GridGain

THE DATAWAREHOUSE IS DEAD- LONG LIVE THE DATAWAREHOUSE

Rachel Pedreschi

Lead Solutions Architect- GridGrain Systems

@rachelpedreschi

Gartner 2014 CoolVendor

AlwaysOn TOP 100 WINNER ONCLOUD 2015

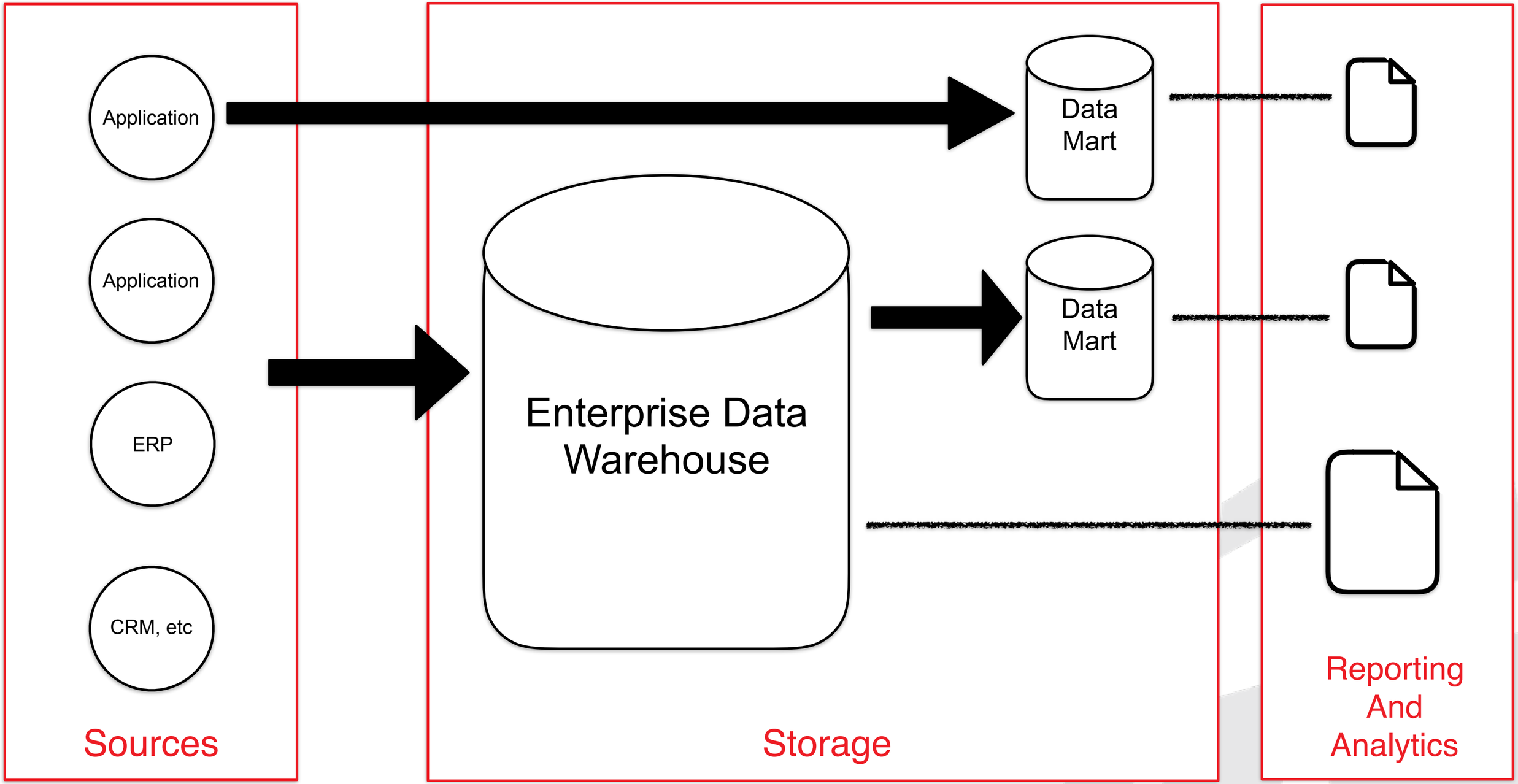
AlwaysOn GLOBAL 250 WINNER SILICON VALLEY INNOVATION SUMMIT 2015

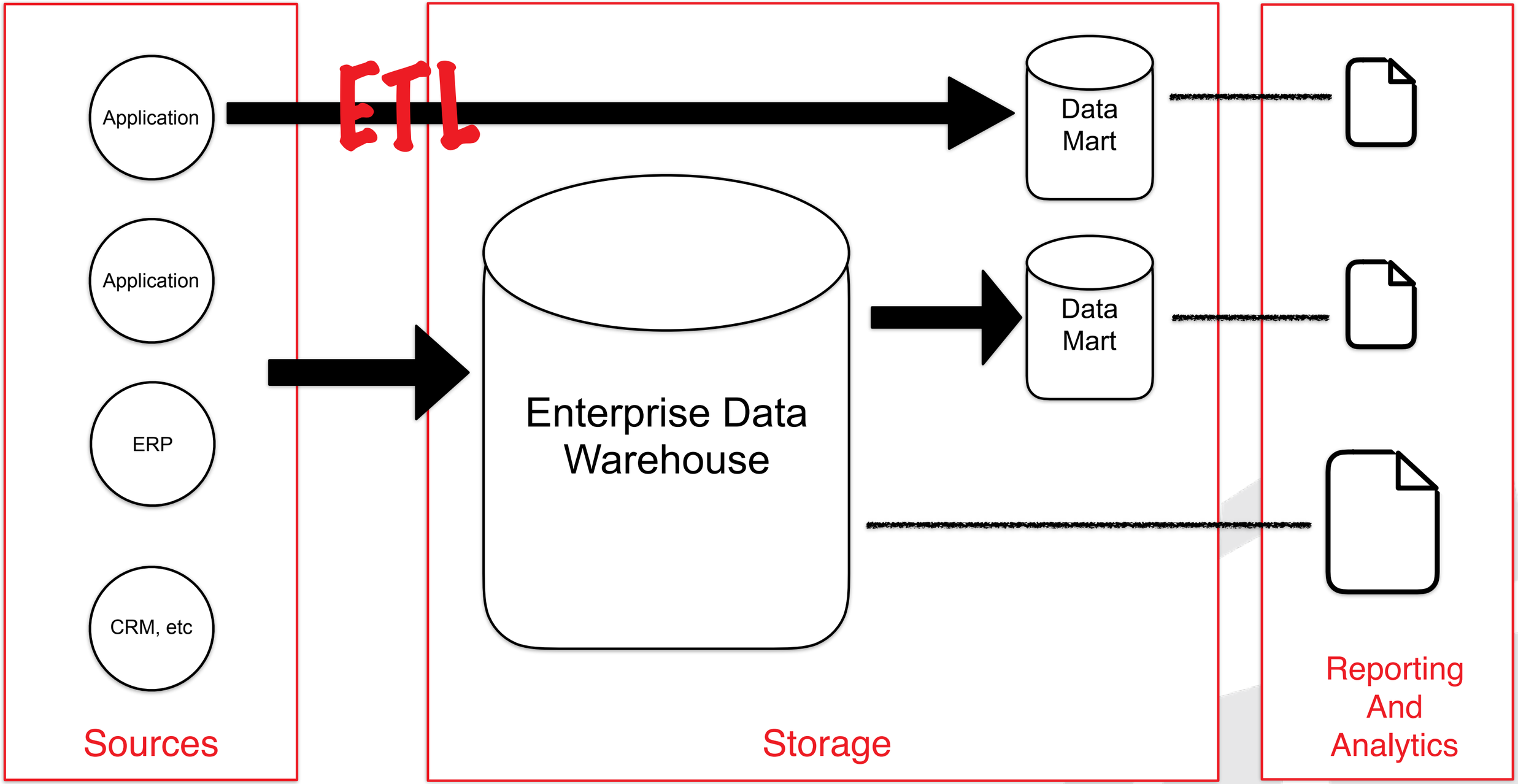
AlwaysOn GLOBAL 250 WINNER SILICON VALLEY INNOVATION SUMMIT 2014

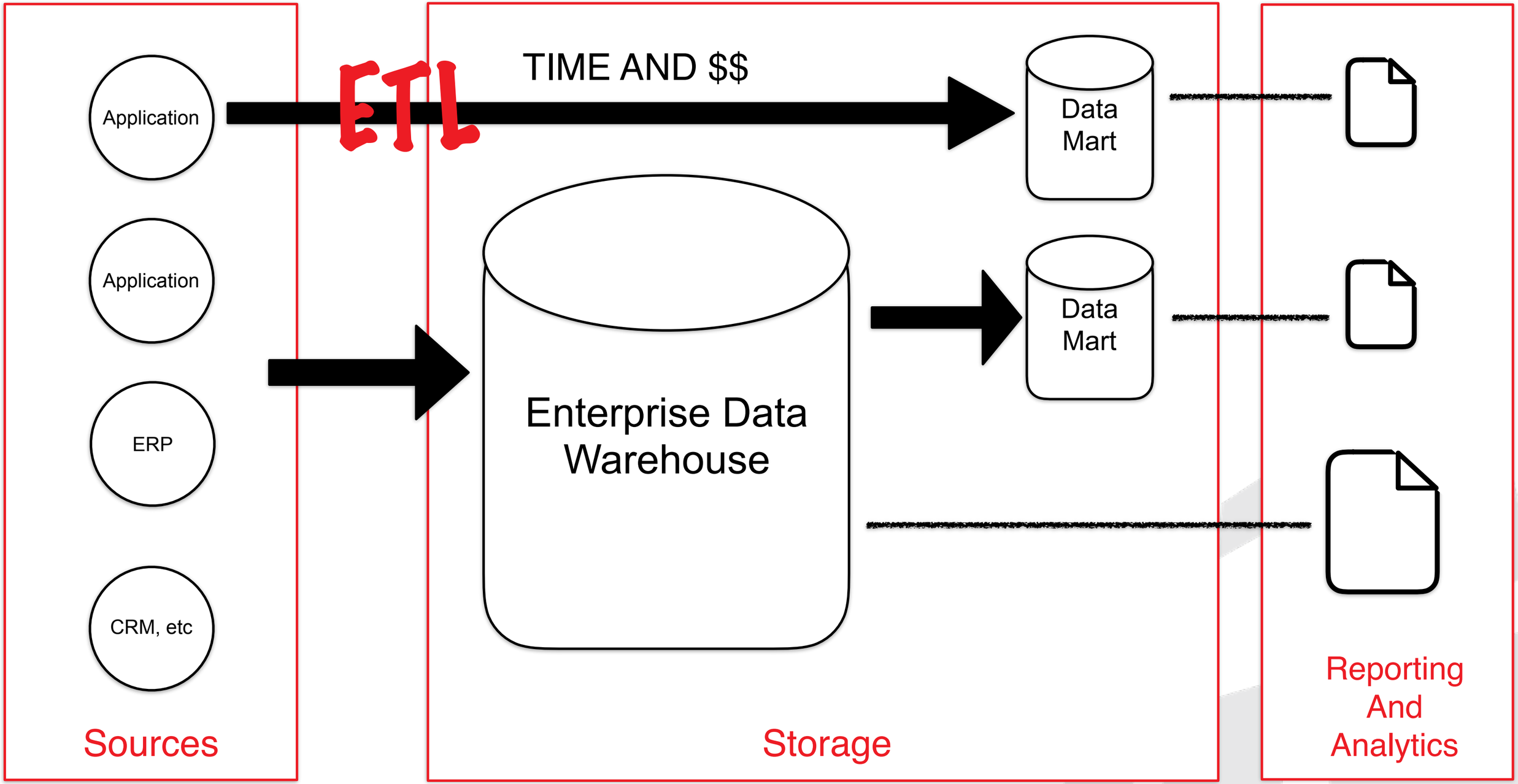
BIG DATA 50

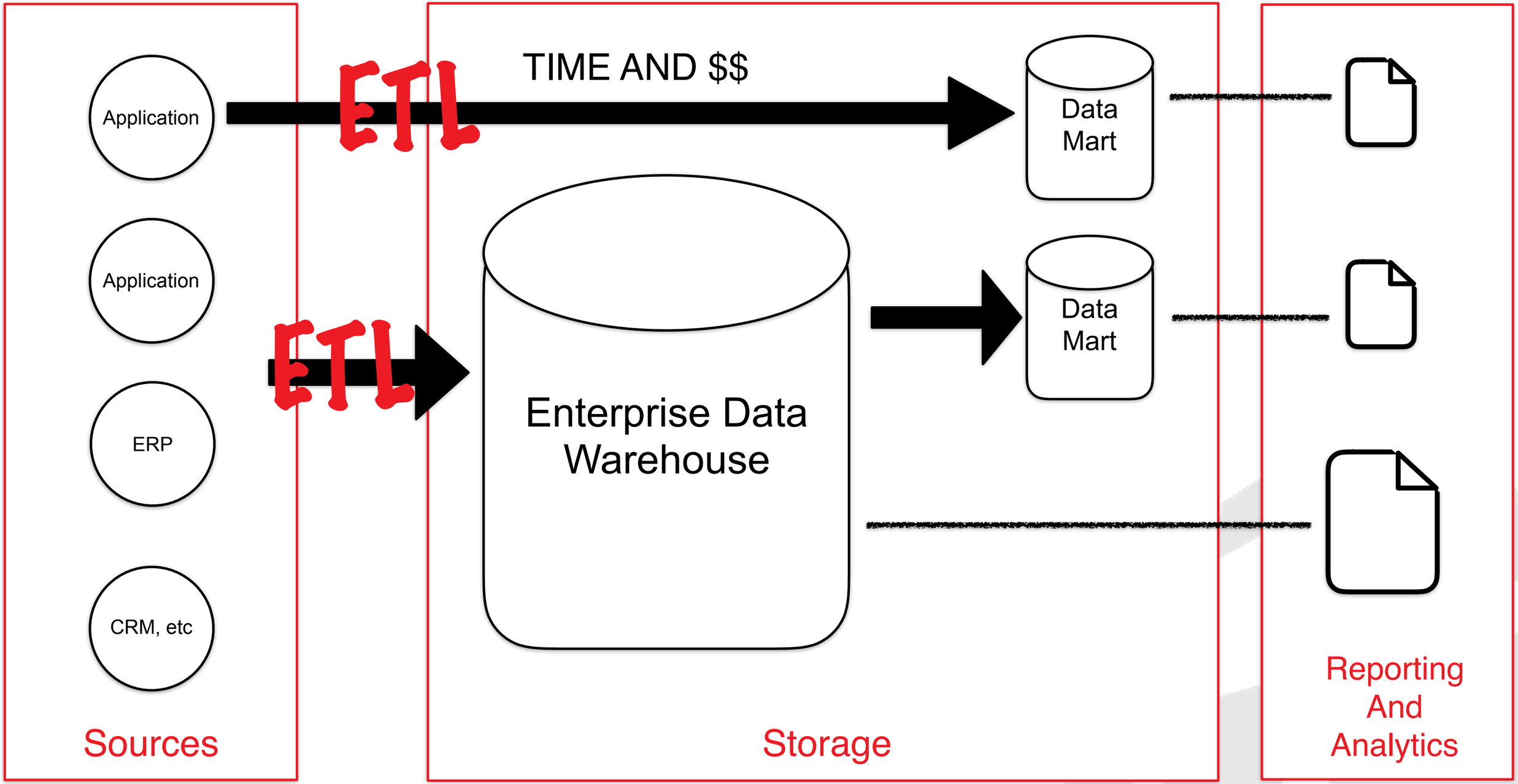
Tell ‘em what you’re gonna tell ‘em

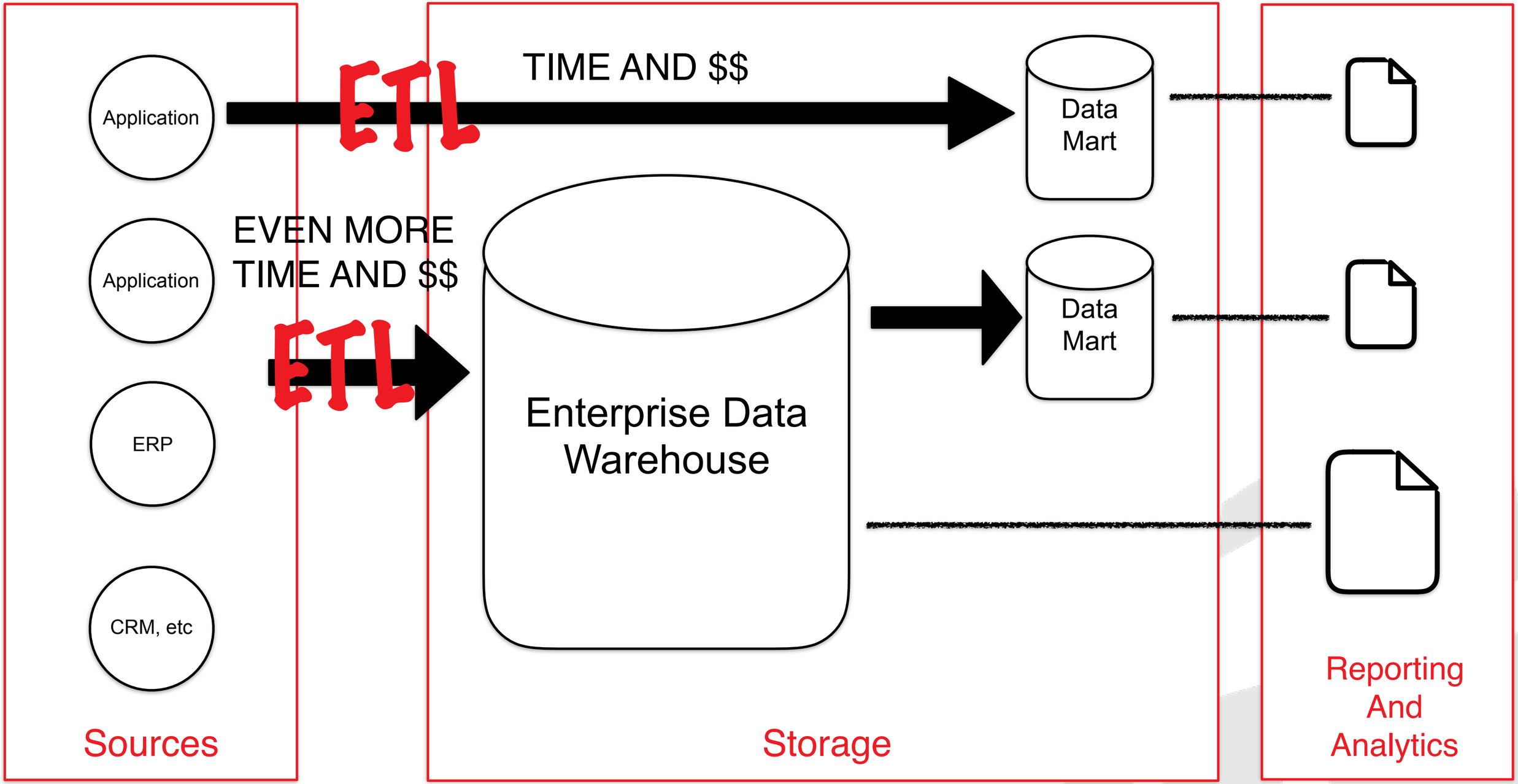
- Review 3 common Data Warehouse architectures
- Why it is time for In-Memory solutions
- Introduction to Apache Ignite and GridGain
- How Ignite fits into DW architectures
- What Next?





