Data Grid

سید مجید عظیمی

@majidazimi

majid.azimi@live.com

پاییز ۱۳۹۲

Grid Computing

Distribute high-volume of computations to computers separated by large geographical distances.

Cluster Computing

Distribute high-volume of computations to computers inside data center.

Data Grid

- Hybrid Technology
- It is mainly Cluster computing technology
- Provides access pattern from different DCs

RAM

is becomming new hard disk

What is Data Grid

Data Grid is mainly a

Distributed Cache

Distributed in-memory RDBMS

- Sharding is hard to achieve
- Distributed Query Processing
  - Cross node join
  - Parrallel query processing

Distributed Key/Value DB

- Too simple data model
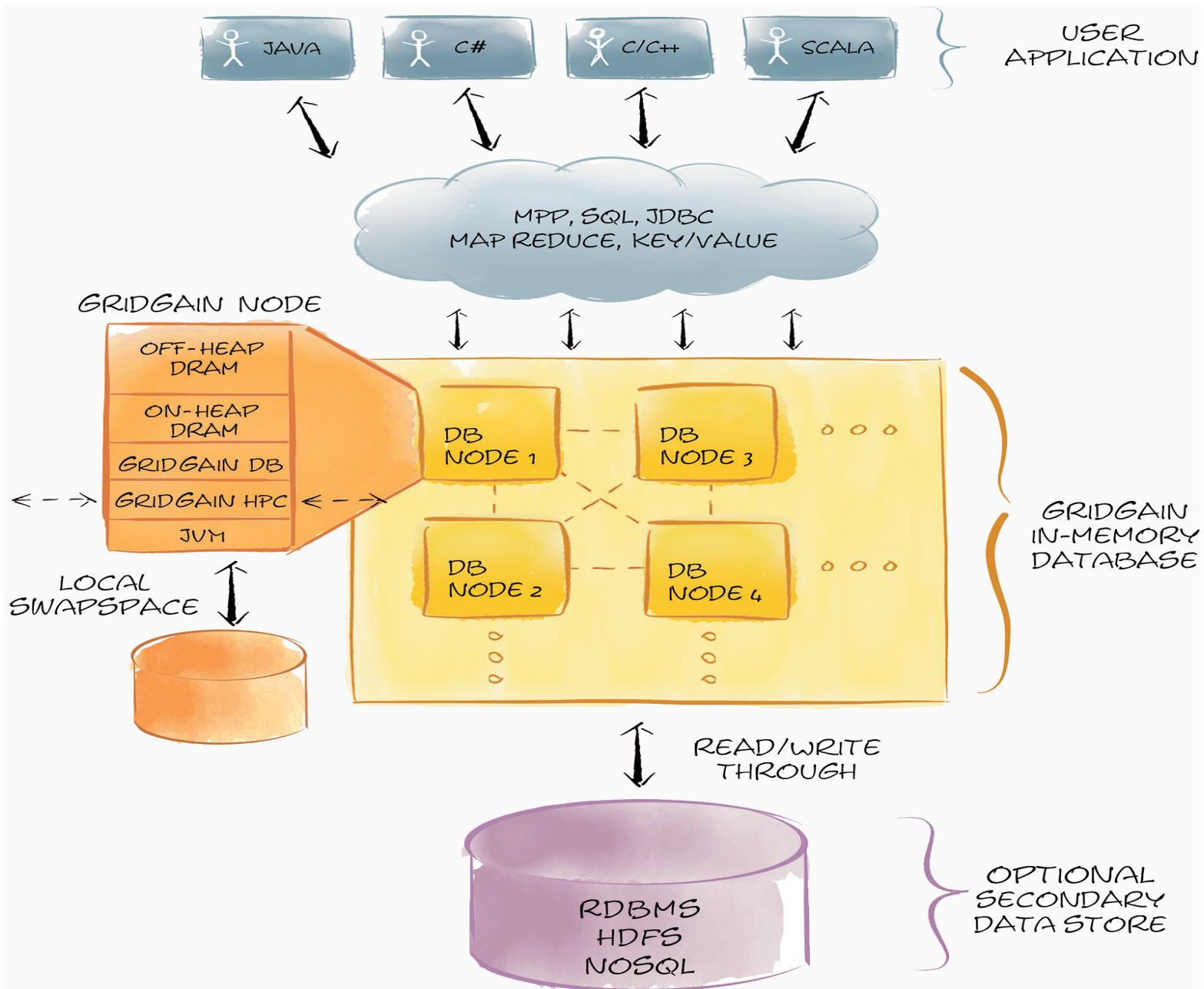- No Parrallel data processing
  - Map/Reduce

