In-Memory Performance

Durability of Disk

© 2017 GridGain Systems, Inc.

Apache Ignite

In-Memory Hammer for Your Data Science Toolkit

Denis Magda

Ignite PMC Chair

GridGain Director of Product Management

© 2017 GridGain Systems, Inc.

Agenda

- Apache Ignite Overview
  - Use Cases
- Data Science Toolkit Box
  - Data Grid
  - Durable Memory
  - Distributed SQL
  - Compute Grid
  - Machine Learning Grid (Beta)
- Q&A

© 2017 GridGain Systems, Inc.

Apache Ignite In-Memory Computing Platform

SQL | Key/Value | Transactions | Compute | Services | Streaming | ML

Memory-Centric Storage

Ignite Native Persistence

(Flash, SSD, Intel 3D XPoint)

Third-Party Persistence

(RDBMS, HDFS, NoSQL)

© 2017 GridGain Systems, Inc.

Apache Ignite Users

Financial Services

FinTech

Software

Telco

Pharma & Healthcare

Logistics & Travel

E-commerce

IoT

Adtech

© 2017 GridGain Systems, Inc.

- Drug Discovery and Network Biology

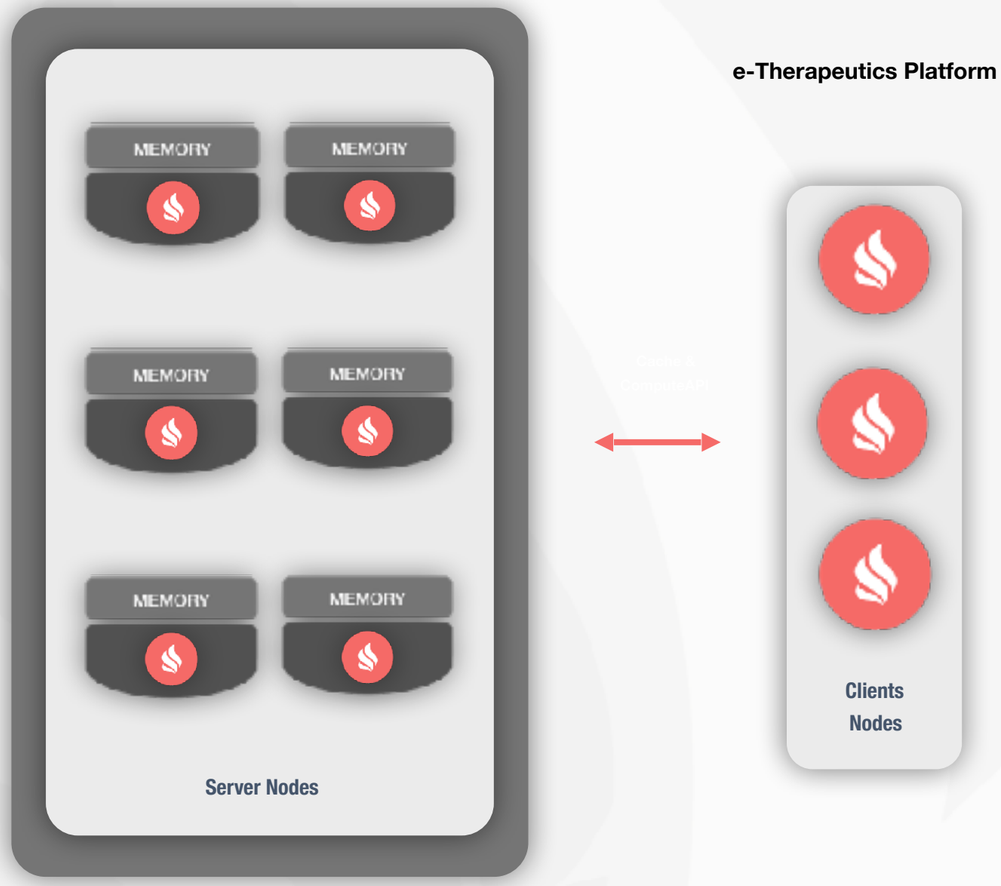
e-Therapeutics provides a computer-based drug discovery platform and a specialized approach to network biology.

