



# Apache Ignite

In-Memory Hammer for Your Data Science Toolkit



Akmal Chaudhri  
Technology Evangelist  
GridGain

# Agenda

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

# Apache Ignite In-Memory Computing Platform

## Applications



Financial  
Services



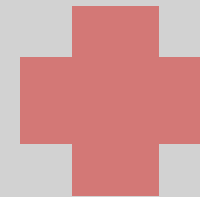
Telco



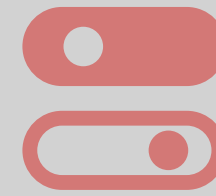
Travel &  
Logistics



E-Commerce



Pharma &  
Healthcare



IoT

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)

# Apache Ignite Use Cases

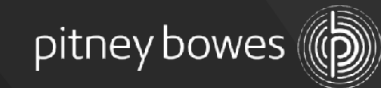
## Financial Services



## Software



## Logistics & Travel



## E-commerce



## FinTech



## Telco



## IoT



## Pharma & Healthcare



## Adtech



# e-therapeutics - **Drug Discovery and Network Biology**

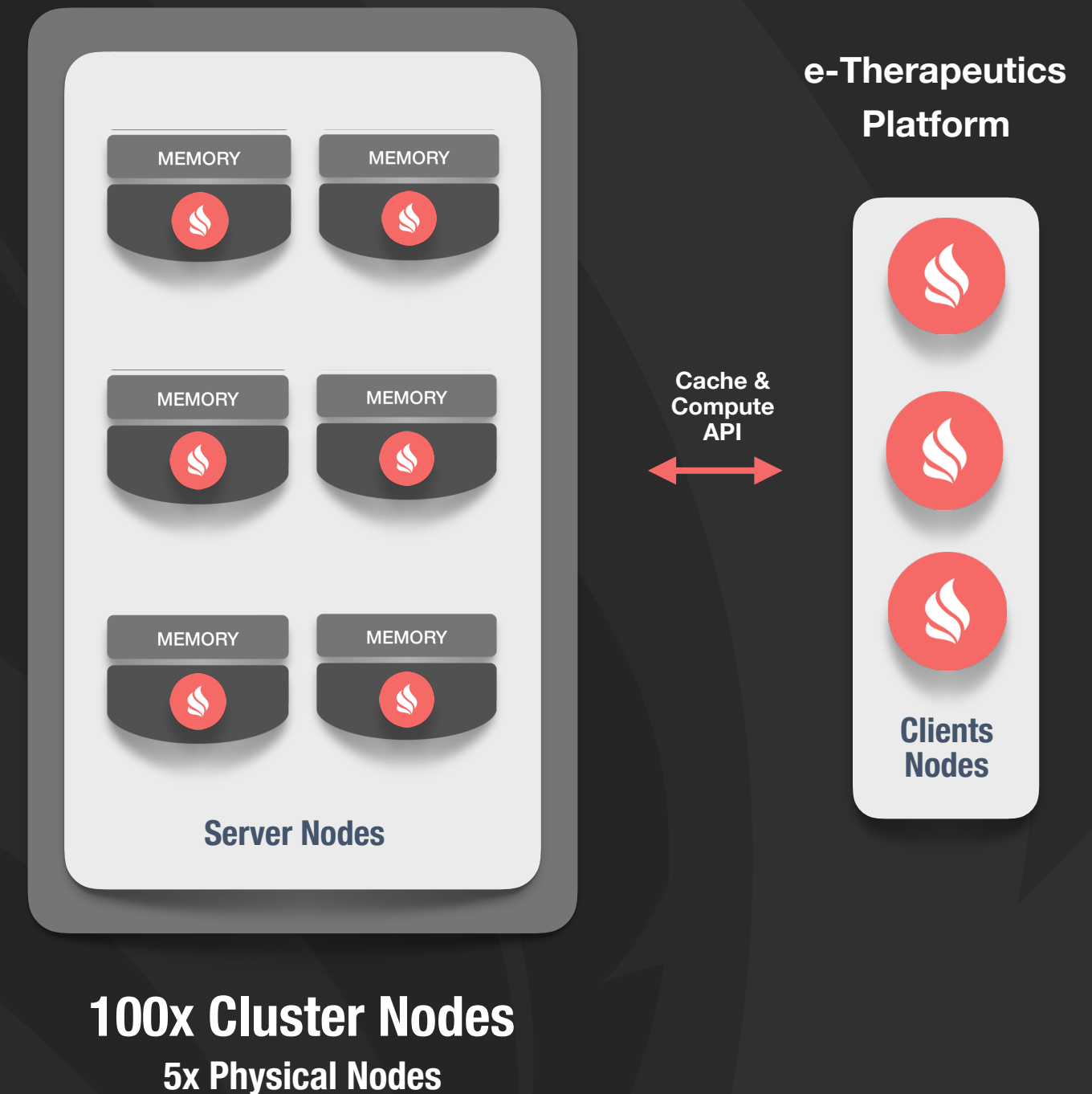
e-Therapeutics provides a computer-based drug discovery platform and a specialized approach to network biology.

