apache Ignite

In-Memory Performance
Durability of Disk

Apache Ignite and Apache Spark

Where Fast Data Meets the IoT

Akmal Chaudhri
GridGain Systems

Agenda

- IoT Demands to Software
- IoT Software Stack
    - Device OS/RTOS
    - Data Collection and Enrichment
    - NewSQL Database
    - Application APIs
- Demo

IoT Demands to Software

Real-time Processing

Analytics (BI, ML)

Simple Scalability

SQL, Geo-Spatial

High-Availability

IoT Software Stack

Application APIs

NewSQL Database

Data Collection and Enrichment

Device OS/Real-Time OS

Apache IoT Software Stack

Application APIs — apache Ignite, Apache Spark

NewSQL Database — apache Ignite

Data Collection and Enrichment — Apache Spark, kafka, Apache Flink

Device OS/Real-Time OS — Apache mynewt

Apache MyNewt

Open Source RTOS
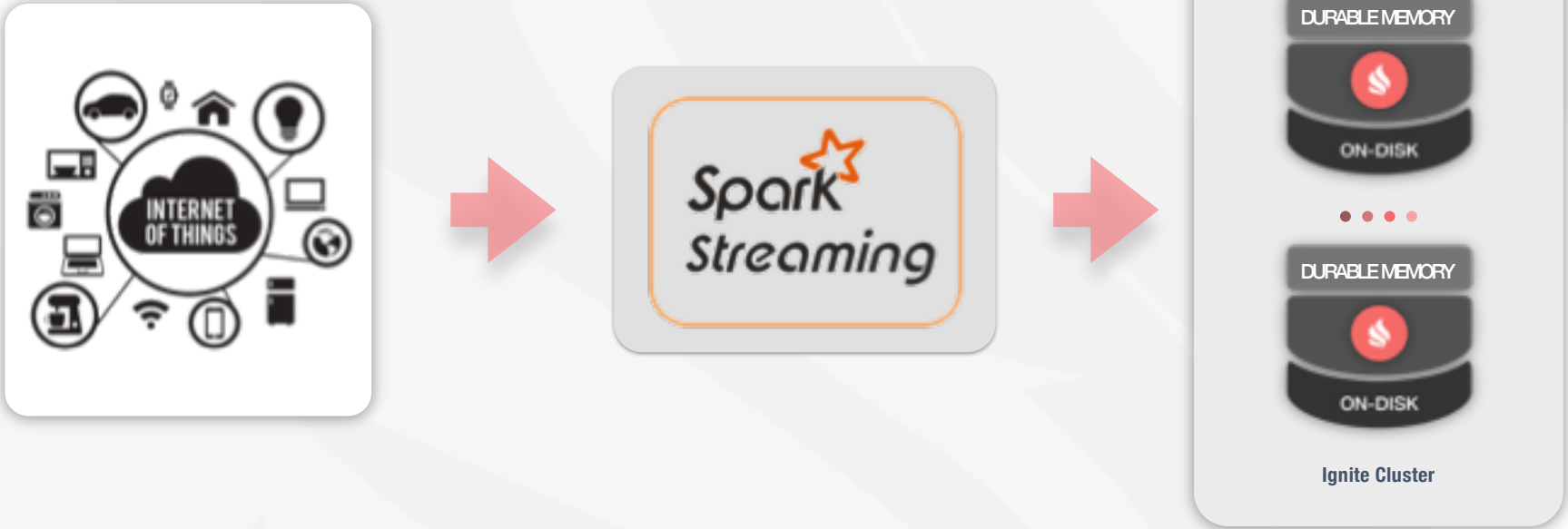Cortex M, MIPS

Secured Bootloader

Bluetooth, Wifi,
TCP/IP

Remote Firmware
Upgrade

Data Collection and Enrichment

Apache Ignite Database, Caching and Processing Platform

| Financial Services | Telco | Travel & Logistics | E-Commerce | Pharma & Healthcare | IoT |