Problem

- Analysis of a network of proteins influencing a disease and drugs discovery could be measured in weeks
- Could not parallelize existing algorithms

Apache Ignite Solution

- 80x speed increase over the non-parallelized environment
- Analysis projects completion in hours and minutes
- Computational resources for abandoned research projects

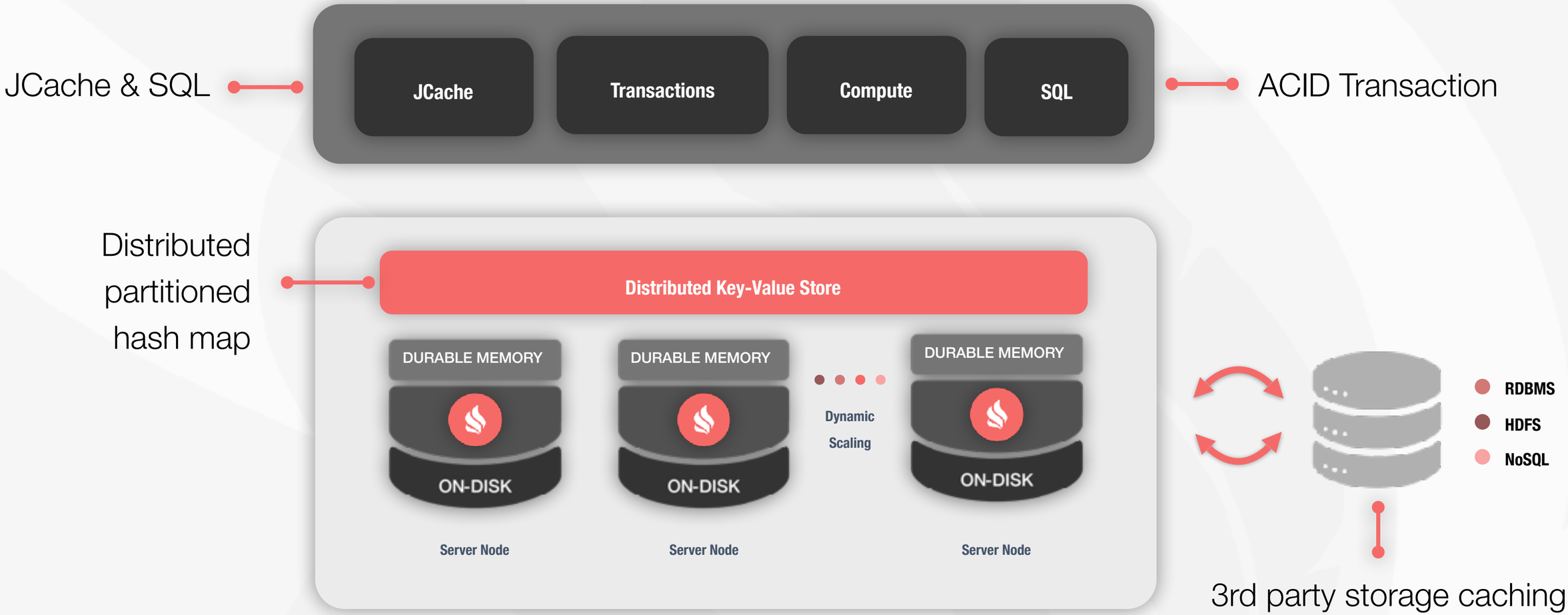


100x Cluster Nodes

5x Physical Nodes

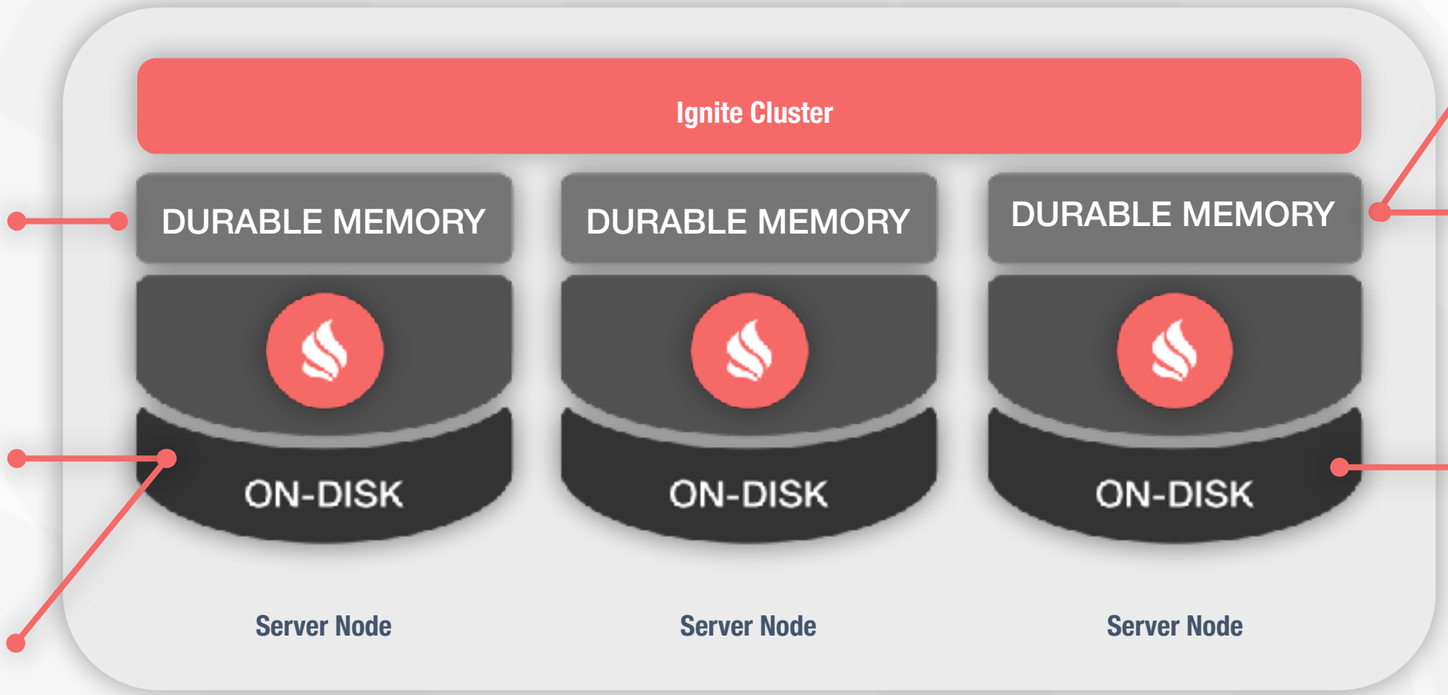
© 2017 GridGain Systems, Inc.

Data Grid



© 2017 GridGain Systems, Inc.

Durable Memory



Predictable memory consumption

Fully Transactional (Write-Ahead Log)