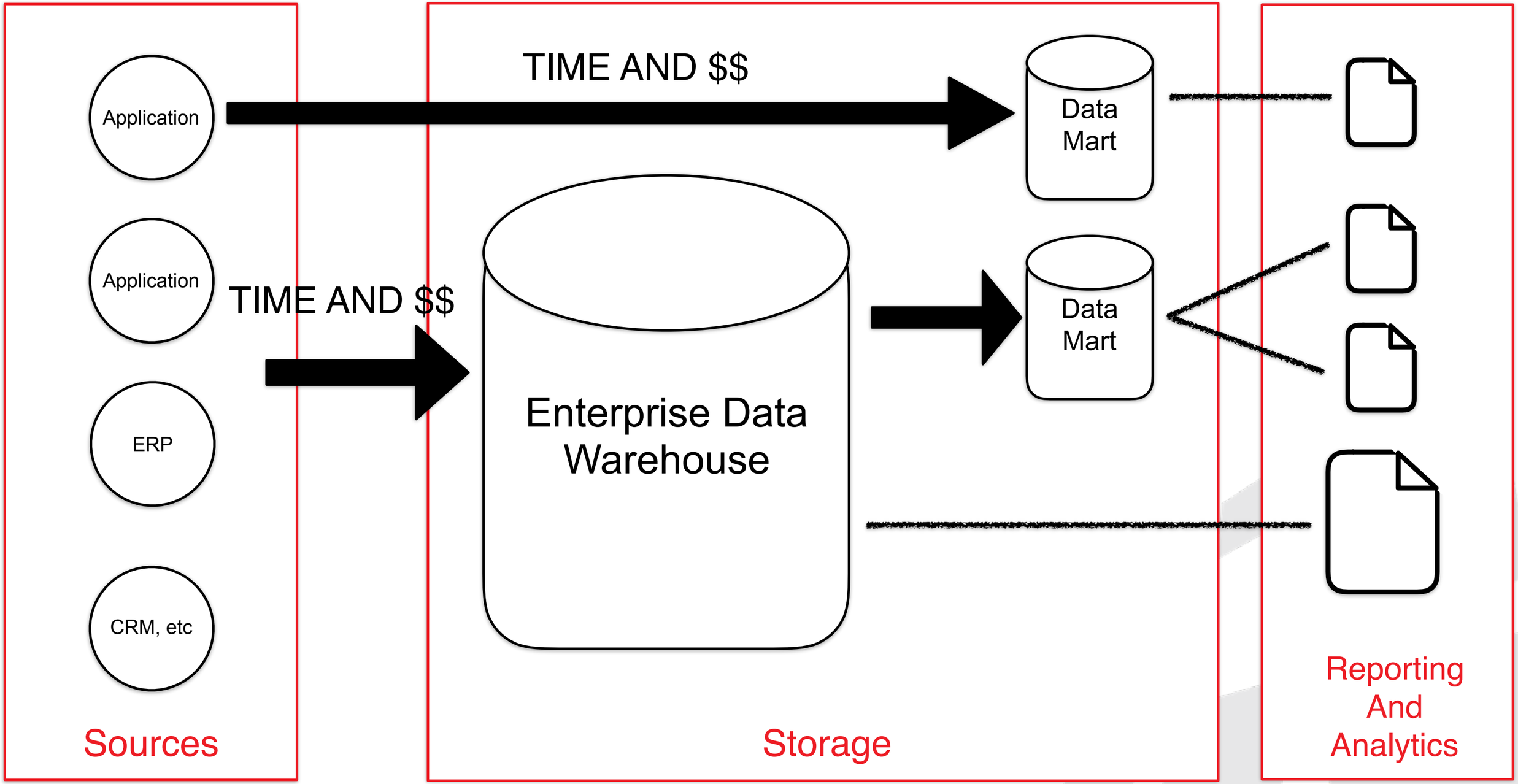


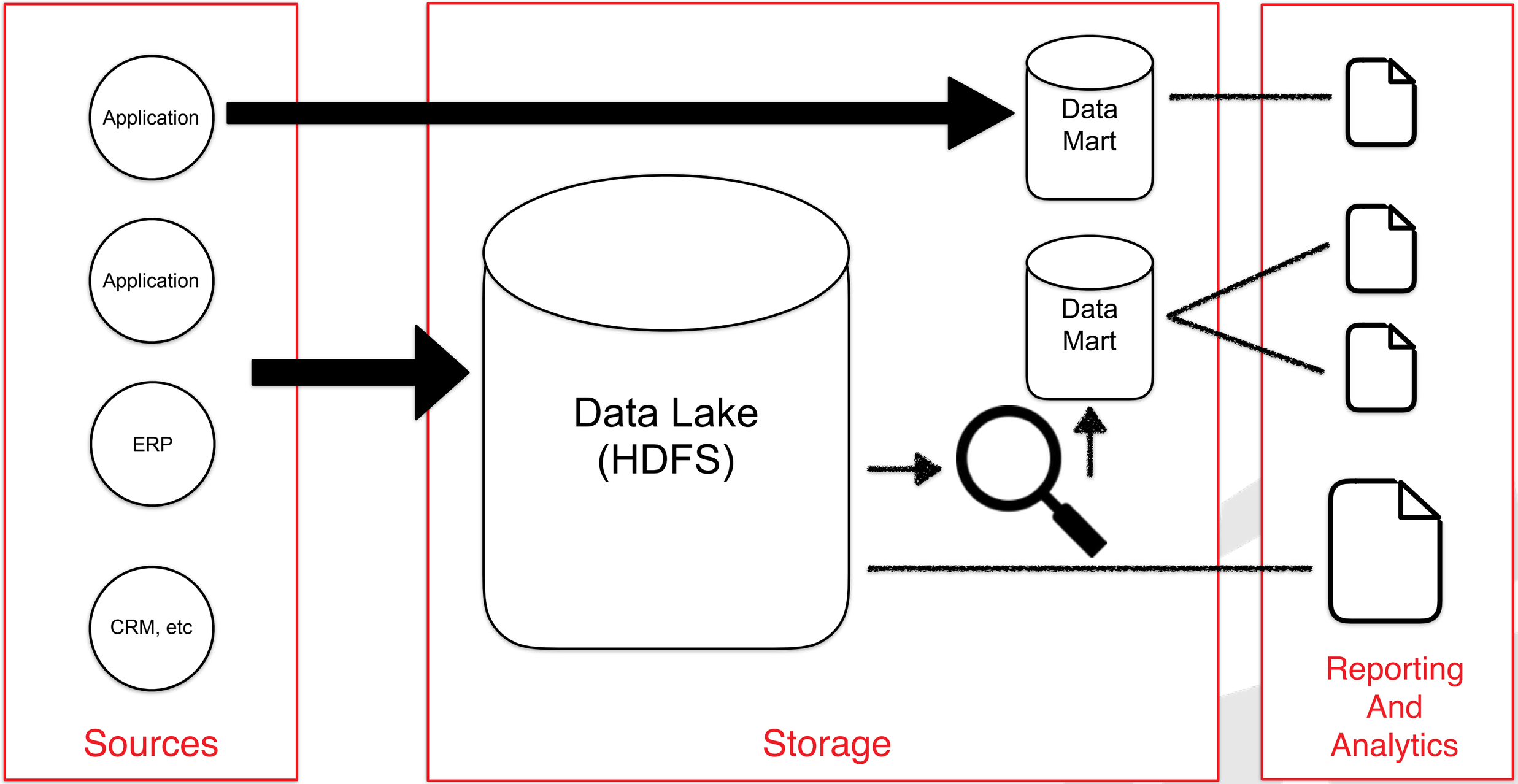


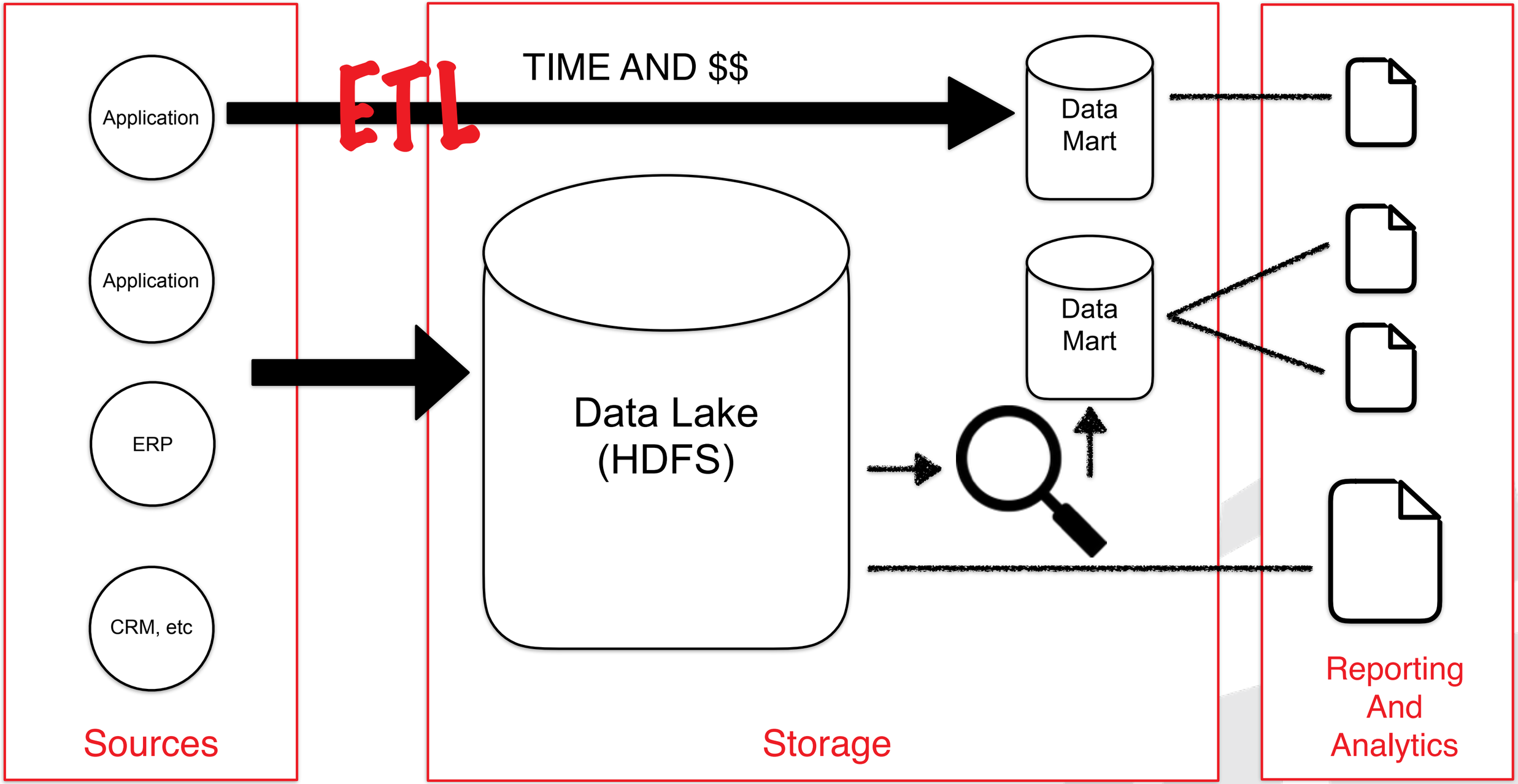


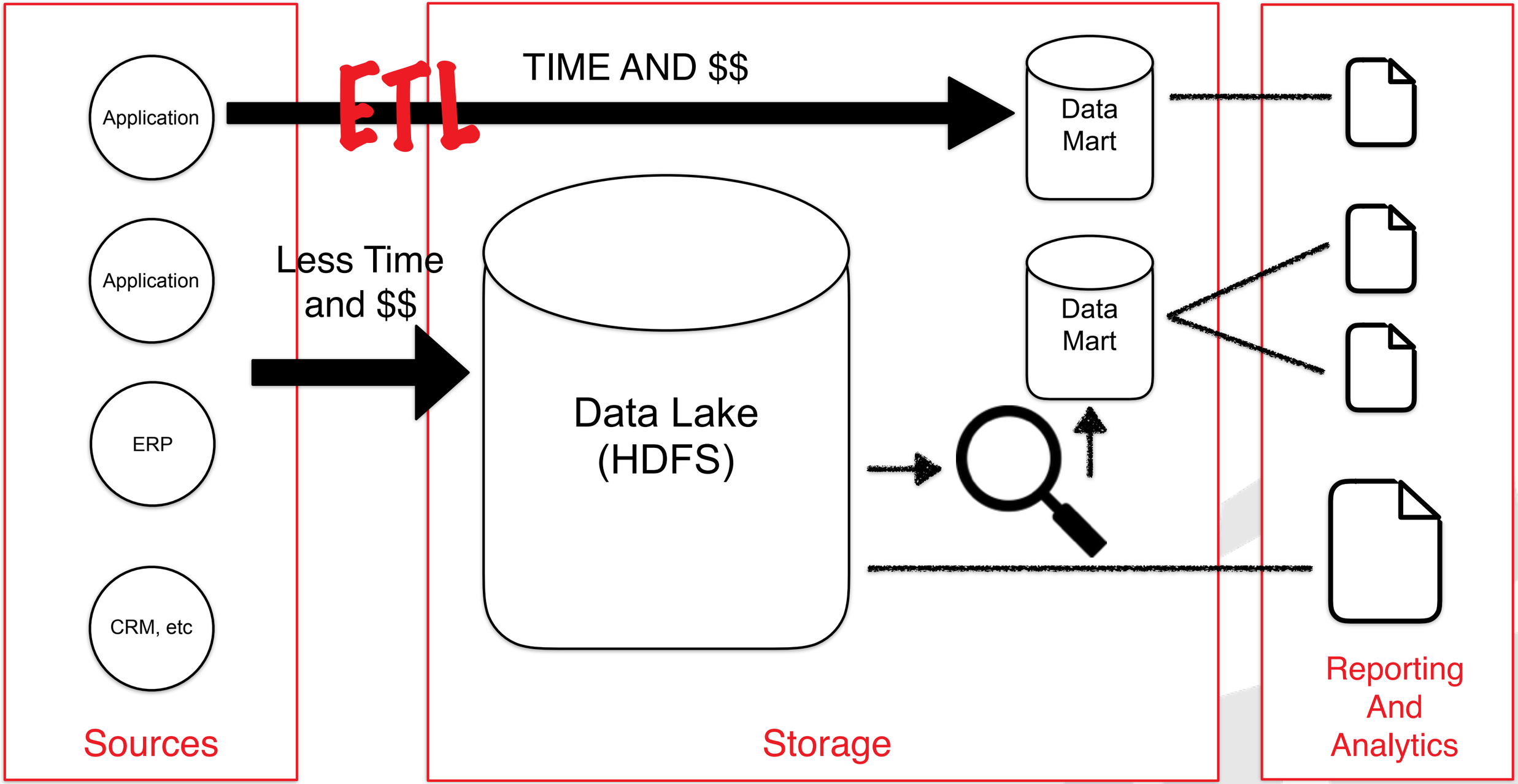
“Analytic” Databases

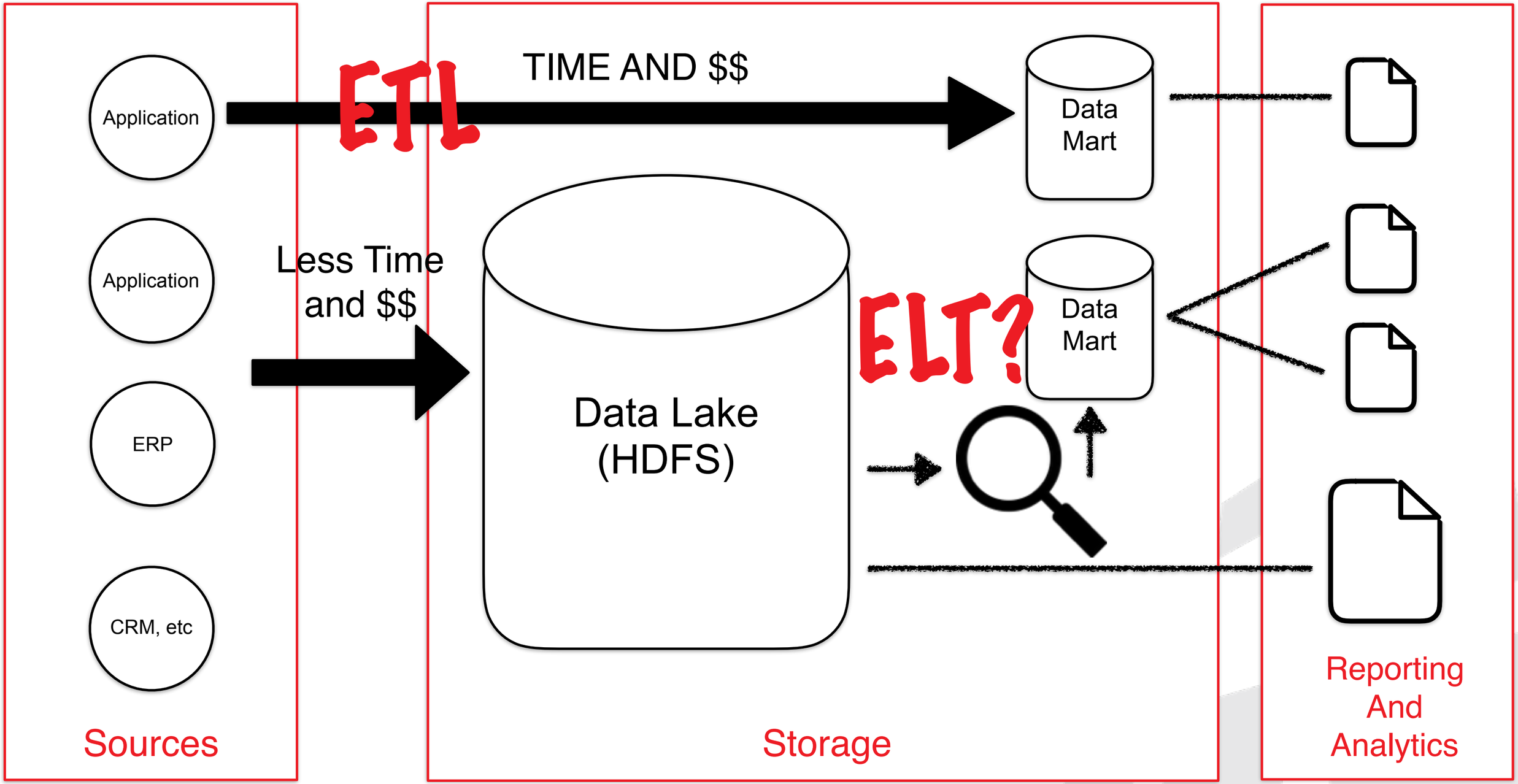
“These aren’t the unicorns you were looking for.”

