MEMORY-CENTRIC ARCHITECTURE

IN-MEMORY PERFORMANCE, DURABILITY OF DISK

Dmitriy Setrakyan

Founder, Chief Product Officer, Apache Ignite PMC

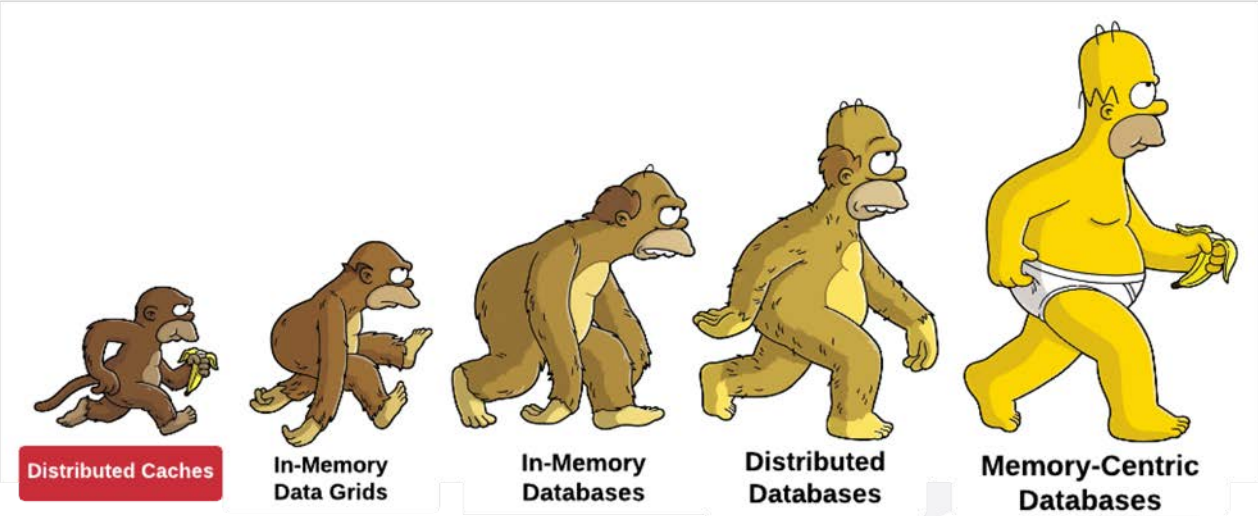
© 2018 GridGain Systems, Inc.

Agenda

- In-Memory Computing Evolution
  - Distributed Caches
  - In-Memory Data Grids
  - In-Memory Databases
  - Distributed Databases
  - Memory-Centric Databases
- Data Storage Ecosystem

© 2018 GridGain Systems, Inc.

Distributed Caches



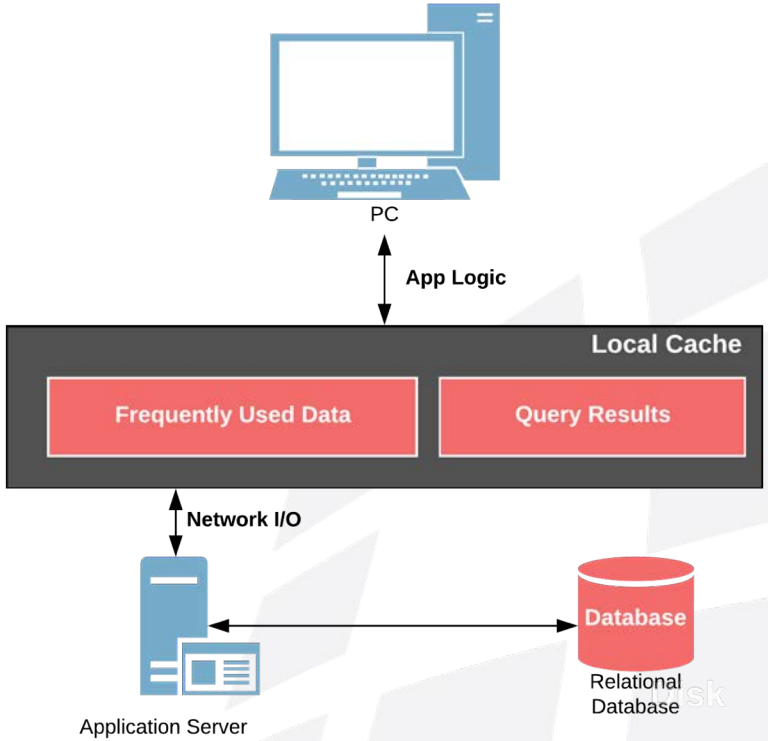
© 2018 GridGain Systems, Inc.

Cache is a hardware or software component that stores data so future requests for that data can be served faster

© 2018 GridGain Systems, Inc.