Instantaneous Restarts

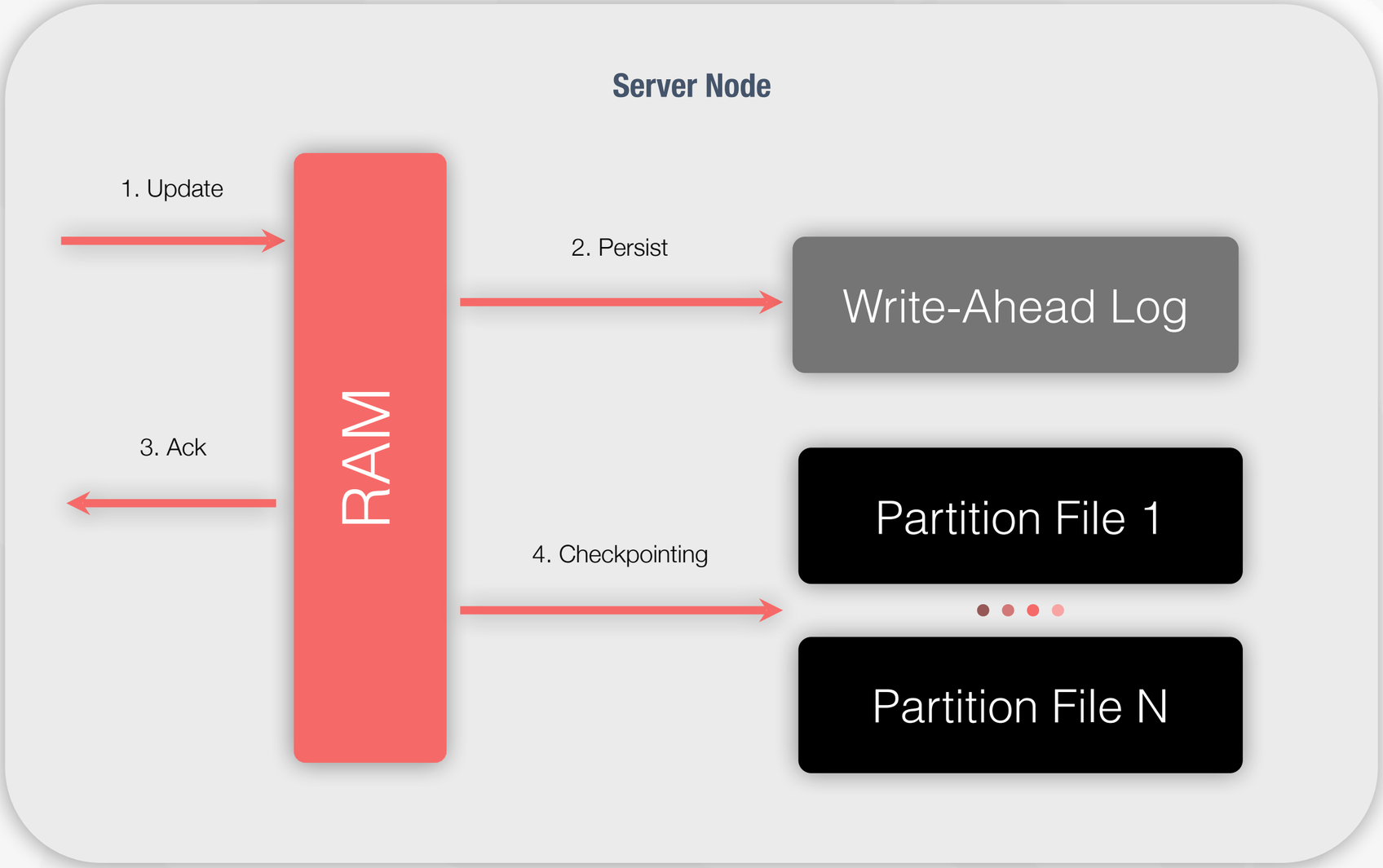
Automatic Defragmentation

Off-heap Removes noticeable GC pauses

Stores Superset of Data

