

MARCH 2018

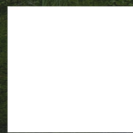


A High Performance Data Platform Using Flash

Super Computing Asia

# iguazio

20+ Deployments



*Industrial IoT*



*Financial Services*



*Telecommunications*



*Cyber Security*

\$48M in Funding



Verizon  
Ventures



DELL  
Technologies



100+ Employees



NA



APAC

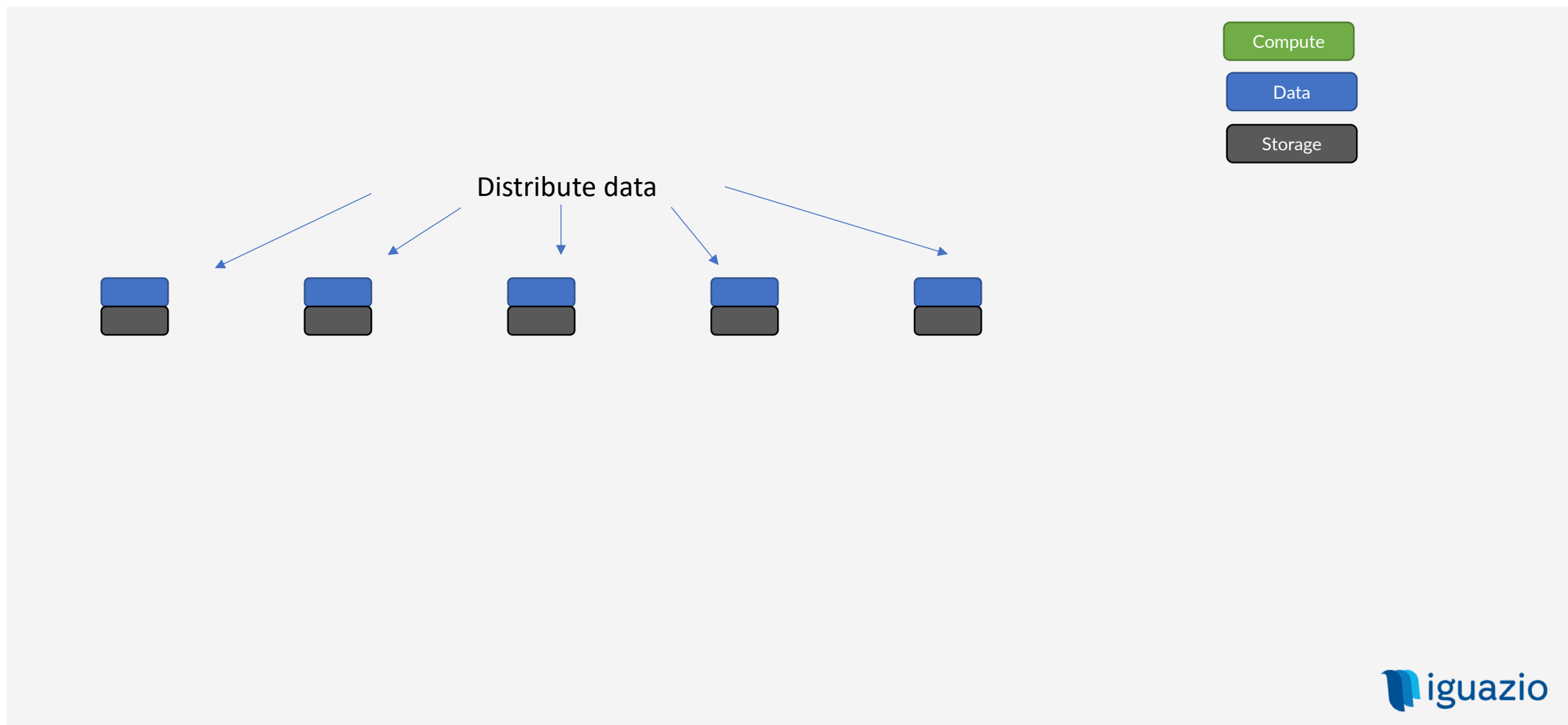


EMEA

# When Hadoop started

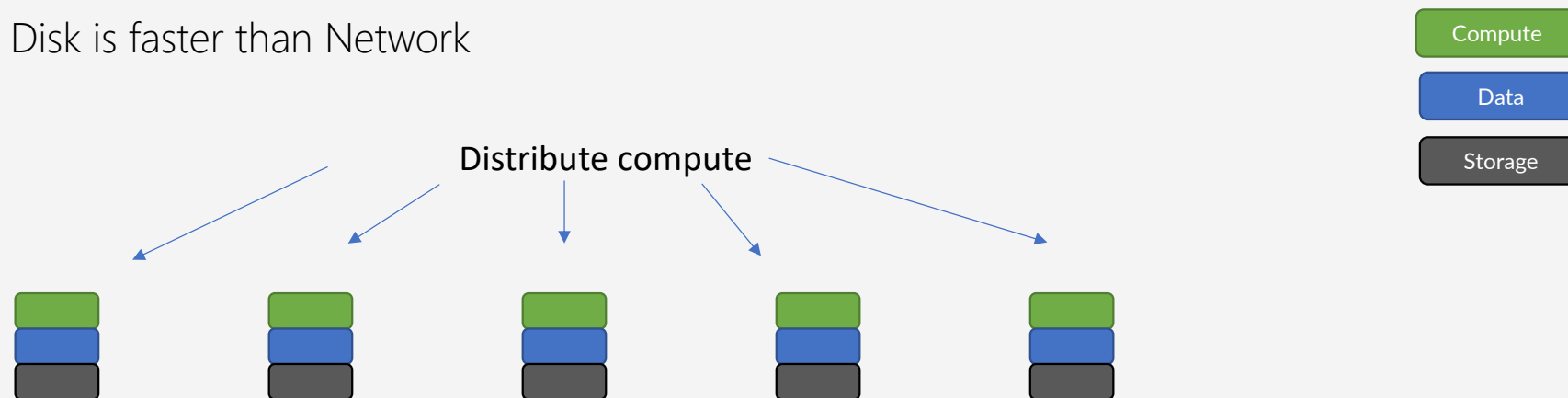
- Disk and Network were still slow
- Network was in general slower than disk

# Data at scale



# Compute is where data is

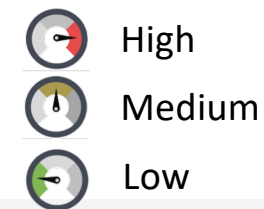
- Disk is faster than Network



# Big Data Workloads – Performance Characteristics

- Scalable but inherently batch and slow
  - Disk and Network I/O (Map, reduce, shuffle ... )
  - Runs on JVM
- Not suitable for all type of workloads (real-time, interactive, iterative M/L)
  - Cannot scale compute independent of storage and vice versa

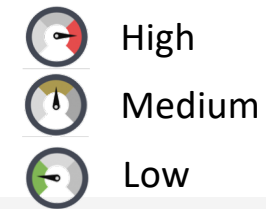