Local Caching

- Application In-Process Caching
  - Querying Results
  - Most Frequently Used Data
- Browser Caching
- Benefits
  - Speed up Applications!
  - Network is slow!

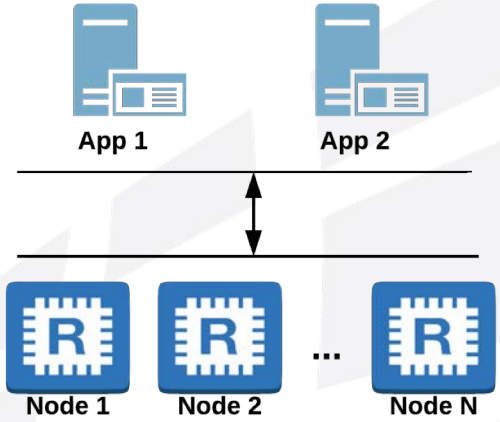


© 2018 GridGain Systems, Inc.

Distributed Caches

- Share-nothing (consistent hashing or sharding)
- Simple client protocol
- Memcached, Redis, AWS ElastiCache

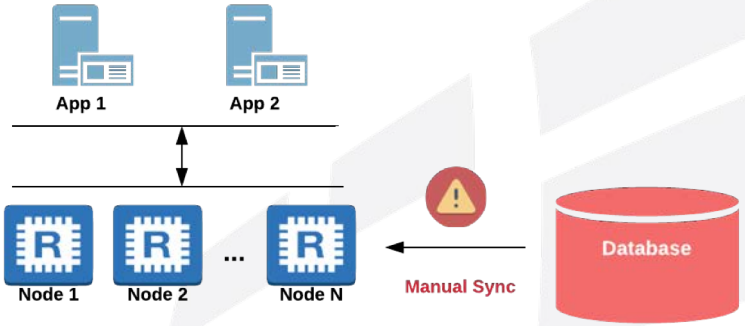
- Benefits
  - Shared cache
  - Beyond local RAM capacity
  - Fault tolerant and scalable



© 2018 GridGain Systems, Inc.

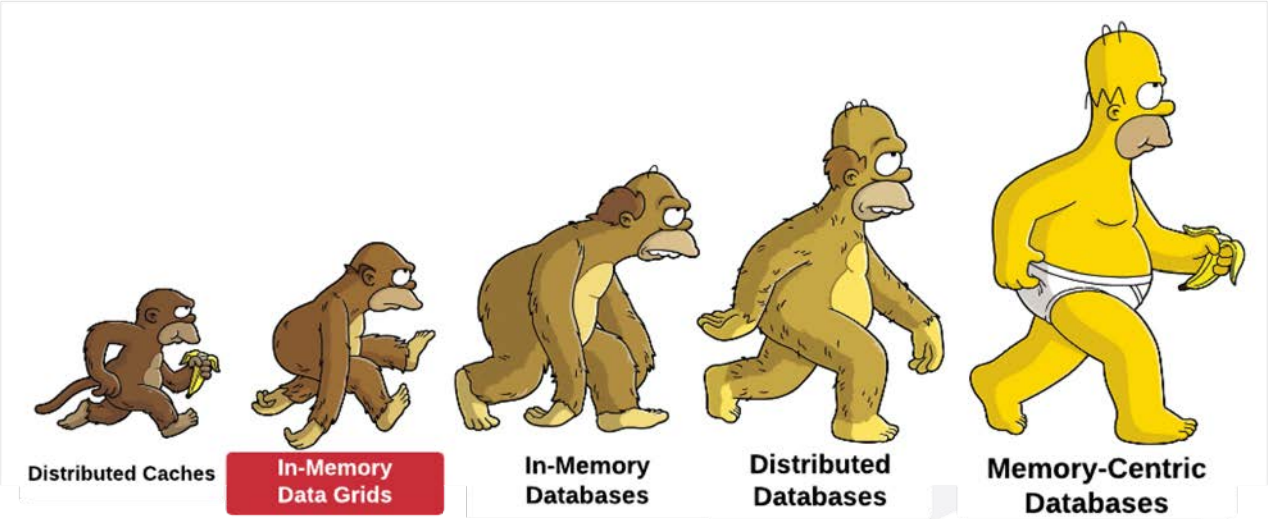
Shortcomings of Distributed Caches

- Lack of queries (SQL or JOINs)
- Lack of ACID transactions
- Lack of collocated processing
- Lack of database synchronization
  - Cache-aside pattern
- Lack of native persistence
  - cannot persist more than fits in RAM
- RAM warm-up on restart



© 2018 GridGain Systems, Inc.

