In-Memory Computing:
From Big Data to Fast Data

NIKITA IVANOV
GridGain CTO/Founder

nivanov@gridgain.com #c64hacker

GridGain

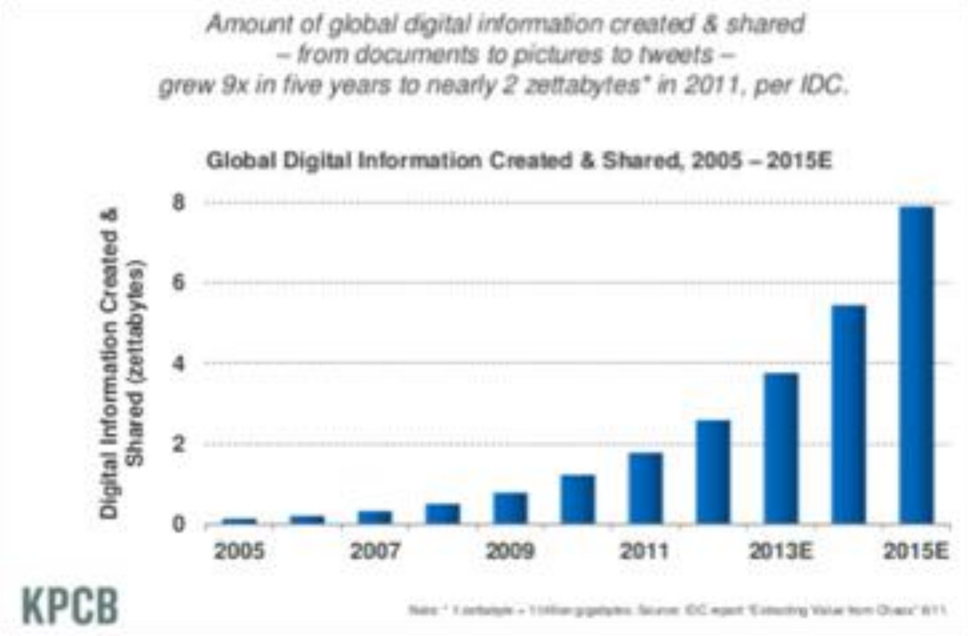
Agenda

- Why In-Memory Computing?
- What is In-Memory Computing?
- Facts & Myths
- Use Cases
- Q & A

Why In-Memory

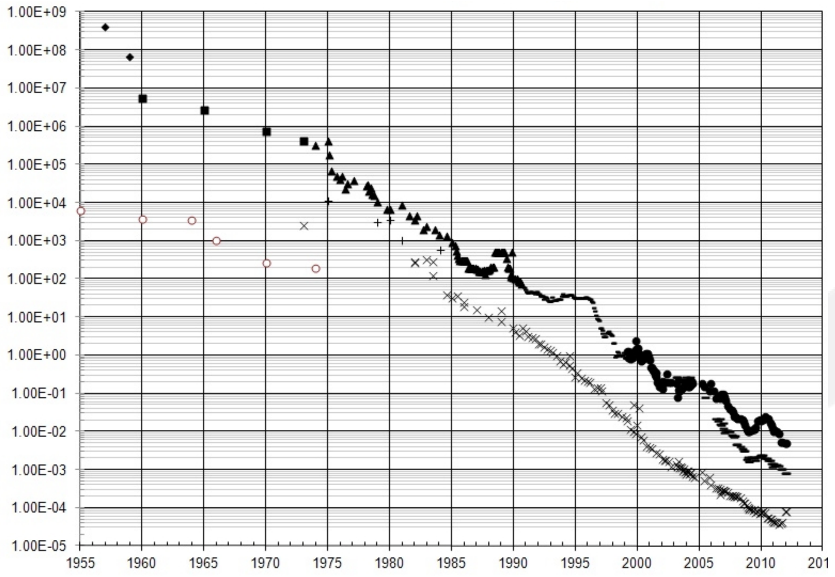
"In-memory computing will have a long term, disruptive impact by radically changing users' expectations, application design principles, product architectures, and vendor strategies." (Gartner)

Data Growth



Less than 2 zetabytes in 2011, 8 in 2015

DRAM Cost, $



Cost drops 30% every 12 months

GridGain

Paradigm Shift à la 1970s

1970 – 2000s: Era of Disk

- IBM released "Winchester" IBM 340 disk
  Tapes start to decline
- SQL
  Era of Structured Data

2010s – … : Era of Memory

- 64-bit CPUs + DRAM prices drop 30% YoY
  HDDs start to decline
- NoSQL + SQL
  Era of Unstructured Data
- **Last frontier for storage**

RAM is the new disk, disk is the new tape (Gartner)

Memory First - Disk Second

Disk First Architecture: 1970-2000s

Disk as primary storage, memory for caching

Reading Record: API call <-> OS I/O <-> I/O controller <-> disk

Latency: milliseconds

Memory First Architecture: since 2000s

Memory is primary storage, disk for backups

Reading Record: API call <-> pointer arithmetic

Latency: nanoseconds to microseconds

Bring Computation to Data

Client-Server 1970 - 2000s

- **Data is moved to application layer:**
  Data not-partitioned
  Data sizes are small