# Big Data Workloads – Performance Characteristics



Typically Disk or Network maxes out during a massively parallel job, before Compute

Big Data Workload	Disk I/O	Network I/O	Compute
Iterative Machine Learning Jobs			
Interactive Analytics at scale			
Lambda & Kappa Architecture at scale			

# Big Data Workloads – Performance Characteristics



Typically Disk or Network maxes out during a massively parallel job, before Compute

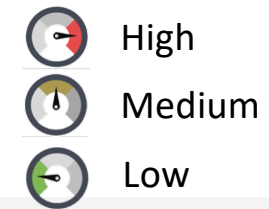
Big Data Workload	Disk I/O	Network I/O	Compute
Iterative Machine Learning Jobs			
Interactive Analytics at scale			
Lambda & Kappa Architecture at scale			

Compare this to traditional RDBBS – CPU or Disk I/O maxes out

<b>Traditional Database Workload</b>			
--------------------------------------	--	--	--



# Big Data Workloads – Performance Characteristics



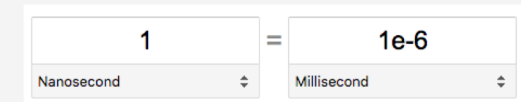
Big Data Workload	Disk I/O	Network I/O	Compute
Iterative Machine Learning Jobs			
Interactive Analytics at scale			
Lambda & Kappa Architecture at scale			

in-memory and high-speed networking ?