© 2017 GridGain Systems, Inc.

Ignite Native Persistence



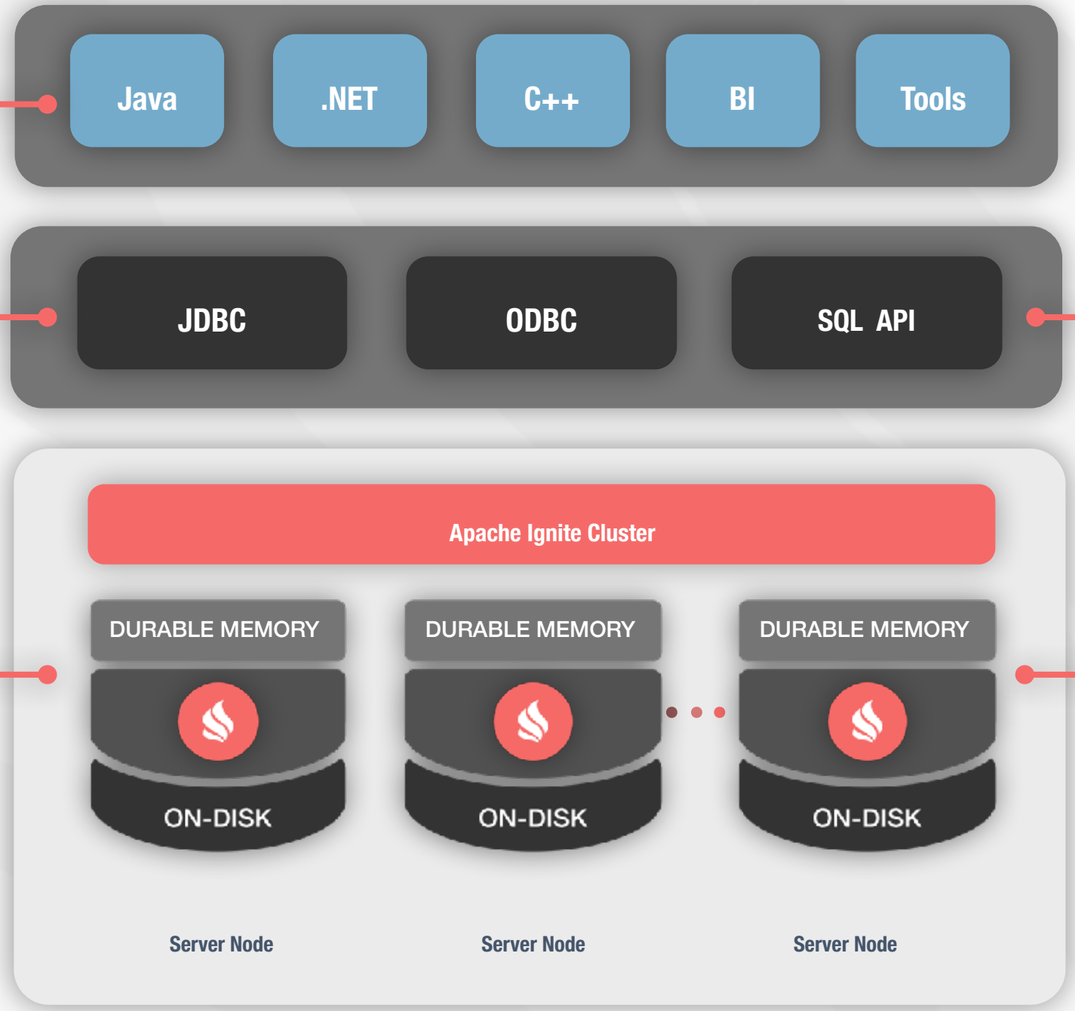
© 2017 GridGain Systems, Inc.

