GridGain vs Hadoop
Why Elephants Can't Fly

GridGain System
1065 East Hillsdale Boulevard
Suite 230
Foster City, CA 94404
www.gridgain.com

GridGain In A Glance

GridGain is Java based open source middleware for transactional **real time big data** processing that scales up from one server to thousands of machines.

Unlike complex, decade-old Hadoop MapReduce systems which use stale data for batch offline analytics, our platform allows companies to harness live data for smarter, faster real time processing.

GridGain History

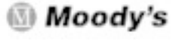
- GridGain Systems founded in 2005
- VC funded
- Headquarter in Foster City, California, USA
- 12 product releases:
  - GridGain **1.x**, Jul 2007
  - GridGain **2.x**, Feb 2008
  - GridGain **3.0**, Aug 2010
- Current release is GridGain **3.6**

GridGain Facts

GridGain starts every **10 seconds** around the globe

Over **8,000,000 starts worldwide**
1000 unique IP/month
400 active projects/month
4000 forum views/month

GridGain Users



GridGain Partners

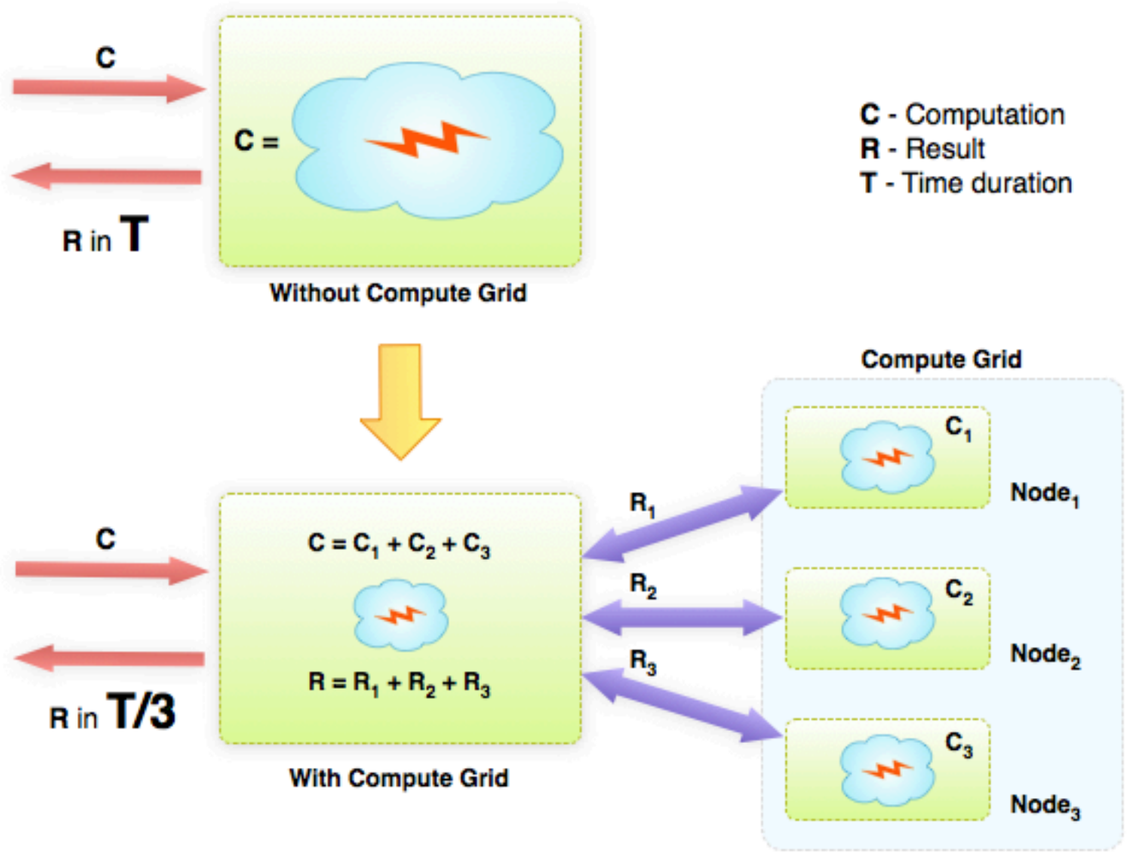
GridGain Technology

- Fully integrated cloud middleware: **Compute Grid** + **Data Grid**
- Real Time Transactional Big Data
- Zero Deployment

- Two editions:
  - **Community Edition**: License: GPLv3 + Basic Features
  - **Enterprise Edition**: Commercial License + Enterprise Features
- Language support:
  - Java 1.6
  - Scala 2.9.1