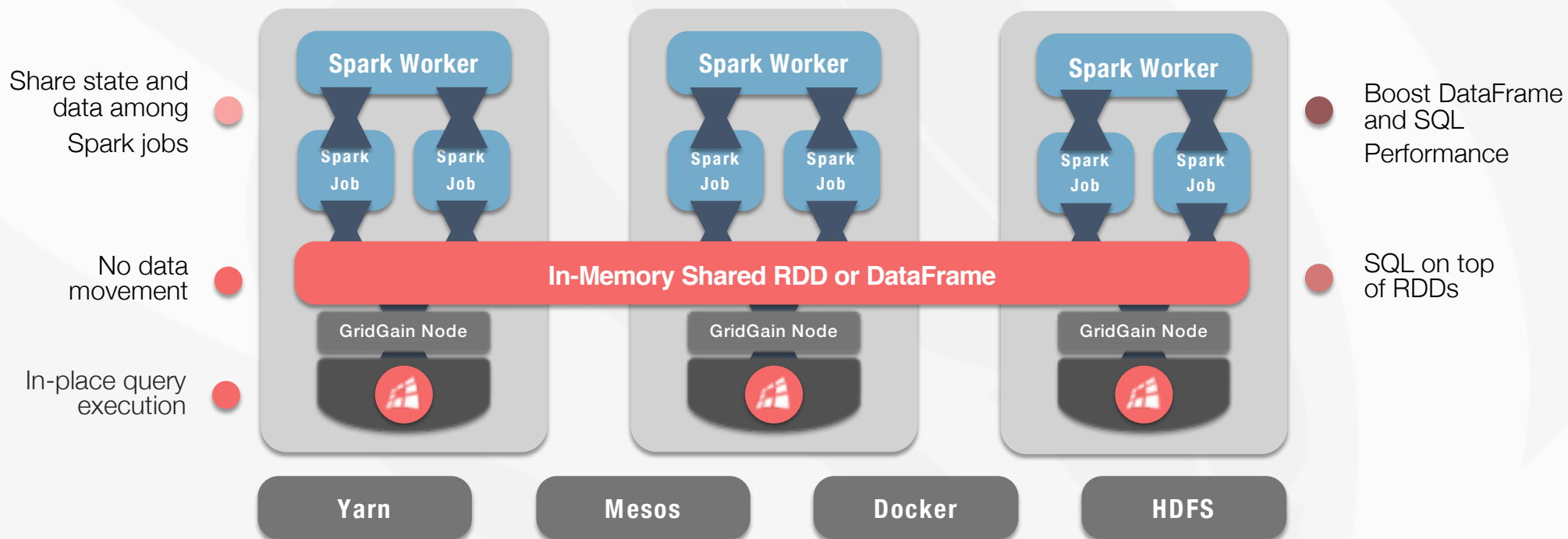| SQL | Key/Value | Transactions | Compute | Services | Streaming | ML |

Memory-Centric Storage

Ignite Native Persistence
(Flash, SSD, Intel 3D XPoint)

Third-Party Persistence
(RDBMS, HDFS, NoSQL)

Ignite and Spark Integration

Spark Application

Share state and data among Spark jobs

No data movement

In-place query execution

Spark Worker	Spark Worker	Spark Worker
Spark Job	Spark Job	Spark Job	Spark Job	Spark Job	Spark Job

In-Memory Shared RDD or DataFrame

GridGain Node   GridGain Node   GridGain Node

Boost DataFrame and SQL Performance

SQL on top of RDDs

Yarn	Mesos	Docker	HDFS

SQL Queries Execution Flow

1

```
SELECT ct.name, count(c.name)
FROM Country as ct
JOIN City as c ON c.countryCode = ct.code
WHERE ct.name IN ('Canada', 'India') GROUP BY (ct.name);
```

2

Ignite Node

Canada

Toronto

Montreal

Ottawa

Calgary

2

Ignite Node

India

Mumbai

New Delhi

3

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

Comparing Ignite and Spark

Apache Ignite	Apache Spark
• Distributed memory-centric **database**	• Ingests data from HDFS or another storage
• Fully fledged compute platform: SQL, transactions, key-value, collocated processing, ML/DL	• Streaming and compute engine
• OLAP and OLTP	• Inclined towards OLAP and focused on MR payloads

Ignite is a memory-centric store for Spark

- No data movement from Ignite to Spark
  - In-place query execution

- Boost DataFrame and SQL performance

- Share state and data among Spark jobs

- Faster data and streaming analytics

DEMO

Any Questions?

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

● ● ● ●

#apacheignite