Distributed SQL

Cross-platform Compatibility

DDL, DML Support

Indexes in RAM or Disk

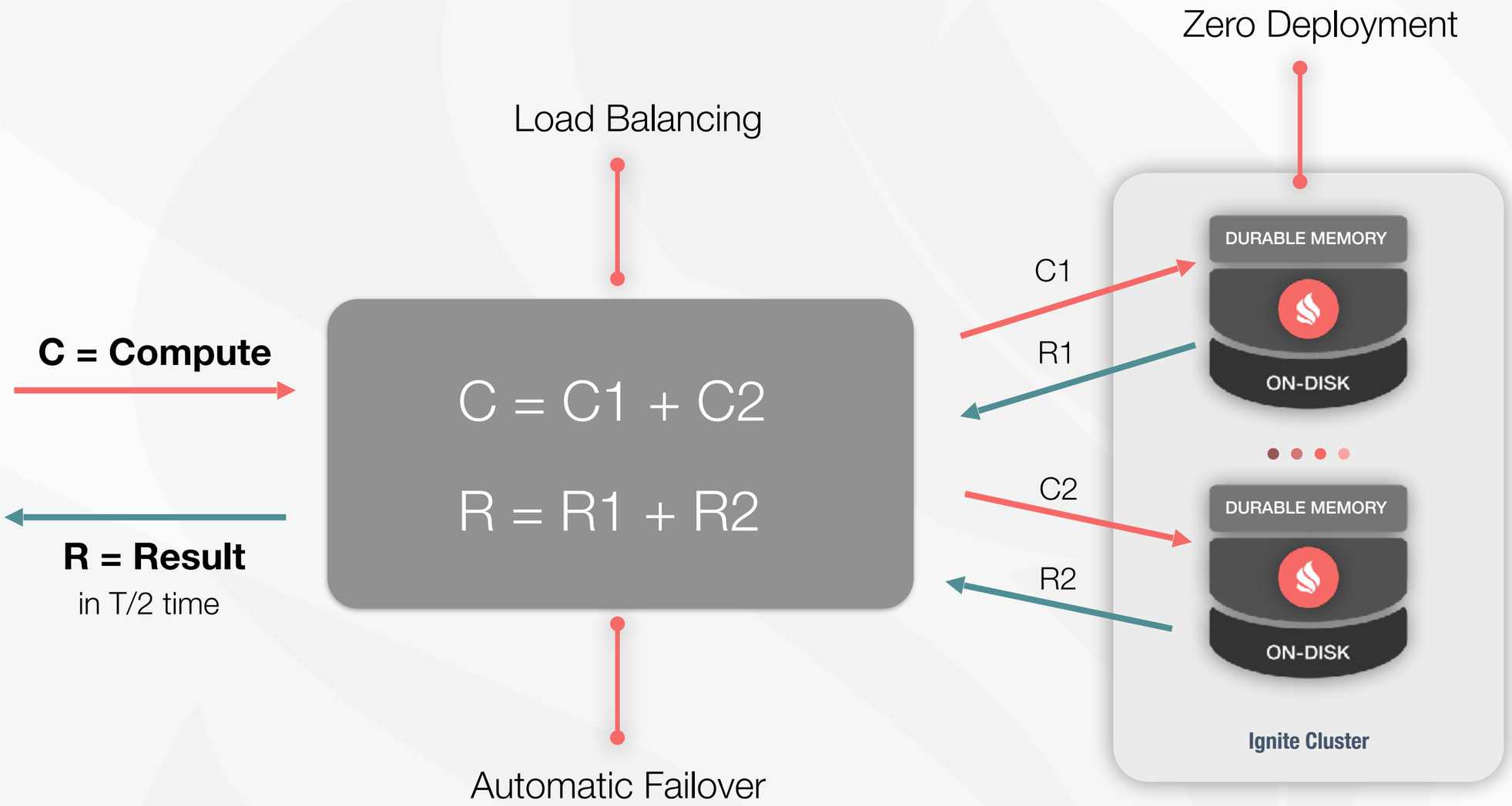


SELECT, UPDATE, INSERT, MERGE, DELETE, CREATE and ALTER