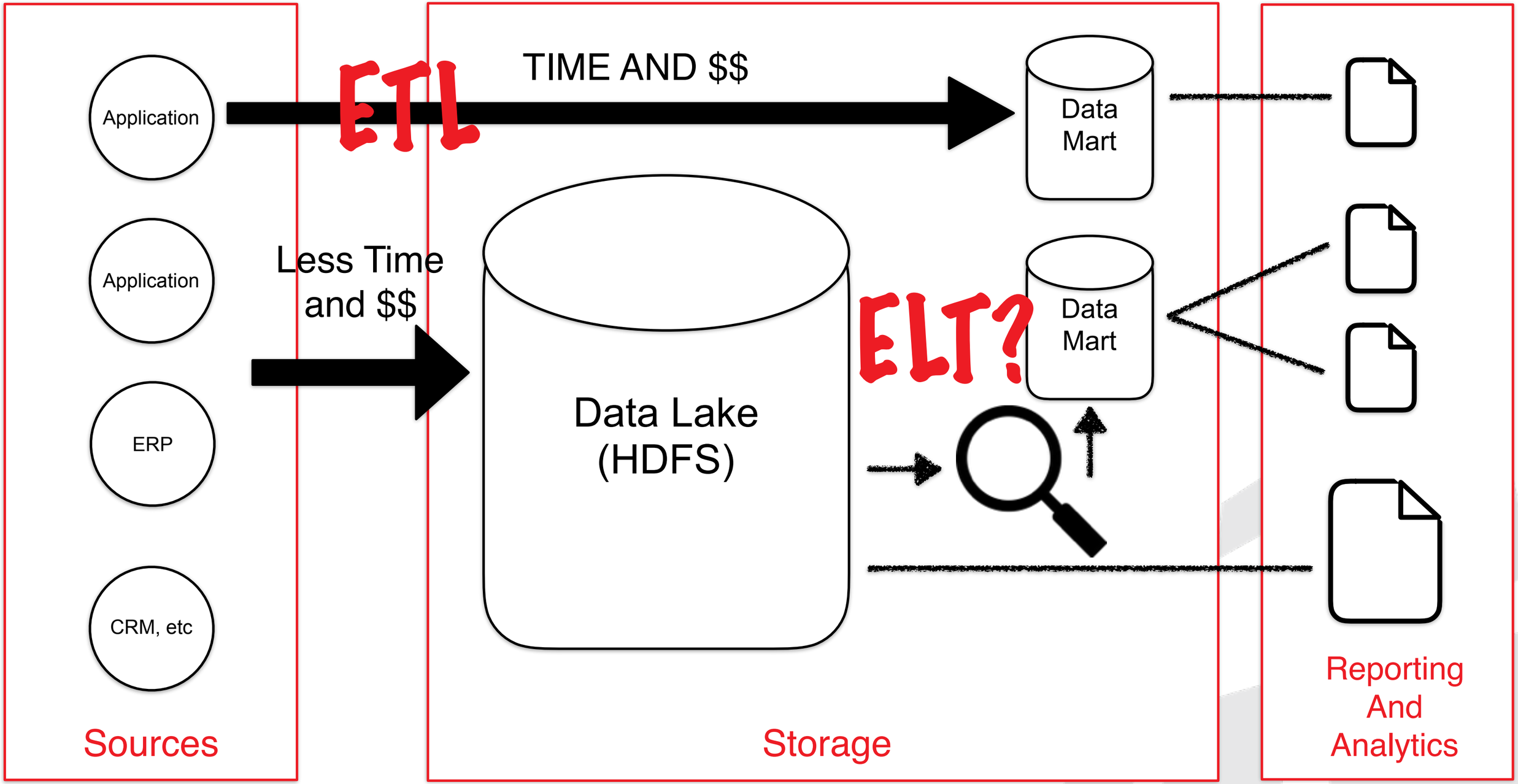






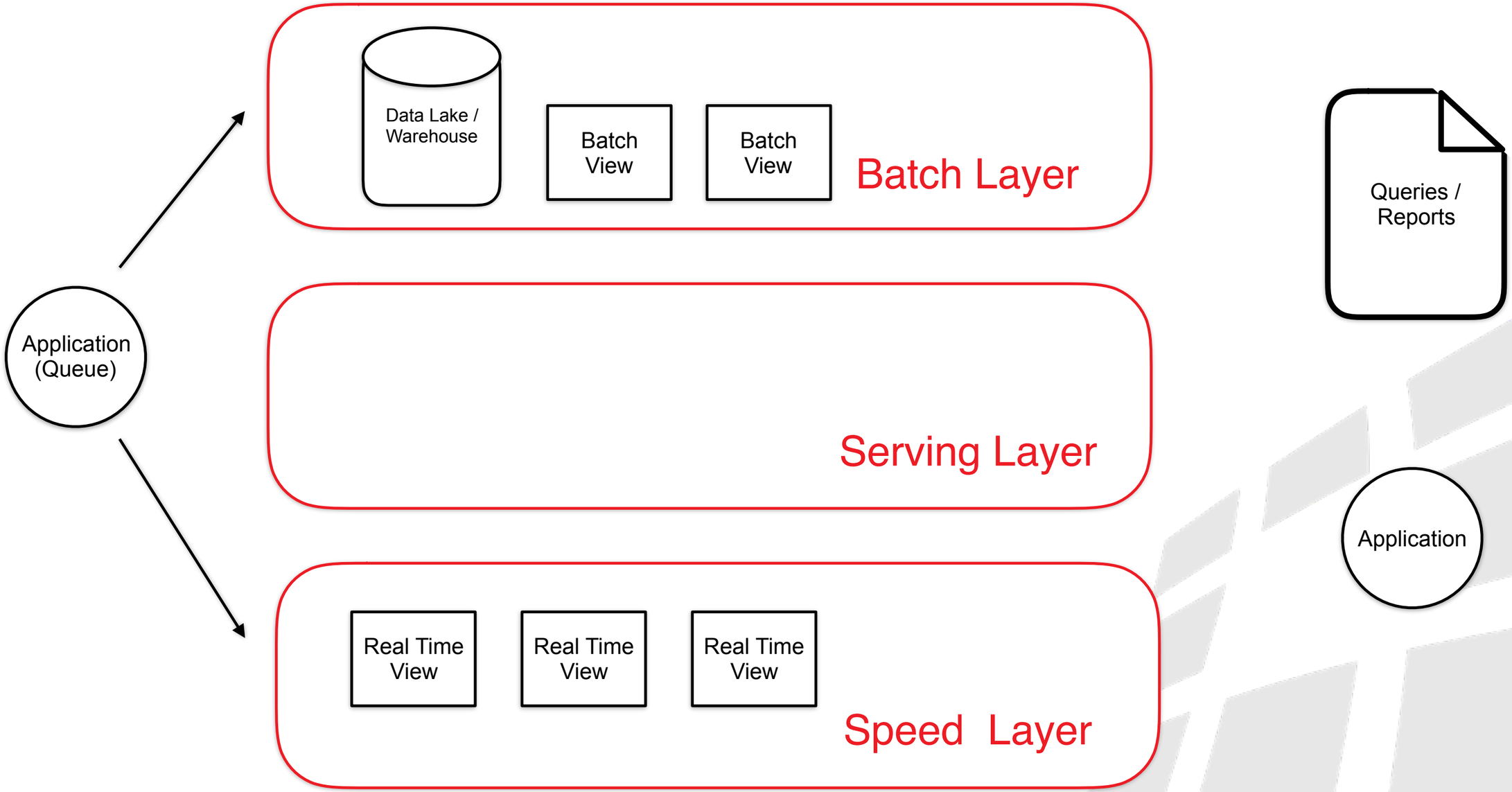




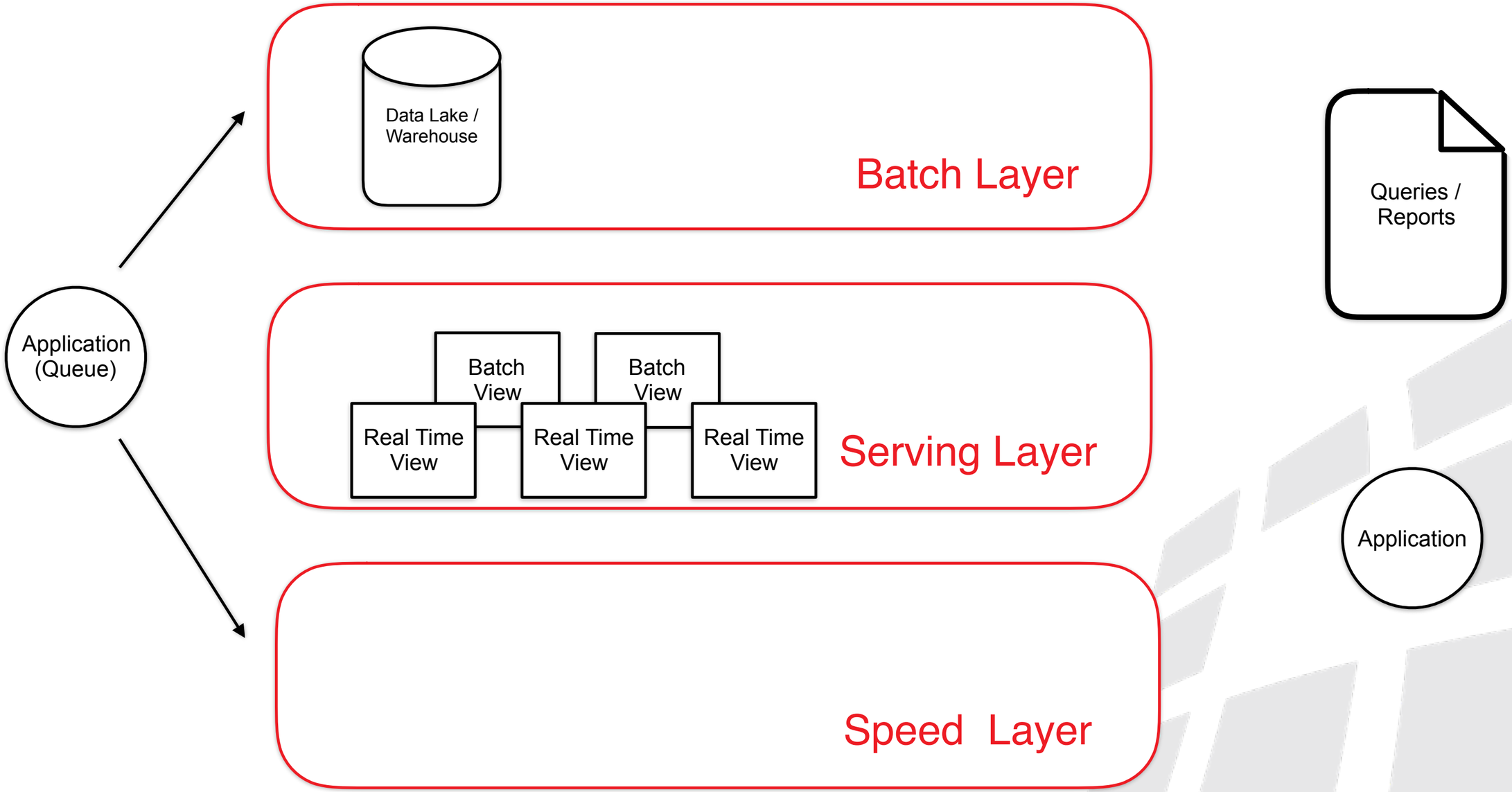


I need fast access to historical data on the fly for predictive modeling with real time data from the stream

