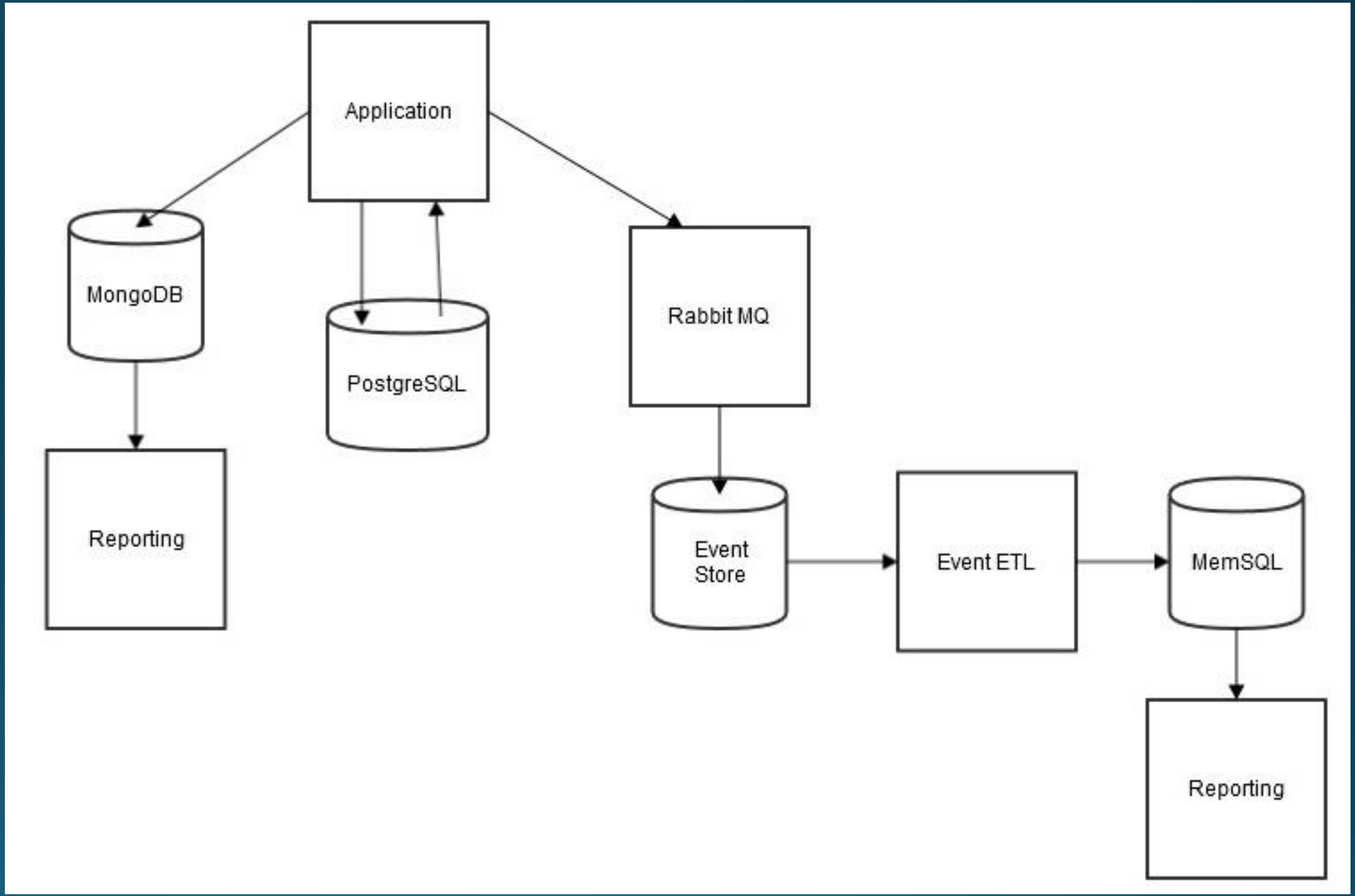


Mass Street Analytics

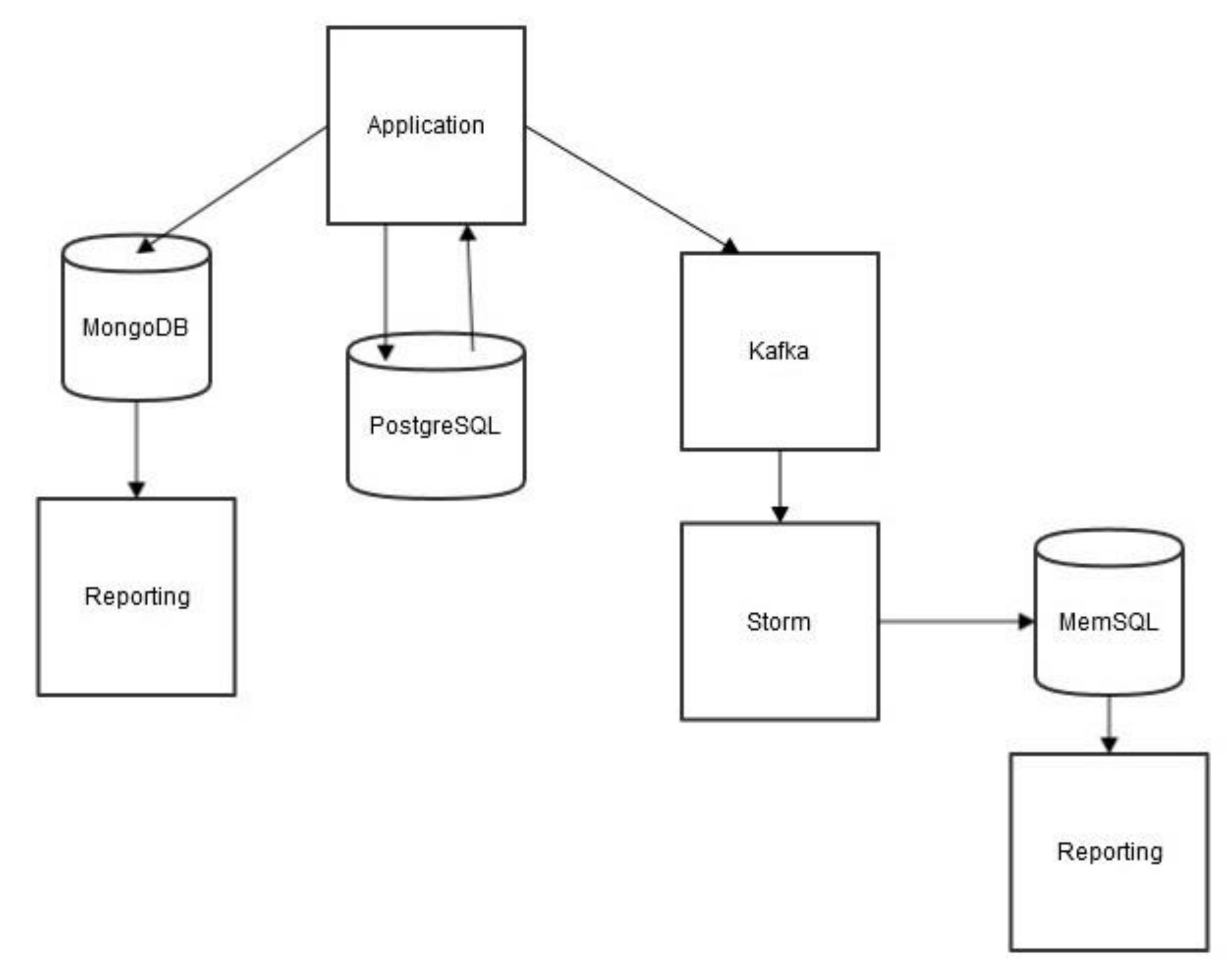
Initial State



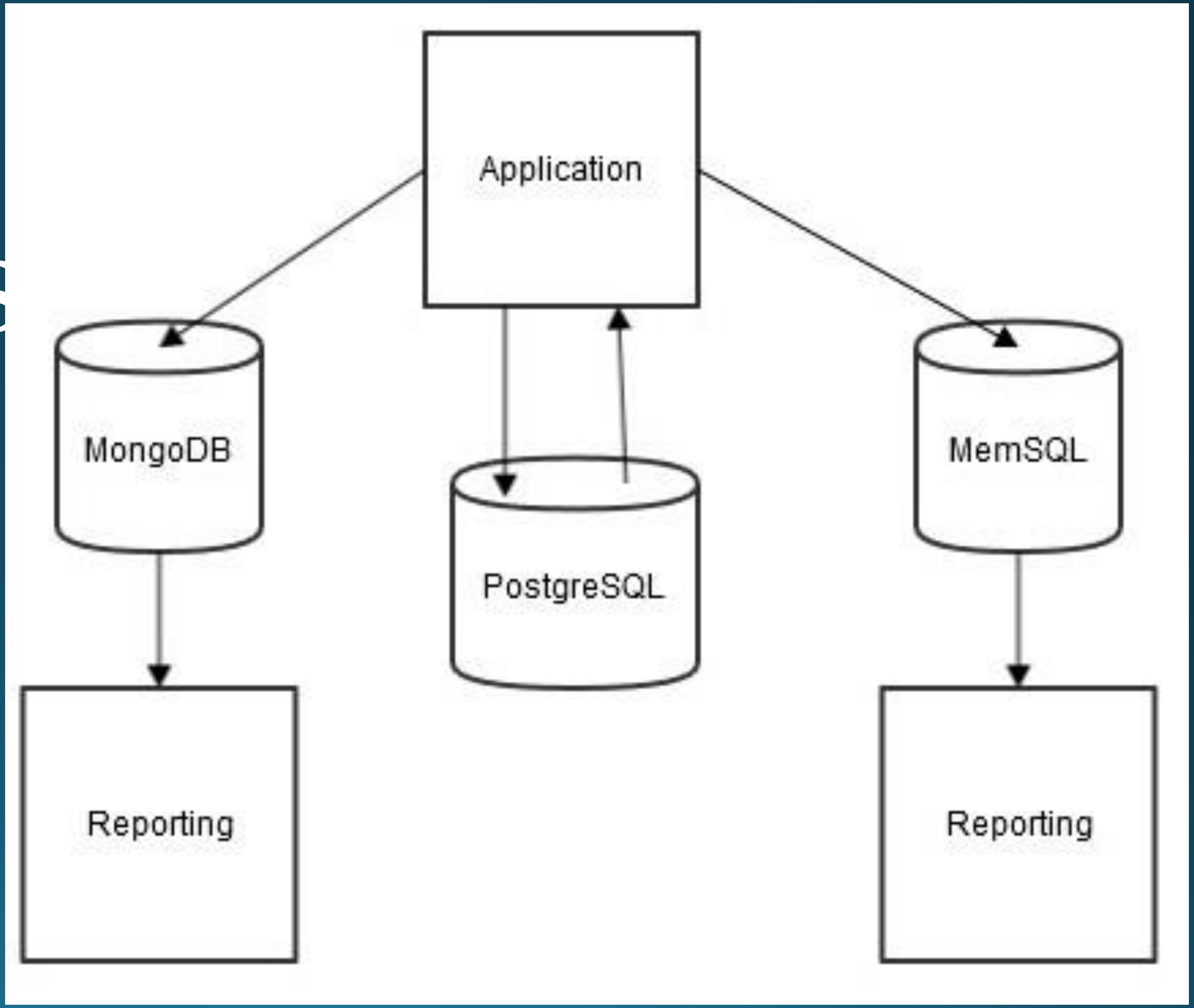
V1 Eliminate Batch Processin g



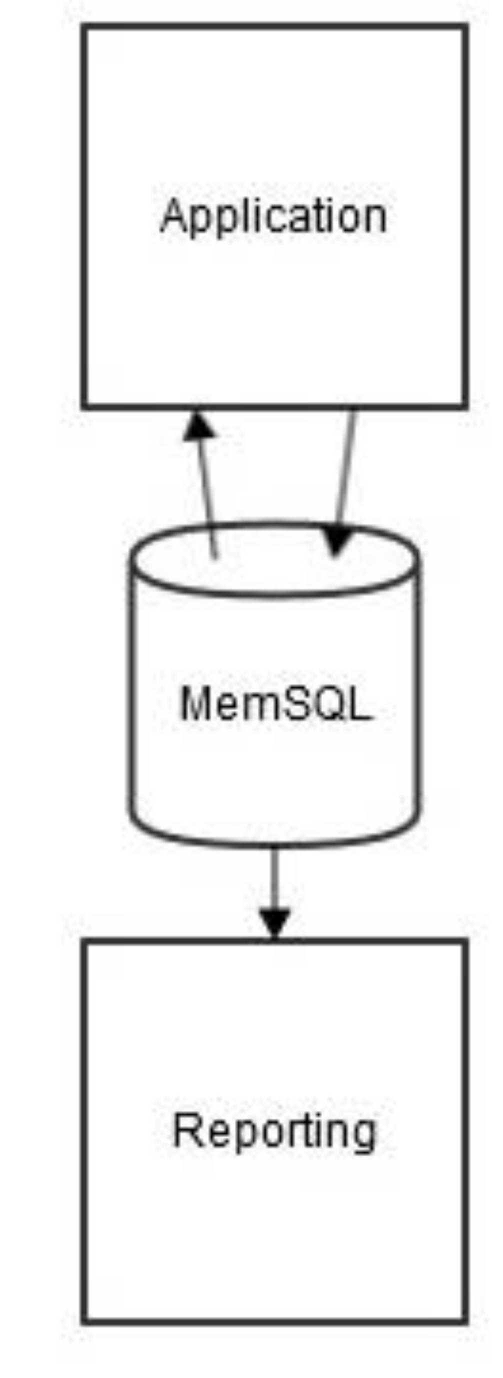