Dynamic Scaling

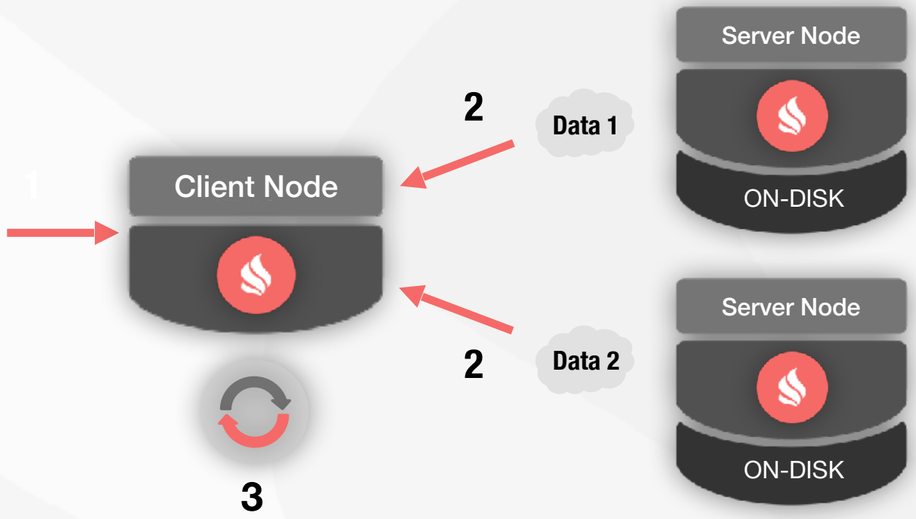
© 2017 GridGain Systems, Inc.

Compute Grid



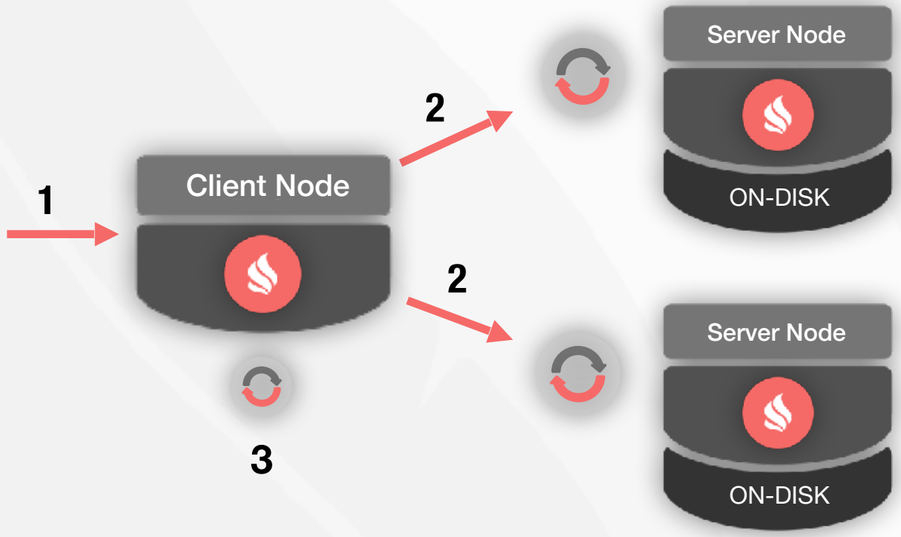
© 2017 GridGain Systems, Inc.