## 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



# Data Grid

JCache & SQL

JCache

Transactions

Compute

SQL

ACID Transaction

Distributed  
partitioned  
hash map

Distributed Key-Value Store

DURABLE MEMORY



ON-DISK

Server Node

DURABLE MEMORY



ON-DISK

Server Node

Dynamic  
Scaling

DURABLE MEMORY



ON-DISK

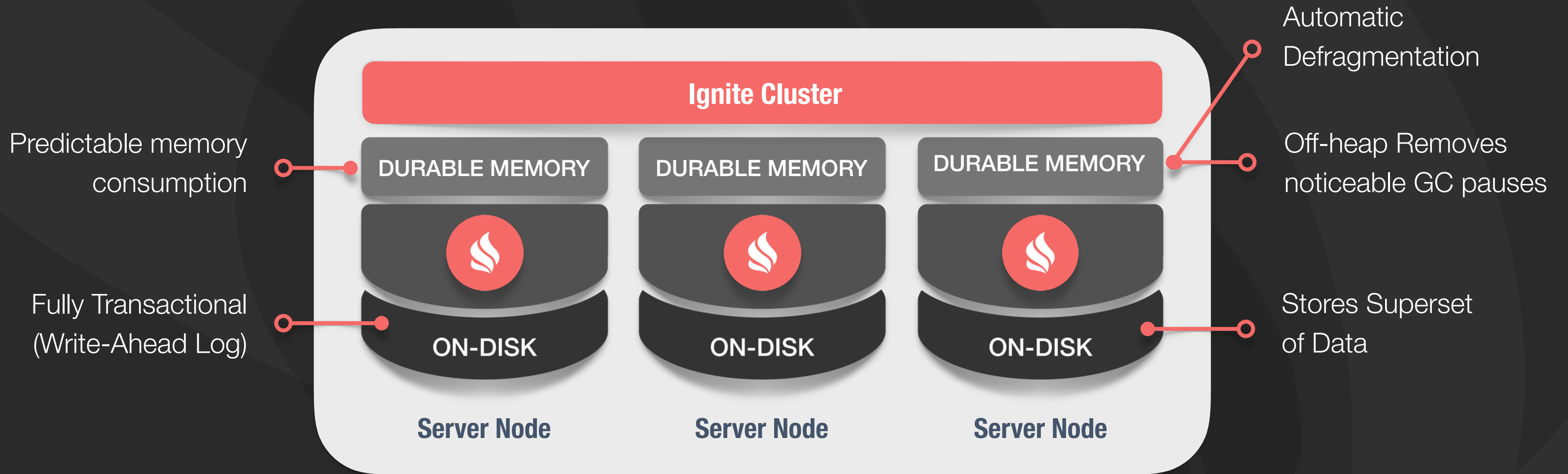
Server Node



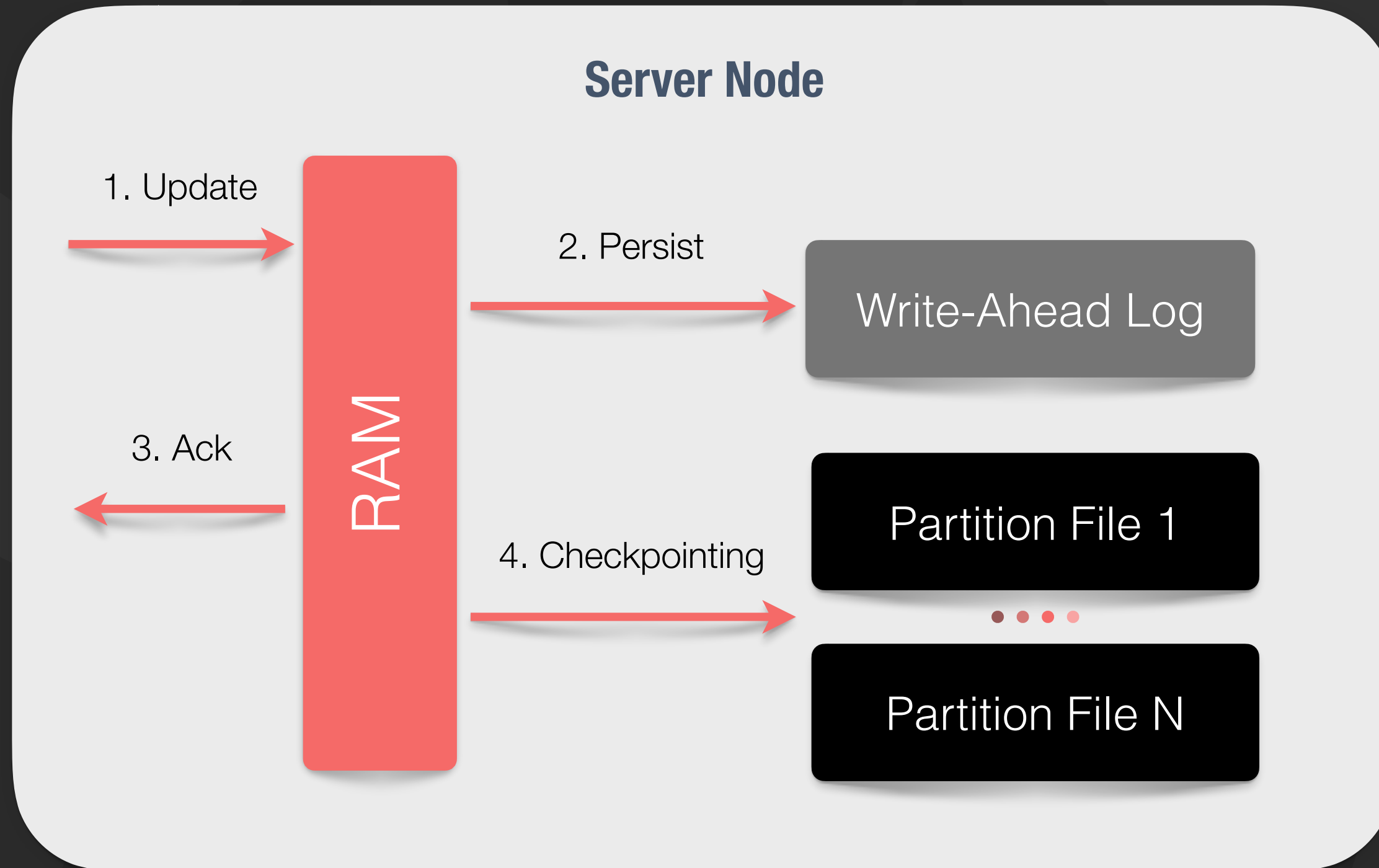
3rd party storage caching

● RDBMS  
● HDFS  
● NoSQL