In-Memory Data Grids



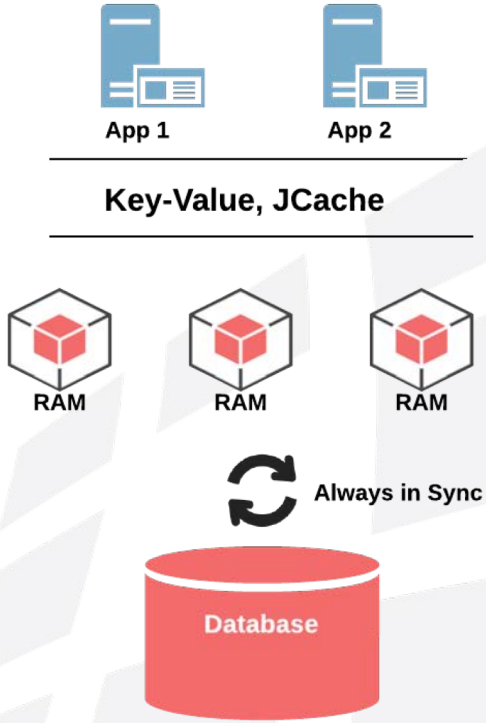
© 2018 GridGain Systems, Inc.

GridGain

In-Memory Data Grids

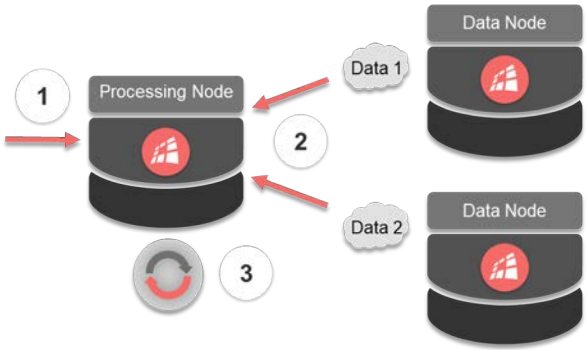
- Intelligent Distributed Caches
- Hazelcast, GigaSpaces, Apache Ignite

- Benefits
  - ACID Transactions
  - Query APIs
  - Event notifications
  - Continuous queries
  - Database Read/Write-Through
  - Collocated Processing



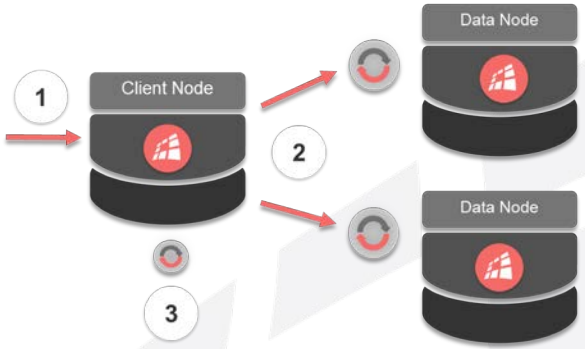
© 2018 GridGain Systems, Inc.

Client-Server Processing



1. Initial Request
2. Fetch data from remote nodes
3. Process the entire data-set

Co-Located Processing



1. Initial request
2. Co-locate processing with data
3. Reduce multiple results into one

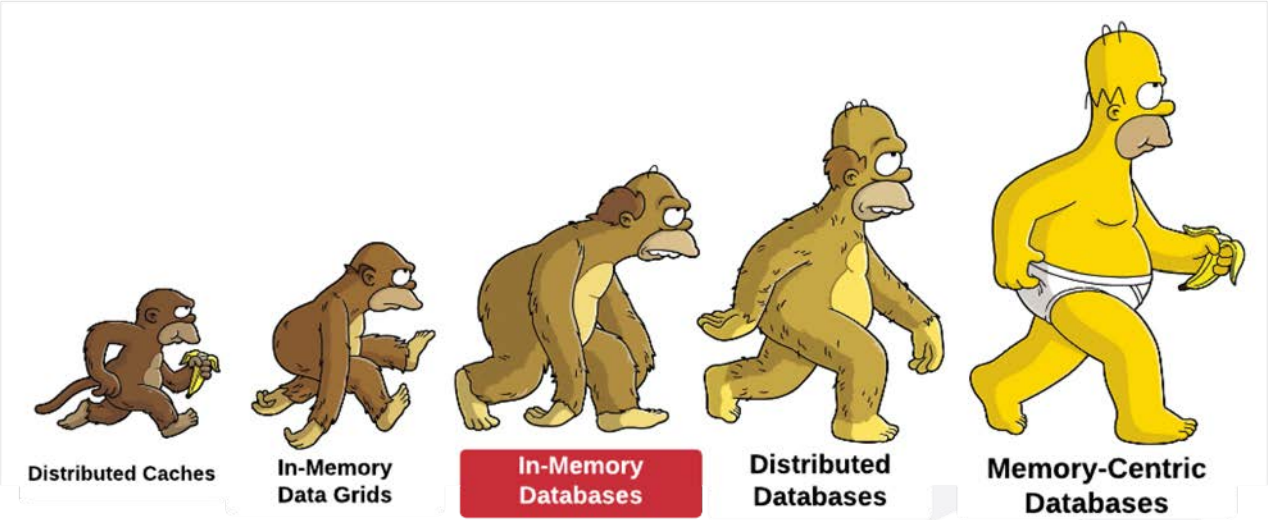
© 2018 GridGain Systems, Inc.