Client-Server Processing



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

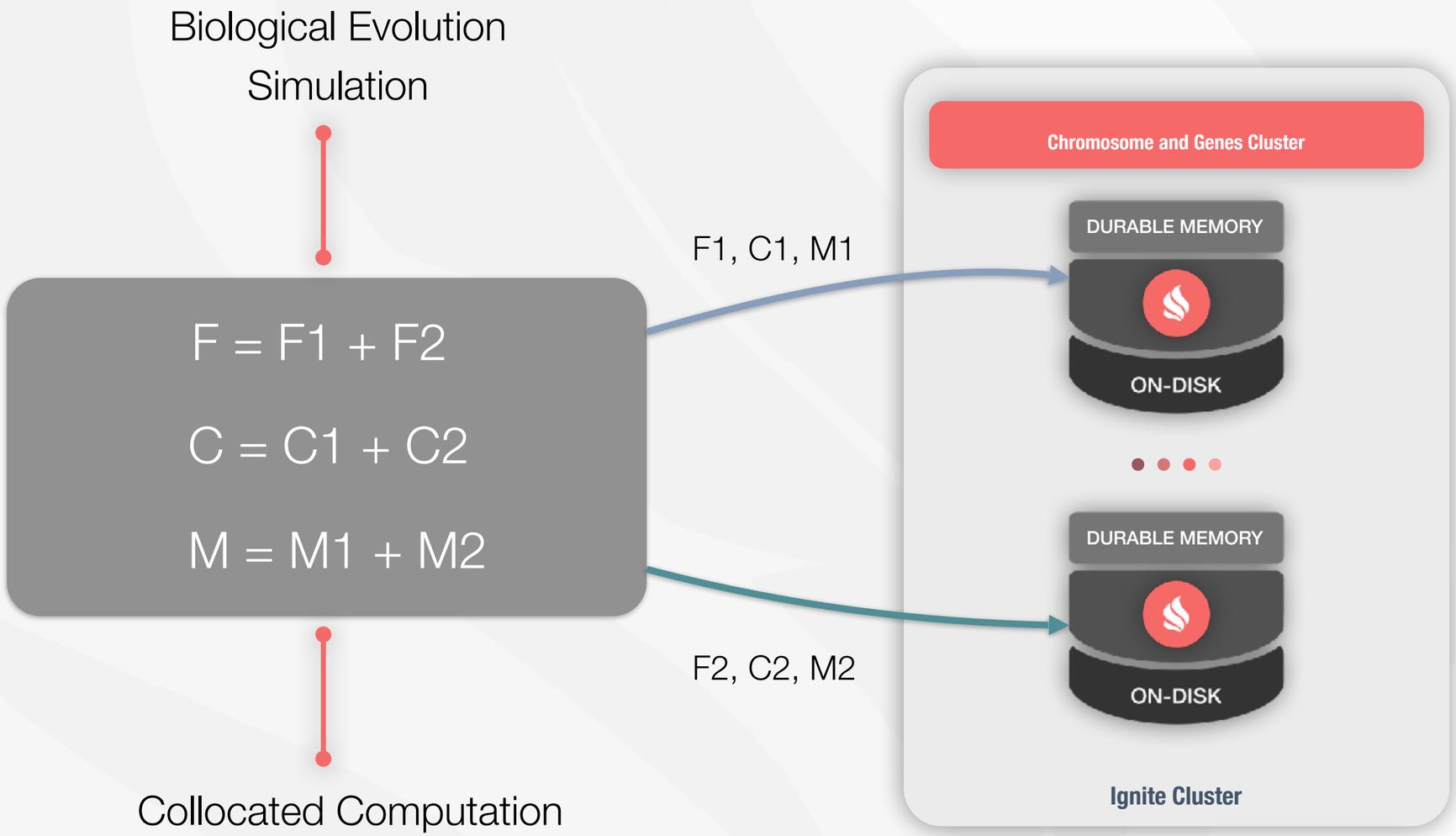
Co-located Processing



1. Initial Request
2. Co-located processing with data
3. Reduce multiple results in one

© 2017 GridGain Systems, Inc.