Traditional Database Workload			
-------------------------------	--	--	--

# Optimizing Big Data Performance

- L1 cache reference 0.5 ns
- Branch mispredict 5 ns
- L2 cache reference 7 ns
- Mutex lock/unlock 100 ns
- Main memory reference 100 ns
- Send 2K bytes over 1 Gbps network 20,000 ns
- SSD seek 80,000 ns
- Read 1 MB sequentially from memory 250,000 ns
- Round trip within same datacenter 500,000 ns
- Disk seek 10,000,000 ns
- Read 1 MB sequentially from network 10,000,000 ns
- Read 1 MB sequentially from disk 30,000,000 ns
- Send packet CA->Netherlands->CA 150,000,000 ns



1 = 1e-6

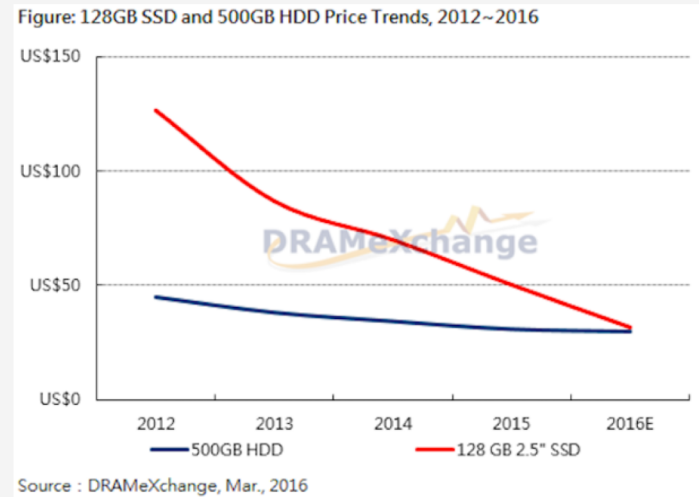
Nanosecond      Millisecond

Designs, Lessons and Advice from Building Large Distributed Systems by Dr Jeff Dean of Google Source <http://www.slideshare.net/ikewu83/dean-keynoteladis2009-4885081>