Shortcomings of In-Memory Data Grids

- Custom Query Languages
  - No SQL, No JOINs
- Lack of native persistence
  - cannot persist more than fits in RAM
- RAM warm-up on restart

© 2018 GridGain Systems, Inc.

In-Memory Databases



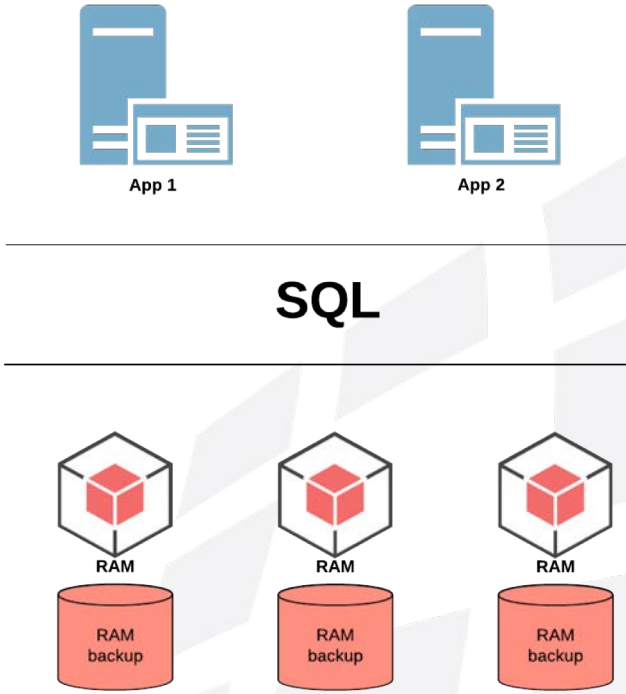
© 2018 GridGain Systems, Inc.

GridGain

In-Memory Databases

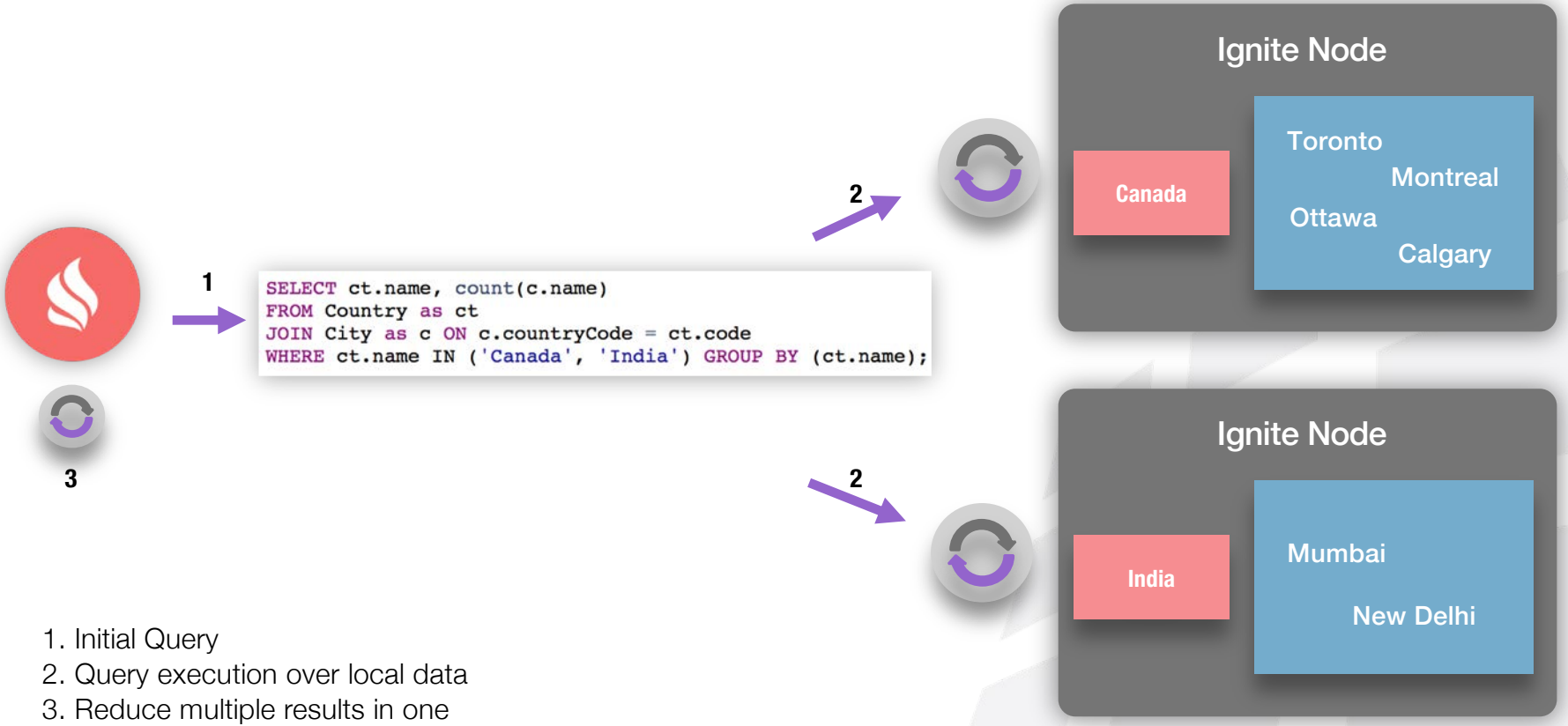
- Primary in-RAM Storage
- In-Memory Data Grids for SQL
- VoltDB, SAP Hana