In-Memory Computing / Hadoop since mid-2000s

- **Computations are moved to data:**
  Data is partitioned
  Data sizes are massive
  Possible to distribute computations to partitioned data

Myth #1: Too Expensive

Facts:

> 2013: 1TB DRAM cluster **$25K**
>
> 2015: 1TB DRAM cluster <**$20K**
>
> Memory Channel Storage (MCS)
>
> Storage Class Memory (SCM)
>
> **Non-Volatile RAM (NVDIMM)**

Myth #2: Not Durable

Facts:

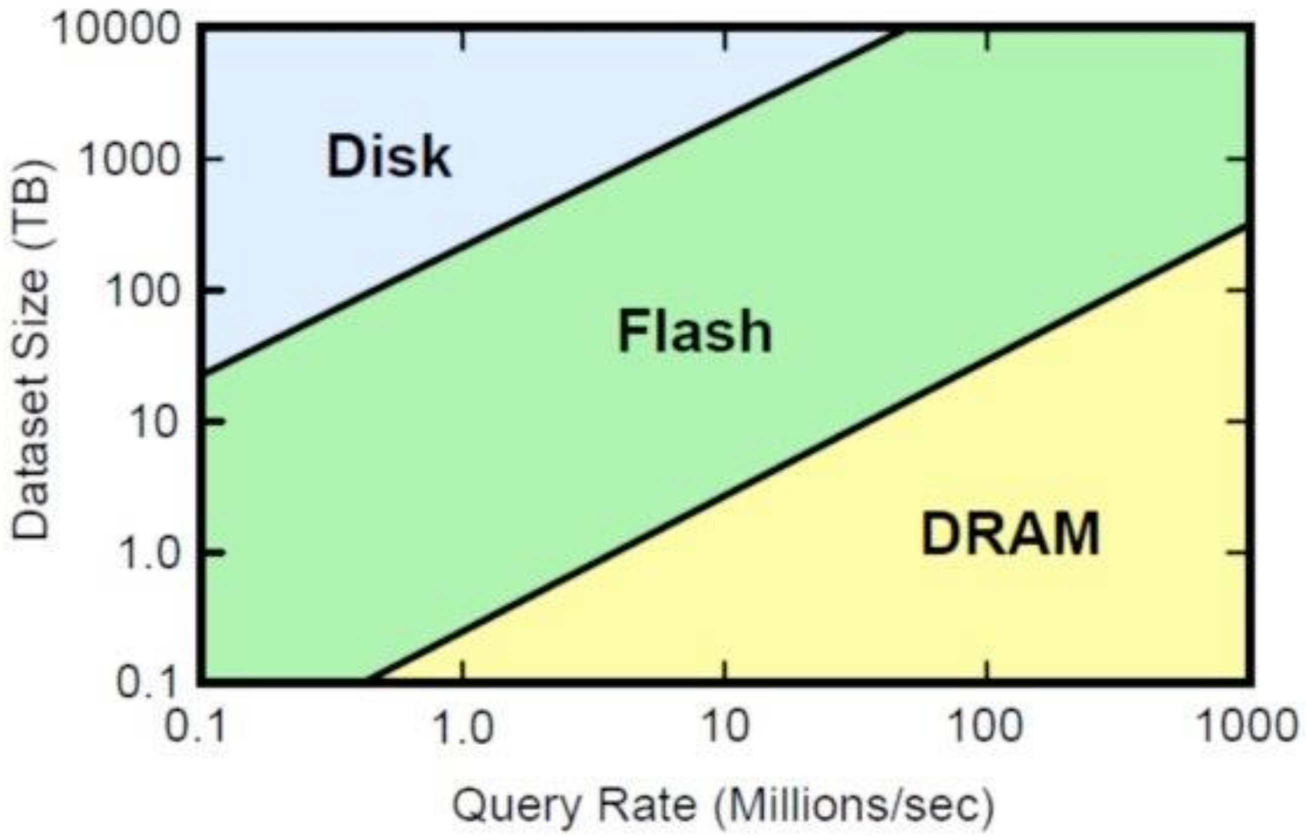
- IMC have durable backups and disk storage
  Active or passive replicas, transactional read-through and write-through
- Mature IMC provide tiered storage
  DRAM - Local Swap - RDBMS/HDFS
- Operational vs. Historical datasets
  99% of operational datasets < 10TB

Myth #3: Flash is Fast Enough

Facts:

Flash on PCI-E is still... **a block device.**

Still going through OS I/O, I/O controller, marshaling, buffering.



Myth #4: Only for Caching

Facts:

- Caching is important use case for **yesterday**
  Easiest adoption and a “low-hanging fruit”
- In-Memory Data Fabrics for **today**
  Main system of records moving to in-memory
- Vertical and PnP products are the **future**
  Minimal integration, maximum benefit

In-Memory Computing: Key Use Cases

> Automated Trading Systems

Real time analysis of trading positions & market risk. High volume transactions, ultra low latencies.

> Hybrid OLAP/OLTP

Fraud Detection, Risk Analysis, Insurance rating and modeling.

> Online & Mobile Advertising

Real time decisions, geo-targeting & retail traffic information.

> Real Time 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.

THANK YOU!

nivanov@gridgain.com #c64hacker