NoSQL Systems

- Good Horizontal scalability

- Fills gap between cache Key/Value and Relational DB

- Offers storage vision

  - Data grid offers processing vision

Data Grid: Concepts

- In-Memory Database

- In-Memory Streaming

- In-Memory HPC

In-Memory Database

In-Memory Database

- Local, Replicated, Partitioned
- MVCC Based Concurrency
- Off-Heap Memory
- Write Behind Cache
- Web Session Cashing
- Continuous Query
- ACID
- Pluggable Persistency
- REST Interface
- Distributed Data Structures
- SQL and Lucene Querying

In-Memory Streaming

In-Memory Streaming

- Event Pattern Detection

- Grouping, aggregation, sorting, filtering, merging, splitting or duplicating

  - Detection

  - Aggregation

In-Memory HPC
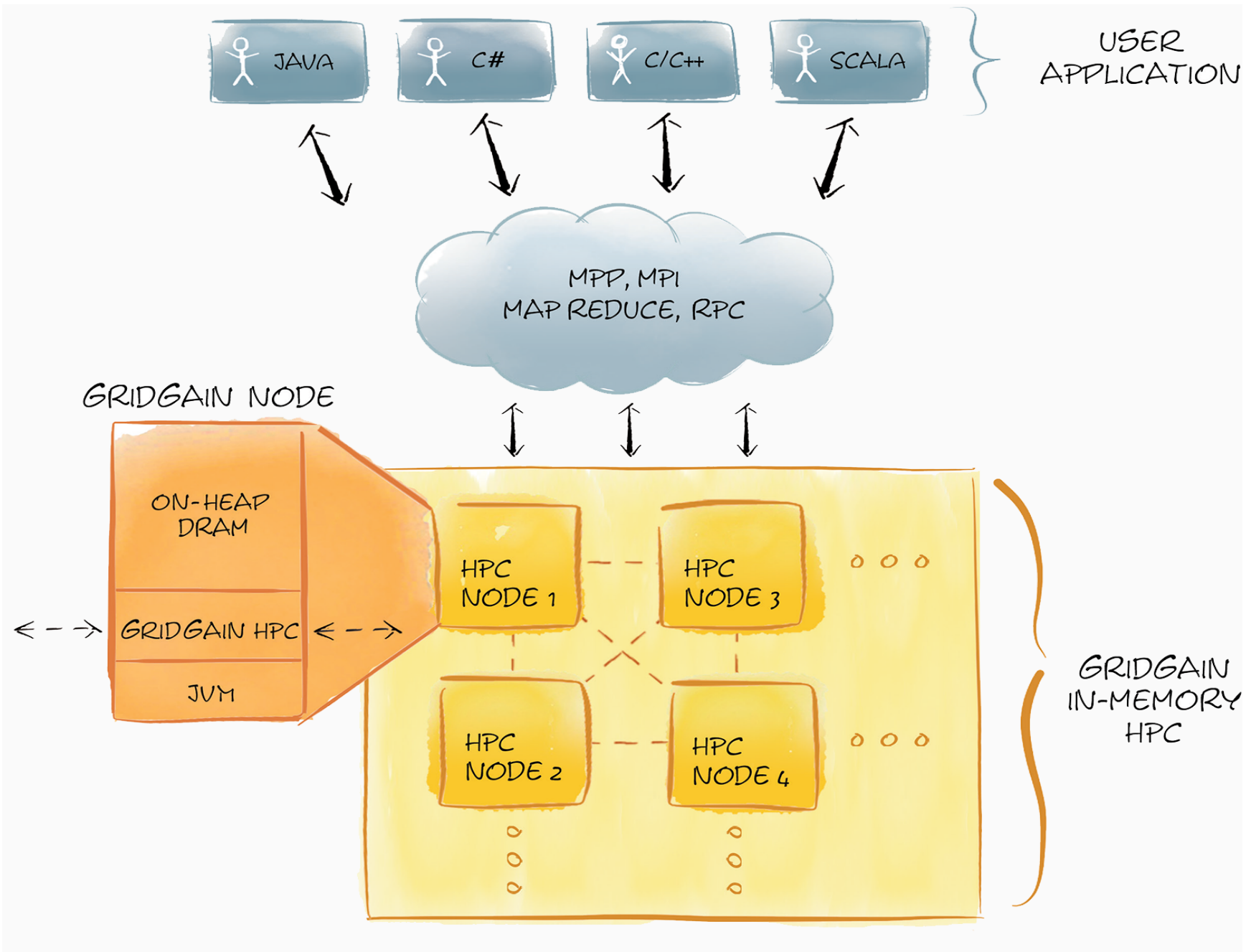
In-Memory HPC

- Map/Reduce
- RPC Processing
- MPI-Style Processing
- Cron-Based Scheduling
- Topology Resolution
- Job Checkpointing

Implementation

- GridGain

- Jboss Infinispan

- Hazelcast

- Oracle Coherence