- Benefits
  - Distributed SQL
  - ACID Transactions
  - Better Scalability



© 2018 GridGain Systems, Inc.

Collocated Joins

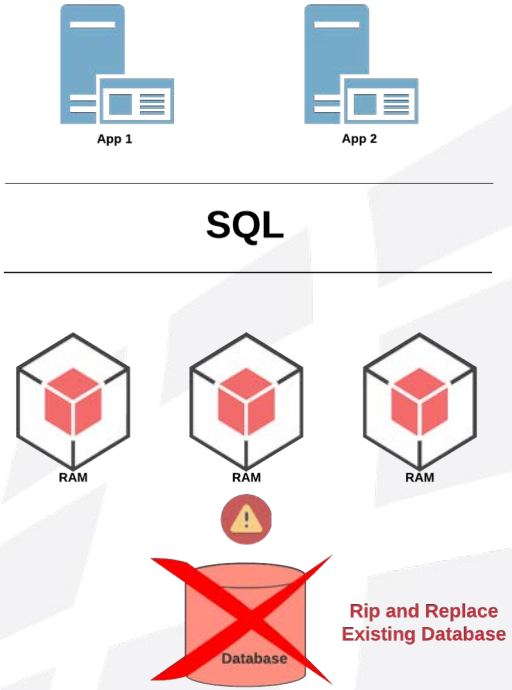


1. Initial Query
2. Query execution over local data
3. Reduce multiple results in one

© 2018 GridGain Systems, Inc.

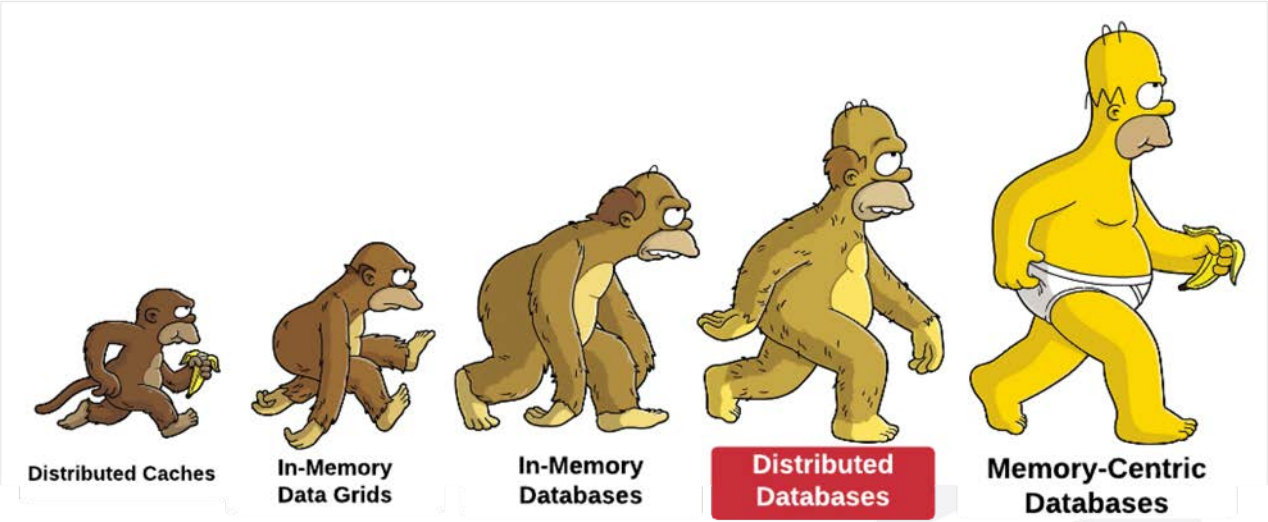
Shortcomings of In-Memory Databases

- Rip-n-Replace existing database
  - cannot keep or reuse existing database
- Lack of collocated processing
- Lack of native persistence
  - cannot scale beyond in RAM
  - basic memory-offload on disk
  - RAM warm-up on restart
  - not a system of record



© 2018 GridGain Systems, Inc.