## RAM is expensive

RAM is 10-20x more expensive than Flash in \$Cost / GB

# SSDs prices are falling



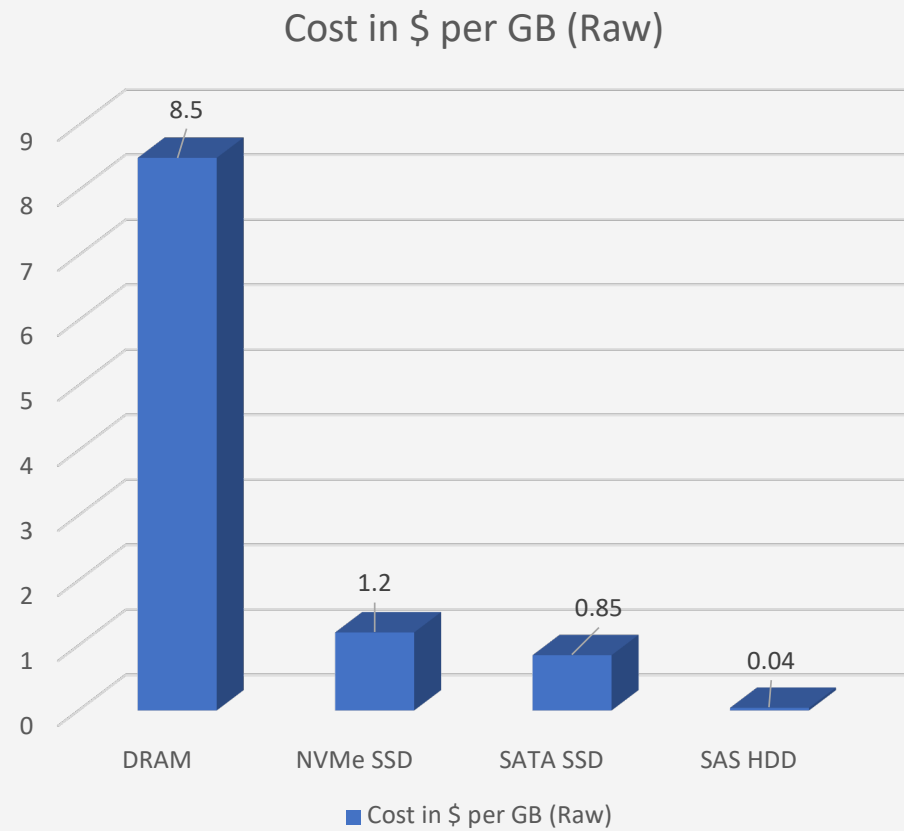
# Optimizing Big Data Performance

- L1 cache reference 0.5 ns
- Branch mispredict 5 ns
- L2 cache reference 7 ns
- Mutex lock/unlock 100 ns
- Main memory reference 100 ns
- **Send 2K bytes over 1 Gbps network 20,000 ns**
- **SSD seek 80,000 ns**
- Read 1 MB sequentially from memory 250,000 ns
- **Round trip within same datacenter 500,000 ns**
- **Disk seek 10,000,000 ns**
- **Read 1 MB sequentially from network 10,000,000 ns**
- **Read 1 MB sequentially from disk 30,000,000 ns**
- Send packet CA->Netherlands->CA 150,000,000 ns

A unit conversion interface with two input fields. The first field contains the number '1' and is labeled 'Nanosecond' below it. An equals sign '=' is between the two fields. The second field contains '1e-6' and is labeled 'Millisecond' below it. Both fields have a small downward arrow on the right side, indicating they are dropdown menus.

Designs, Lessons and Advice from Building Large Distributed Systems by Dr Jeff Dean of Google Source <http://www.slideshare.net/ikewu83/dean-keynoteladis2009-4885081>

# Price of Memory Technology 2016/17

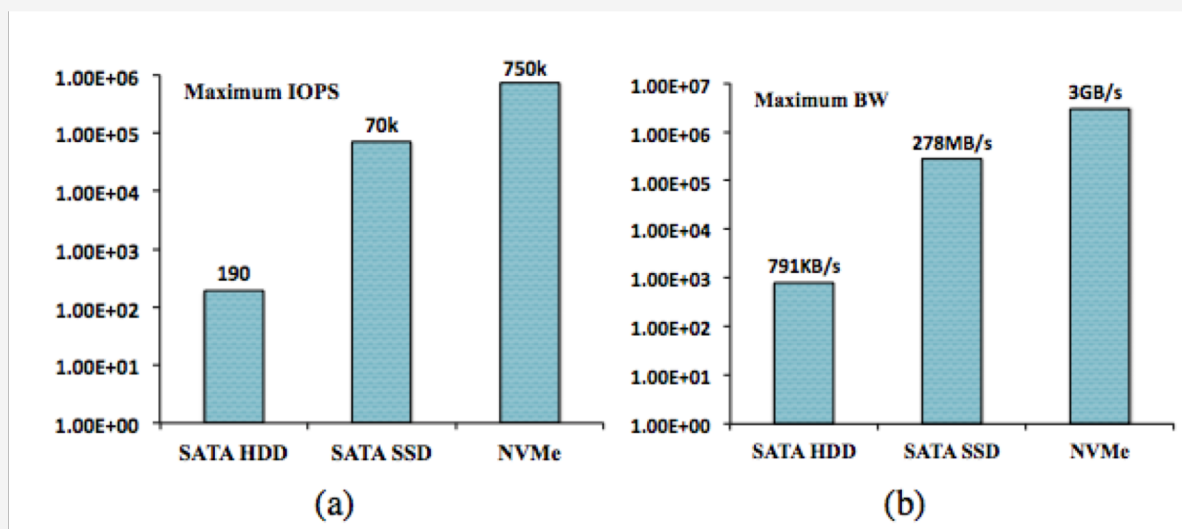


# What is NVMe?

- NVMe is a software based standard that was specifically optimized for SSDs connected through the PCIe interface.

## Why is NVMe faster (than SATA) ?

- Shorter hardware data access path – directly connected via PCIe, Faster compared to SATA
- NVMe Completely redesigned software - bypasses conventional block layer request queue –
  - Asynchronous Submission Queue for requests
  - Asynchronous Completion Queue