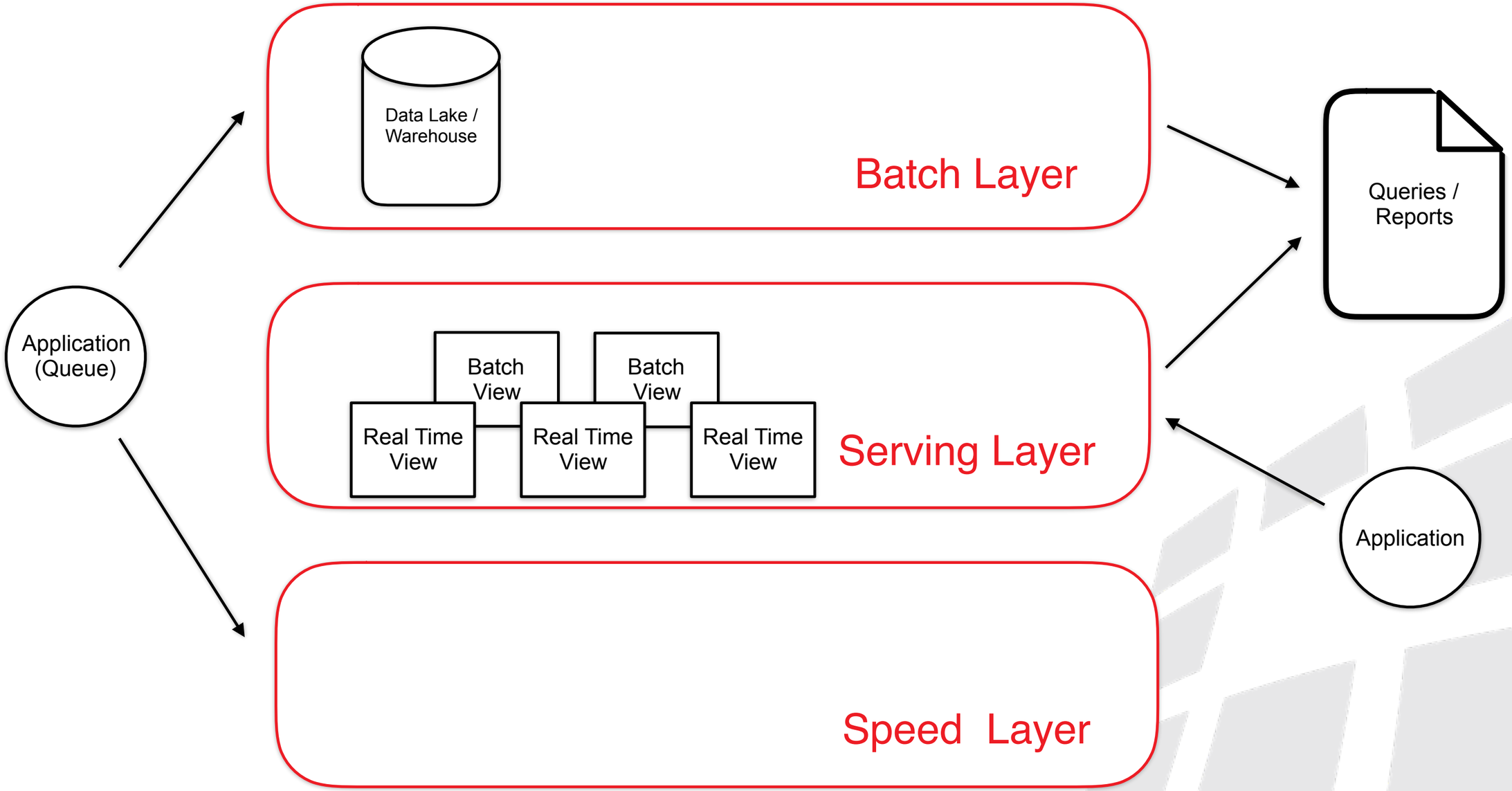
Pushing Data Towards Real Time- Lambda



Pushing Data Towards Real Time- Lambda

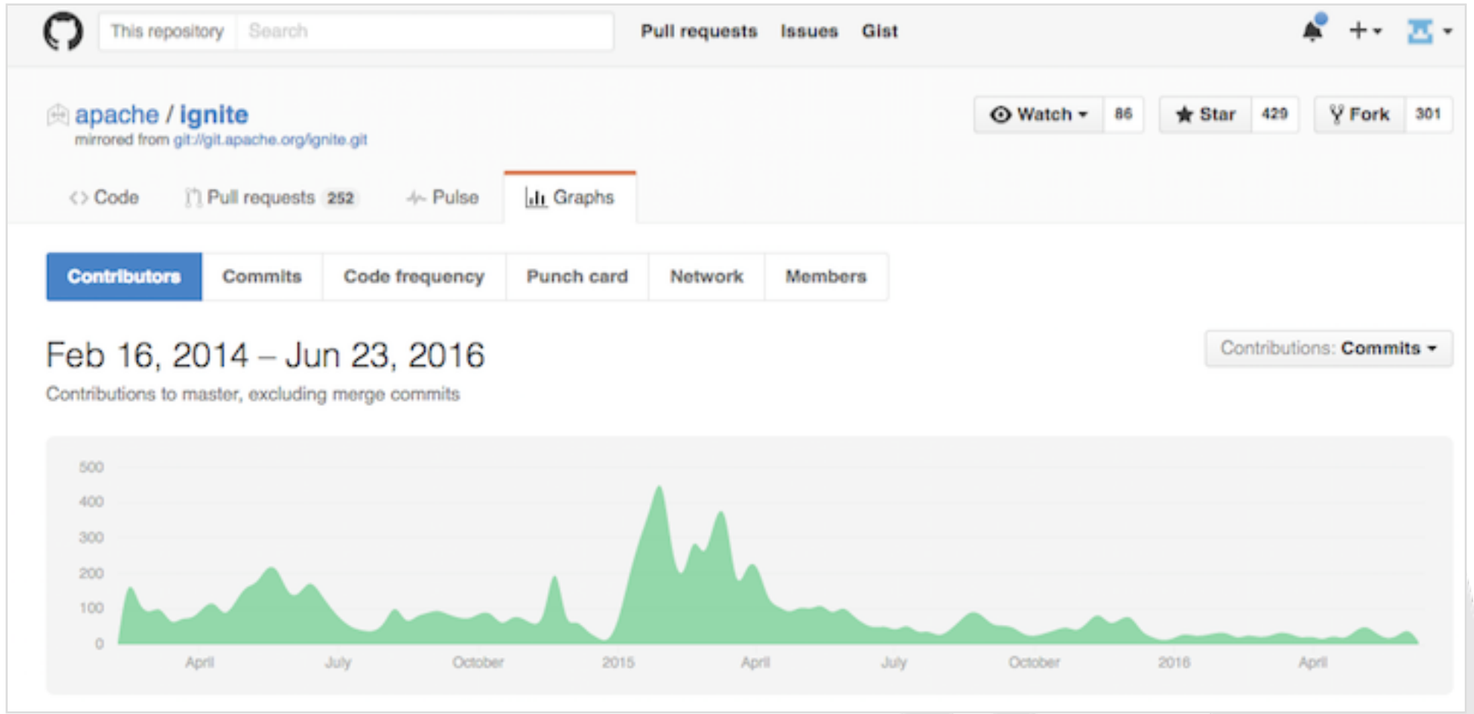


Pushing Data Towards Real Time- Lambda

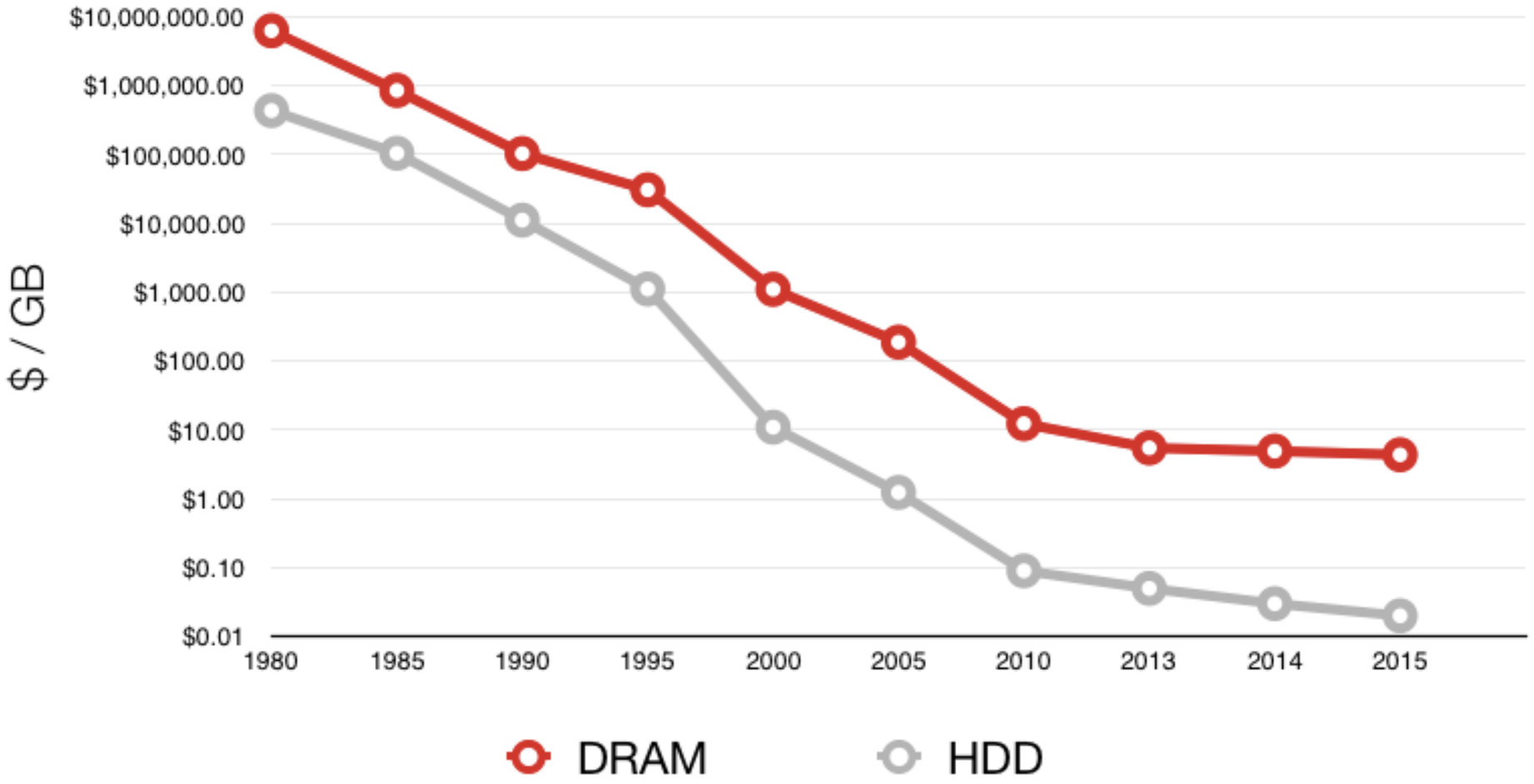


Apache Ignite Project

- **2007:** First version of GridGain
- **Oct. 2014:** GridGain contributes Ignite to ASF
- **Aug. 2015:** Ignite is the second fastest project to | graduate after Spark
- **Today:**
  - 60+ contributors and rapidly growing
  - Huge development momentum - Estimated 192 years of effort since the first commit in February, 2014 [Openhub]
  - Mature codebase: 1M+ lines of code

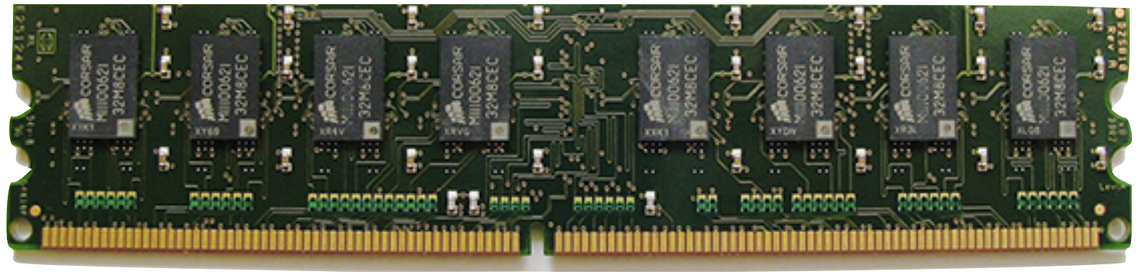


Cost of Memory



“In-memory will have an industry impact comparable to web and cloud.”

“RAM is the new disk, and disk is the new tape.” Gartner



Disk First Architecture

Disk as primary storage, memory for caching

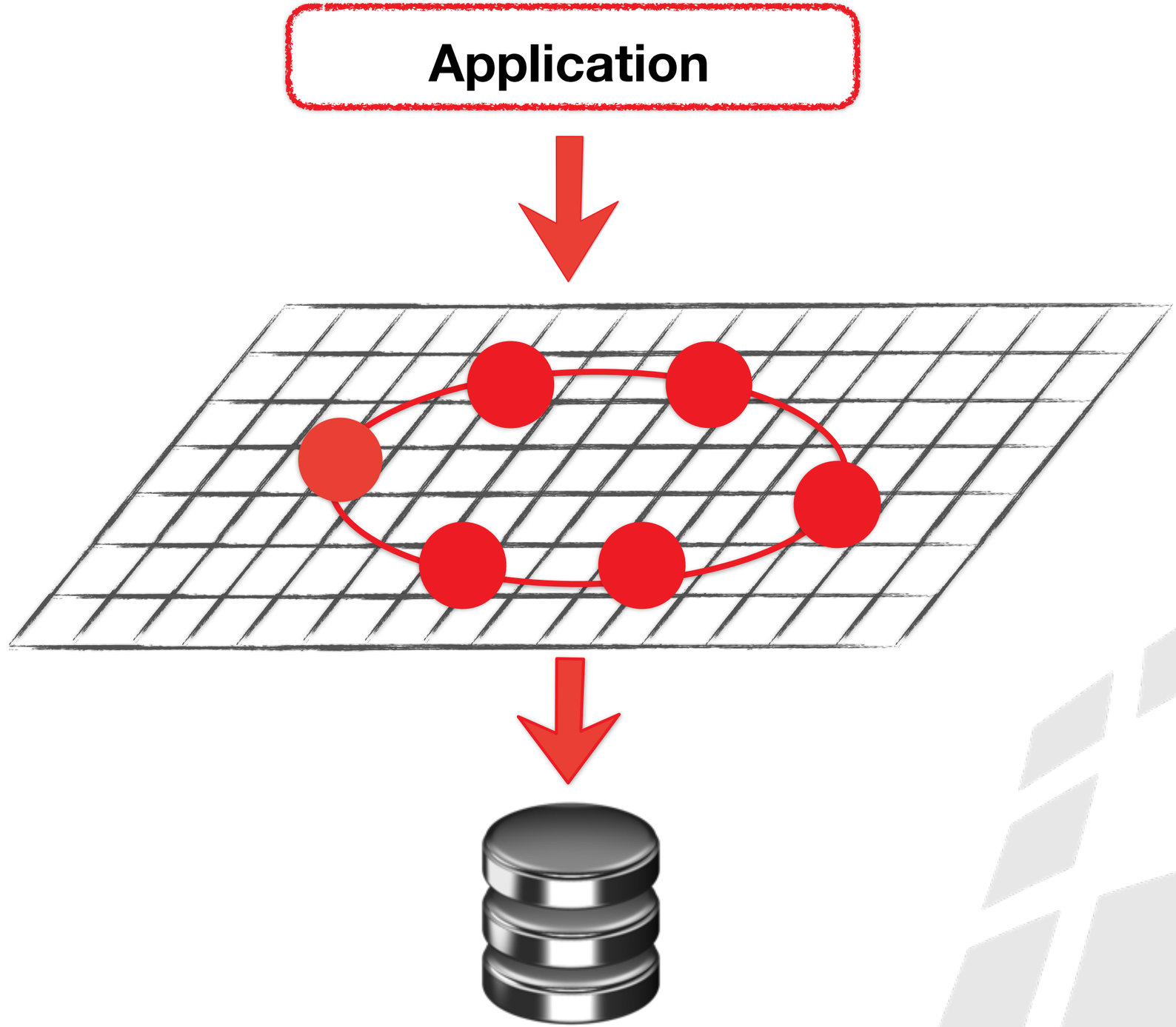
- Access chain: API call <> demarshalling <> OS I/O <> I/O controller <> disk
- Latency: milliseconds

Memory First Architecture

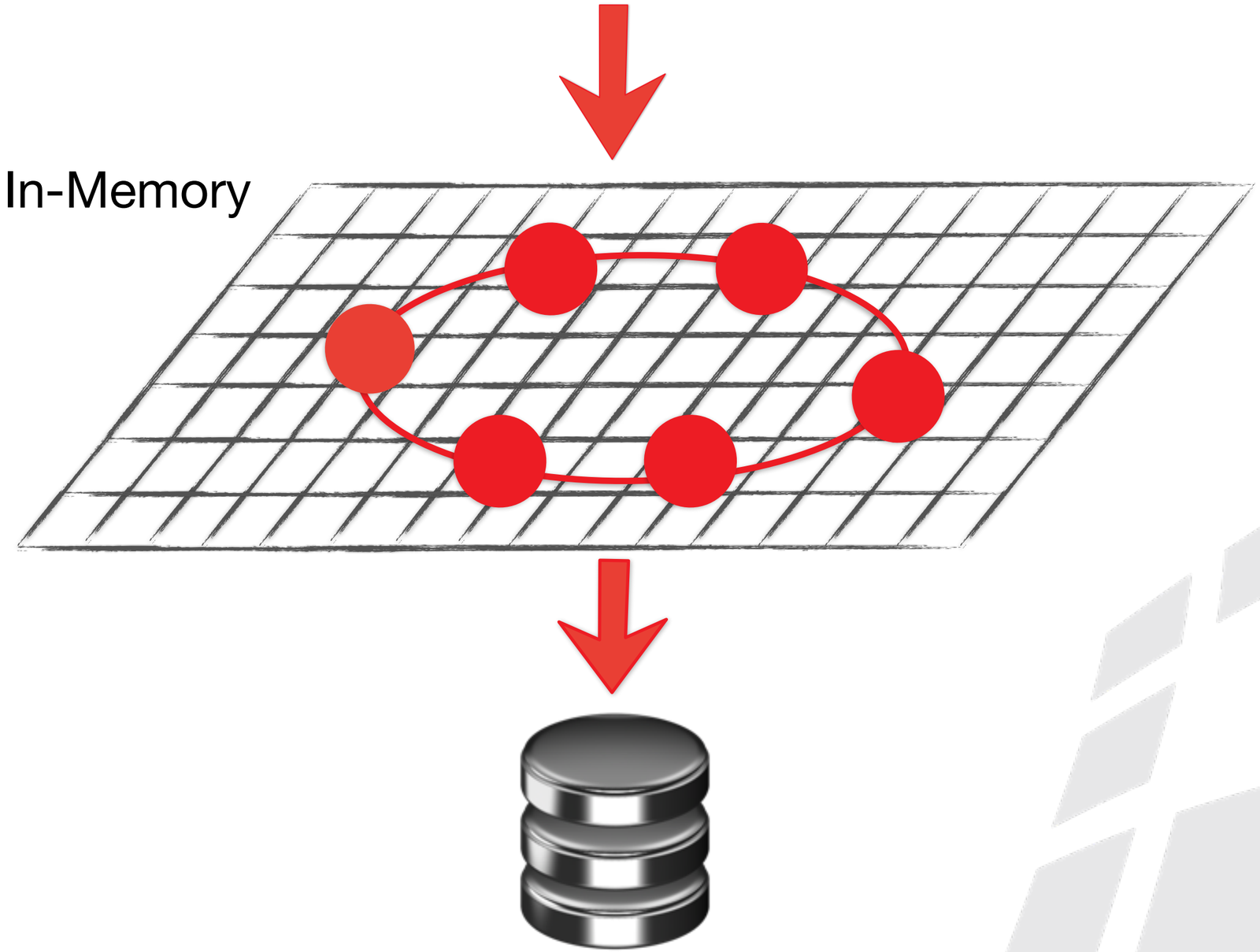
Memory as primary storage, disk as backup

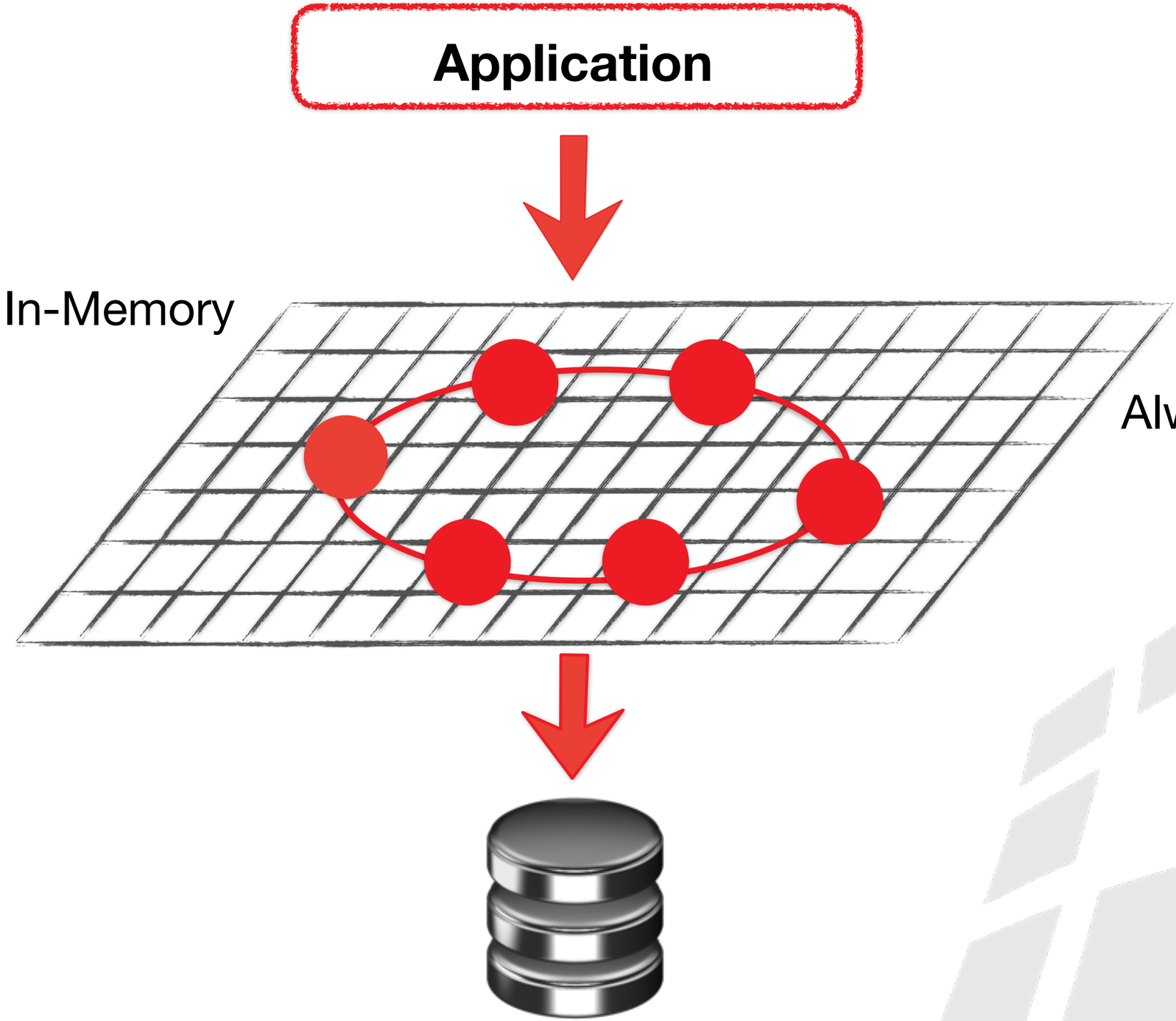
- Access chain: API call <> pointer arithmetic
- Latency: nanoseconds to microseconds