Distributed Databases

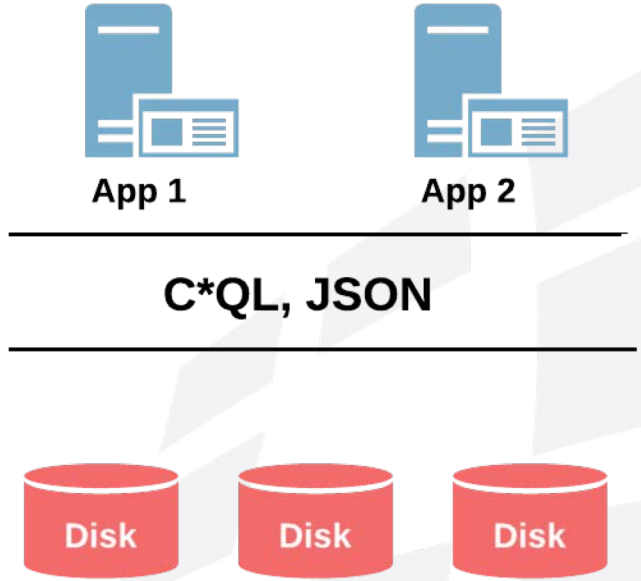


© 2018 GridGain Systems, Inc.

GridGain

Distributed NoSQL Databases

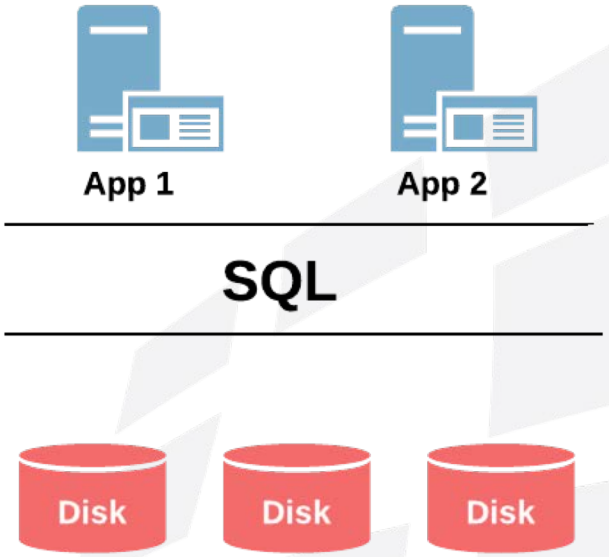
- Cassandra, Mongo
  - Distributed persistence
  - Fault tolerant and scalable
- Shortcomings
  - Lack of SQL (no JOINs)
  - Lack of distributed transactions
  - Lack of in-memory processing



© 2018 GridGain Systems, Inc.