Source

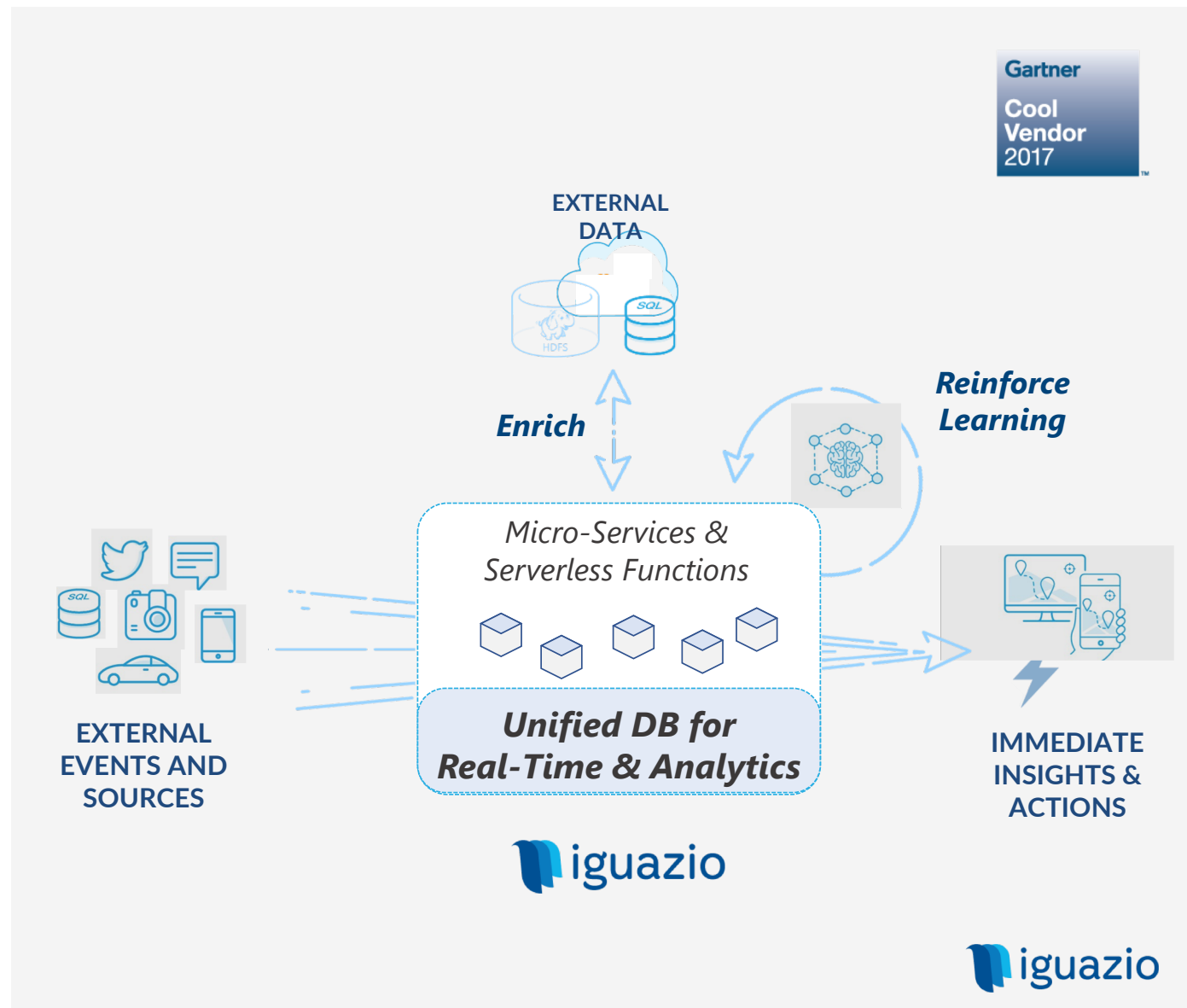
<http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=2CA58A08BFB2821F59A7996DAC8742F1?doi=10.1.1.697.1493&rep=rep1&type=pdf>



# iguazio

## *Unified DB for Real-Time & Analytics*

- Combined fresh data
- and historical data
- Immediate Insights
- Rapid time to production
- Supporting common APIs
- Deployment Everywhere