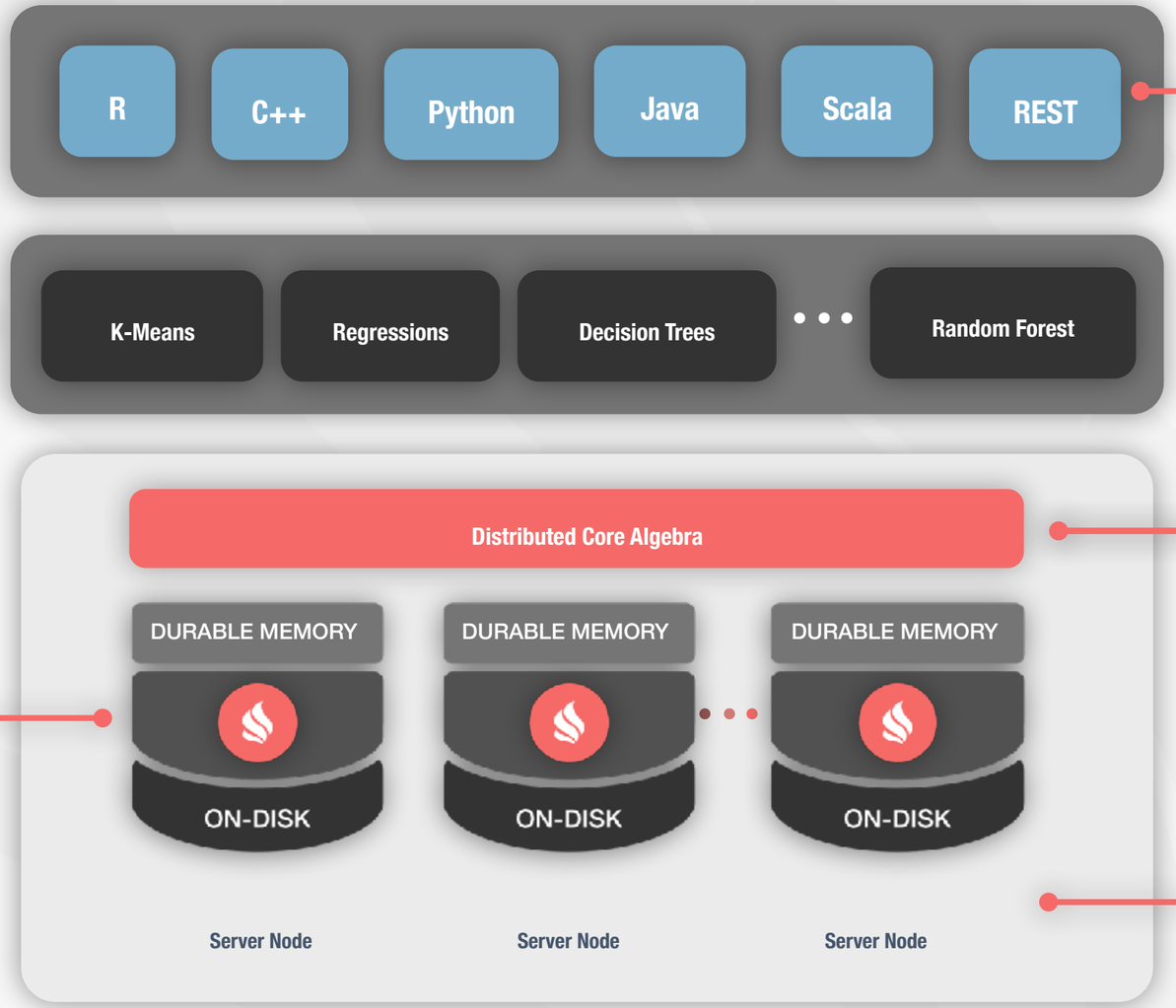
Genetic Algorithm Grid



F = Fitness Calculation
C = Crossover
M = Mutation

© 2017 GridGain Systems, Inc.

Machine Learning Grid



Multi-Language Support

Distributed Algorithms

Dense and Sparse Algebra

Large Scale Parallelization

No ETL

© 2017 GridGain Systems, Inc.

Any Questions?

Thank you for joining us. Follow the conversation.
http://ignite.apache.org

#apacheignite

#denismagda

© 2017 GridGain Systems, Inc.