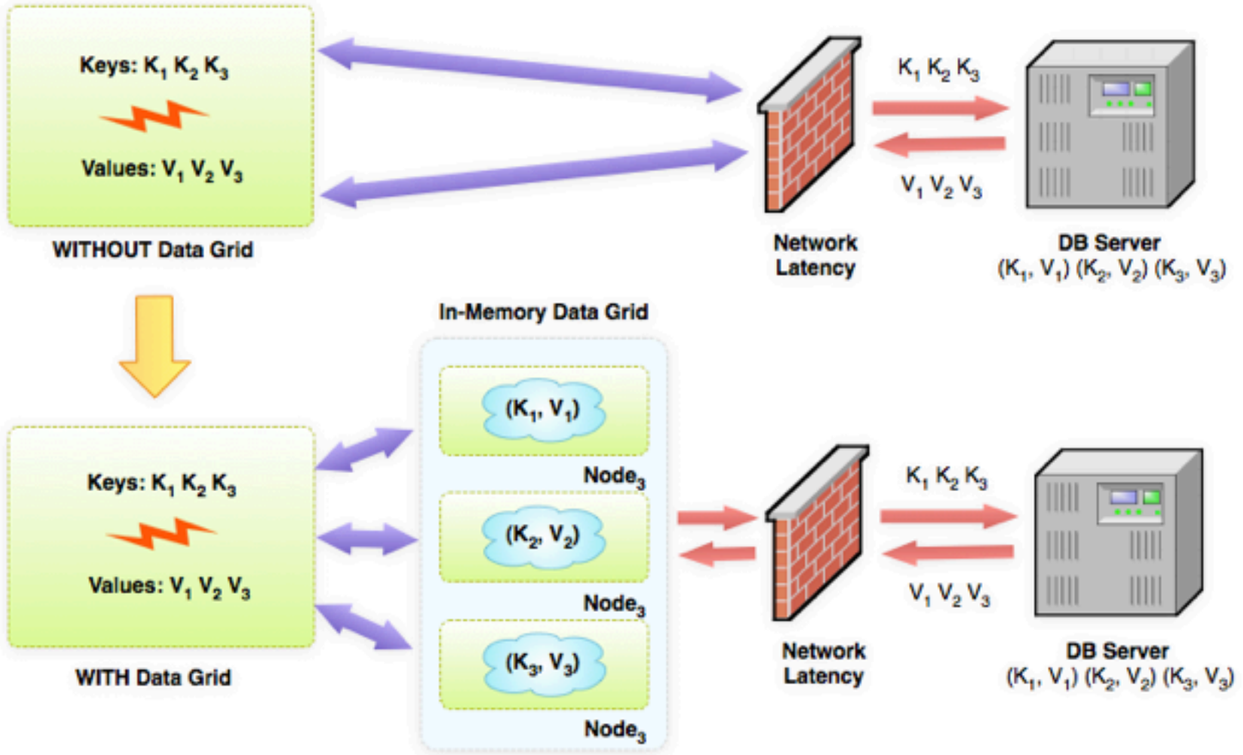
GridGain - Compute Grid

- Direct support for MapReduce
- Auto discovery
- Checkpoints for long running tasks
- Load Balancing
- Affinity co-location with data grids
- Automatic fault tolerance



