GridGain In-Memory Computing Platform:
Empowering Data Analytics With In-Memory Computing

Eric Karpman
Independent Consultant
E-mail: emkarpman@gmail.com
25 years in Finance

Matt Sarrel
Director of Technical Marketing, GridGain
E-mail: matt.sarrel@gridgain.com
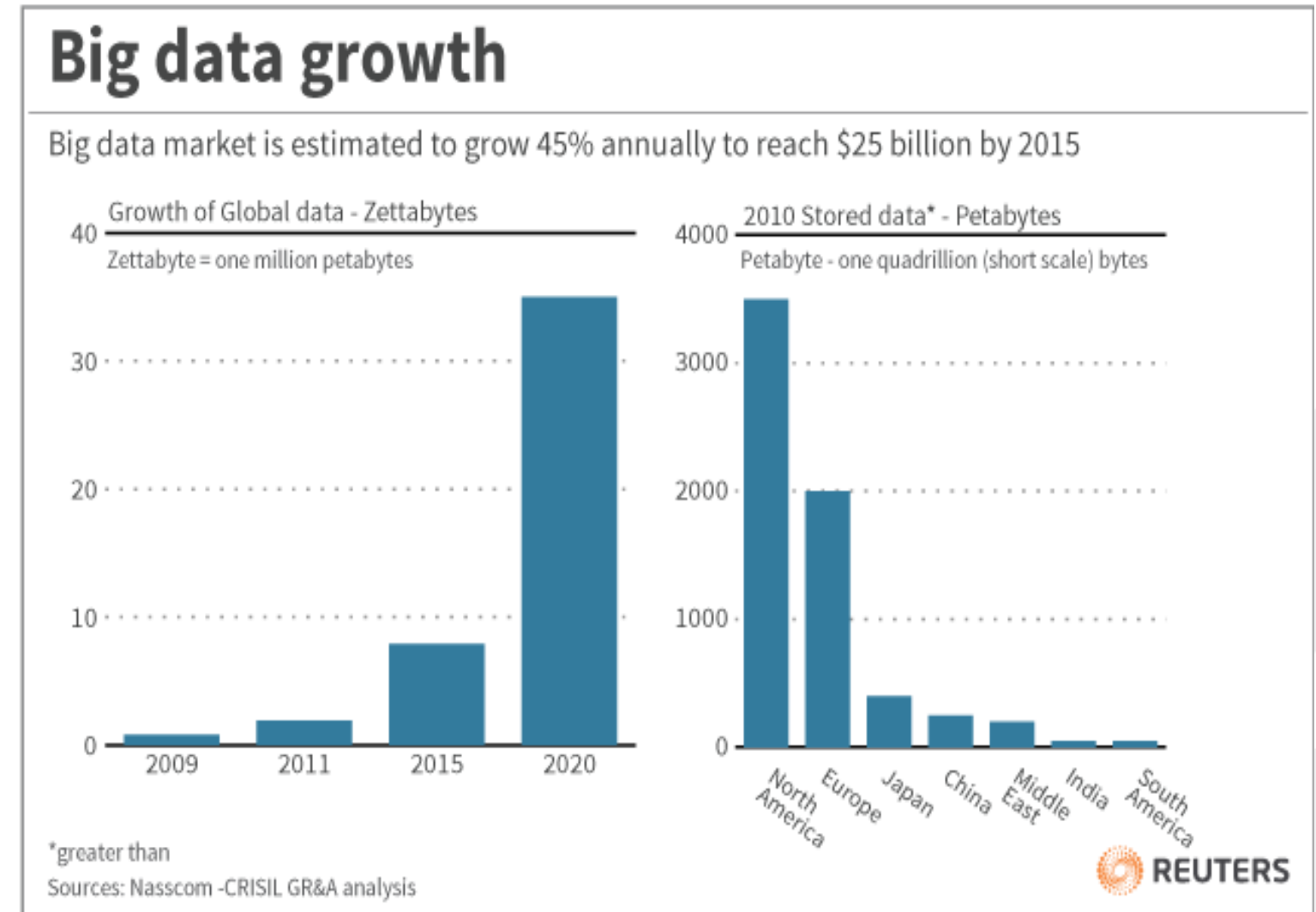30 years in Tech

www.gridgain.com   #gridgain

Data Is Growing

- Continuous flow of data
- Real-time, 24/7 streaming updates
- More than 2.5 quintillion bytes of data added daily
- Data is always available
- Democratization of data
- Main source for business decisions
- Establishing a Chief Data Officer
- Innovation labs
- Regulatory requirements in data capture
- Shift to digital and STP
- Affordable technology
- Innovative services
- Better and faster analytics
- Business Intelligence
- Cloud and subscription-based computing
- Outsourcing

Big data growth

Big data market is estimated to grow 45% annually to reach $25 billion by 2015

Growth of Global data - Zettabytes

Zettabyte = one million petabytes

2010 Stored data* - Petabytes

Petabyte - one quadrillion (short scale) bytes

*greater than
Sources: Nasscom -CRISIL GR&A analysis

REUTERS

Reuters graphic/Catherine Trevethan   05/10/12

GridGain

Data Types

Structured Data

- Transactional data
- Market data
- Reference data
- Pricing
- Risk
- Databases

Unstructured Data

- E-mails
- Documents
- Spreadsheets
- Images
- Telephone conversations
- Social media
- News

Big Data
Volume, Variety, Velocity

GridGain

Challenges

- Noise: separate irrelevant data from useful data
- Data discovery
- Data quality
- Data governance
- Disruptive technologies
- Reputational risk
- Cyber risk
- Regulations
- Performance
- Limitations of legacy technology
- Security concerns

GridGain

Data Use Cases

- Making better investment decisions with consistent results
- Algorithmic trading: real time + historical data
- Post trade analytics (TCA)
- Risk management
- Regulatory compliance
- Fraud detection
- Master Data Management
- Sentiment analysis
- Improving customer intelligence

GridGain

Evolving Regulations

- New reporting regulations
- Model testing and real-time simulations
- Data privacy
- Data breach notification
- Accountability framework
- Fraud prevention
- Additional data analytics requirements
- Data and cyber security

GridGain

Technology Trends

- Open source
- Move from data silos to "data lakes"
- Enterprise Data Warehouse
- Software as a critical component of business
- Businesses make data-driven decisions
- Open architecture
- Open APIs
- Artificial Intelligence
- Predictive analytics
- Machine learning
- Performance, stability, security, scalability
- Distributed systems
- Containers
- Microservice based architectures
- Machine learning
- Complex event processing
- Decision Support Systems
- Hadoop, HDFS and Map Reduce

GridGain

Why In-Memory Now?

Digital Transformation is Driving Companies Closer to Their Customers

- Driving a need for real-time interactions

## Internet Traffic, Data, and Connected Devices Continue to Grow

- Web-scale applications and massive datasets require in-memory computing to scale out and speed up to keep pace
- The Internet of Things generates huge amounts of data which require real-time analysis for real world uses

The Cost of RAM Continues to Fall

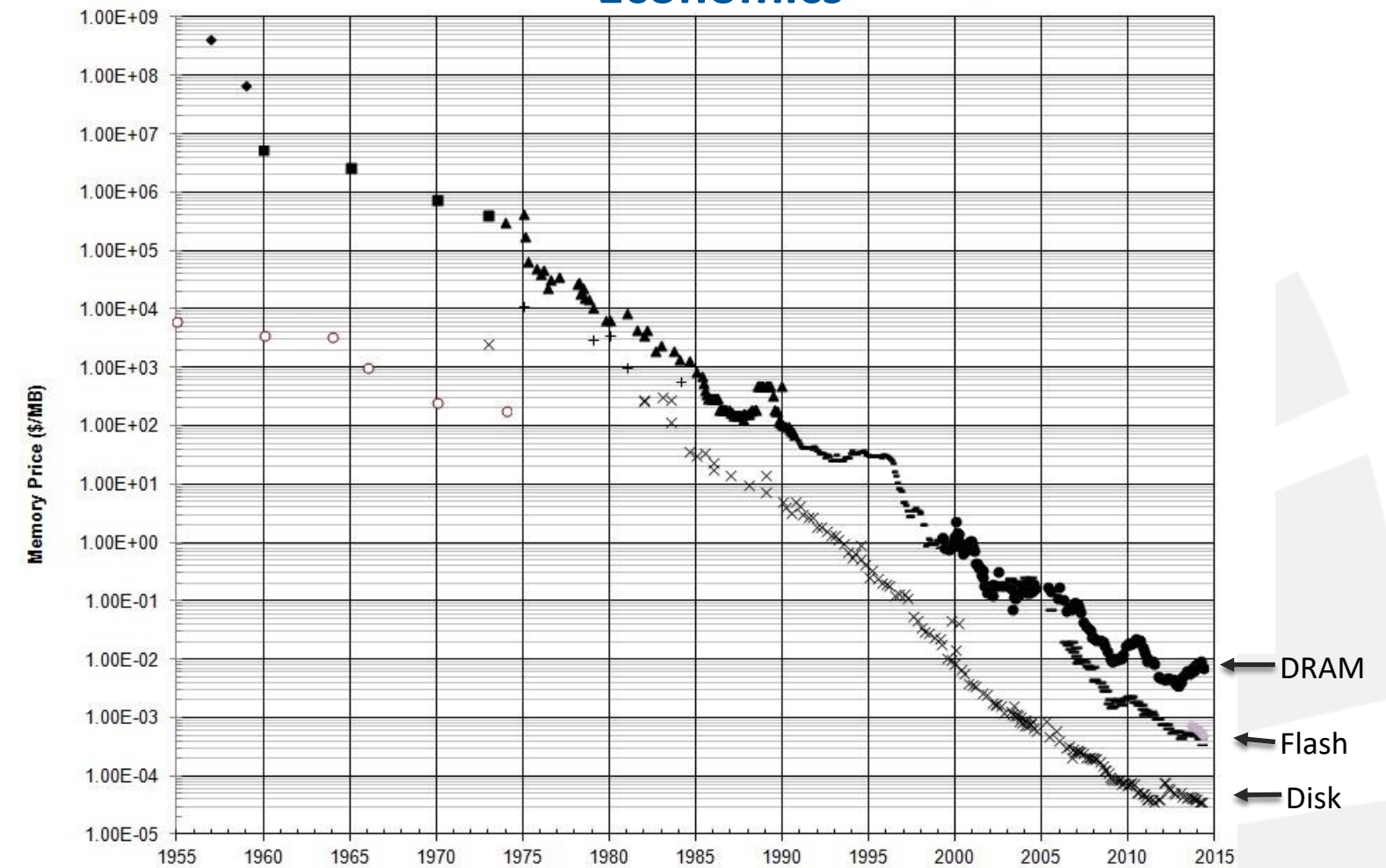- In-memory solutions are increasingly cost effective versus disk-based storage for many use cases

GridGain

Why Now?