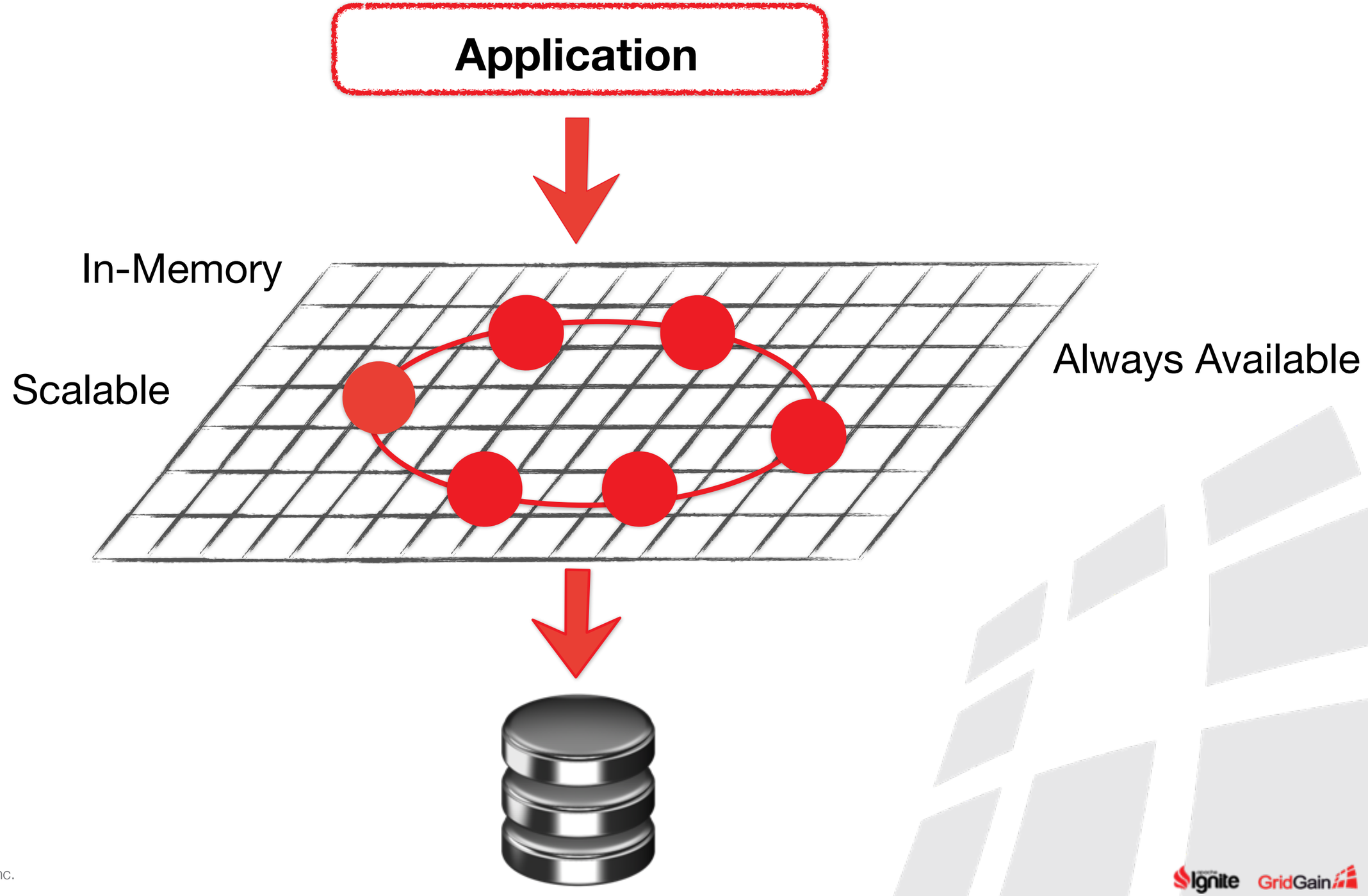
Application

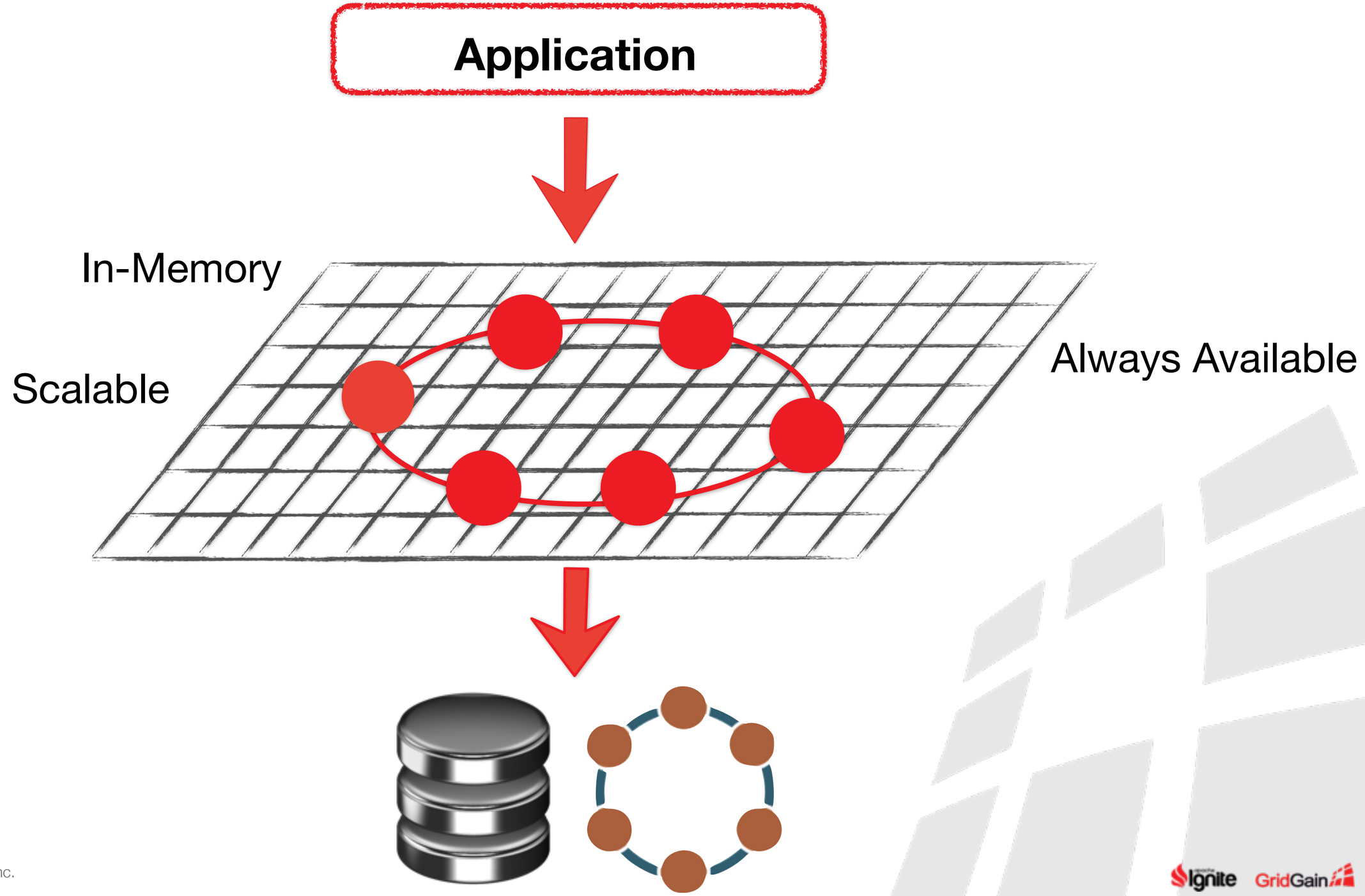


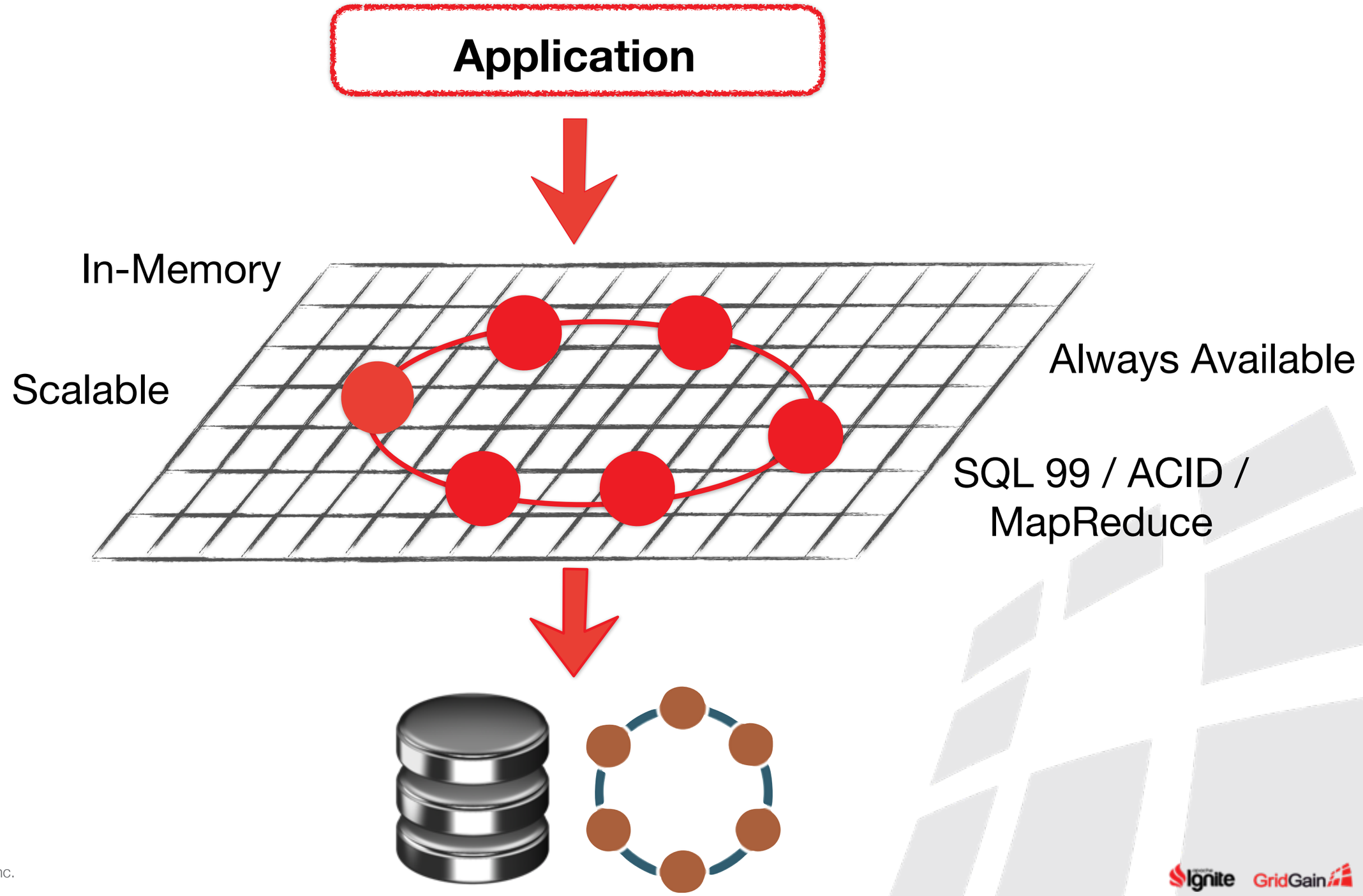
Application



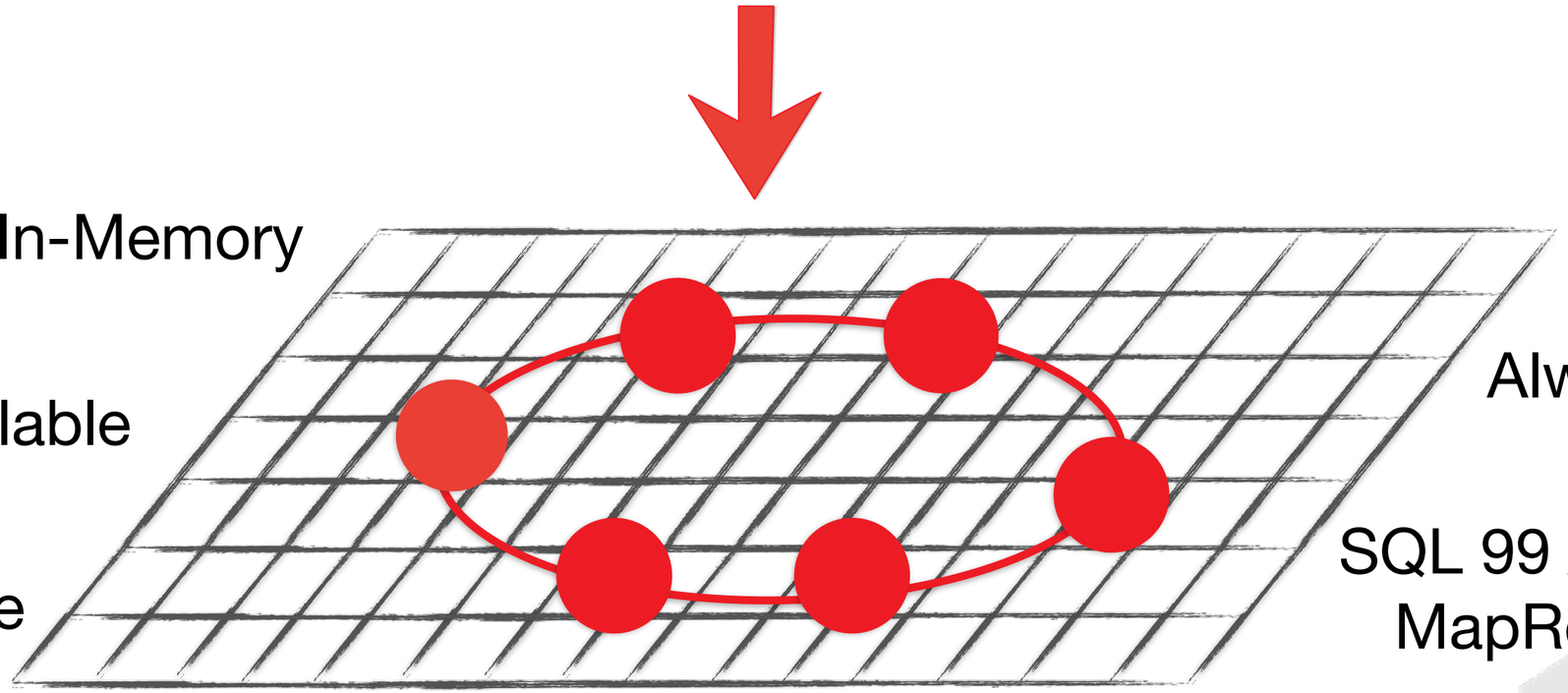








Application



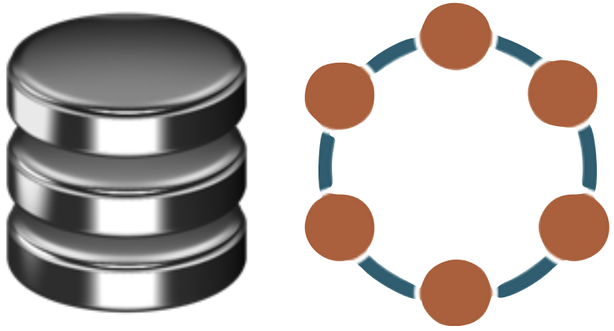
In-Memory

Scalable

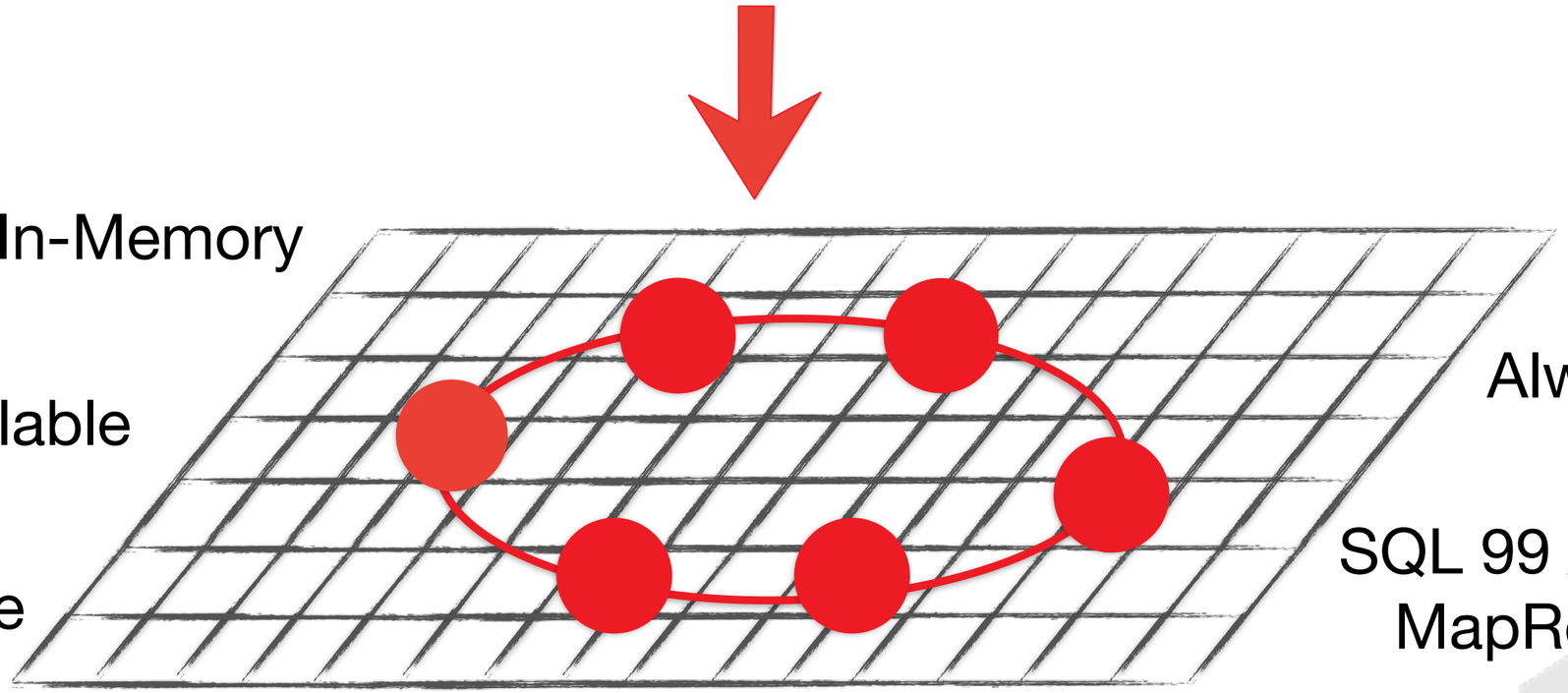
No Rip & Replace

Always Available

SQL 99 / ACID / MapReduce



Application



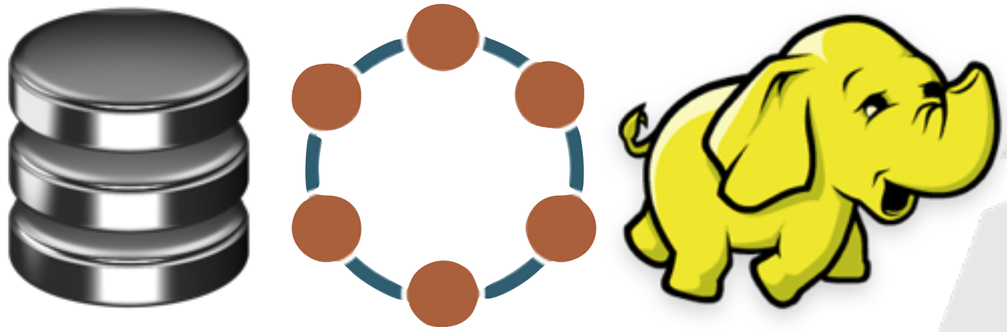
In-Memory

Scalable

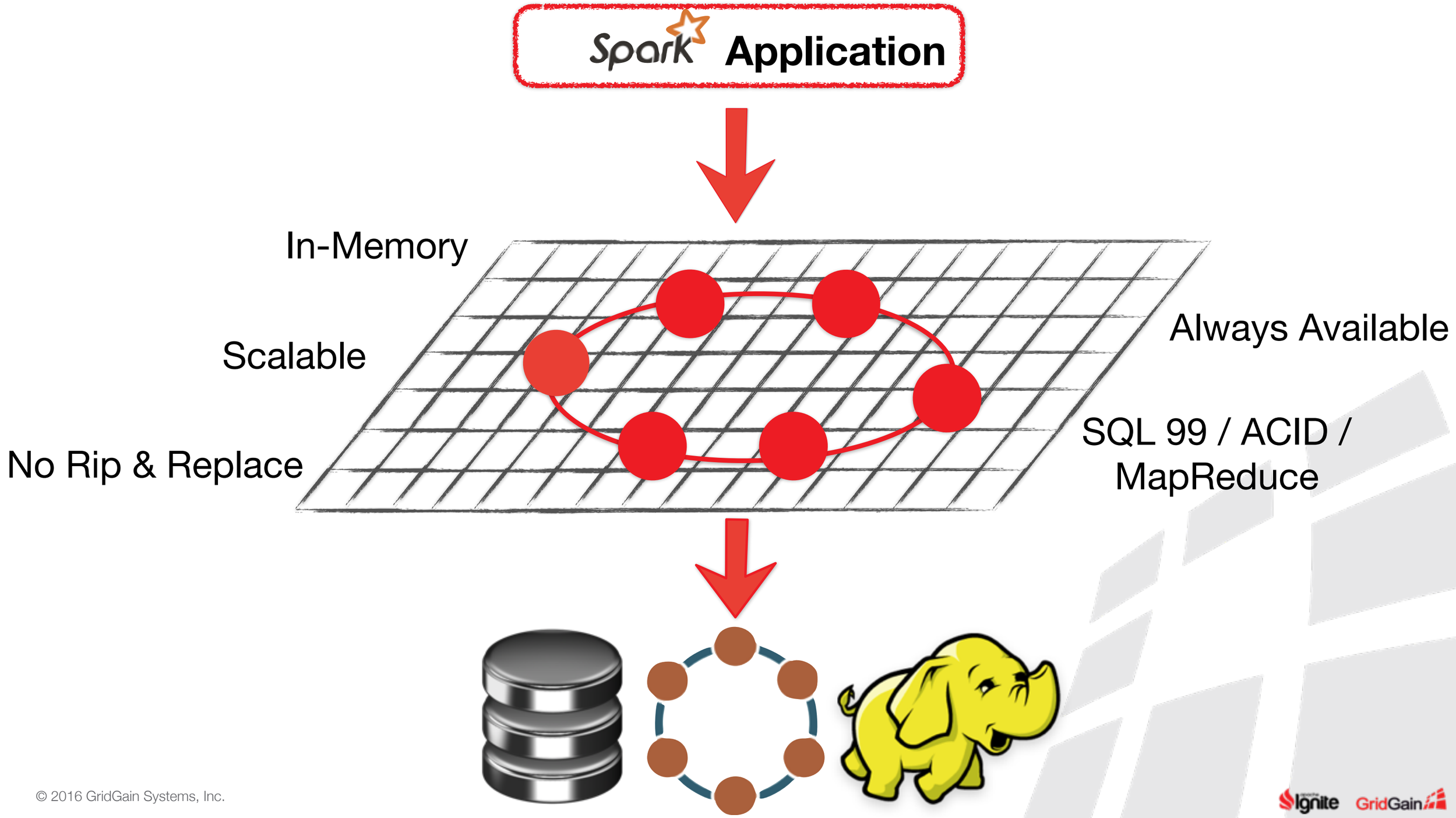
No Rip & Replace

Always Available

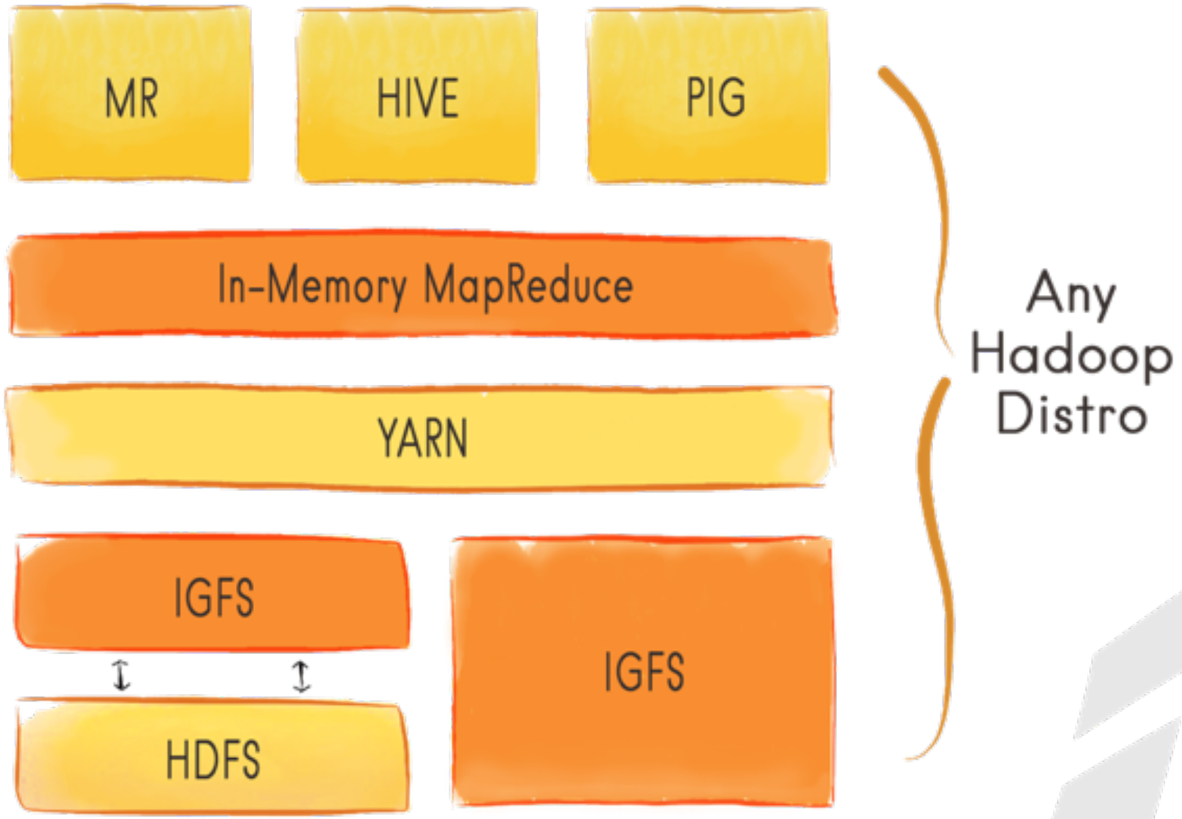
SQL 99 / ACID / MapReduce