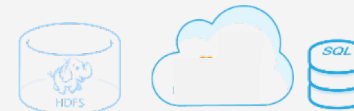
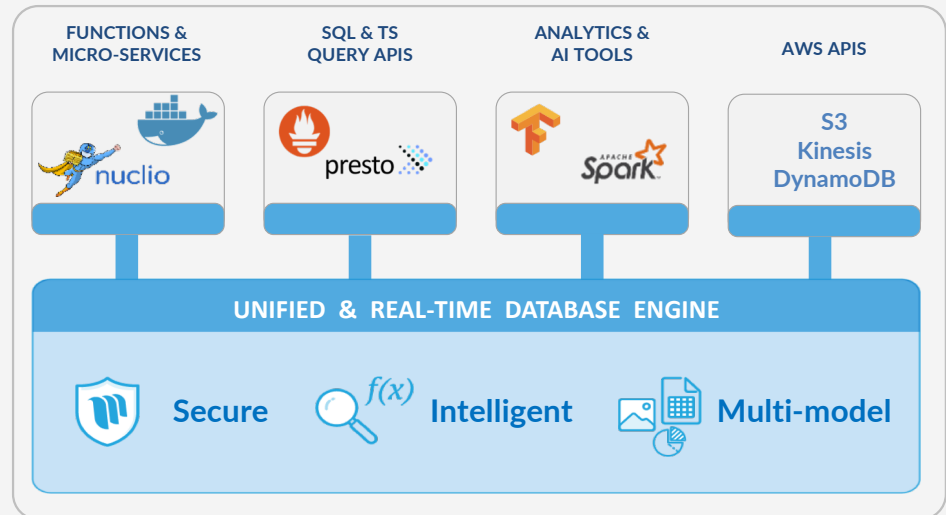
# Architecture

- In-mem DB performance with Flash economies and density
- Access data concurrently through multiple standard APIs
- Fully integrated PaaS



**Single platform  
All in One**

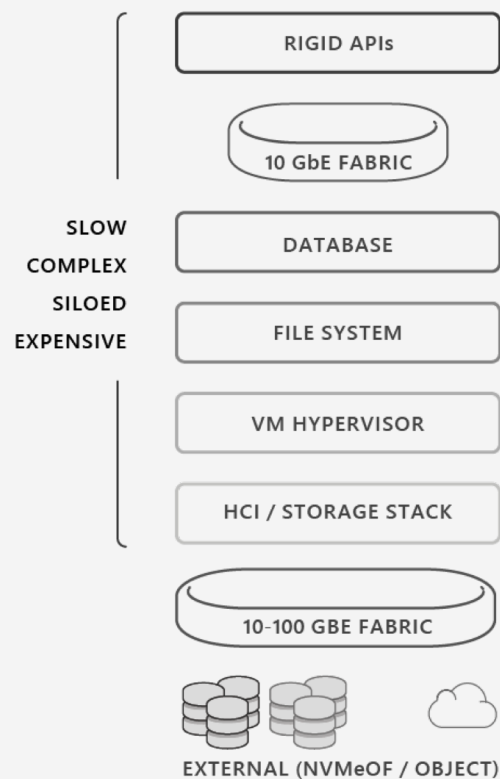
## iguazio Data platform for real time and analytics