Data Growth and Internet Scale Driving Demand

Growth of Global Data
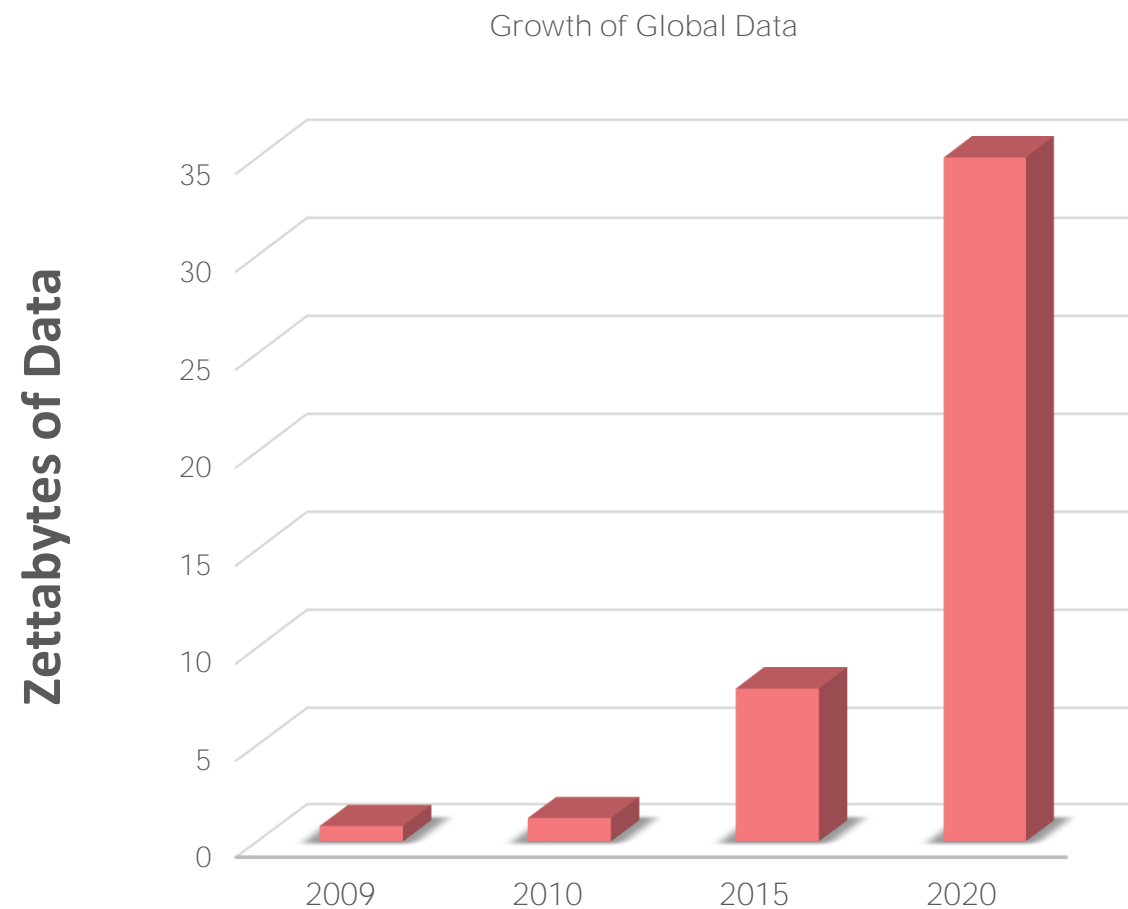
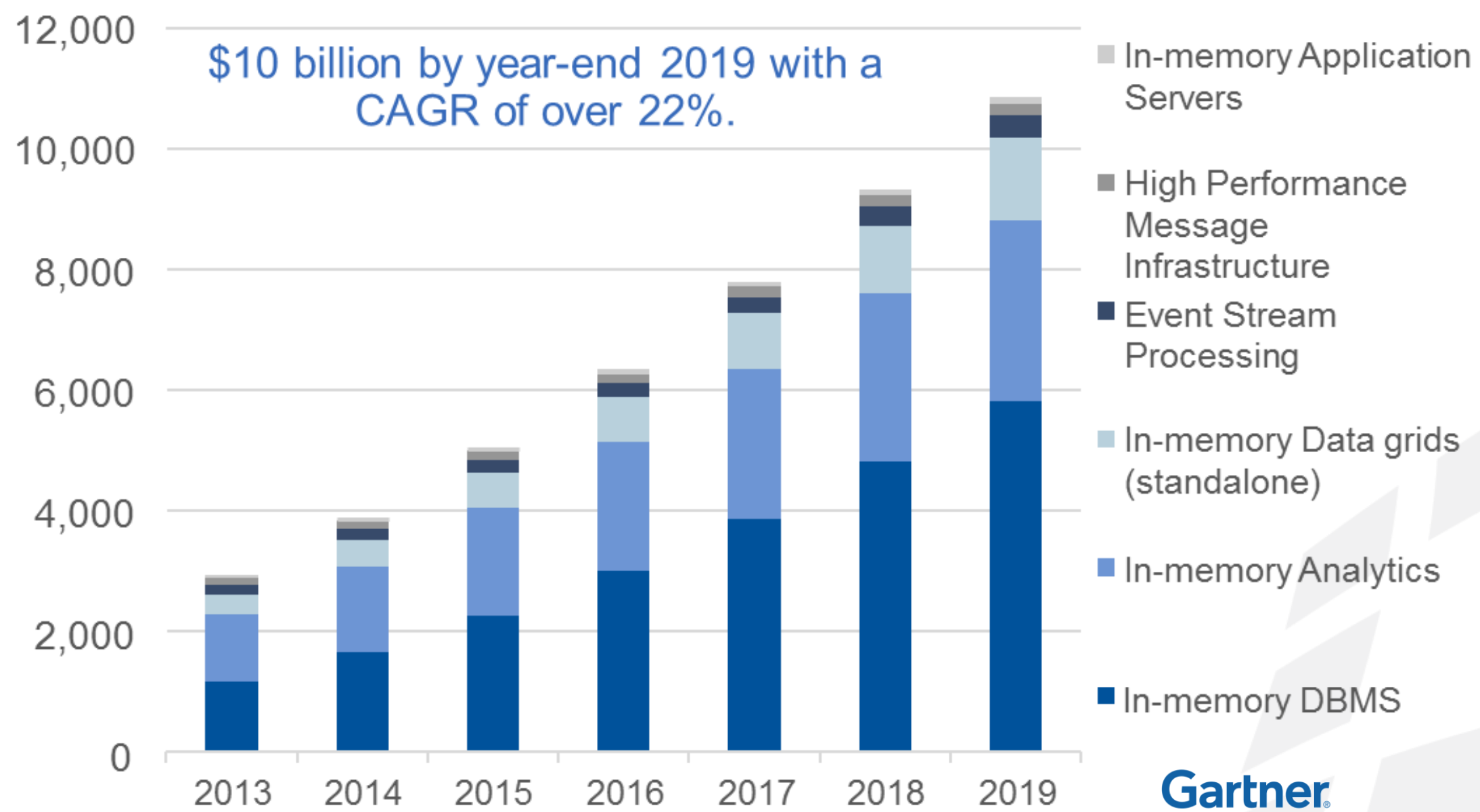8 zettabytes in 2015 growing to 35 in 2020

Declining DRAM Cost Driving Attractive Economics

Cost drops 30% every 12 months

GridGain

The In-Memory Computing Technology Market Is Big — And Growing Rapidly

IMC-Enabling Application Infrastructure ($M)

$10 billion by year-end 2019 with a CAGR of over 22%.

Legend:
- In-memory Application Servers
- High Performance Message Infrastructure
- Event Stream Processing
- In-memory Data grids (standalone)
- In-memory Analytics
- In-memory DBMS

Gartner

GridGain

Evolution of In-Memory Grid Computing

- Move from Disk to 100% In-Memory (RAM)
- Leverage Clustered Memory and Parallel Distributed Processing
- Results: 1000x Faster, 10x ROI Improvement
- Making "Big Data" Fast

"In-memory will have an industry impact comparable to web and cloud."

"RAM is the new disk, and disk is the new tape."

Gartner.

In-Memory Computing Market:
- $10B in 2019
- CAGR 22%

Gartner.

GridGain

What is an In-Memory Computing Platform?

Multi-Featured Solution

- Supports data caching, massive parallel processing, in-memory SQL, streaming and much more

Does Not Replace Existing Databases

- Slides in between the existing application and data layers

Supports OLTP and OLAP Use Cases

- Offers ACID compliant transactions as well as analytics support

Multi-Platform Integration

- Works with all popular RDBMS, NoSQL and Hadoop databases and offers a Unified API with support for a wide range of languages

Deployable Anywhere

- Can be deployed on premise, in the cloud, or in hybrid environments

GridGain

The GridGain In-Memory Computing Platform

- A high-performance, distributed, in-memory platform for computing and transacting on large-scale data sets in real-time
- Built on Apache® Ignite™

Features

- Data Grid
- Compute Grid
- SQL Grid
- Streaming
- Service Grid
- Hadoop Acceleration

Architecture

- Advanced Clustering
- In-Memory File System
- Messaging
- Events
- Data Structures

Apache Ignite Project

- 2007: First version of GridGain

- Oct. 2014: GridGain contributes Ignite to ASF

- Aug. 2015: Ignite is the second fastest project to | graduate after Spark

- Today:

  - 60+ contributors and rapidly growing

  - Huge development momentum - Estimated 192 years of effort since the first commit in February, 2014 [Openhub]

