IDF13

Beyond Hadoop* MapReduce: Processing Big Data

Jeff Parkhurst – Program Director, University Collaboration Office, Intel Corporation

Theodore Willke – Principal Engineer and General Manager, Graph Analytics Operation, Intel Corporation

Christian Black – Datacenter Solutions Architect, Intel Corporation

ACAS001

intel Look Inside.

Agenda

- Introduction
- The Map Reduce Framework and Beyond Hadoop*
- Big Data & Solid State Storage
- Addressing Gaps through University Research

Consumers / Suppliers of Big Data

social

finance

health / medical

Education

These are the Big Data Volume Items

transportation

science

retail

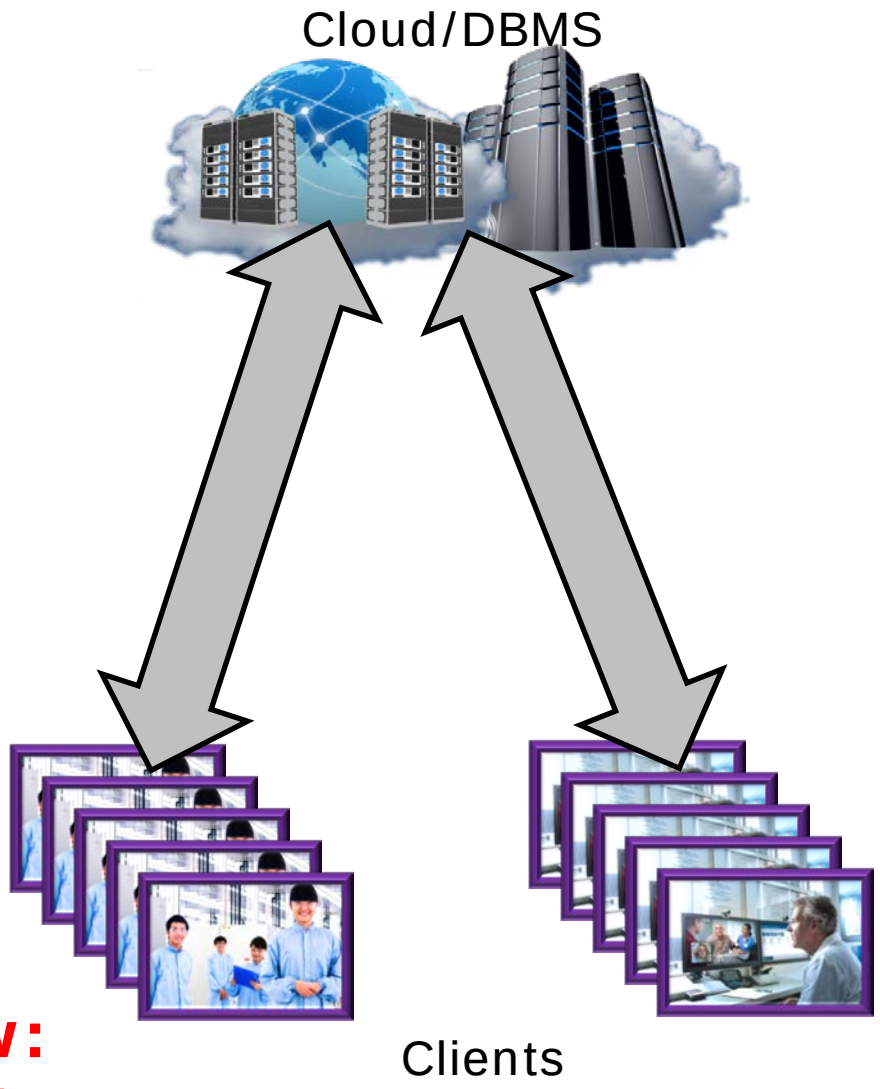
government

entertainment

IDF13

The Quest for More Information

- These users/consumers of BD are non CS
  - Need DM/Info extraction tools for the masses
- Big Data usage sets will continue to grow
- No. of users of big data will grow
  - Along with more users comes more usages
  - Cloud resources and networks will be "taxed"



Network Demand will grow: Need DM/ML at "the edge"

Mind the “Little Things”!!

- Consider future connected devices and uses (IOT)
  - Energy/environmental sensors and home energy mgmt.
  - Traffic cameras and flow optimization
  - Transmission line power sensors and smart grid
  - Human sensors/identification and responsive store/digital signage
  - Communication monitoring and network optimization
- Sensor research has critical mass; systems support growing

