Apache Ignite - In-Memory Data Fabric
Ultimate Speed and Scale for Transactions and Analytics

NIKITA IVANOV
Founder, PMC

www.ignite.incubator.apache.org  #apacheignite
Agenda

• Why In-Memory Computing?
• In-Memory Data Fabric
  • Advanced Clustering
  • In-Memory Compute Grid
  • In-Memory Data Grid
  • In-Memory Service Grid
  • In-Memory Streaming & CEP
  • Plug-n-Play Hadoop Accelerator
• Customer Use Cases
• Q & A
“In-memory computing will have a long term, disruptive impact by radically changing users’ expectations, application design principles, product architectures, and vendor strategies.”

Data Growth

- Amount of global digital information created & shared – from documents to pictures to tweets – grew 9x in five years to nearly 2 zetabytes* in 2011, per IDC.


KPCB

Less than 2 zetabytes in 2011, 8 in 2015

DRAM Cost, $

- Cost drops 30% every 12 months
In-Memory Data Fabric

Strategic Approach to IMC

- Supports Applications of various types and languages
- Open Source – Apache 2.0
- Simple Java APIs
- 1 JAR Dependency
- High Performance & Scale
- Automatic Fault Tolerance
- Management/Monitoring
- Runs on Commodity Hardware
- Supports existing & new data sources
- No need to rip & replace
In-Memory Data Fabric

Main Capabilities

**Performance**
- High Throughput
- Low Latencies

**Scalability**
- Add Cluster Members (cores)
- Add Memory (RAM)

**High Availability**
- Data Backups
- Datacenter Replication

**Transactions**
- Fully ACID Compliant
- Optimistic & Pessimistic

**Persistence**
- SQL, NoSQL, Hadoop

**Security**
- Authentication
- Authorization
- Tracing & Auditing
In-Memory Data Fabric

Advanced Clustering

• Ease of Getting Started
  – Automatic Discovery

• Any Environment
  – Public Cloud
  – Private Cloud
  – Hybrid Cloud
  – Local Laptop

• Zero-Deployment
  – Auto-Deploy Code

• Full Cluster Management

• Pluggable Design
In-Memory Compute Grid

- Direct API for MapReduce
- Zero Deployment
- Cron-like Task Scheduling
- State Checkpoints
- Load Balancing
- Automatic Failover
- Full Cluster Management
- Pluggable SPI Design
In-Memory Data Grid

- Distributed In-Memory Key-Value Store
- Replicated and Partitioned data
- TBs of data, of any type
- On-Heap and Off-Heap Storage
- Highly Available In-Memory Replicas
- Automatic Failover
- Distributed ACID Transactions
- SQL queries and JDBC driver
- Collocation of Compute and Data
In-Memory Service Grid

- Distribute Any Data Structure
  - Available Anywhere on the Grid
  - Automatic Remote Access via Proxies
- Controlled Deployment
  - Support for Cluster Singleton
  - Support for Node Singleton
  - Support for Custom Topology
  - Load Balanced
- Guaranteed Availability
  - Auto Redeployment in Case of Failures
In-Memory Streaming and CEP

- Streaming Data Never Ends
- Branching Pipelines
- Pluggable Routing
- Sliding Windows for CEP/Continuous Query
- Real Time Analysis
In-Memory Hadoop Accelerator

- Plug and Play installation
- 10x to 100x Acceleration
- In-Memory Native MapReduce
- In-Process Data Colocation
- GGFS In-Memory File System
- Pure In-Memory
- Read-Through from HDFS
- Write-Through to HDFS
- Sync and Async Persistence
In-Memory Hadoop Accelerator

- In-Memory Native Performance
- Zero Code Change
- Use existing MR code
- Use existing Hive queries
- No Name Node
- No Network Noise
- In-Process Data Colocation
- Eager Push Scheduling
Management & Monitoring

✴ Enterprise Edition Only
Customer Use Cases

> **Automated Trading Systems**
  Real time analysis of trading positions & market risk. High volume transactions, ultra low latencies.

> **Financial Services**
  Fraud Detection, Risk Analysis, Insurance rating and modeling.

> **Online & Mobile Advertising**
  Real time decisions, geo-targeting & retail traffic information.

> **Big Data Analytics**
  Customer 360 view, real-time analysis of KPIs, up-to-the-second operational BI.

> **Online Gaming**
  Real-time back-ends for mobile and massively parallel games.

> **SaaS Platforms & Apps**
  High performance next-generation architectures for Software as a Service Application vendors.
Use Case: SBERBANK

Largest bank in Eastern Europe, and the third largest in Europe

- Open tender won by GridGain
  - Goal: Real-time risk and leverage reporting on their global financial trading portfolio
  - Performed a detailed evaluation and software assurance test
  - Delivered best performance, scale and high availability

1 Billion Transactions per Second

10 Dell R610 servers < $25K
1 TB Memory
ANY QUESTIONS?

www.ignite.incubator.apache.org  #apacheignite