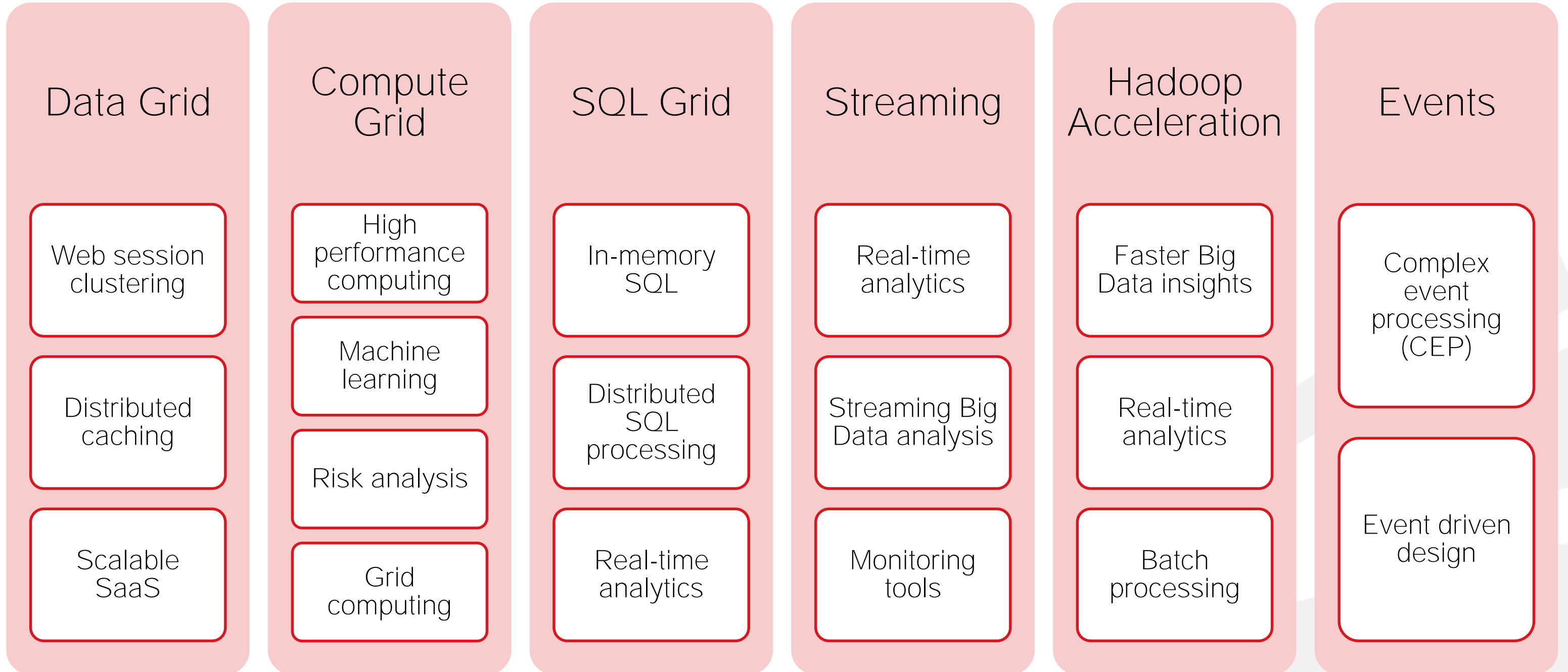 - Mature codebase: 1M+ lines of code
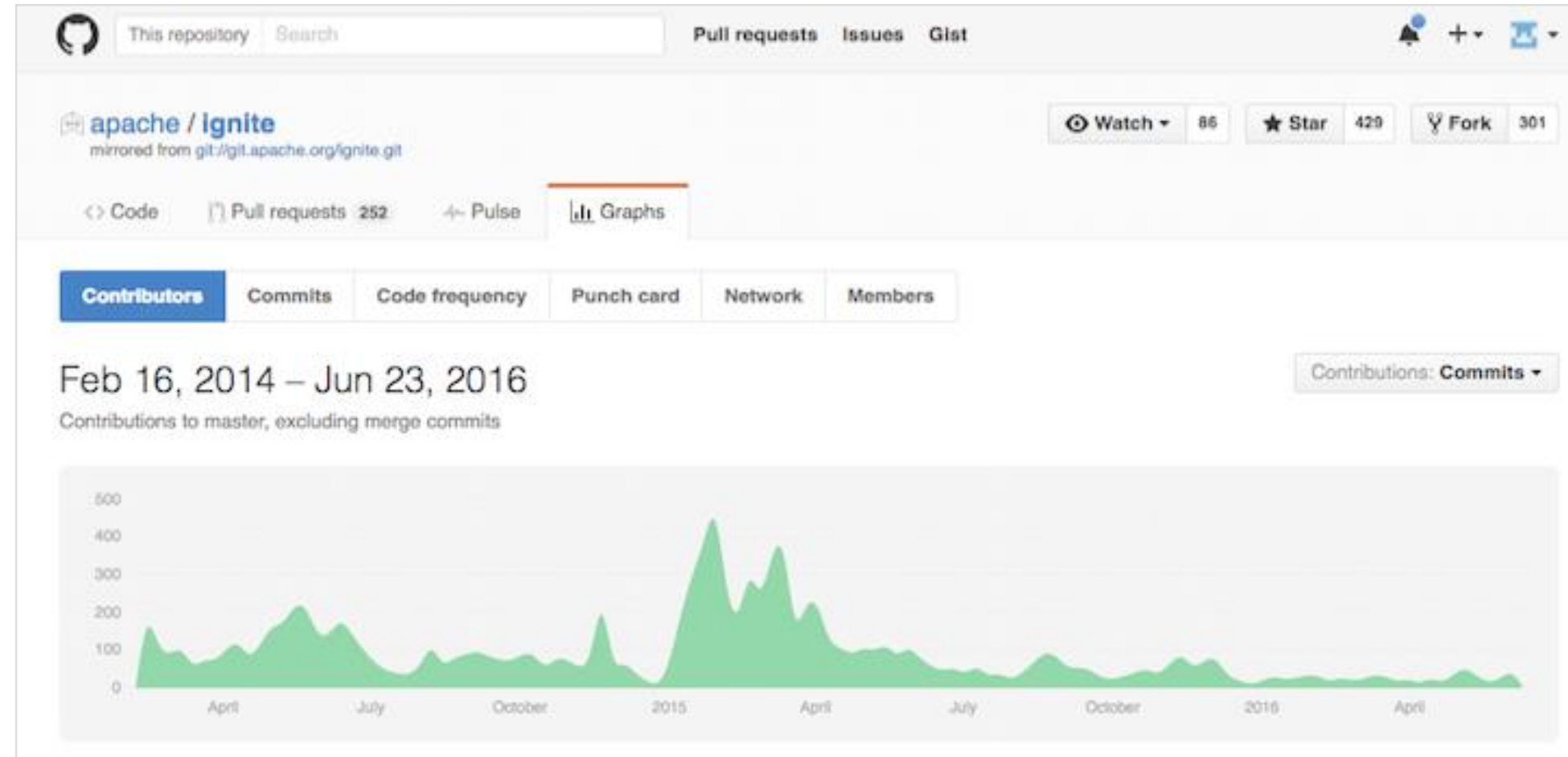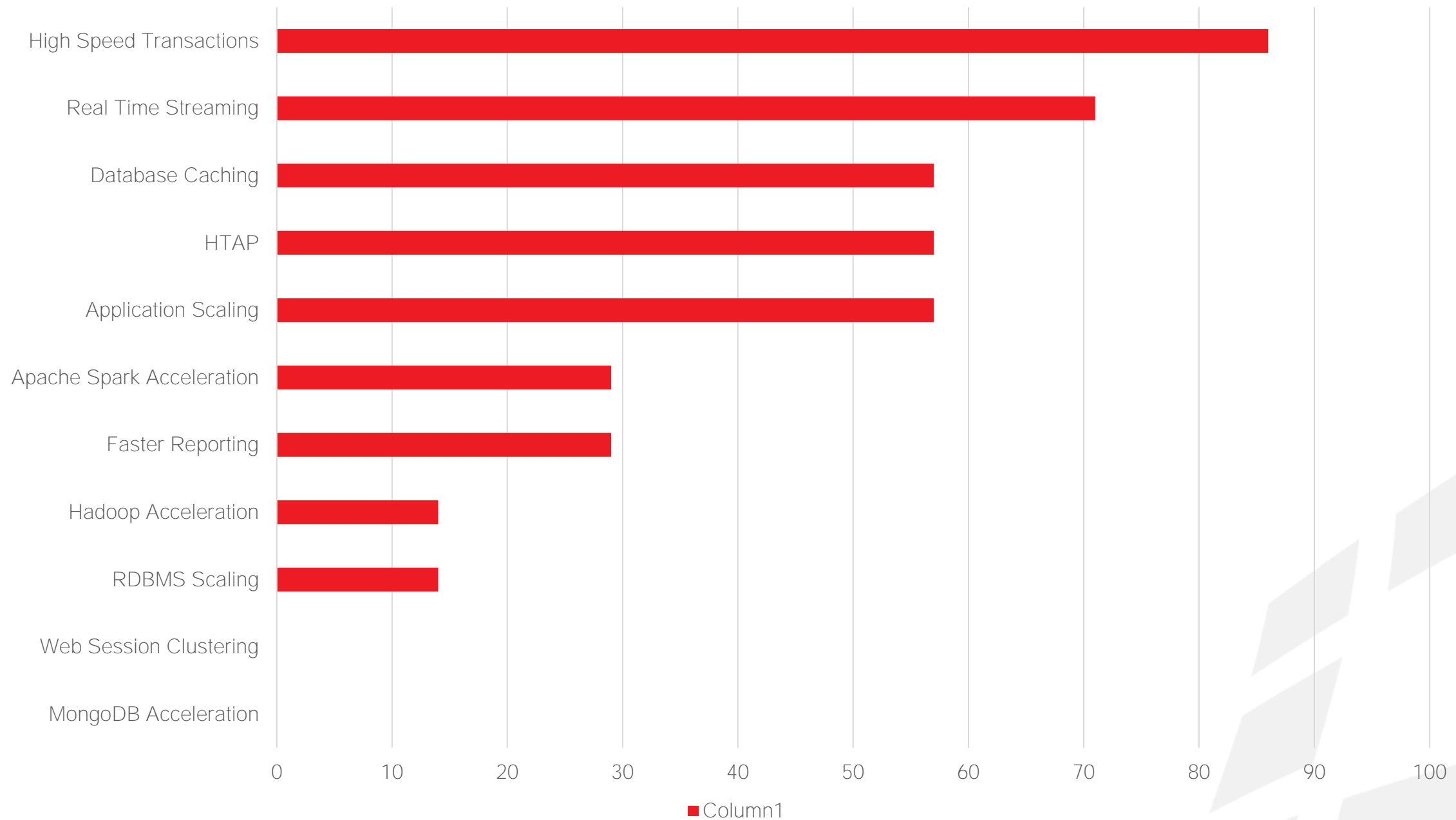
What is an In-Memory Computing Platform?
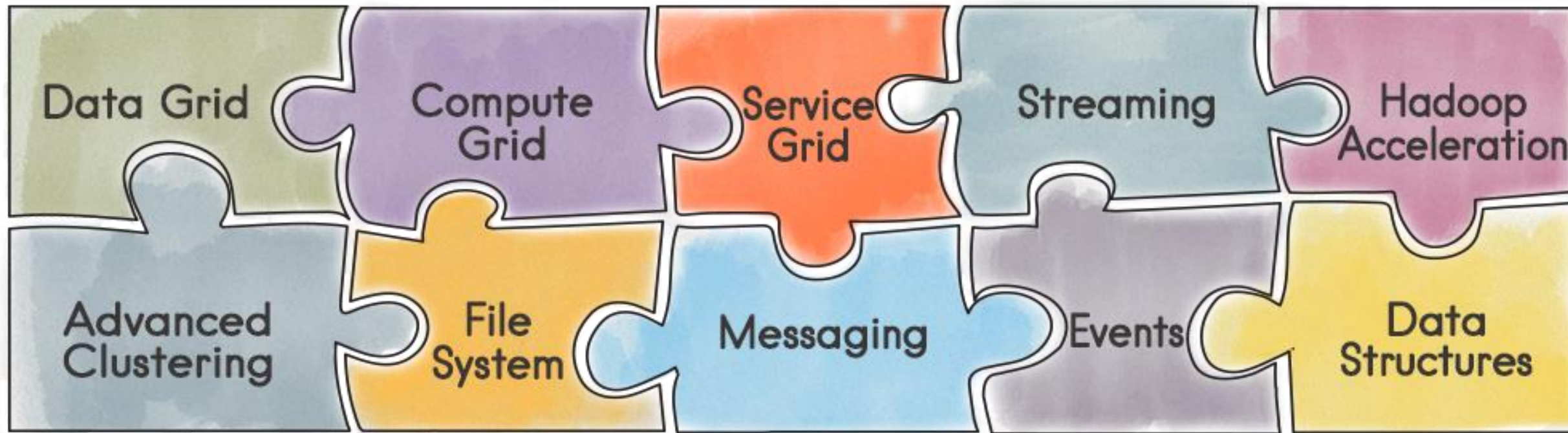
High-performance distributed in-memory platform for computing and transacting on large-scale data sets in near real-time.