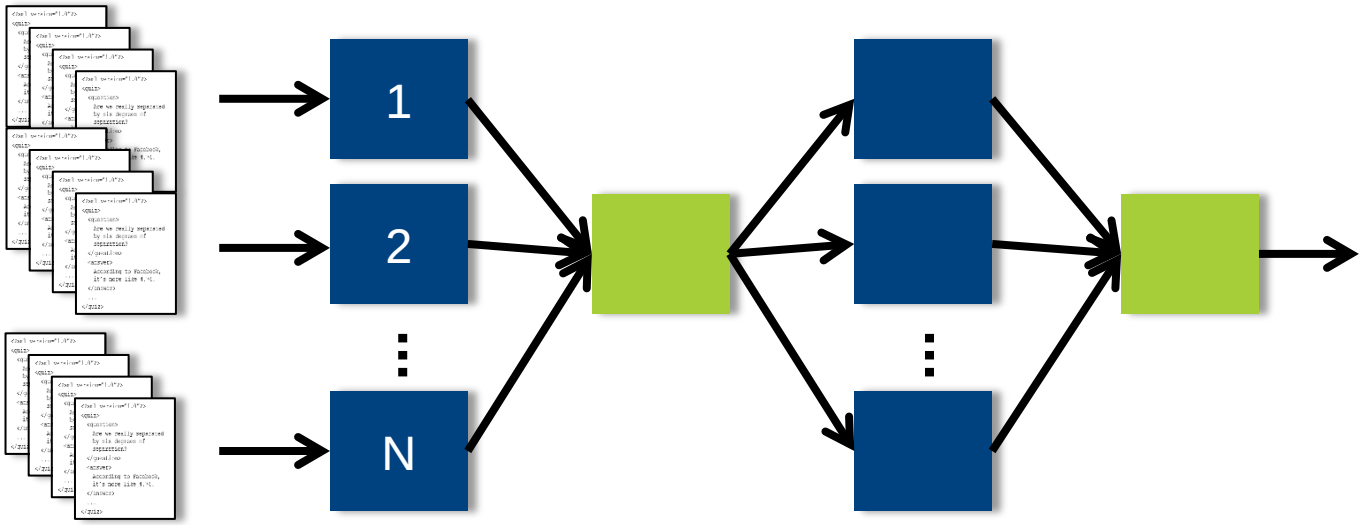
But then there’s Cell Phones & Tablets

Phablets…Oh My!!

IDF13

Throughput-Intensive Workloads

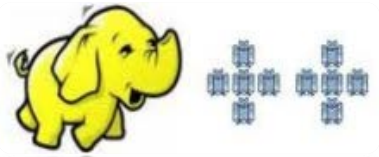
- Have large quantities of course-grained (task) parallelism
- Have inputs that are typically much larger than memory I ›› M x N



- Model: Load → Execute → Shuffle → Execute → Store → (Repeat)
- Metrics: Records/sec, images/sec, edges/sec, etc.

Application Frameworks

- Don’t waste your time developing **new programming paradigms** and **cluster services** (others do this for a living ☺)
- Frameworks solve problems that you’d otherwise solve over, and over, and over again
- Framework selection is driven by the application’s **algorithms** and **processing pipeline**



Progression of Throughput Computing Architectures

Beyond today's implementations of MapReduce

Data-parallel

▼

Data-parallel + Iterative

▼

Graph-parallel + Iterative + …

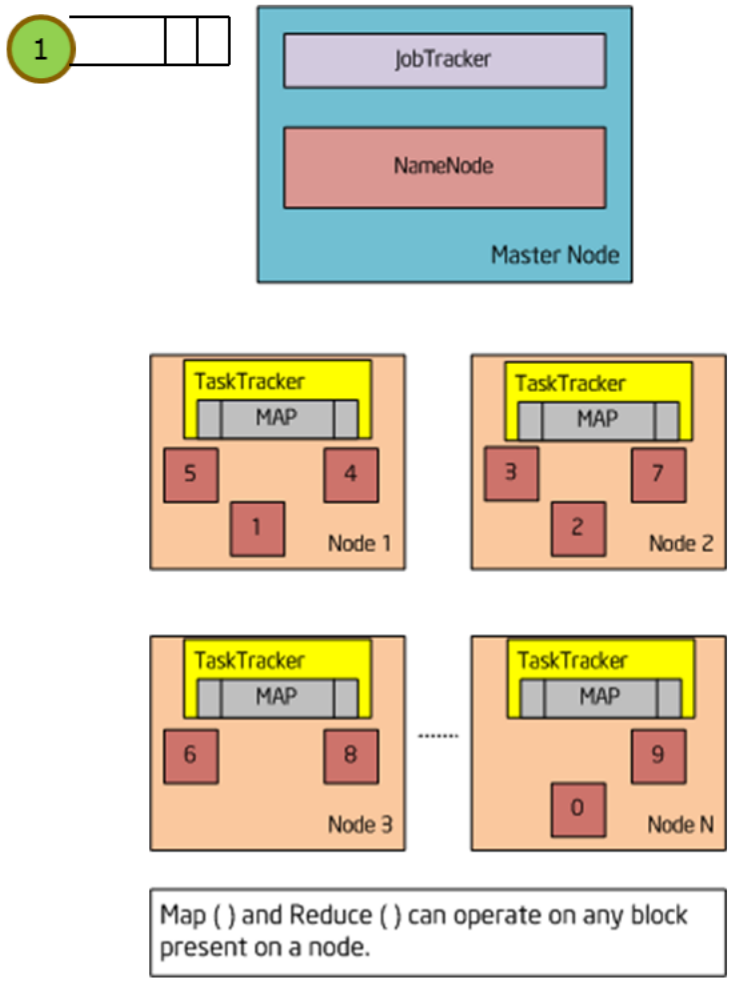
Hadoop* MapReduce Architecture

- Distributed processing of large datasets
- Good for batch processing and data-parallel processing
- Bulk Synchronous Parallel