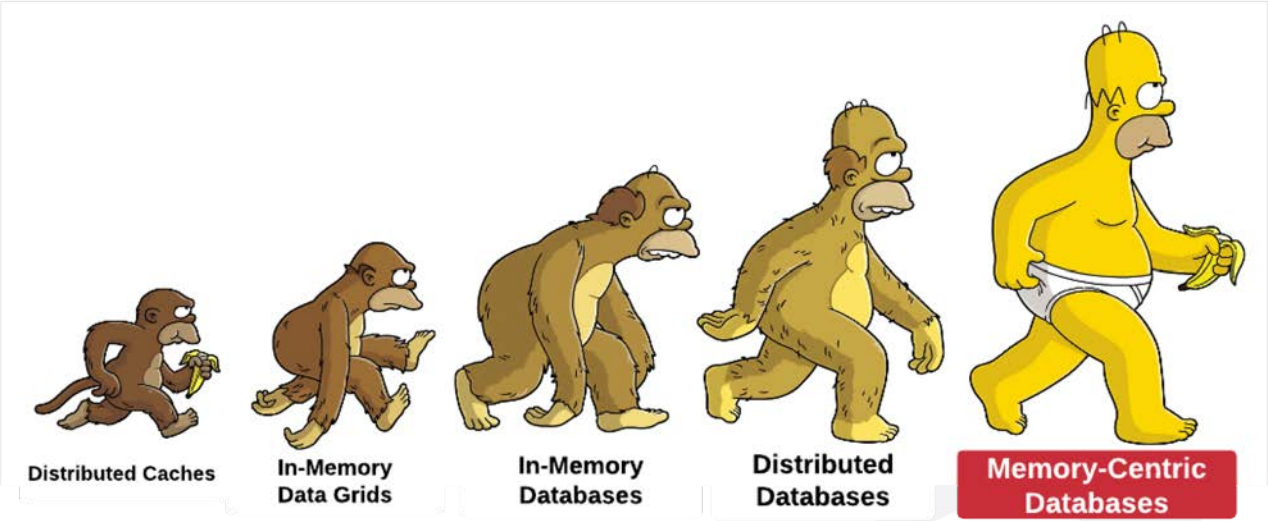
Distributed SQL Databases

- CockroachDB, Cloud Spanner
  - Cloud-native
  - Fault tolerant and scalable
  - SQL (distributed JOINS)
- Shortcomings
  - Lack of in-memory processing
  - Lack of key-value API
  - Lack of event notifications



© 2018 GridGain Systems, Inc.

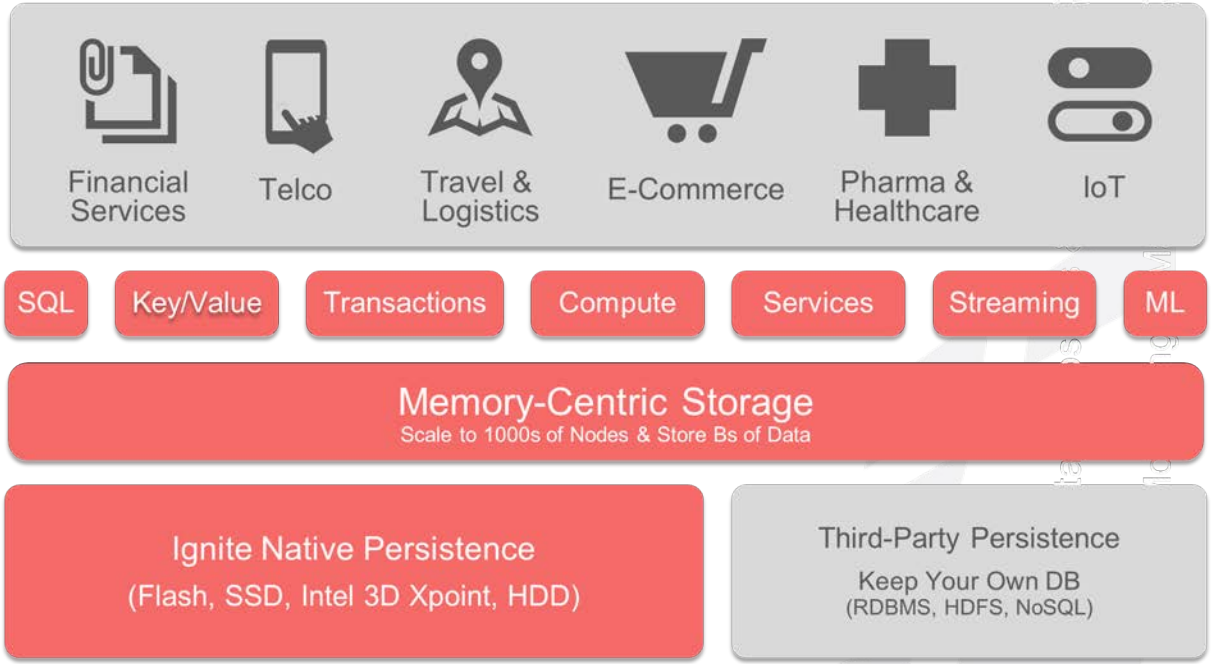
Memory-Centric Databases



© 2018 GridGain Systems, Inc.