GridGain

GridGain In-Memory Computing Use Cases

Data Grid
- Web session clustering
- Distributed caching
- Scalable SaaS

Compute Grid
- High performance computing
- Machine learning
- Risk analysis
- Grid computing

SQL Grid
- In-memory SQL
- Distributed SQL processing
- Real-time analytics

Streaming
- Real-time analytics
- Streaming Big Data analysis
- Monitoring tools

Hadoop Acceleration
- Faster Big Data insights
- Real-time analytics
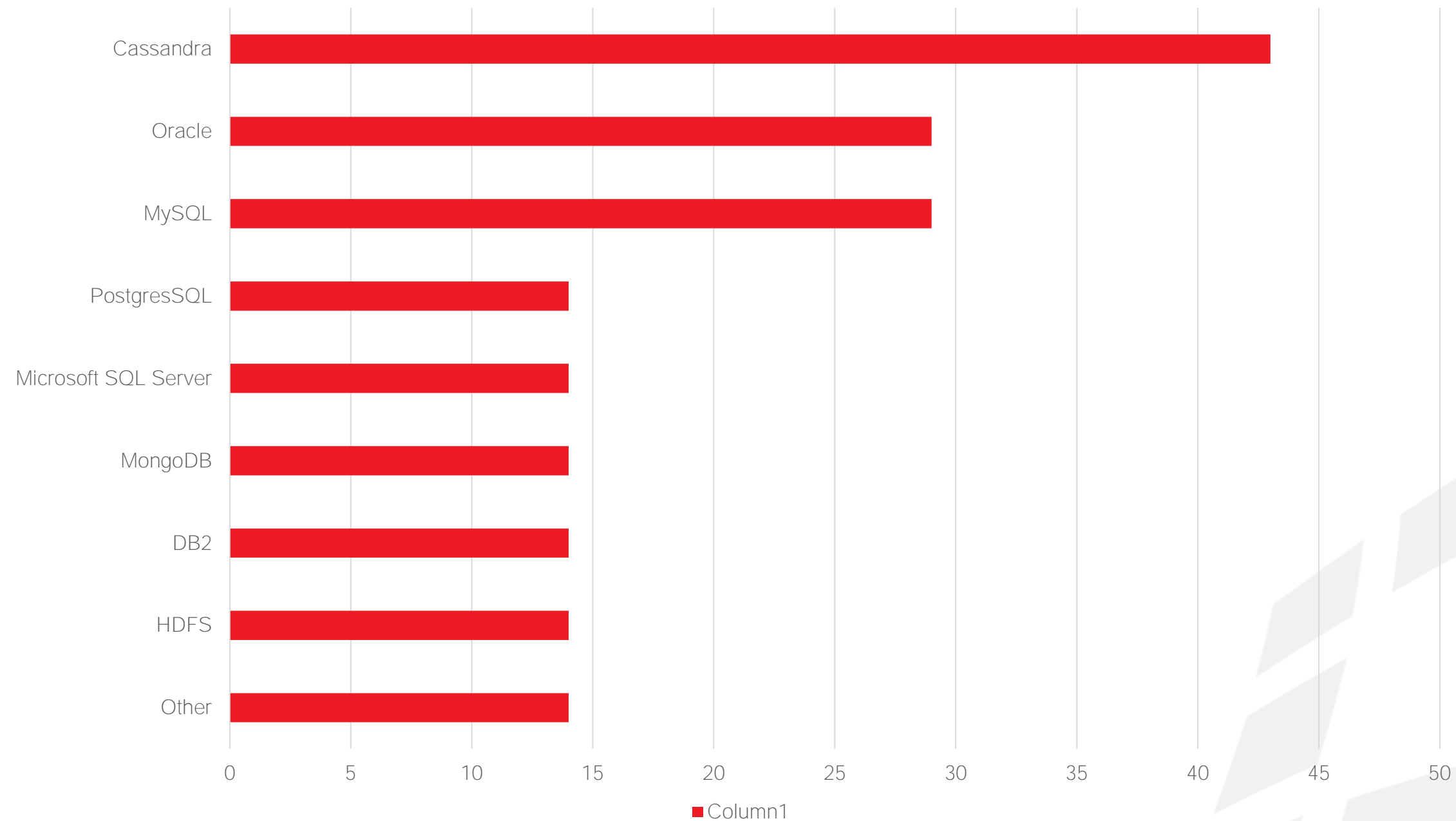- Batch processing

Events
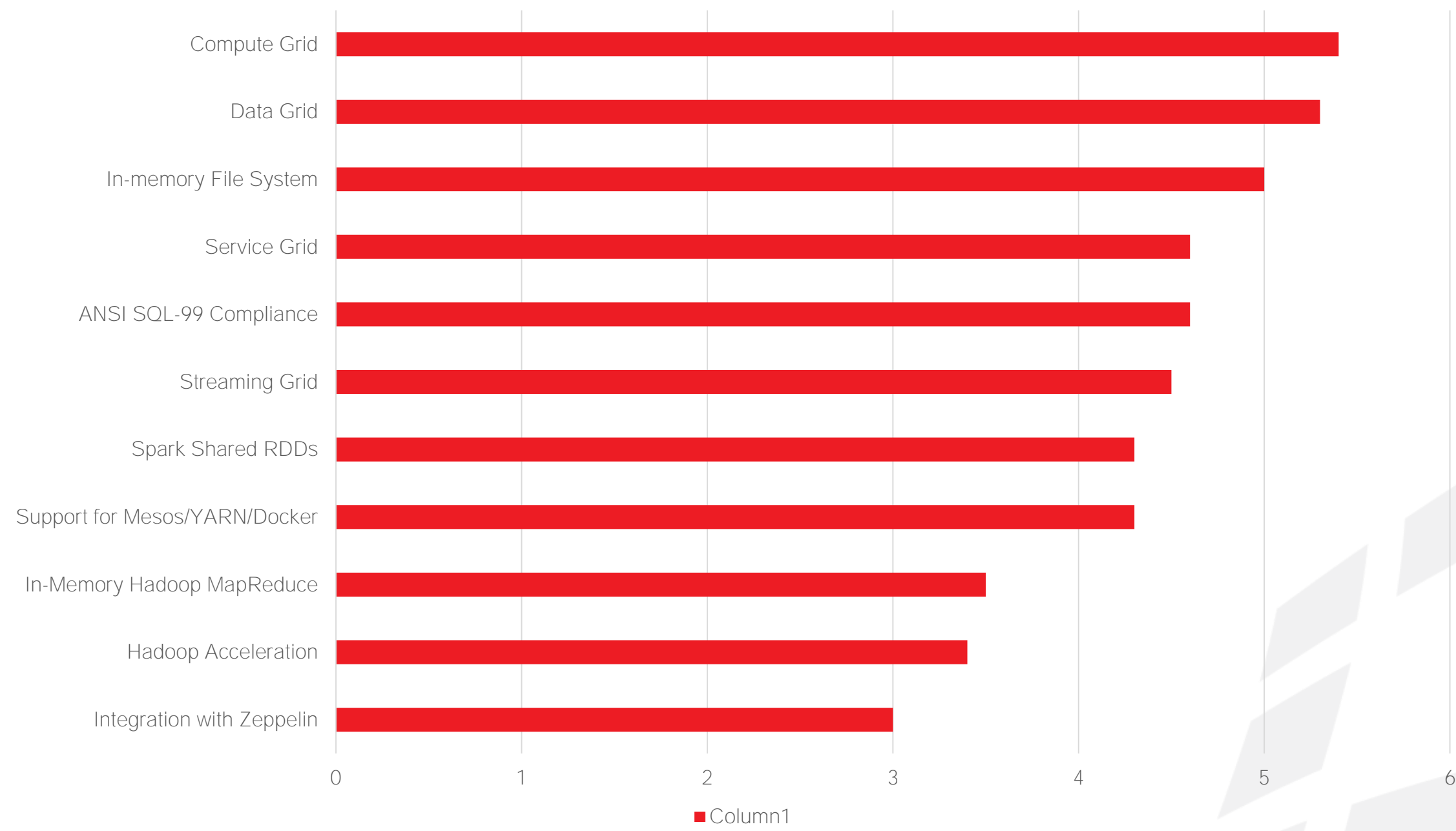- Complex event processing (CEP)
- Event driven design

GridGain

Survey Results: What uses were you considering for in-memory computing

Survey Results: Which data stores are you/would you likely use with GridGain/Apache Ignite?

Horizontal bar chart showing survey results. Legend: Column1.