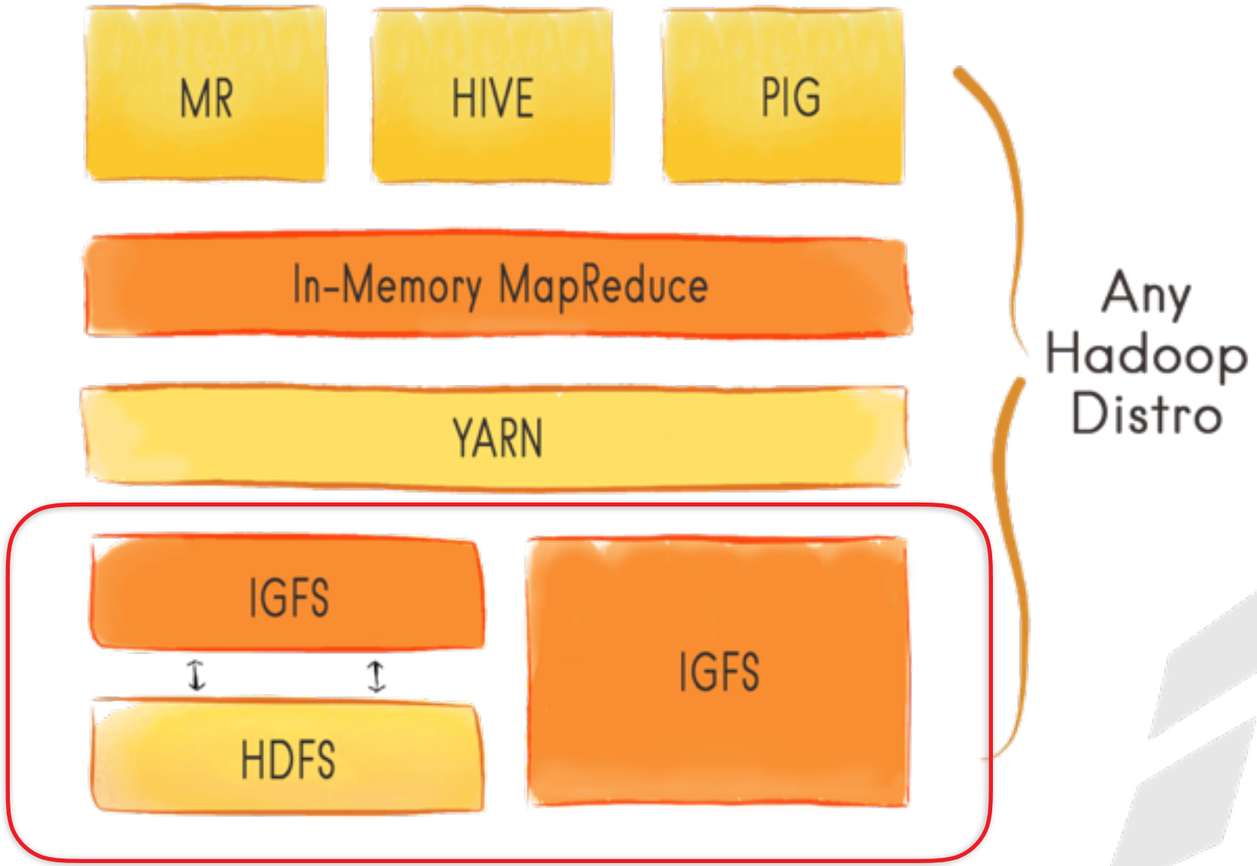
Apache Ignite GridGain



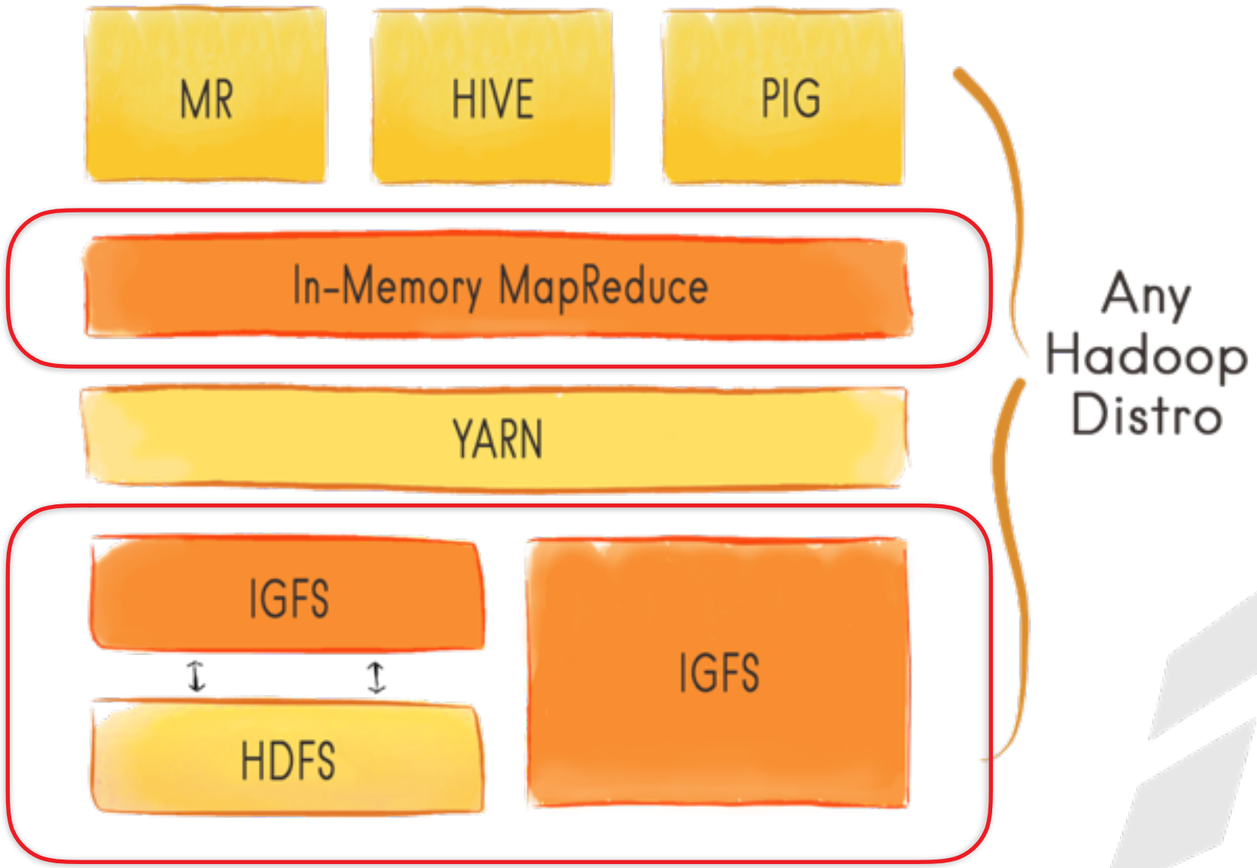
Hadoop Accelerator



Hadoop Accelerator

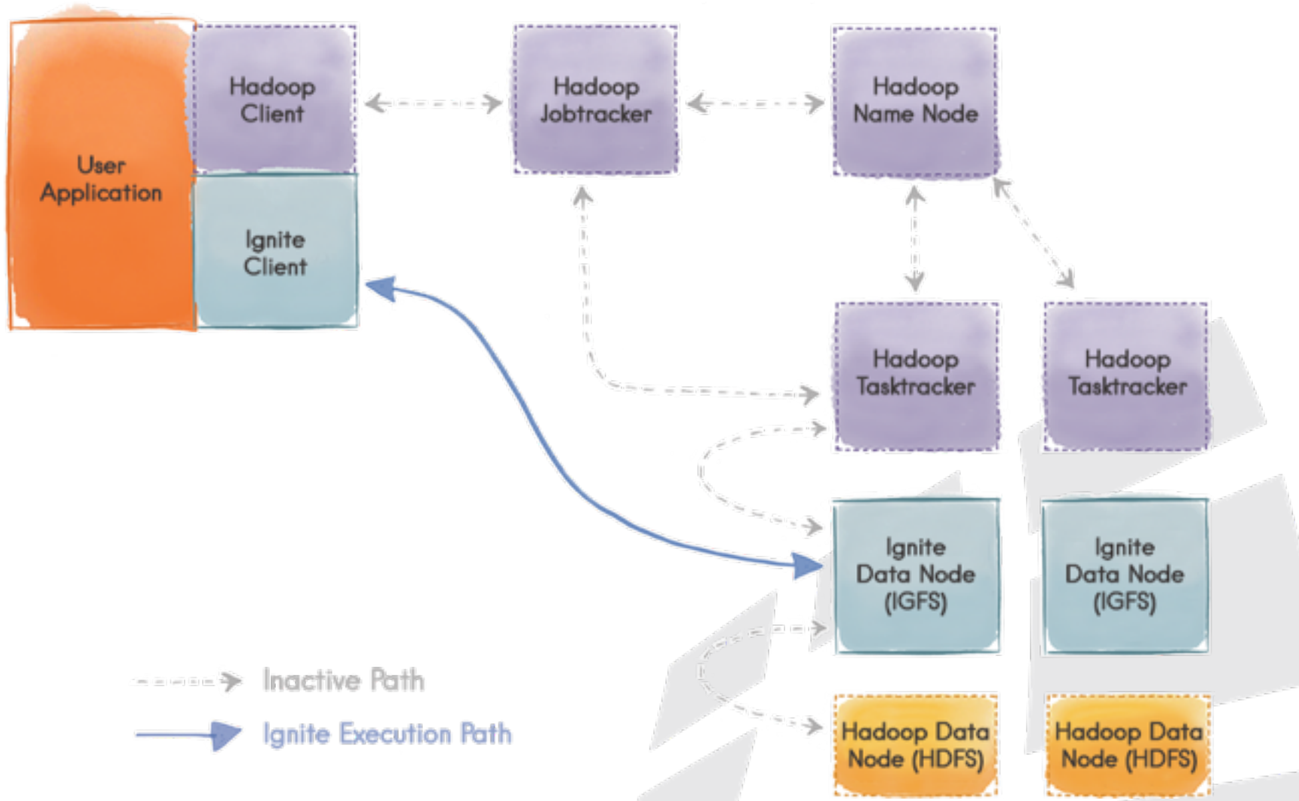


Hadoop Accelerator



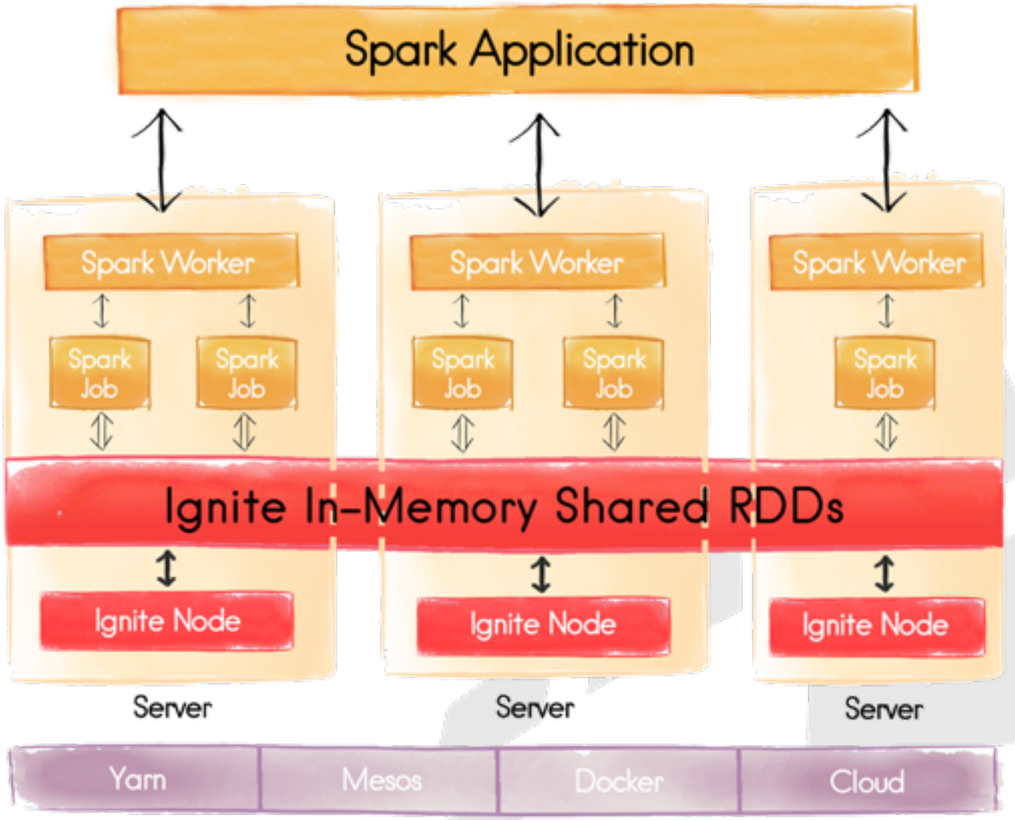
In-Memory Map Reduce

- In-Memory Native Performance
- Zero Code Change
- Use existing MR code
- Use existing Hive queries
- No Name Node
- No Network Noise
- In-Process Data Colocation
- Eager Push Scheduling



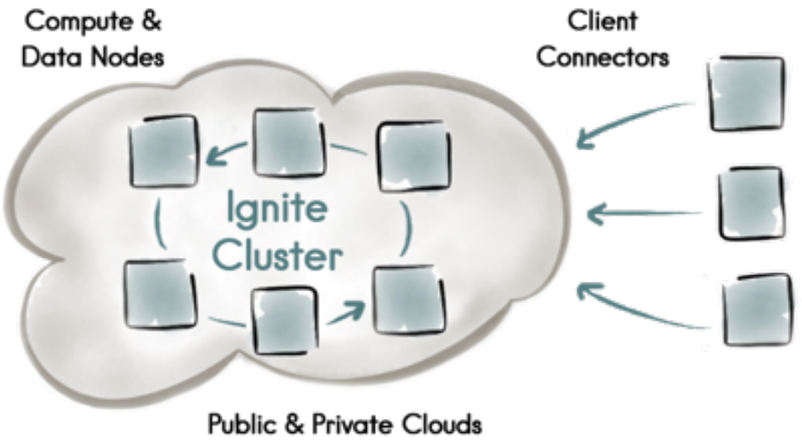
Share RDDs Across Spark Jobs

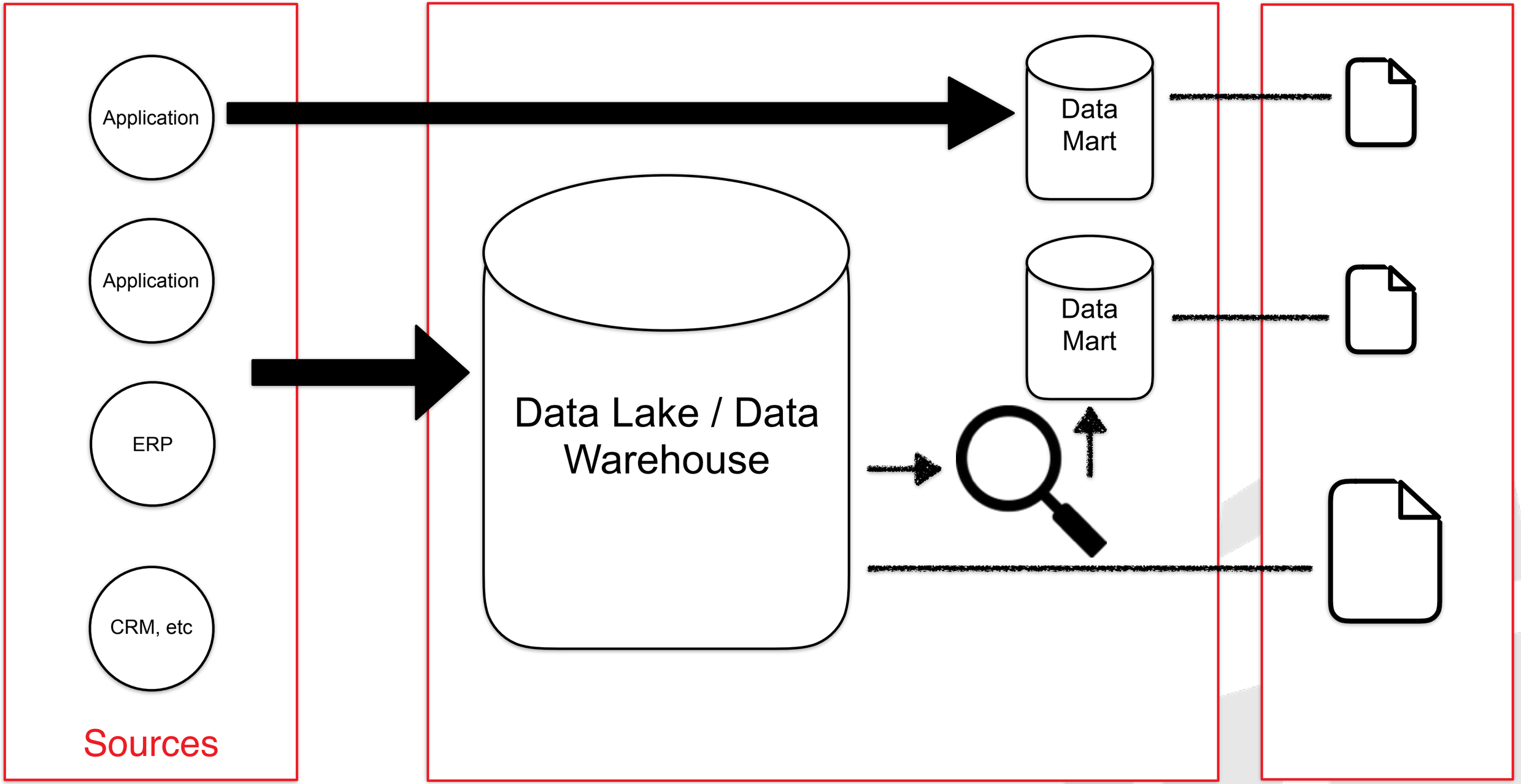
- "Ignite" RDD