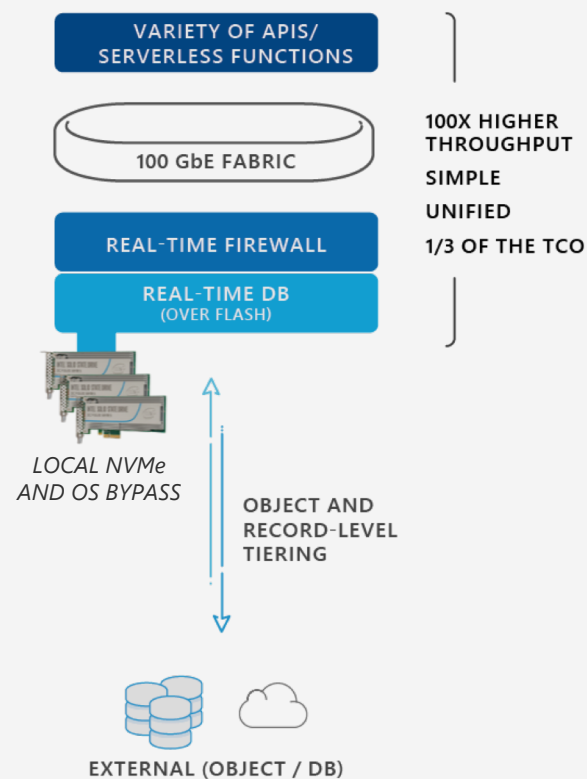
EXTERNAL DATA LAKES & CLOUDS

# How to take full advantage of NVMe?

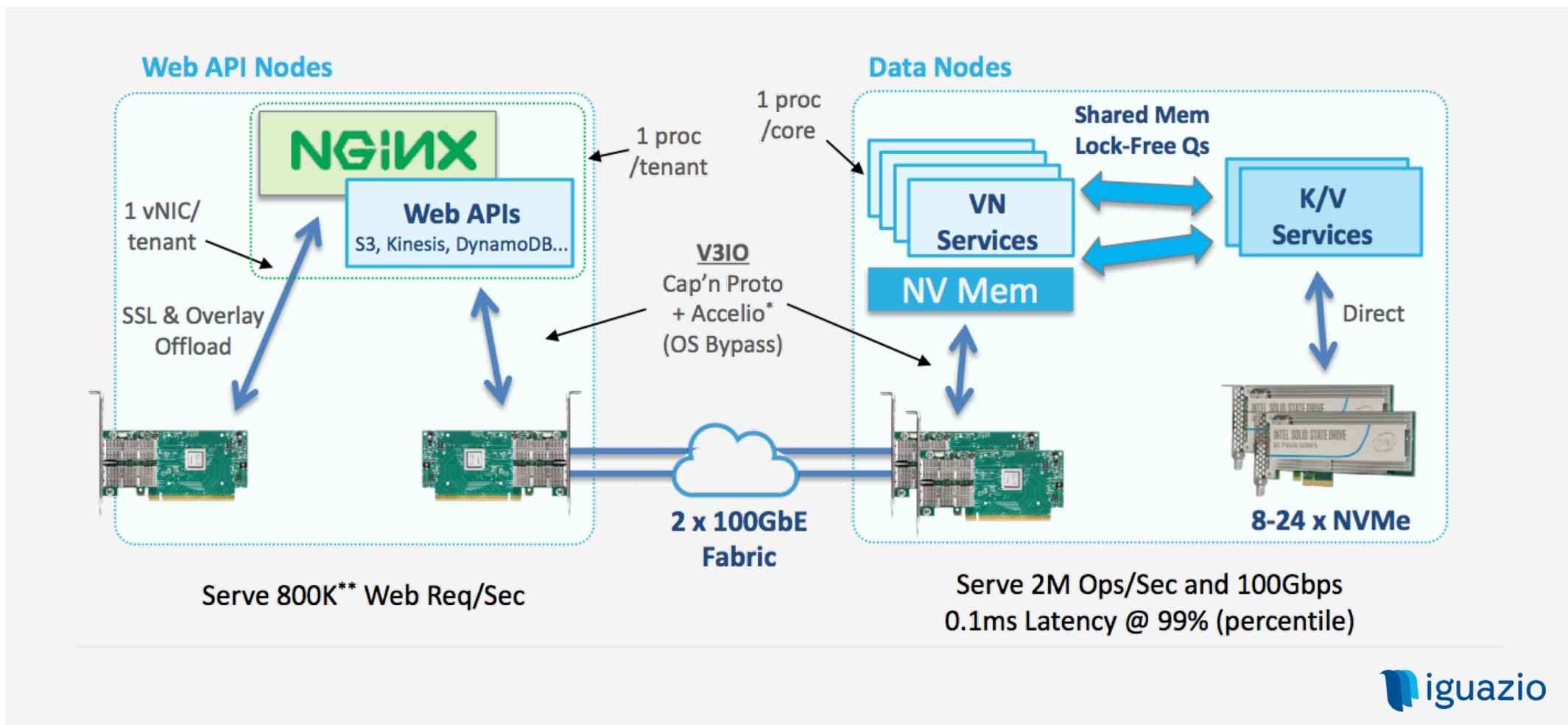
## Traditional Layered Approach



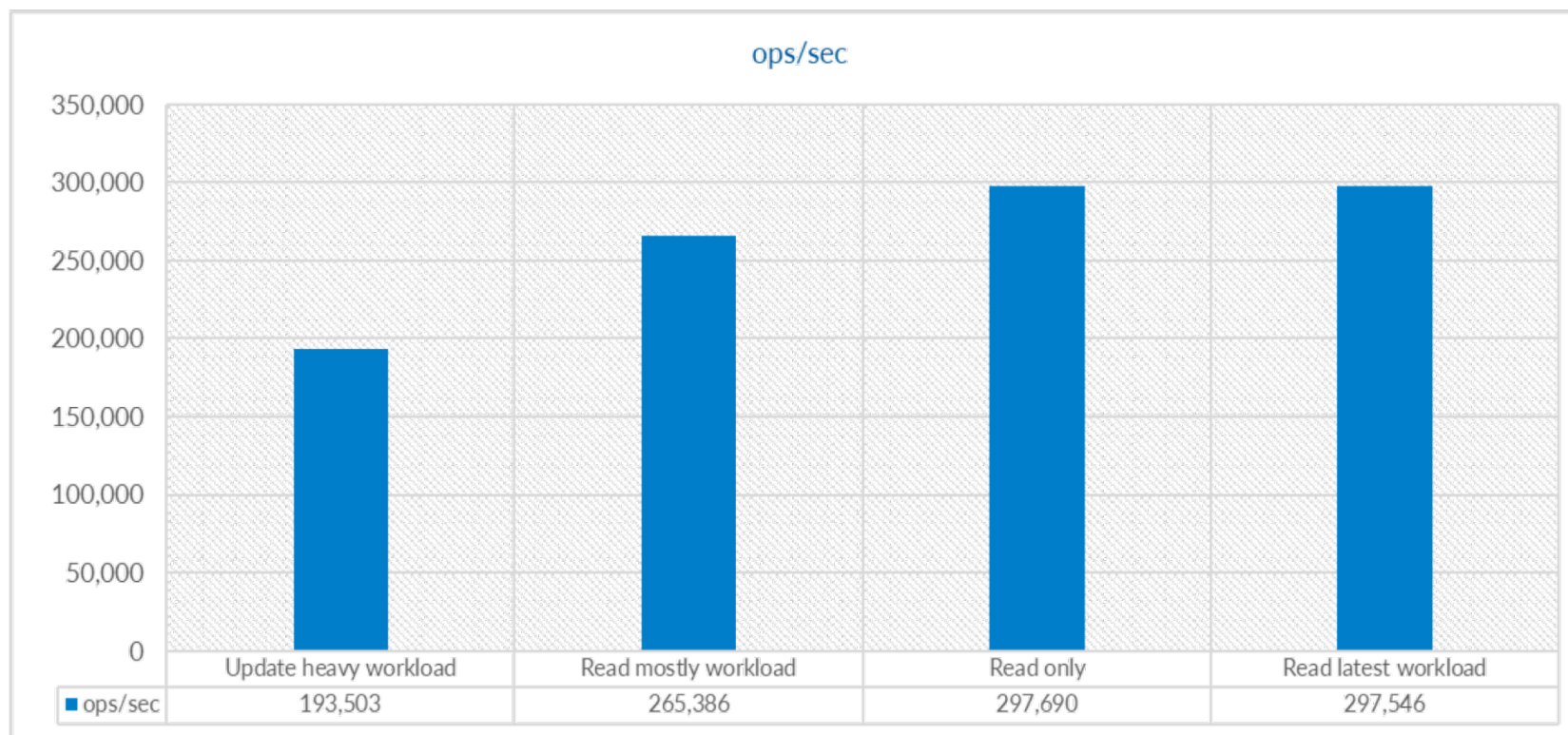
## iguazio



# And Advanced Networking



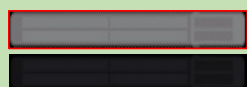
# Performance Results



*The tests were performed using the Yahoo! Cloud Serving Benchmark (YCSB), an open source framework for evaluating and comparing the performance of multiple types of NoSQL database management systems, the de facto industry standard for this purpose.*

# Independently scaling compute and storage

## Scale Up System



Processing



Storage

CPU/Drive ratio	-
Performance	+
Capacity	+
Smart Rebuild	-

## Scale Out System



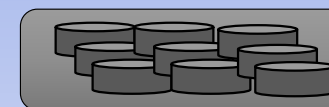
Processing + Storage

CPU/Drive ratio	-
Performance	+
Capacity	-
Smart Rebuild	+

## iguazio Scale Out<sup>x2</sup>








Data nodes  
(Processing)



Storage

CPU/Drive ratio	+
Performance	+
Capacity	+
Smart Rebuild	+

# Why iguazio?

-  Real time analytics on both fresh and historical data
-  Low latency for faster dashboards – in memory speed at the cost of SSD
-  Very high ingestion rate
-  Scale out architecture – enabling real time analytics on large data sets
-  Platform as a service providing cloud experience

# Business benefits - what's in it for me?



## Increase operational efficiency

*Optimize business processes*



## Faster time to market for new services

*Bring new services on-line in less time with greater reliability & security*



## Reduce TCO

*Reduce cost of traditional data center operations while constraining growth of expensive cloud services such as Amazon DynamoDB, Kinesis , S3 , redshift and EMR*



## Increase data engineering efficiency

*Simplify the overall data pipe-line helping engineering to focus on building applications*



*Thank You*

*@Santanu\_Dey*