Data store	Value
Cassandra	~43
Oracle	~29
MySQL	~29
PostgresSQL	~14
Microsoft SQL Server	~14
MongoDB	~14
DB2	~14
HDFS	~14
Other	~14

X-axis: 0, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50

GridGain

Survey Results: How important are each of the following product features to your organization?

Horizontal bar chart (Column1) showing importance ratings:

Feature	Rating
Compute Grid	~5.4
Data Grid	~5.3
In-memory File System	~5.0
Service Grid	~4.6
ANSI SQL-99 Compliance	~4.6
Streaming Grid	~4.5
Spark Shared RDDs	~4.3
Support for Mesos/YARN/Docker	~4.3
In-Memory Hadoop MapReduce	~3.5
Hadoop Acceleration	~3.4
Integration with Zeppelin	~3.0

Column1

GridGain

Survey Results: Where do you run GridGain and/or Apache Ignite?
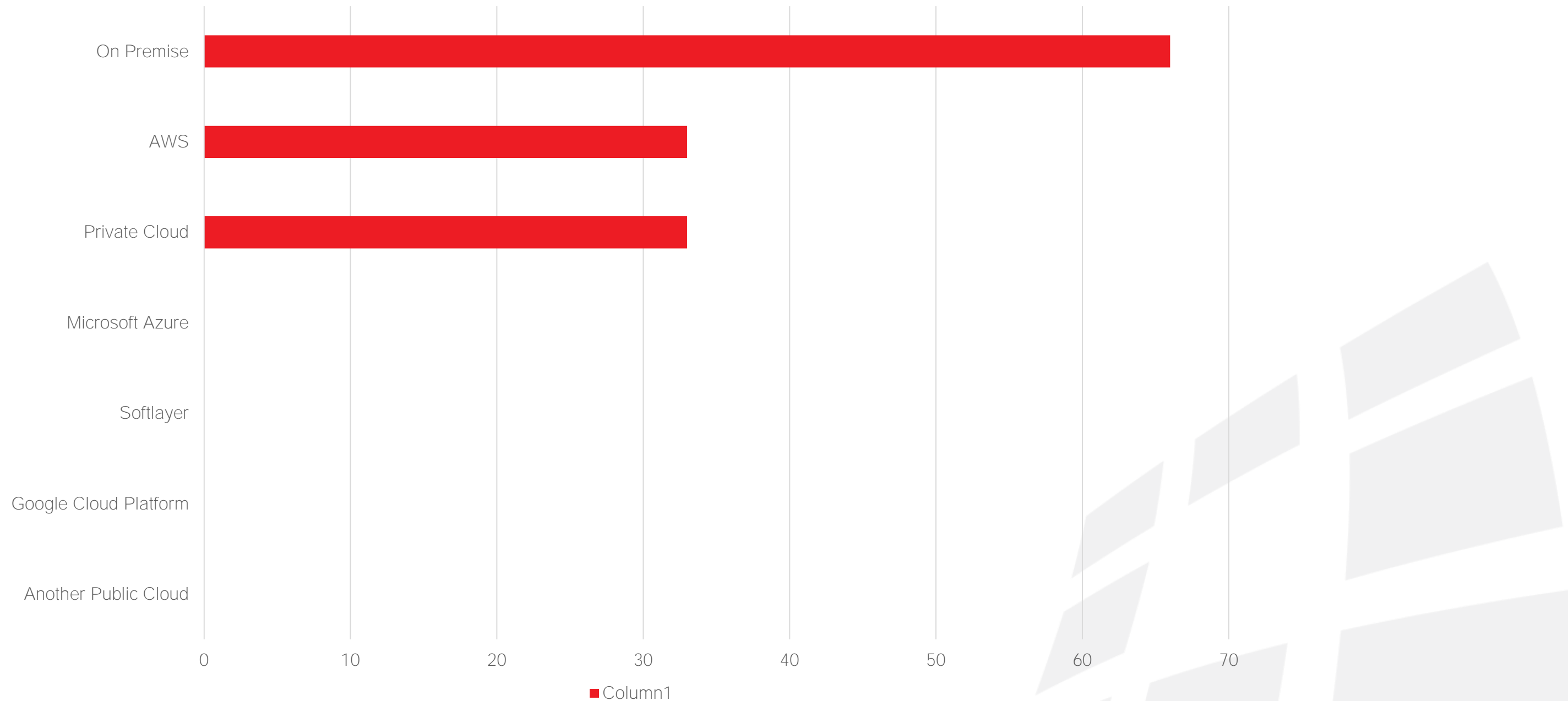
Survey Results: Which of the following languages do you use to access your data?

Horizontal bar chart showing survey results:
- SQL: 100
- Java: 86
- C++: 57
- Scala: 29
- Groovy: 29
- Node.js: 29
- .NET: 14
- PHP: 14
- MapReduce: 0

Legend: ■ Column1

GridGain
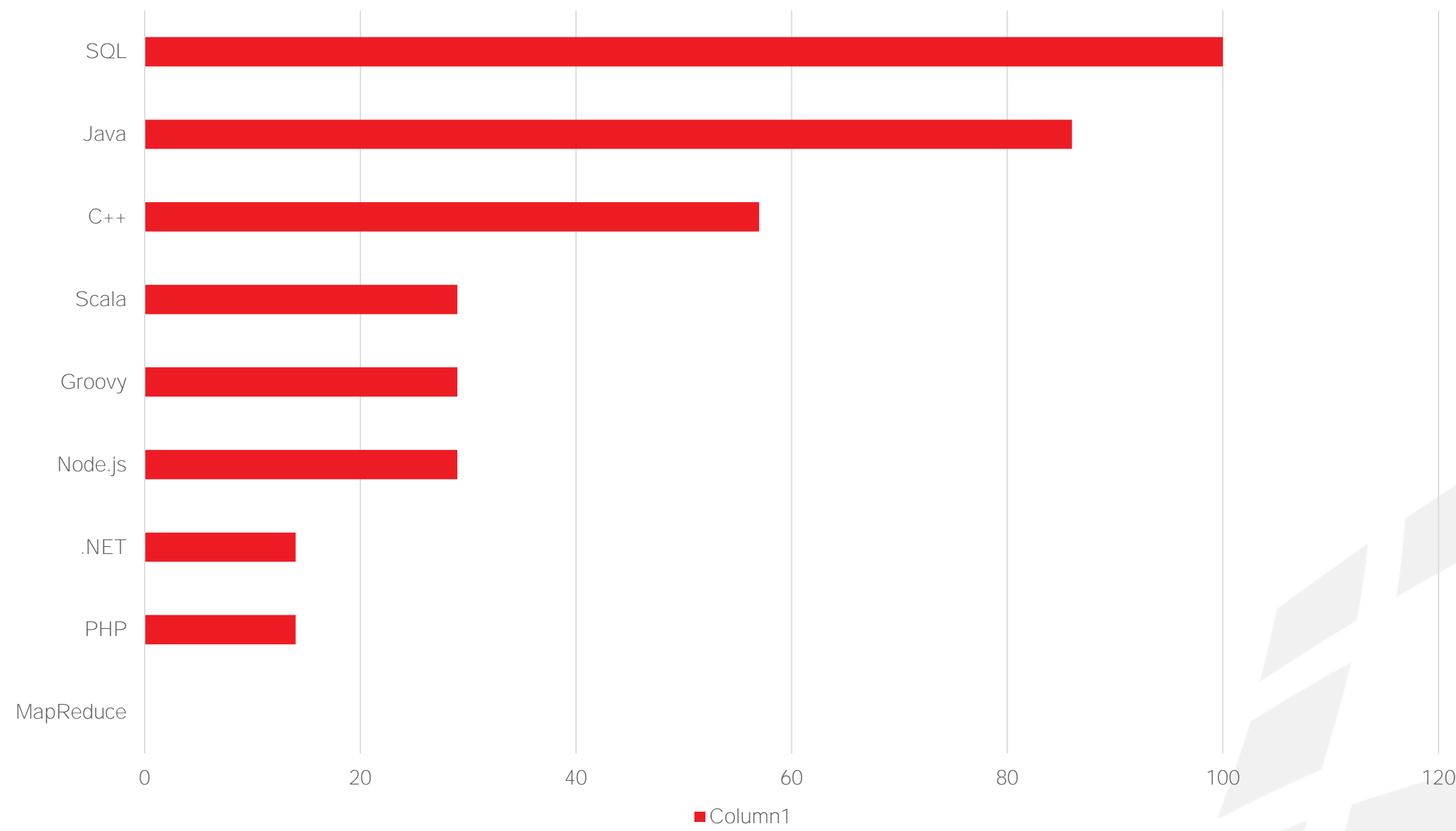
Financial Customer Use Cases

Data Velocity, Data Volume, Data Consistency, Real-Time Performance and Analysis

Core Banking and Trading Platforms
Treasury systems, payment hubs, order management systems, algorithmic trading, high volume transactions, ultra low latencies.

Risk Management
Modeling, financial engineering, pricing, hedging, what-if analysis, reporting.

Financial Analytics
Real time analysis of trading positions, trending, market data analysis, sentiment analysis, complex event processing, hedging, transaction cost analysis, time series, volatility analysis, Monte Carlo simulations, Black-Scholes, derivatives pricing.

Big Data Analytics
Customer and counter party 360 view, master data management, securities masters, reference data, real-time analysis of P&L, up-to-the-second operational BI.

Compliance and Monitoring
Fraud, AML, KYC, market manipulation and abuse, pre and post trade compliance modeling.

Financial SaaS Platforms
High performance next-generation architectures for Software as a Service Application vendors.