# Durable Memory

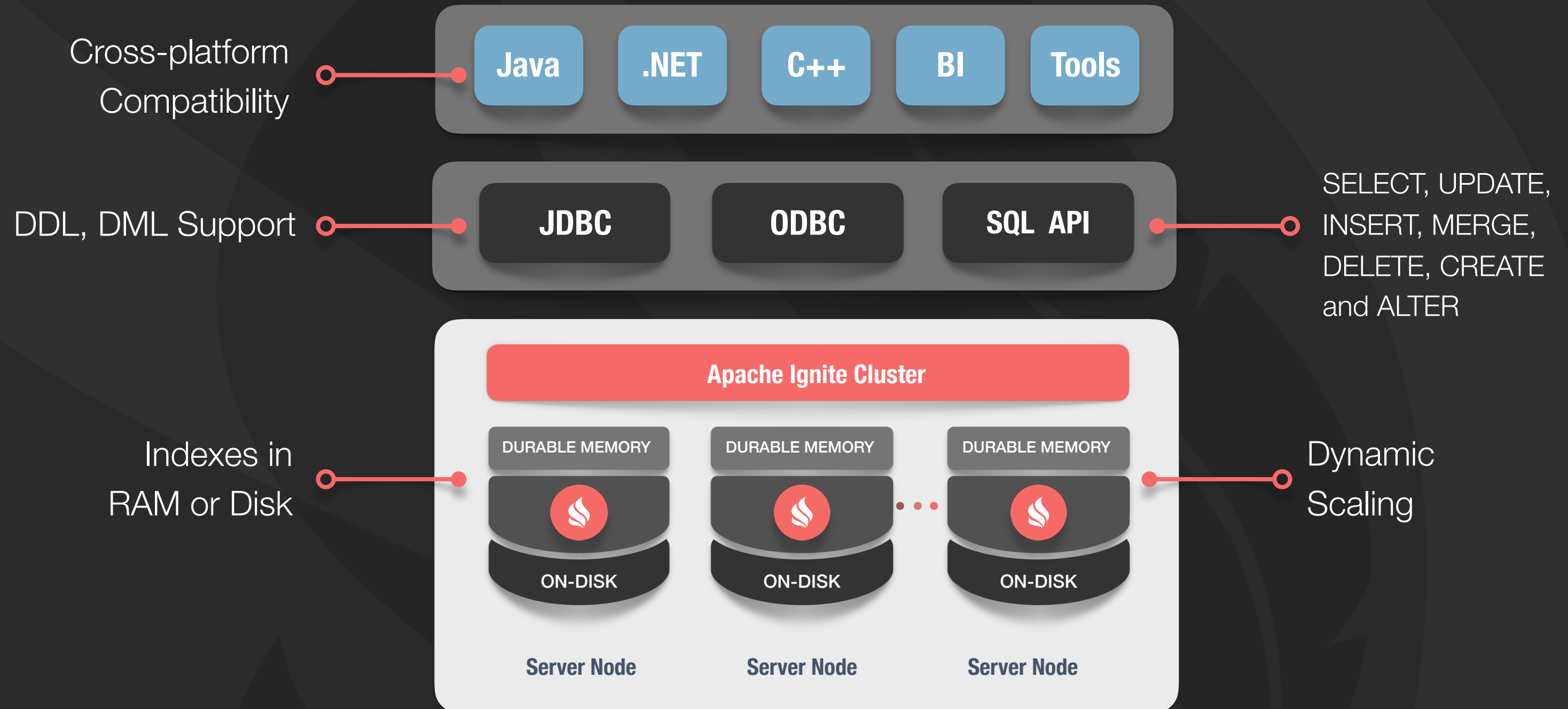


# Ignite Native Persistence

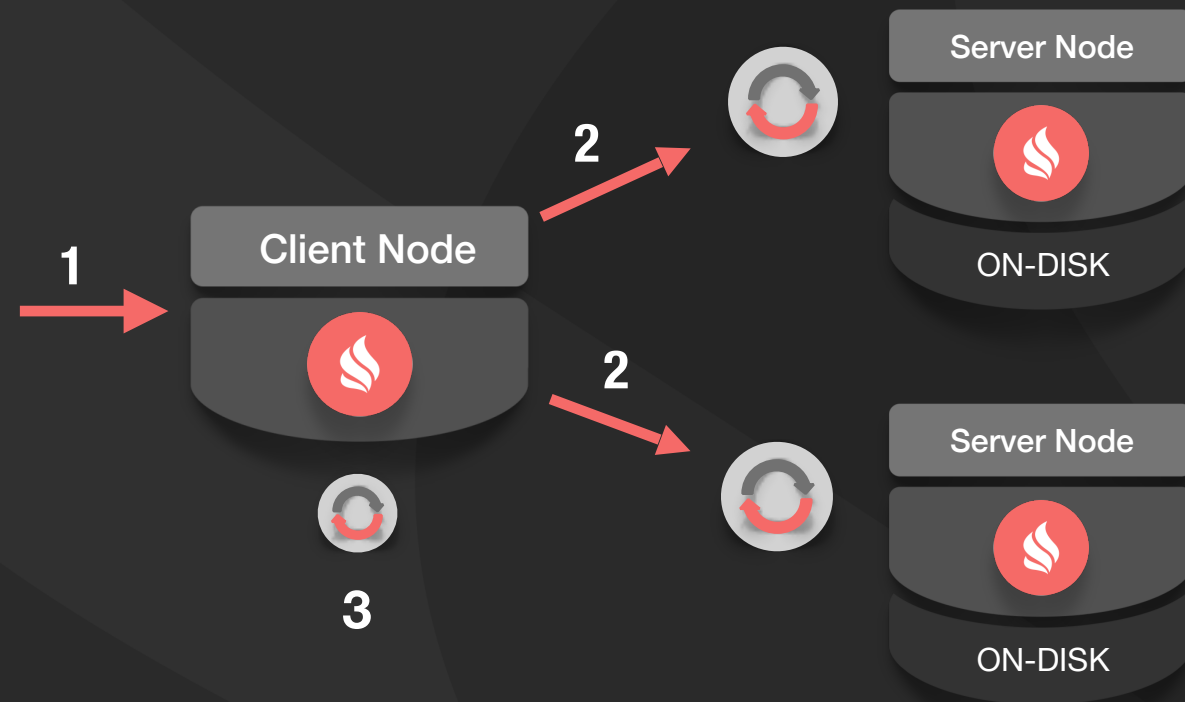




# Distributed SQL

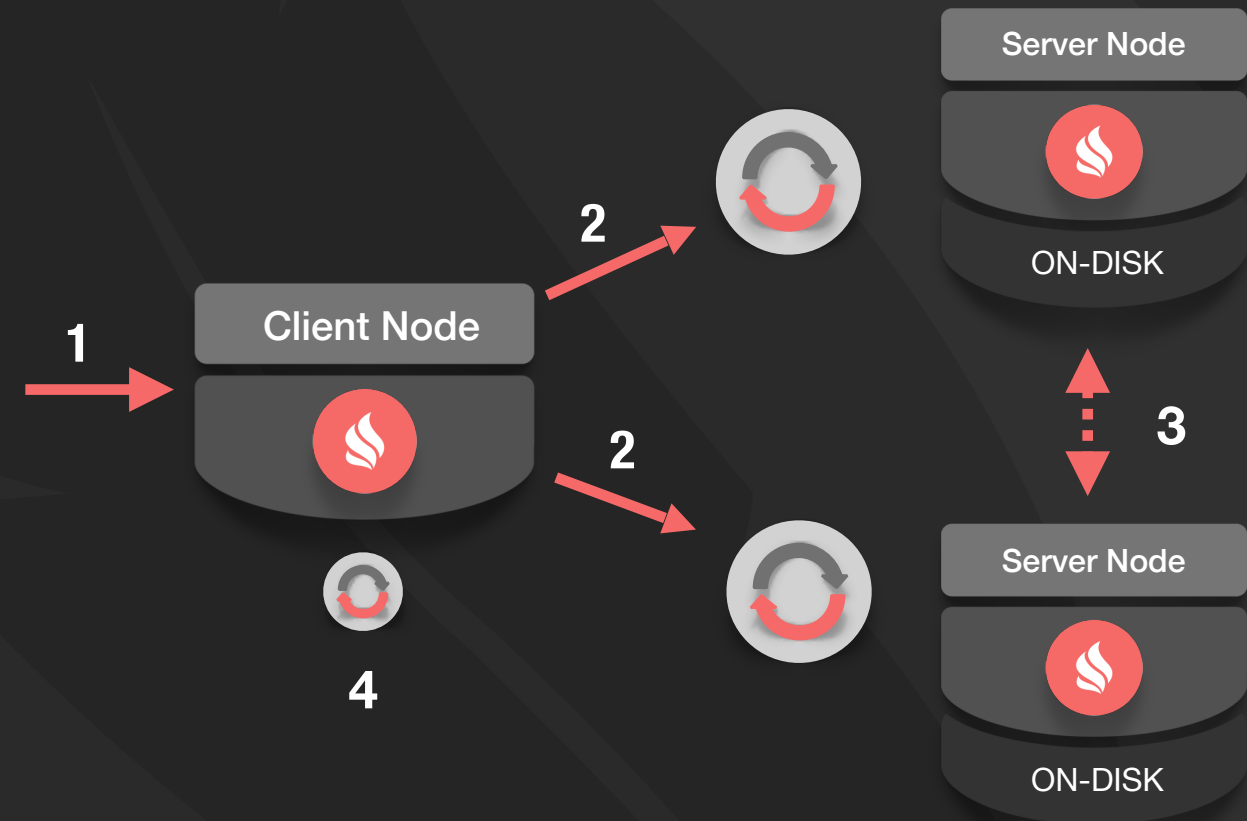