  - Share RDD across jobs on the host
  - Share RDD across jobs in the application
  - Share RDD globally
- Faster SQL
  - In-Memory Indexes
  - SQL on top of Shared RDD

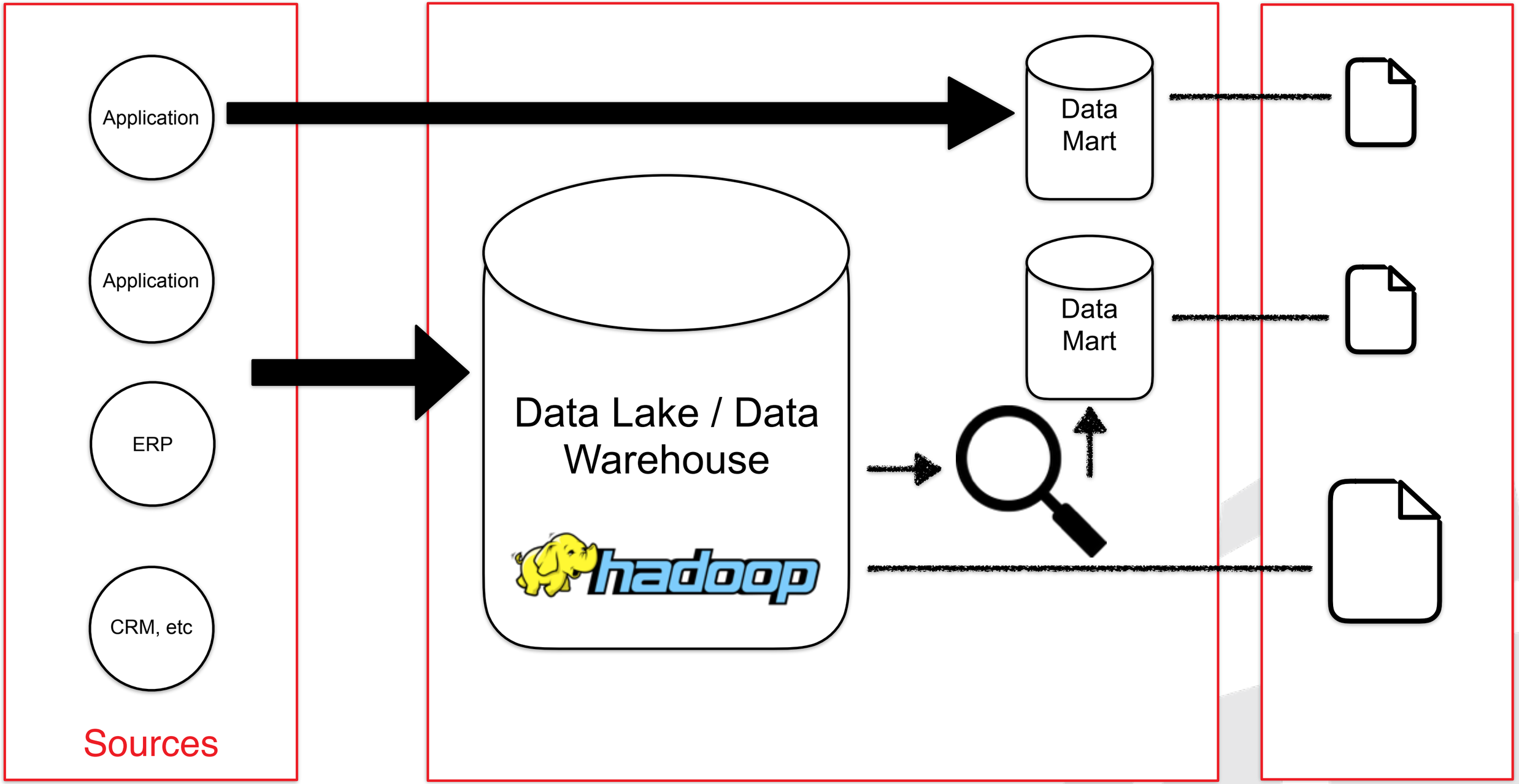


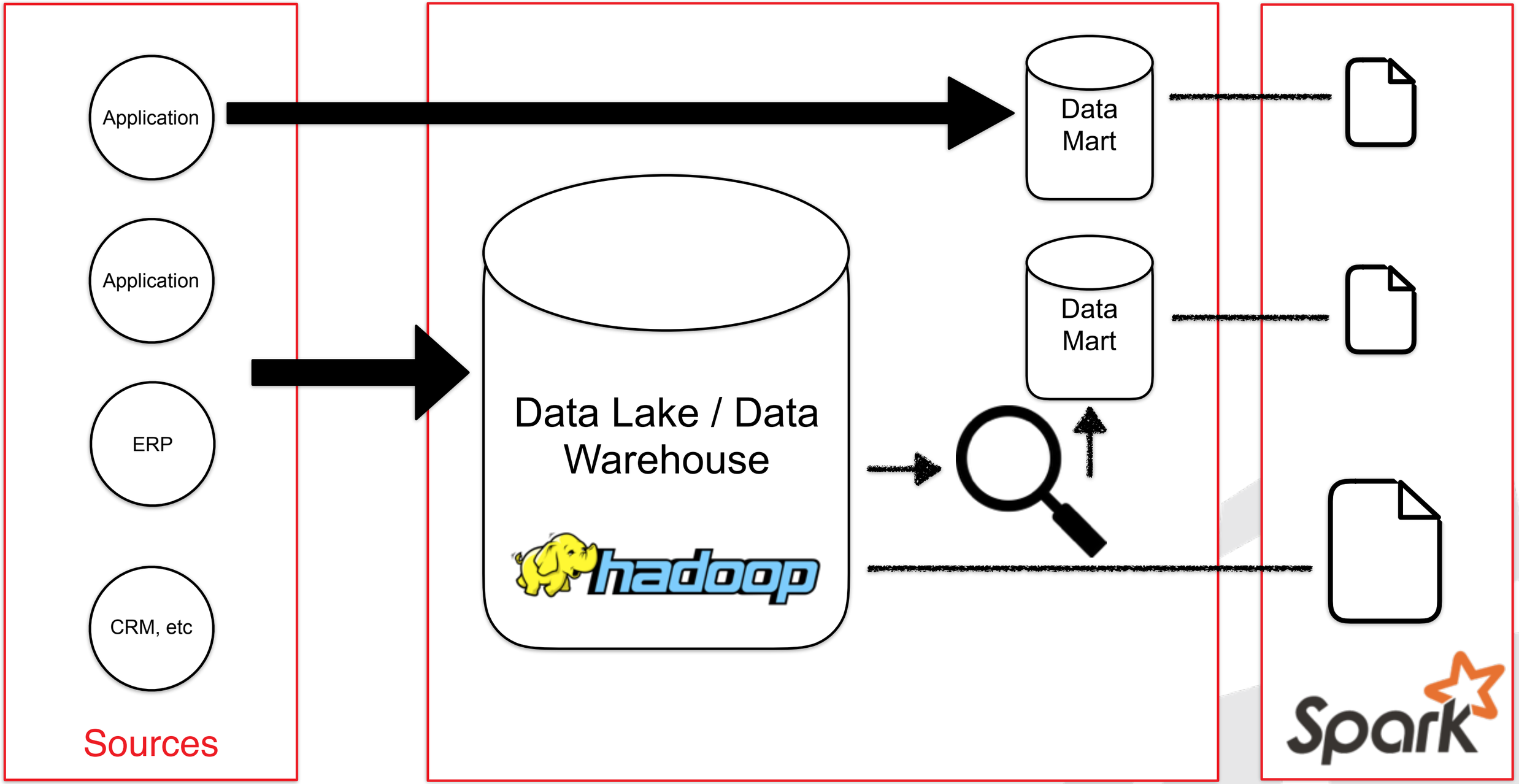
Clustering & Deployment

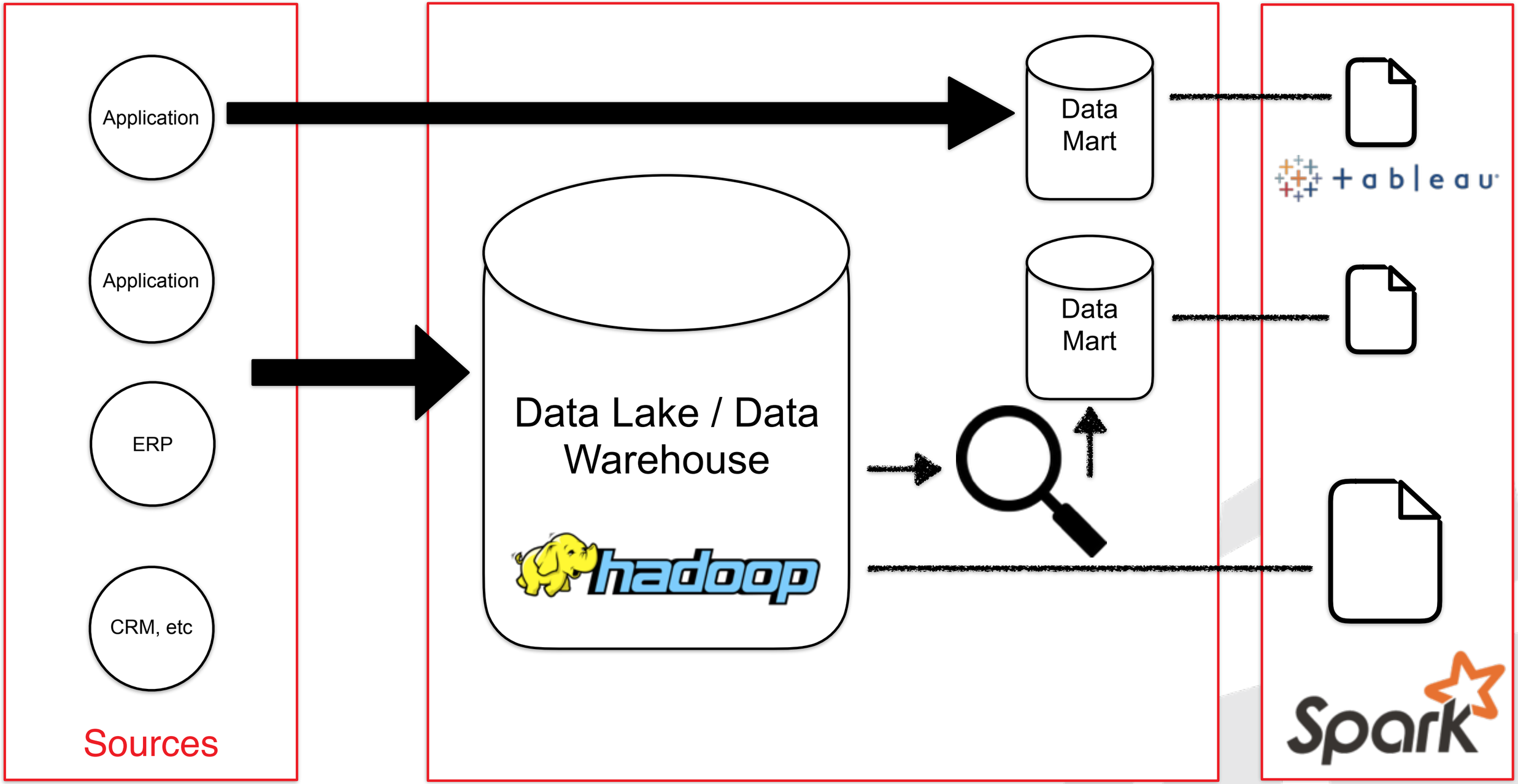
- Any Environment
  - Simple Configuration
  - AWS/Azure/GCE/OpenStack
  - Integration with JClouds
  - Hybrid Cloud
  - Local Laptop
- Zero-Deployment
  - Auto-Deploy Code
- Docker Support
  - Automatically Build and Deploy
- Pluggable Design
- Integration with Yarn and Mesos