Case Study:

MISYS
FINANCIAL SOFTWARE

- Financial services software
  - Retail and corporate banking
  - Lending
  - Treasury
  - Capital markets
  - Investment management
  - Enterprise risk
- More than 2,000 customers in 130 countries
- Used by 48 of the world's 50 largest banks

- The Challenge: Eliminating Data Processing Bottlenecks
  - Huge amounts of trade and accounting data
  - Customers need
    - High-speed transactions
    - Real-time reporting
  - New Java-based IT stack with data lake support
  - Global regulatory compliance

GridGain

Case Study:

MISYS
FINANCIAL SOFTWARE

- Commodity servers (256GB RAM)
- Data stored in memory
  - Transactions
  - Market data
- Parallel processing across cluster
  - Calculation heavy reporting for regulatory compliance

GridGain

Case Study:

MISYS
FINANCIAL SOFTWARE

FusionFabric.cloud

- Integrates trading systems with cloud-based components
  - OTC derivatives
  - Exchange traded derivatives
  - Inflation
  - Fixed income
  - FX/MM
  - Hybrids
  - Developing additional modules

"With GridGain, we have achieved real-time processing of massive amounts of trade and transaction data, eliminating bottlenecks and enabling us to offer next-generation financial services to our customers."
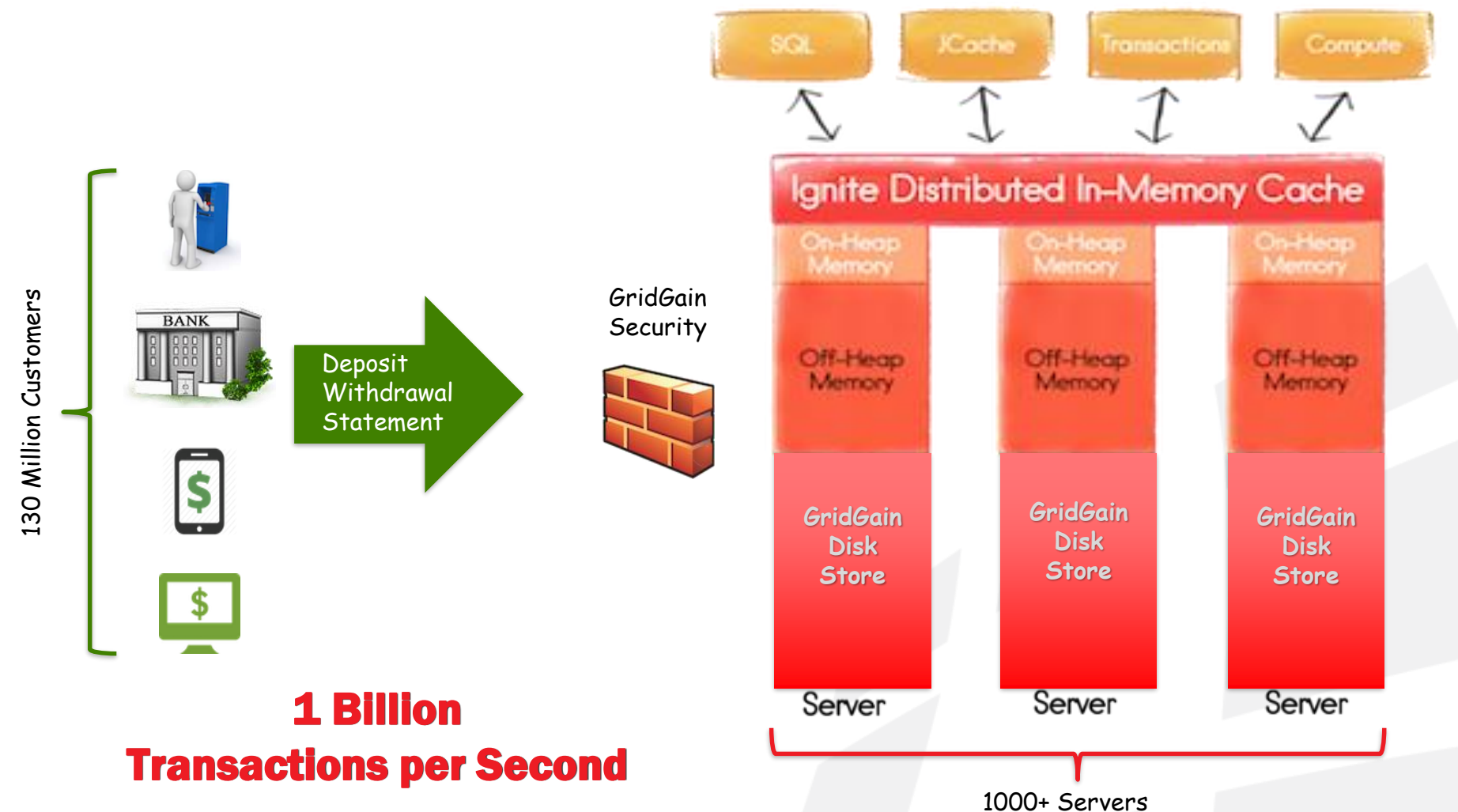-Felix Grevy, Director of Product Management for FusionFabric.cloud at Misys

Use Case: SBERBANK

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

- Sberbank Requirements
  - Expect significant transactional volume growth
  - Migrate to data grid architecture to build next generation platform
  - Minimize dependency on Oracle
  - Move to open source

- Why GridGain Won
  - Best performance
    - 10+ competitors evaluated
  - Demonstrated best
    - Fault tolerance & scalability
    - ANSI-99 SQL Support
    - Transactional consistency
  - Strict SLAs
    - Less then 5 min cluster restart (regulatory requirement)
    - Fully Operational from disk & memory
    - Compliance with personal data law and cyber-security regulations

1 Billion Transactions per Second

10 Dell R610 blades 1 TB Memory = **$25K**

130 Million Customers

Deposit Withdrawal Statement

GridGain Security

Ignite Distributed In-Memory Cache

SQL · JCache · Transactions · Compute

On-Heap Memory / Off-Heap Memory / GridGain Disk Store

Server · Server · Server

1000+ Servers

GridGain Company Confidential

GridGain

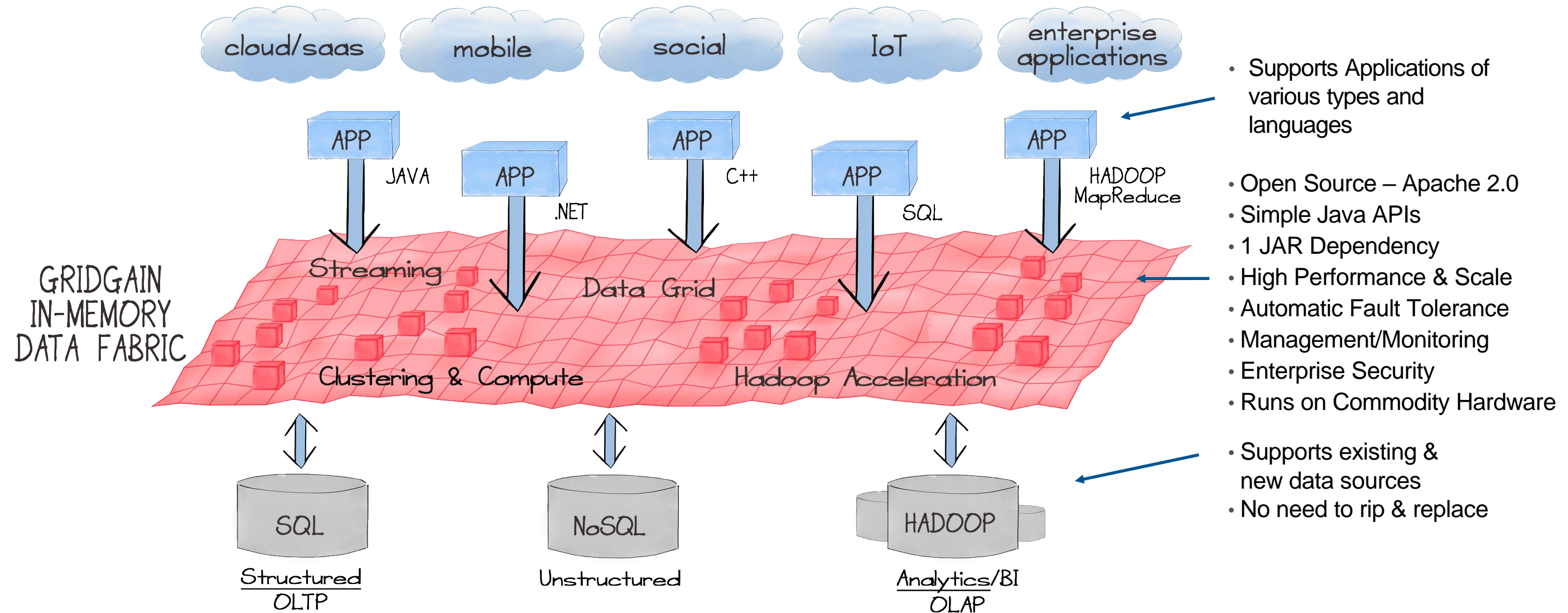# From ◢RBC article – January, 2016

Herman Gref
CEO & Chairman, Sberbank

"The new Sberbank IT plan is to create a platform that enables the bank to introduce new products in hours, not weeks. The platform will have virtually unlimited performance and very high reliability. It will be much cheaper and will significantly reduce human interaction during customer transactions. The system will use machine-learning, flexible pricing, and artificial intelligence," said Herman Gref, head of Sberbank.

"The new system will use technology from GridGain, which won the tender from Oracle, IBM and others, and turned out to deliver an order of magnitude higher performance than those of the largest companies," he added.

GridGain

Flexibility and Enterprise Breadth of In-Memory Computing Platform

- 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
- Enterprise Security
- Runs on Commodity Hardware