# Collocated Joins



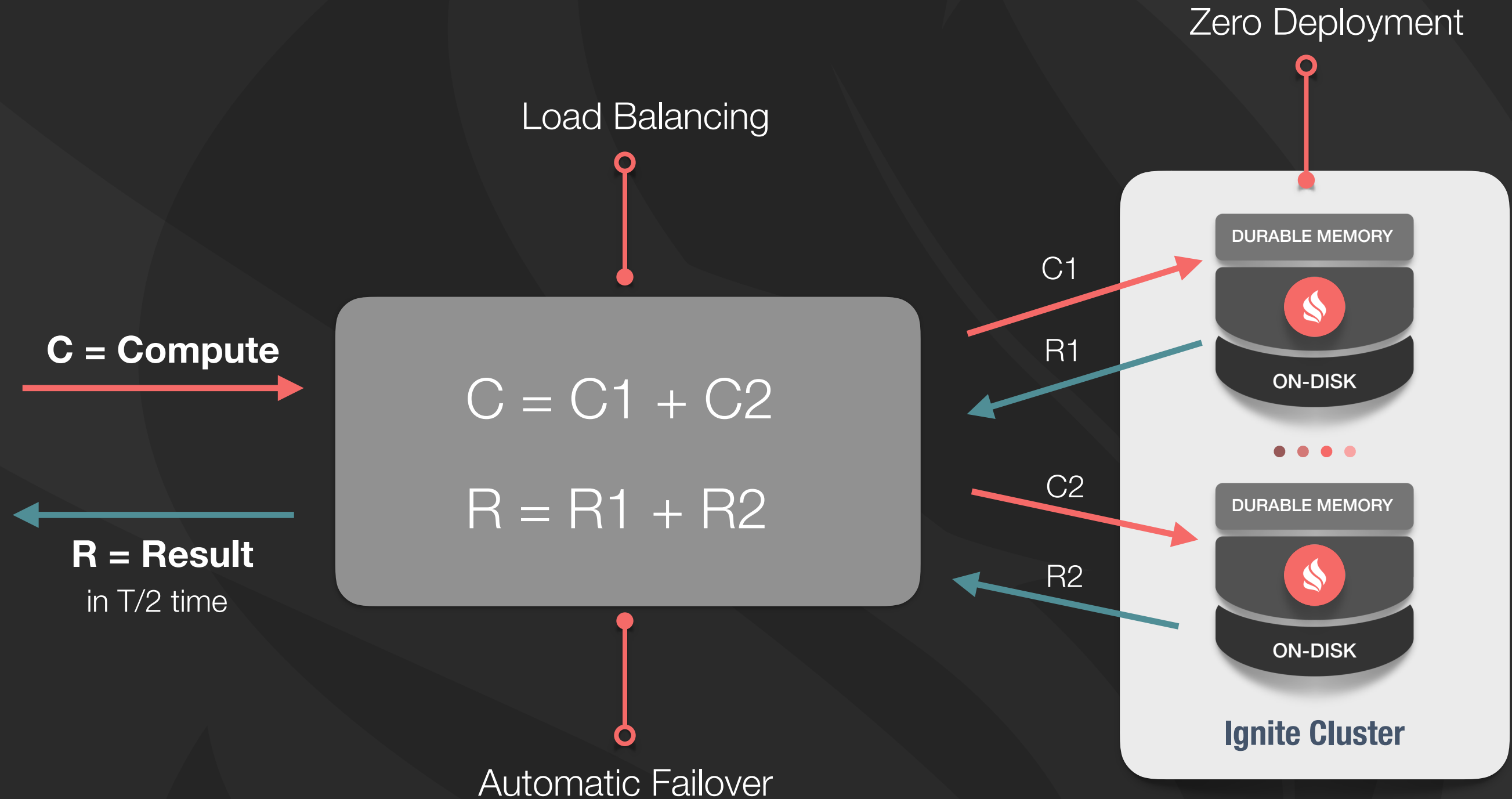
1. Initial Query
2. Query execution over local data
3. Reduce multiple results in one