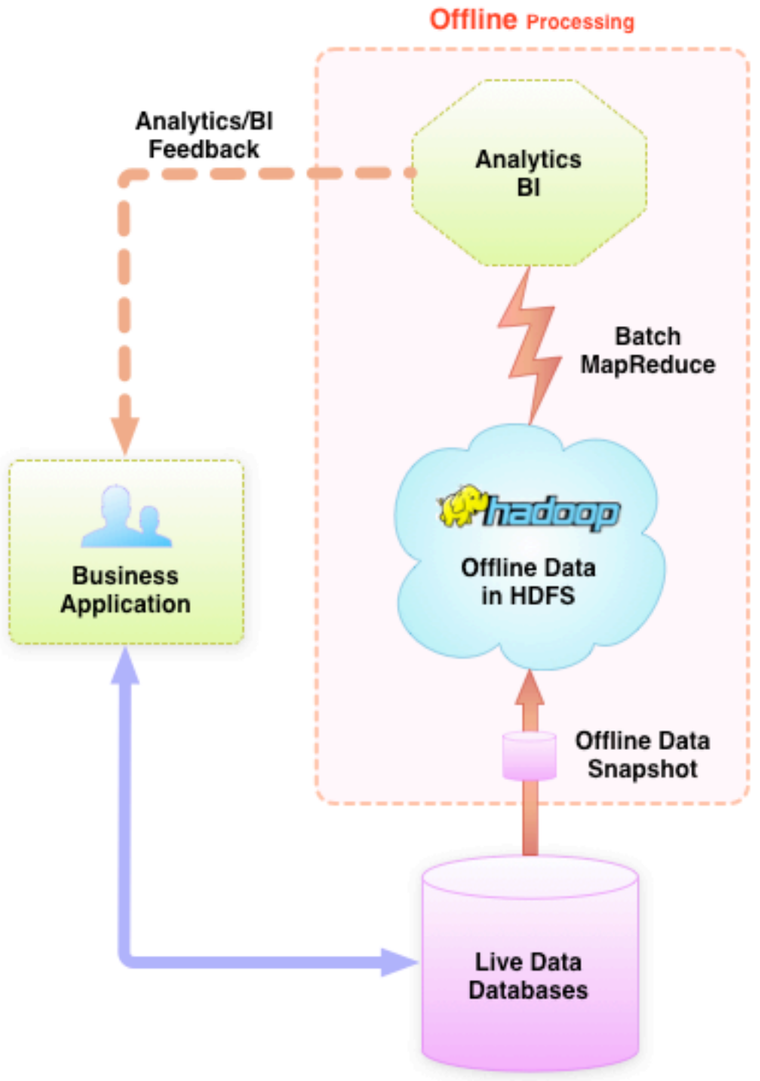
GridGain - Data Grid

- Replication & Partitioning
- Pessimistic & Optimistic Tx
- Read-Through and Write-Through
- Pluggable data overflow storage
- Distributed Queries
- Distributed Queues and Latches
- Distributes Java Atomics



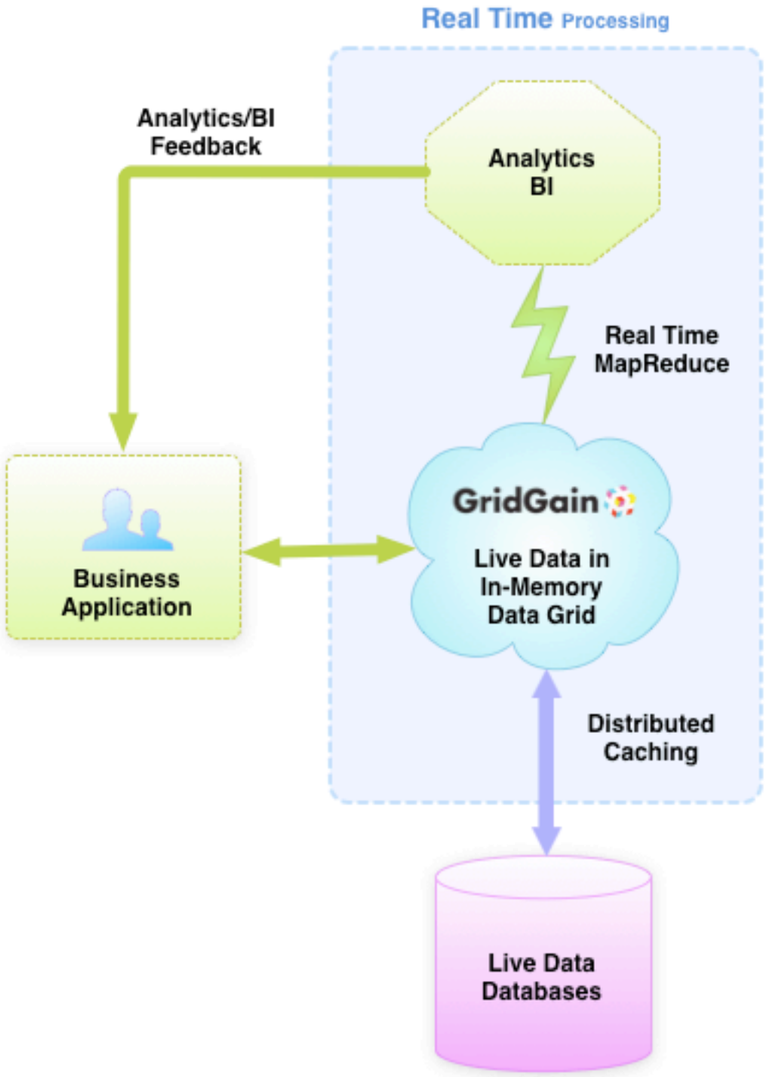
Hadoop Processing

- Very large data sets, BUT...
- Not Real Time
- Mandatory data snapshots
- HDFS instead of live databases
- Analytics based on offline data



GridGain Processing

- Large data sets
- Near Real Time Processing
- Online databases
- In-memory data caching
- Co-location of analytics and data
- Business analytics on Live Data



Live Coding - Real Time Word Count

- **Real time** uploading of books into Cache
- **Real time** updates of word counts
- **Real time** SQL queries for popular words
- **Real time** print-outs of most popular words
- ... using Scala & GridGain

Thank You!