- Supports existing & new data sources
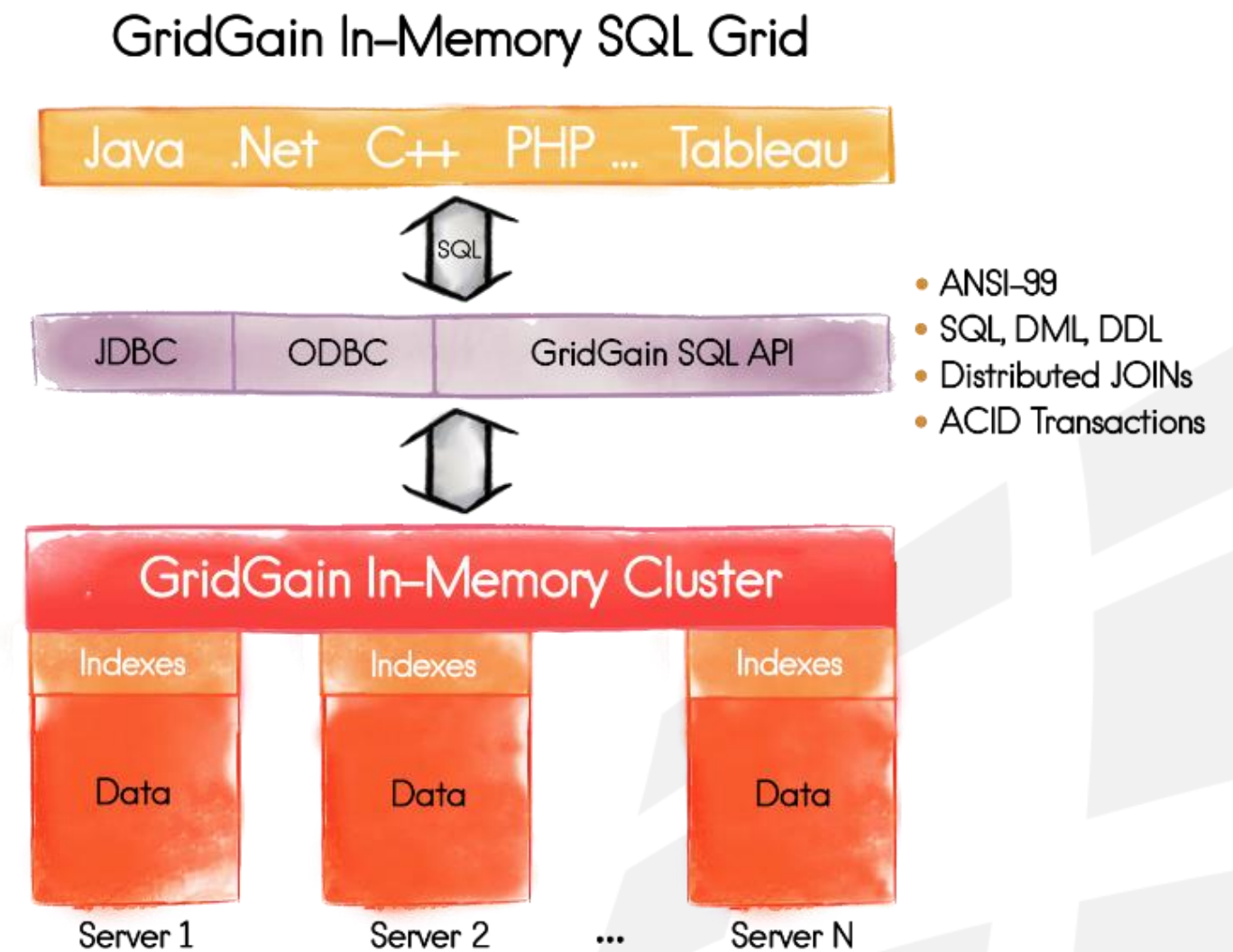- No need to rip & replace

GridGain

In-Memory Data Grid

- Inserted between the application and data layers. Moves disk-based data from RDBMS, NoSQL or Hadoop databases into RAM

- **Features:**
  - Distributed In-Memory Key-Value Store
  - Replicated and Partitioned Data Caches
  - Lightning Fast Performance
  - Elastic Scalability
  - Distributed In-Memory Transactions (ACID)
  - Distributed In-Memory Queue and Other Data Structures
  - Web Session Clustering
  - Hibernate L2 Cache Integration
  - On-Heap and Off-Heap Storage
  - Distributed SQL Queries with Distributed Joins
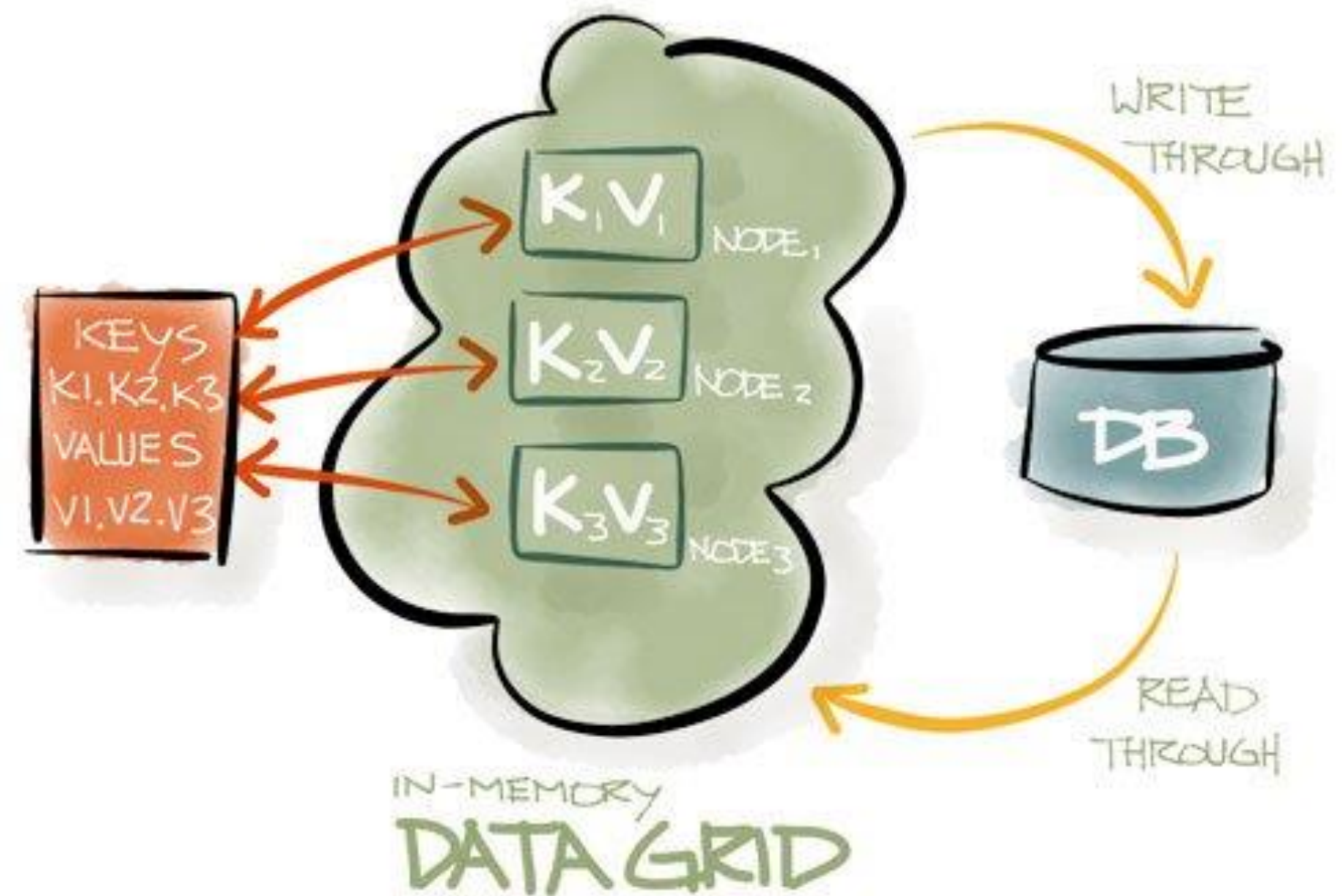
In-Memory SQL Grid

- Horizontally scalable, fault tolerant, ANSI SQL-99 compliant, and fully supports all SQL and DML commands

- **Features:**
    - Supports SQL and DML commands including SELECT, UPDATE, INSERT, MERGE and DELETE Queries
    - Distributed SQL
    - Geospatial Support
    - SQL Communications Through the GridGain ODBC or JDBC APIs Without Custom Coding
    - ANSI SQL-99 Compliance

GridGain In–Memory SQL Grid

Java .Net C++ PHP ... Tableau

SQL

JDBC | ODBC | GridGain SQL API

- ANSI-99
- SQL, DML, DDL
- Distributed JOINs
- ACID Transactions

GridGain In–Memory Cluster

Indexes | Indexes | Indexes
Data | Data | Data
Server 1 | Server 2 ... | Server N

GridGain

In-Memory Compute Grid

- Enables parallel processing of CPU or otherwise resource intensive tasks
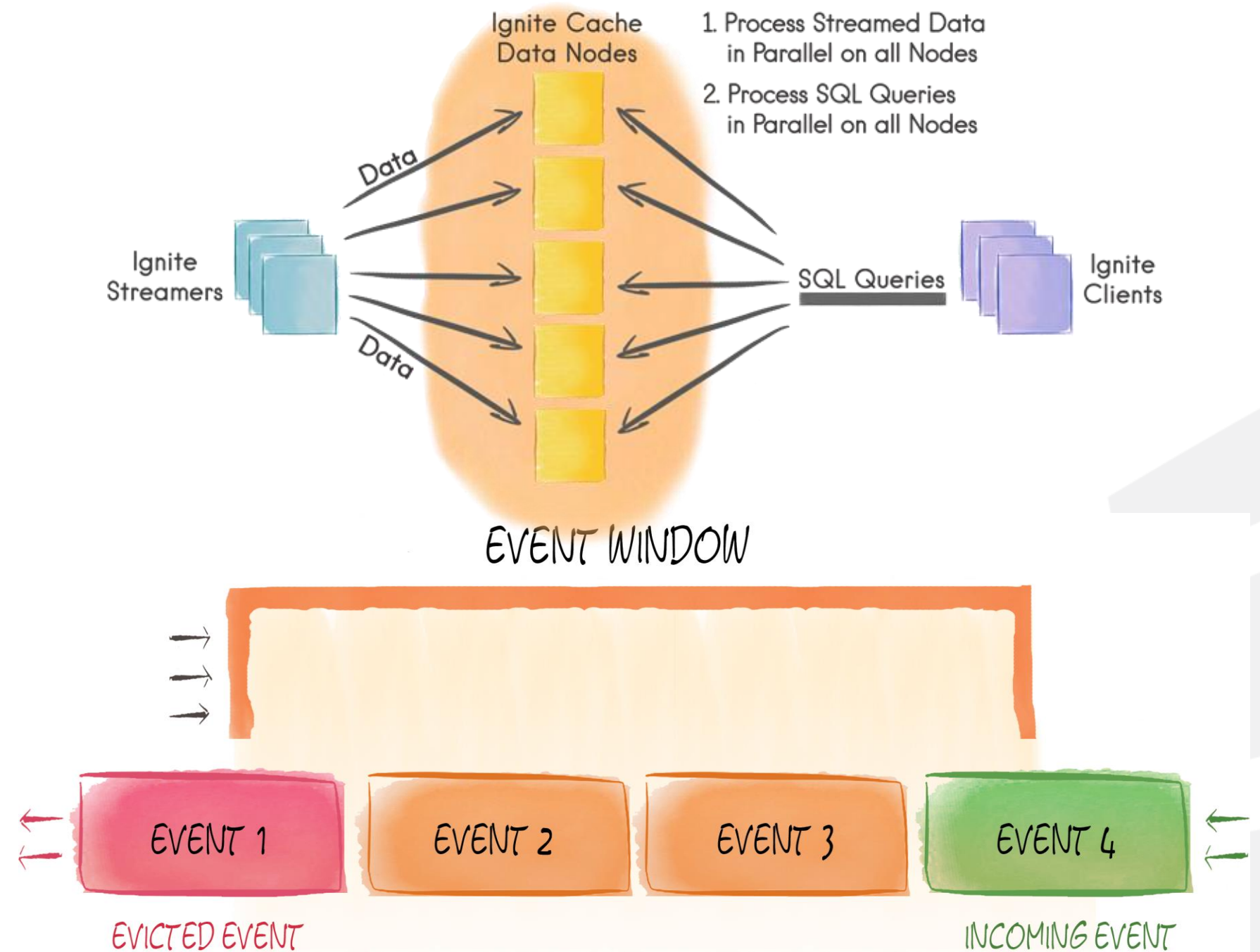
- **Features:**
  - Dynamic Clustering
  - Direct API for Fork-Join & MapReduce Processing
  - Distributed Closure Execution
  - Adaptive Load Balancing
  - Automatic Fault Tolerance
  - Linear Scalability
  - Custom Scheduling
  - State Checkpoints for Long Running Jobs
 - Pluggable SPI Design