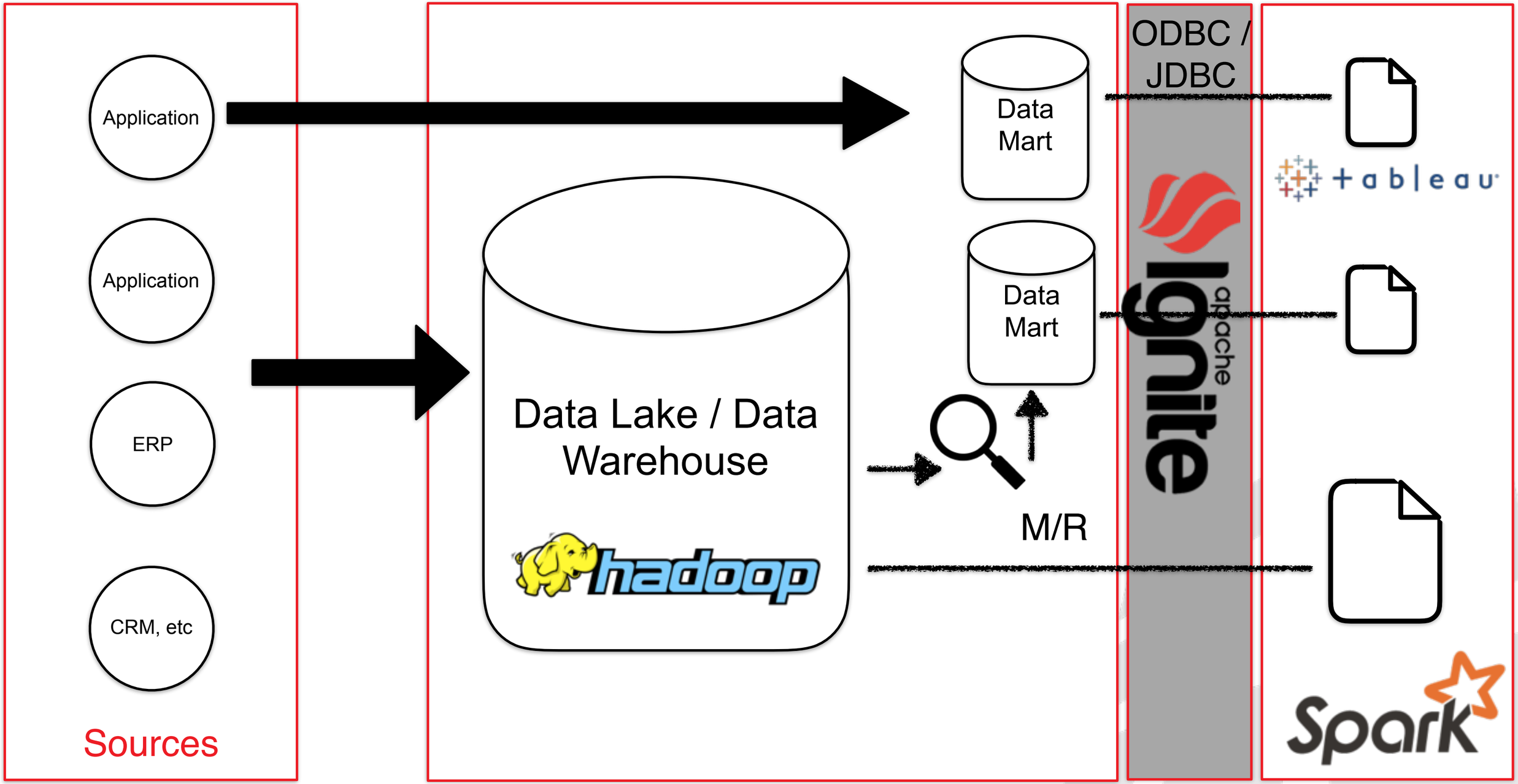




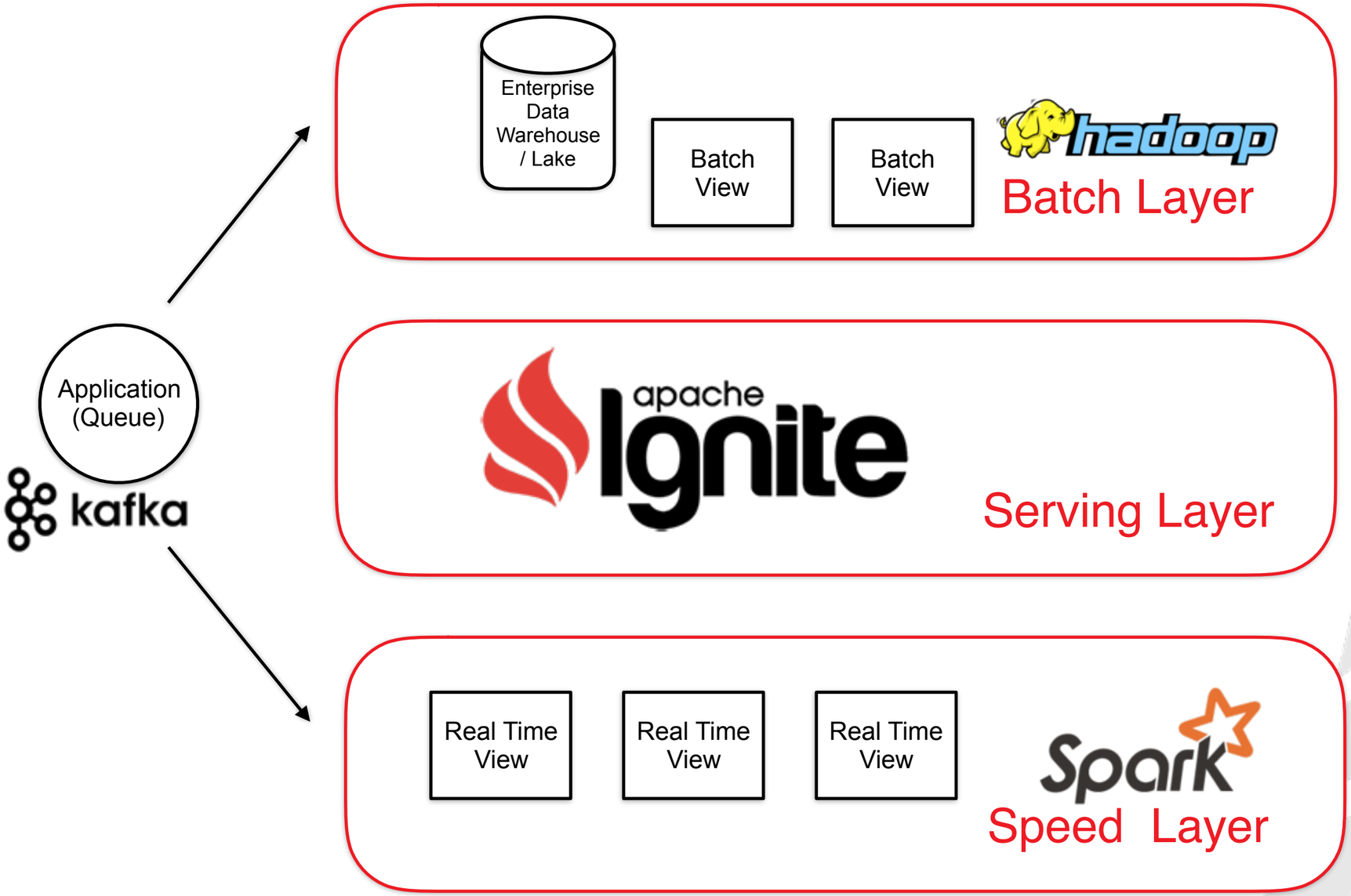




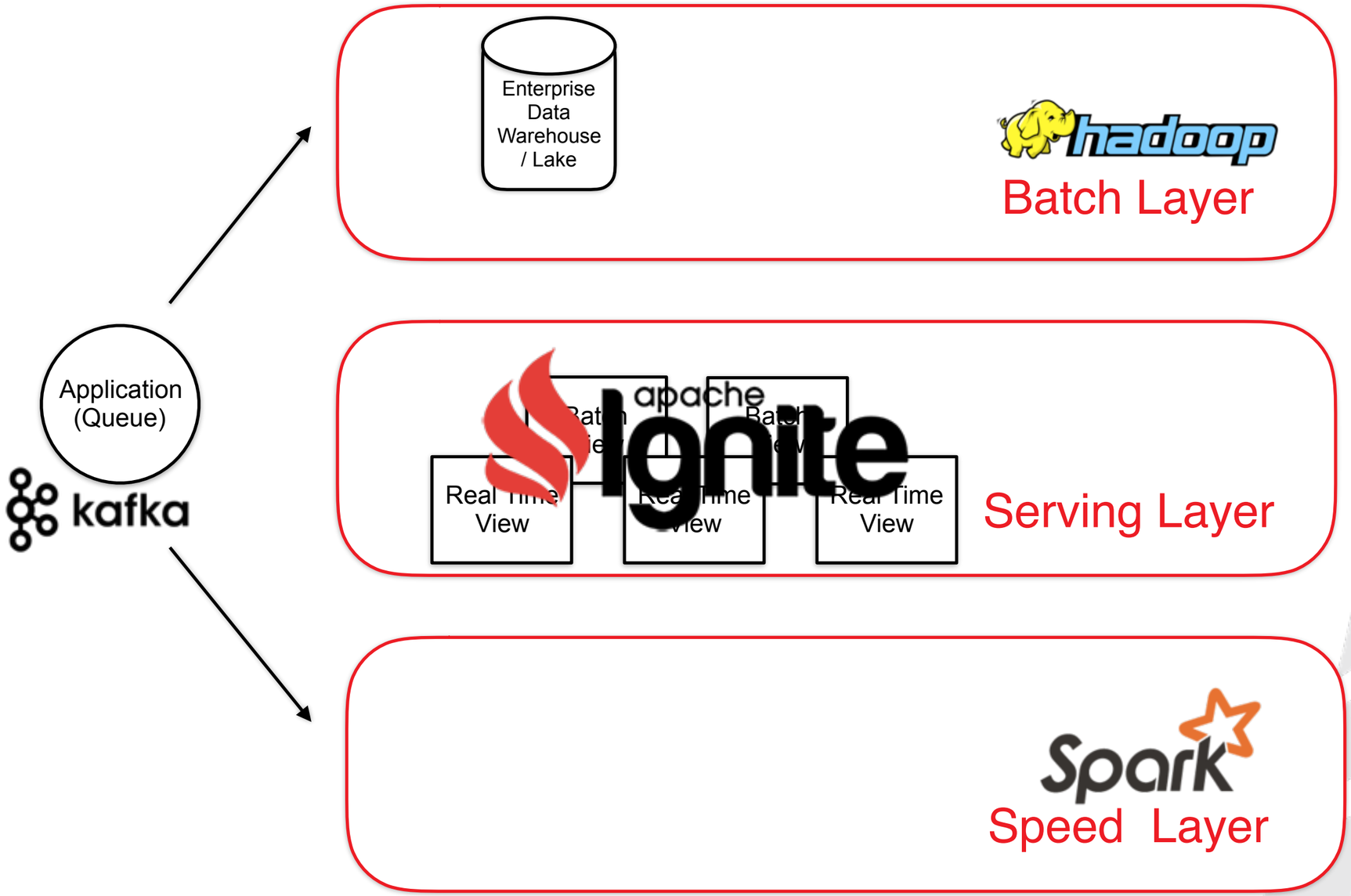




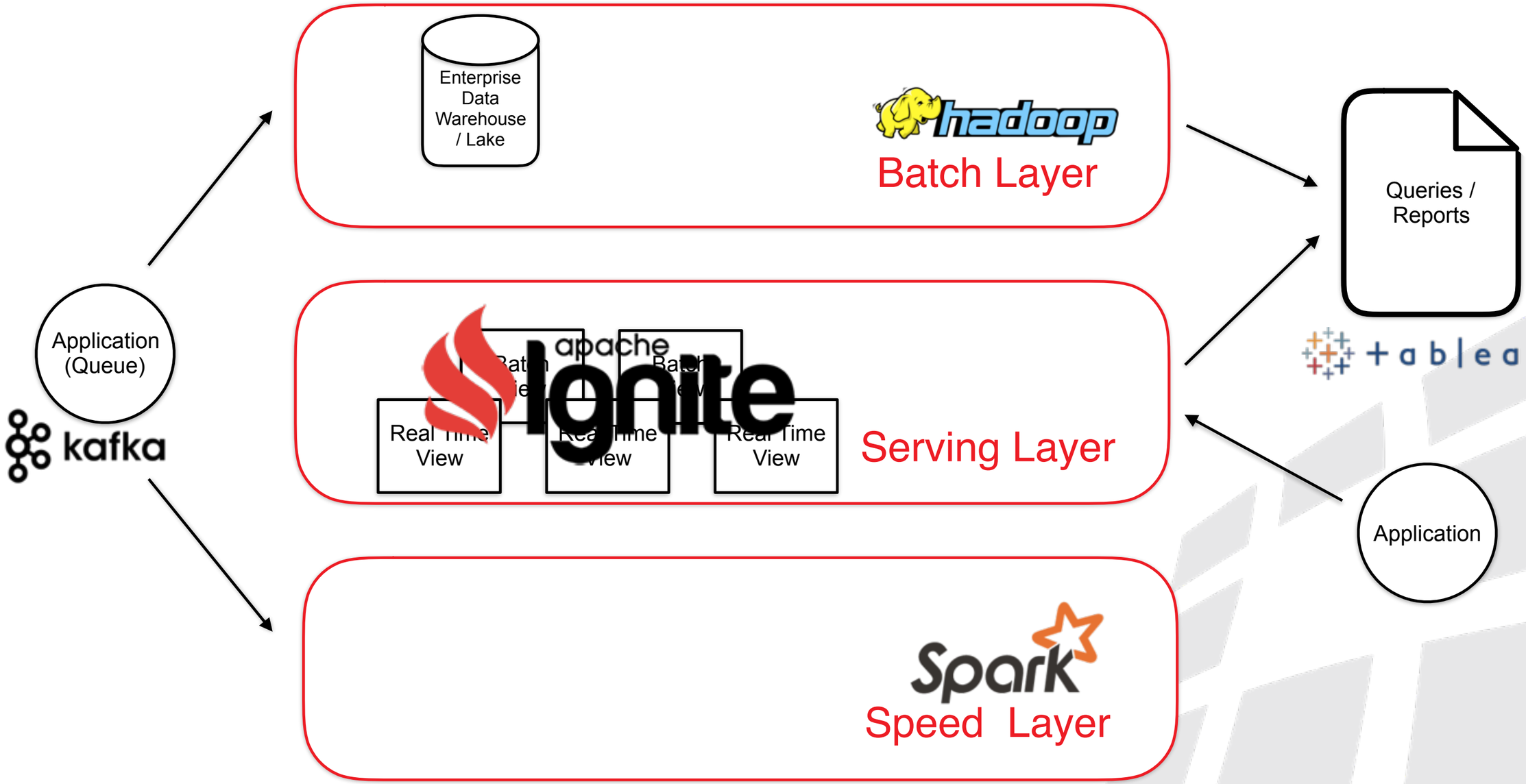
Lambda Revisited



Lambda Revisited



Lambda Revisited



Better Together: Fast Data with Apache Spark™ and Apache Ignite™

Wednesday, September 7, 2016

Nikita Ivanov

Apache Spark™ and Apache Ignite™ are two powerful solutions for high-performance Big Data and Fast Data. Using Spark and Ignite together is an easy way to boost performance by orders of magnitude for your next generation real-time applications. With Spark plus Ignite, you can share state across Spark jobs, applications, and workers and your Spark queries will also run much faster.

Read More →

ANY QUESTIONS?

Thank you for joining us. Follow the conversation.

www.gridgain.com

@gridgain
#gridgain