GridGain

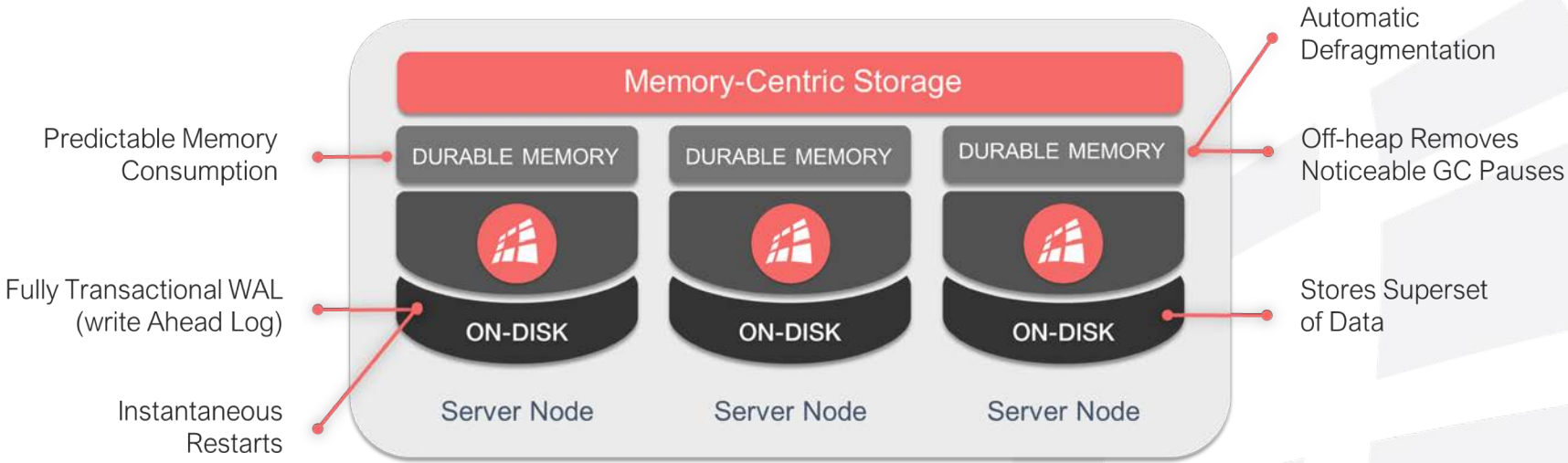
Ignite Memory-Centric Architecture



© 2018 GridGain Systems, Inc.

GridGain

Memory Centric Storage



© 2018 GridGain Systems, Inc.

GridGain

Memory & Disk Utilization

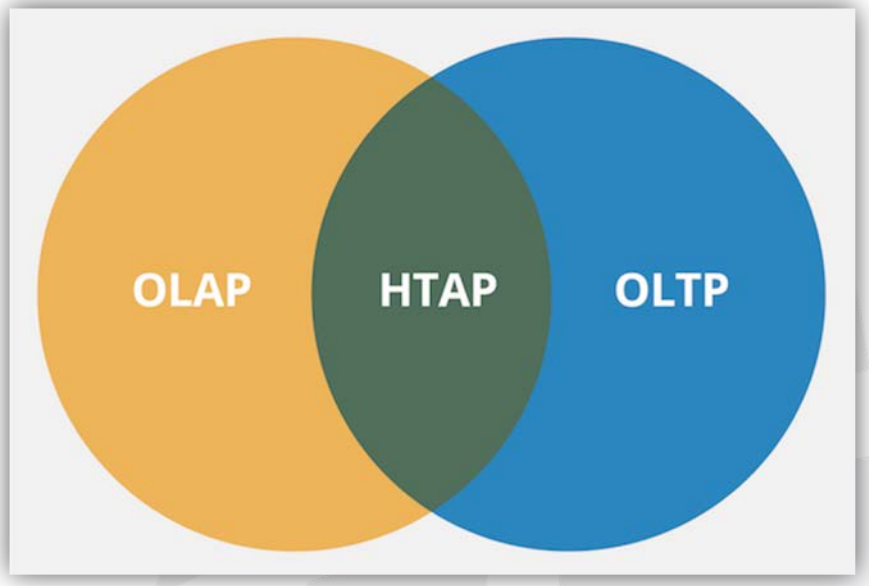
Mode	Description	Major Advantage
In-Memory	Pure In-Memory Storage	Maximum perfomance possible (data is never written to disk)
In-Memory + 3rd Party DB	Caching layer (aka. in-memory data grid) for existing databases – RDBMS, NoSQL, etc	Horizontal scalability Faster reads and writes
In-Memory + Full Copy on Disk	The whole data set is stored both in memory and on disk	Survives cluster failures
100% on Disk + In-Memory Cache	100% of data is in Ignite native persistence and a subset is in memory	Unlimited data scale beyond RAM Collocated memory and storage

© 2018 GridGain Systems, Inc.

GridGain Company Confidential

Supports HTAP Principles

- One Platform, Multiple Workloads
  - OLTP and OLAP
  - Real-Time Streaming and Batching
  - No ETL
- Distributed Everything
  - Scalability
  - No Single Point of Failure
  - SQL, Calculations, ML, etc.
  - Cloud Native



© 2018 GridGain Systems, Inc.

Data Storage Ecosystem

Feature	RDBMS	NoSQL	IMDG	IMDB	GridGain and Ignite
Scale Out	X	✓	✓	✓	✓
Availability	X	✓	✓	✓	✓
Consistency	✓	X	✓	✓	✓
In-Memory	✓	X	✓	✓	✓
Persistence	✓	✓	X	X	✓
SQL	✓	X	X	✓	✓
Key-Value	X	✓	✓	X	✓
Collocated Processing	X	X	✓	X	✓

© 2018 GridGain Systems, Inc.

Any Questions?

https://www.gridgain.com
https://ignite.apache.org

© 2018 GridGain Systems, Inc.