GridGain
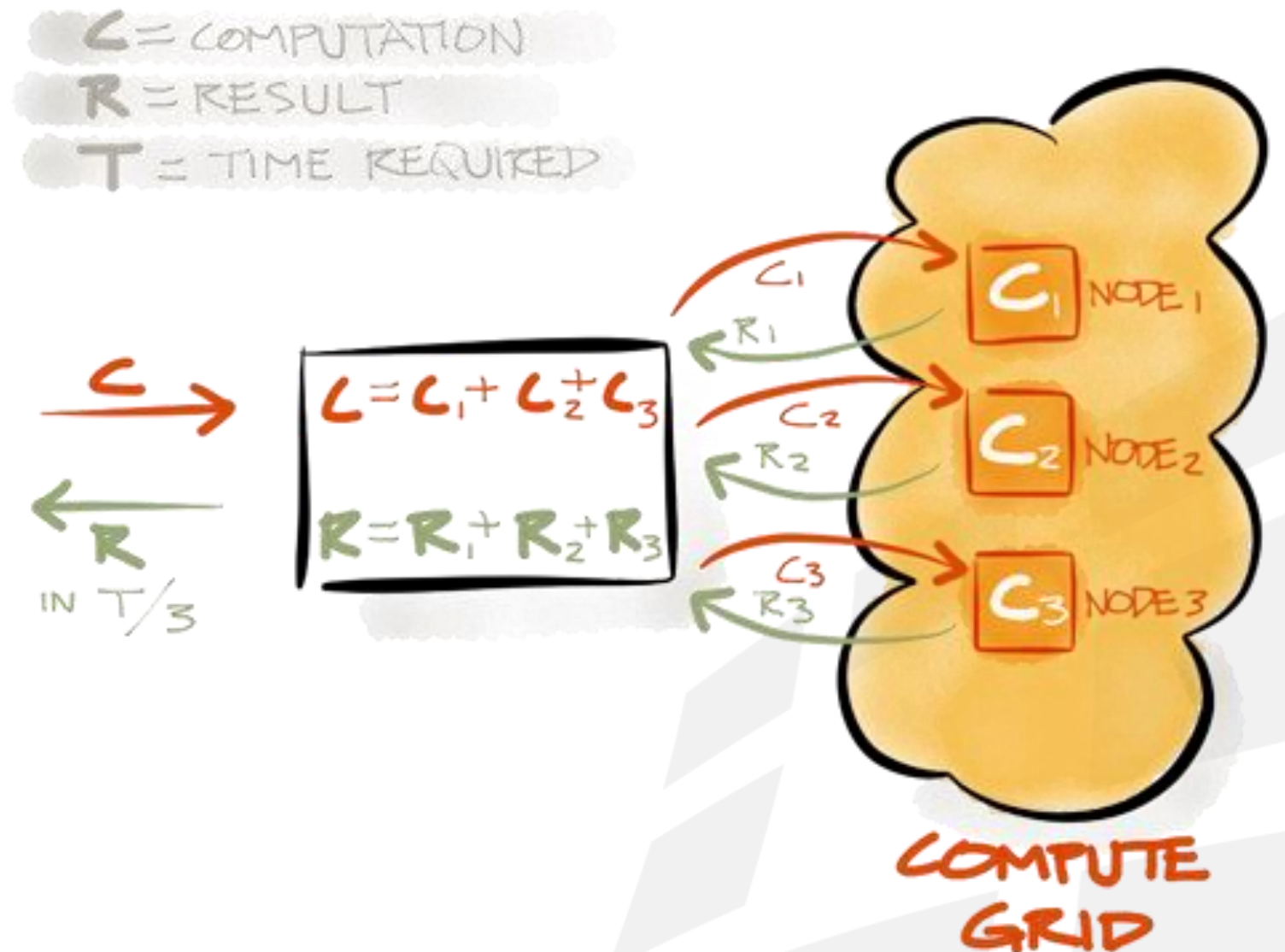
In-Memory Streaming and CEP

- Streaming Data Never Ends
- Sliding Windows for CEP/Continuous Query
- Customizable Event Workflow
- Branching Pipelines
- Pluggable Routing
- Real Time Analysis
- Data Indexing
- Distributed Streamer Queries

In-Memory Hadoop Acceleration

- Provides easy to use extensions to disk-based HDFS and traditional MapReduce, delivering up to 10x faster performance

- **Features:**
  - Use existing MapReduce / Pig / Hive
  - 10x Faster Performance
  - In-Memory MapReduce
  - Highly Optimized In-Memory Processing
  - Standalone File System
  - Optional Caching Layer for HDFS
  - Read-Through and Write-Through with HDFS

MR · HIVE · PIG

IN-MEMORY MAPREDUCE

YARN

GGFS

GGFS

HDFS

ANY HADOOP DISTRO
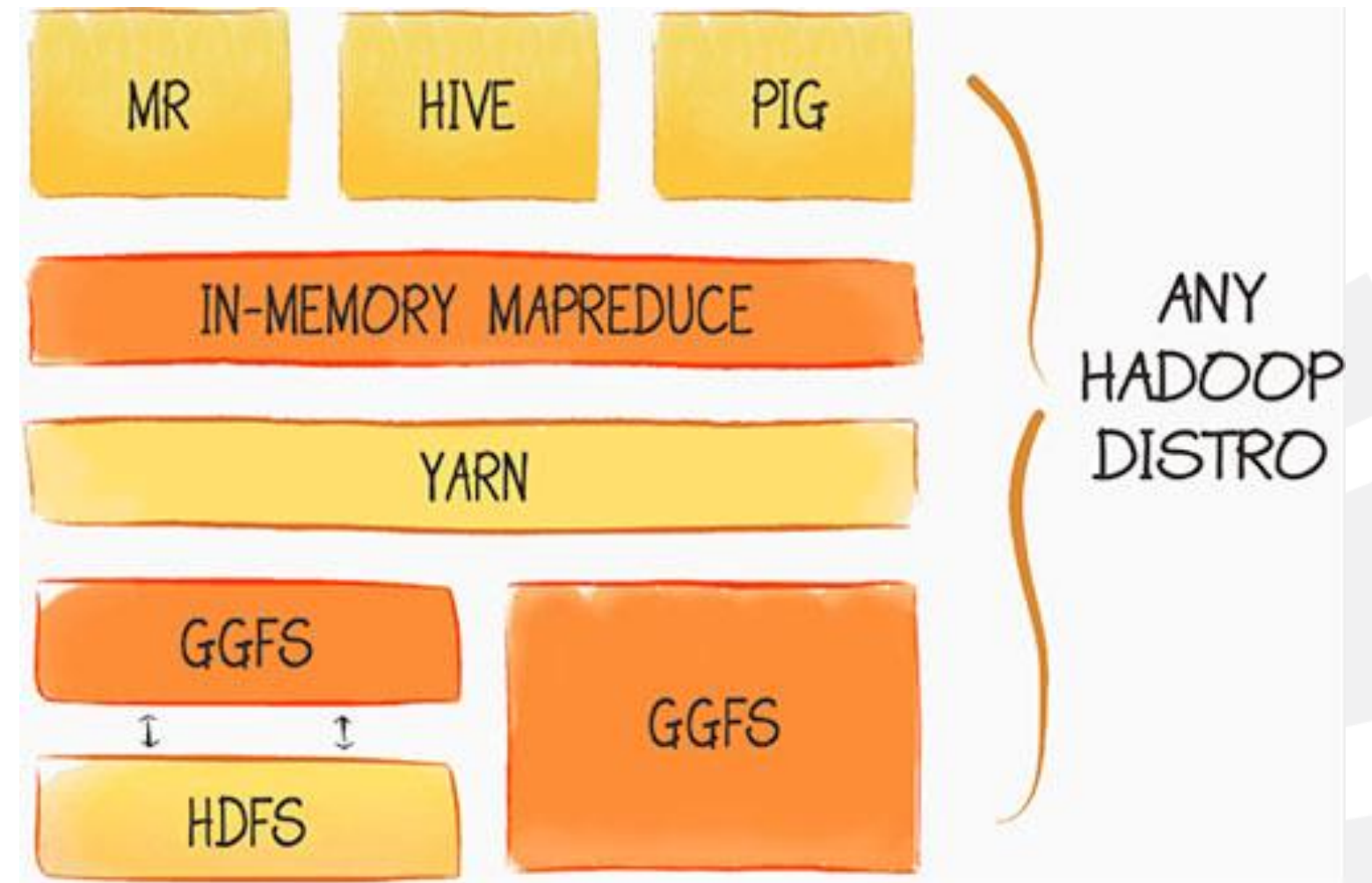
GridGain

GridGain's In-Memory Computing Platform Enterprise Edition

GridGain Enterprise Subscriptions include:

> Right to use GridGain Enterprise Edition

> Bug fixes, patches, updates and upgrades

> 9x5 or 24x7 Support

> Training and Consulting Services from GridGain

Features	Apache Ignite	GridGain Enterprise
In-Memory Data Grid	√	√
In-Memory Compute Grid	√	√
In-Memory Service Grid	√	√
In-Memory Streaming	√	√
In-Memory Hadoop Acceleration	√	√
Distributed In-Memory File System	√	√
Advanced Clustering	√	√
Distributed Messaging	√	√
Distributed Events	√	√
Distributed Data Structures	√	√
Portable Binary Objects	√	√
Management & Monitoring GUI		√
Enterprise-Grade Security		√
Network Segmentation Protection		√
Recoverable Local Store		√
Rolling Production Updates		√
Data Center Replication		√
Integration with Oracle GoldenGate		√
Basic Support (9×5)	√	√
Enterprise Support (9×5 and 24×7)		√
Security Updates		√
Maintenance Releases & Patches		√
	Free w/ optional Paid Support	Annual License Subscription

GridGain

THANK YOU