# Non-Collocated Joins

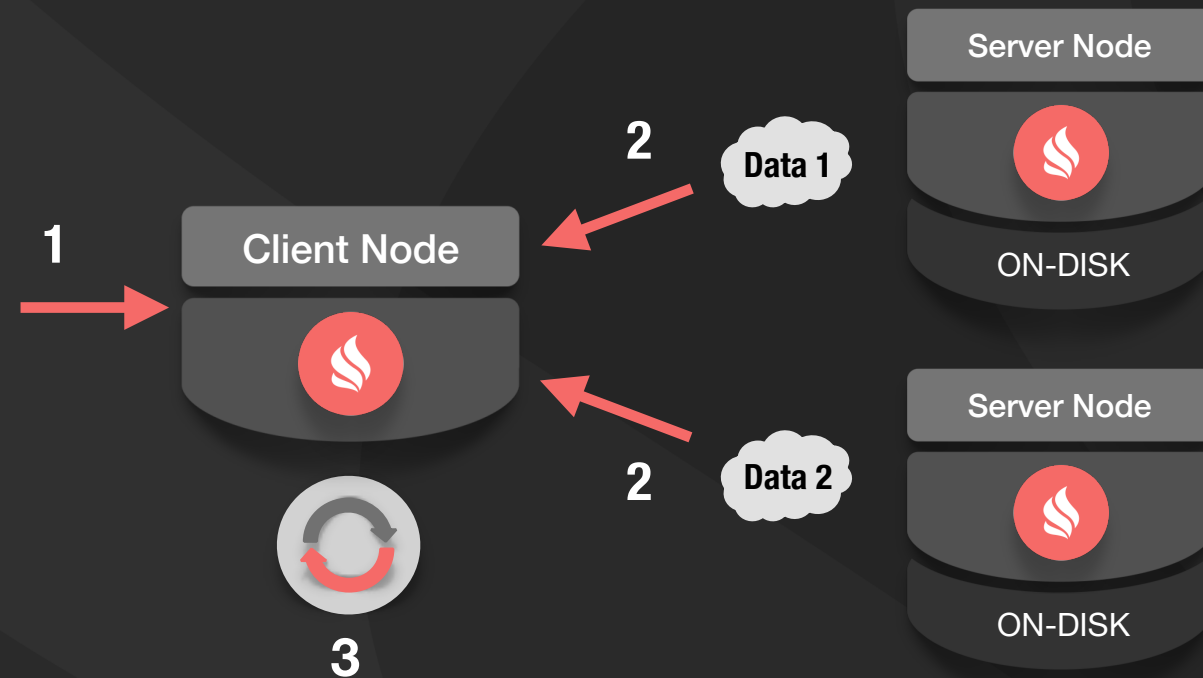


1. Initial Query
2. Query execution (local + remote data)
3. Potential data movement
4. Reduce multiple results in one

# Compute Grid

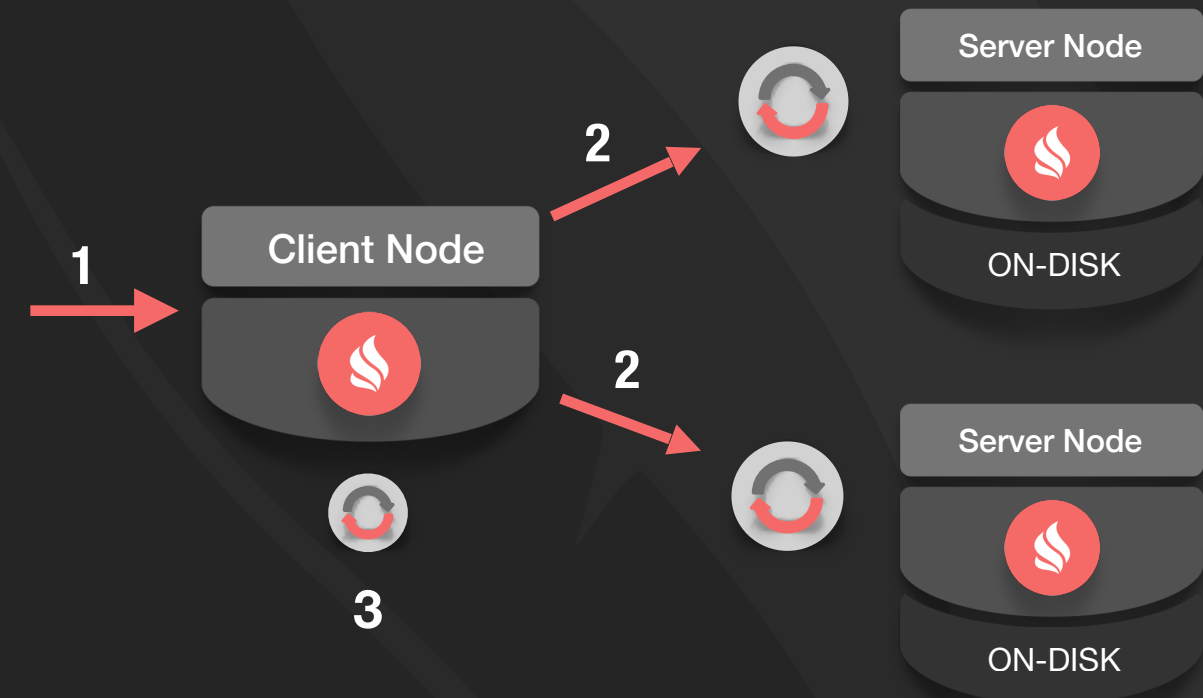


# Client-Server Processing



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

# Genetic Algorithms Grid

Biological Evolution  
Simulation