V2
Replace transport elements with Apache software



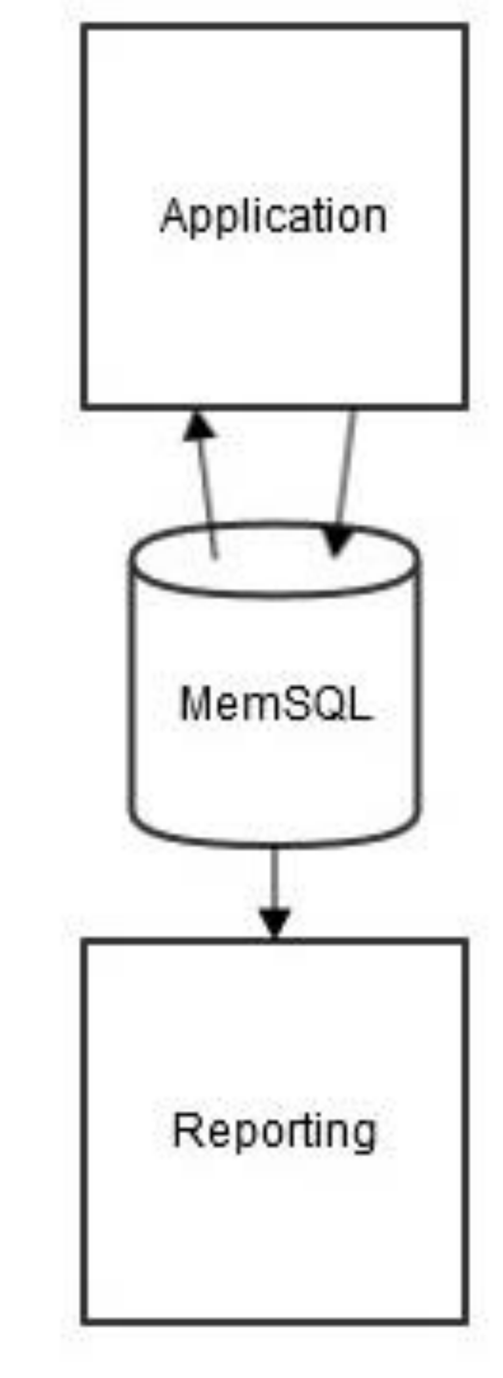
V3
Remove elements that are theoretically unnecessary



V4
The ultimate incarnation of HTAP



V4
The ultimate incarnation of HTAP



Engineering Food For Thought

- Everyone says HTAP needs to be done in memory. What if you have more data than RAM? Is that even a big deal?
- Do you just keep the warm data then send the cold stuff to Hadoop then federate it with virtualization?
- If disk I/O is a bottle neck, could the problem be alleviated with SSDs?

Mass Street Analytics

If you want more technical info

- http://developers.memsql.com