$$\begin{aligned} F &= F1 + F2 \\ C &= C1 + C2 \\ M &= M1 + M2 \end{aligned}$$

Collocated Computation

F = Fitness Calculation  
C = Crossover  
M = Mutation

F1, C1, M1

F2, C2, M2

Chromosome and Genes Cluster

DURABLE MEMORY



ON-DISK

...

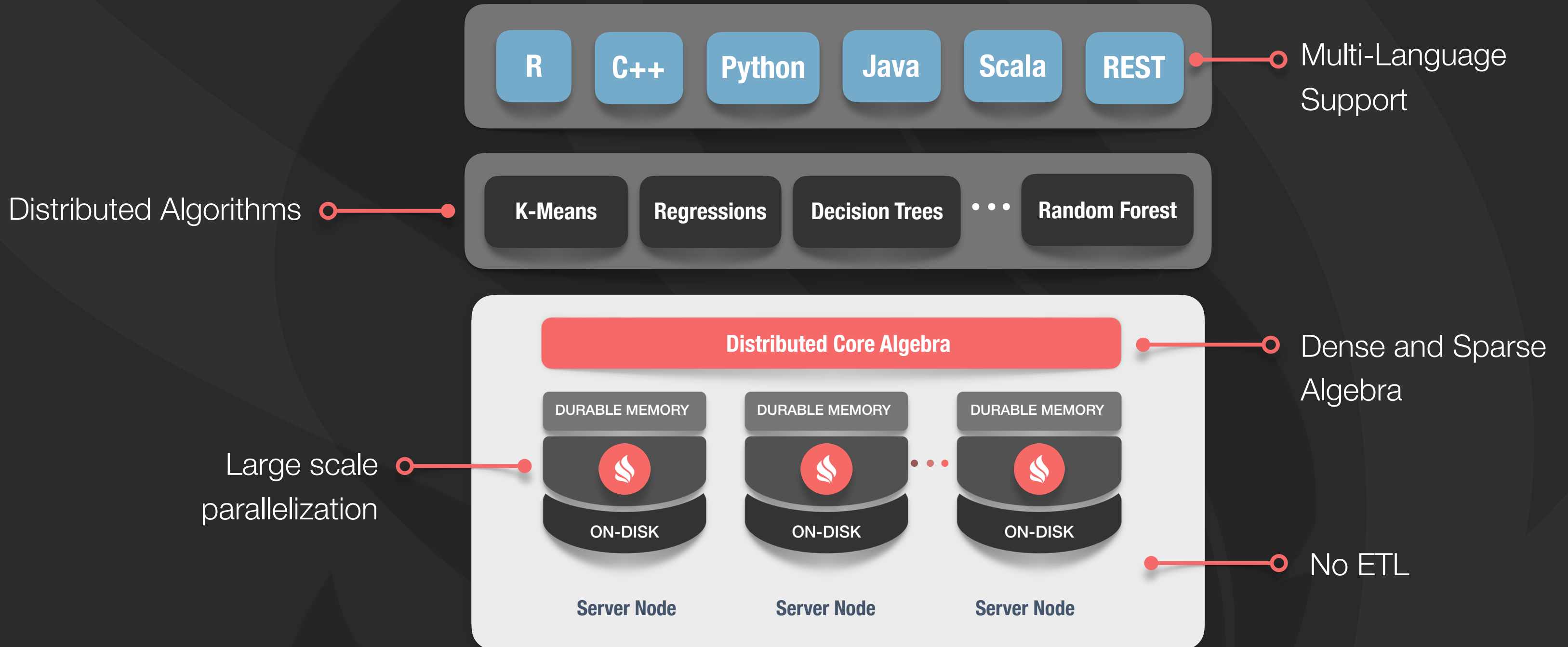
DURABLE MEMORY



ON-DISK

Ignite Cluster

# Machine Learning Grid





# Any Questions?

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



#apacheignite  
#denismagda