Hadoop* MapReduce Architecture

- Distributed processing of large datasets
- Good for batch processing and data-parallel processing
- Bulk Synchronous Parallel

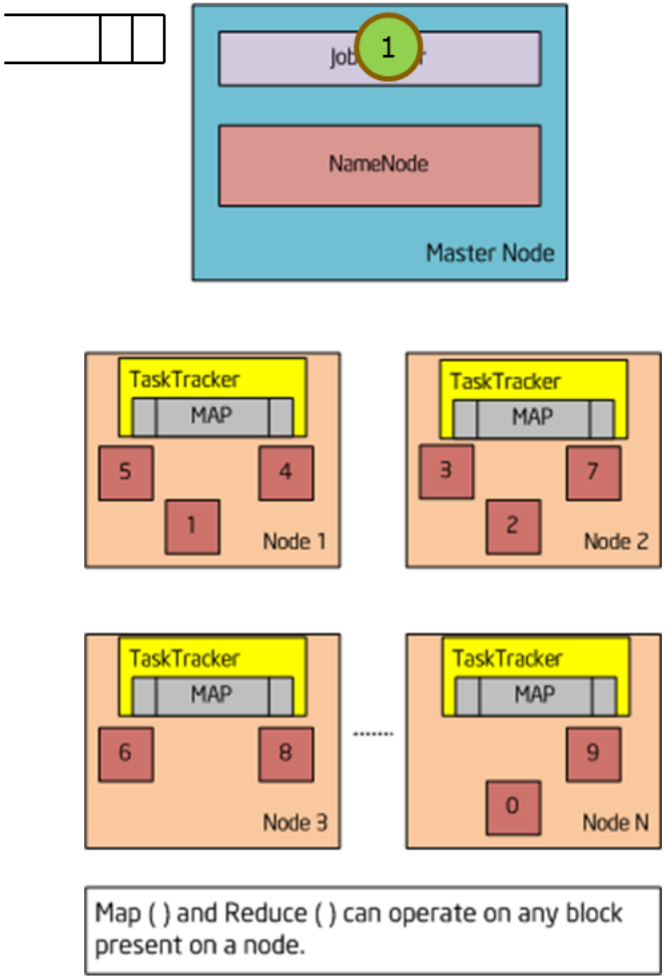
- JobTracker schedules and manages jobs on the NameNode
- TaskTracker executes individual map( ) and reduce( ) tasks on each DataNodes



Hadoop* MapReduce Architecture

- Distributed processing of large datasets
- Good for batch processing and data-parallel processing
- Bulk Synchronous Parallel

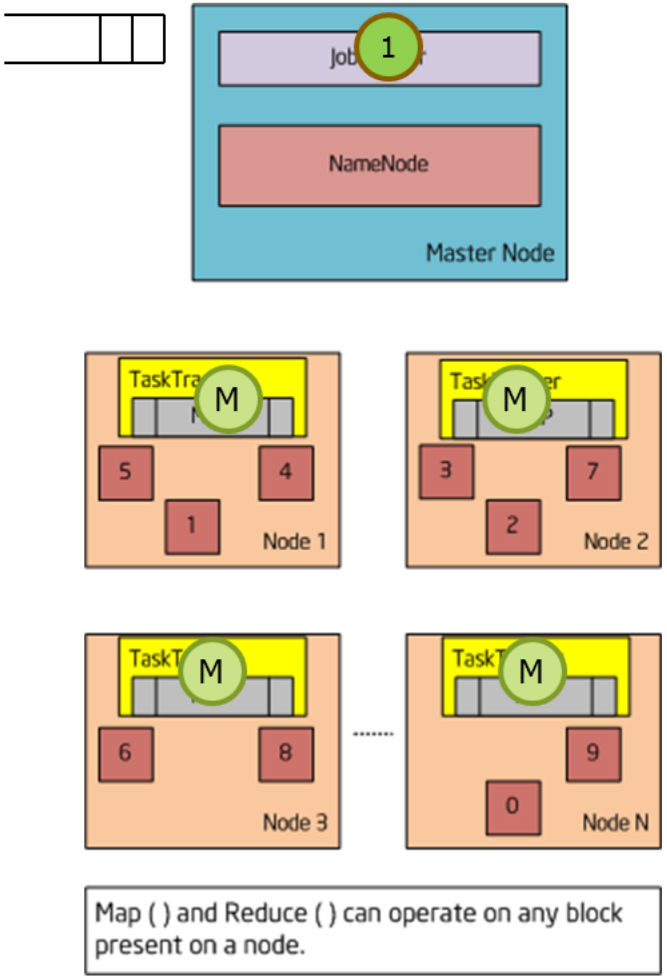
- JobTracker schedules and manages jobs on the NameNode
- TaskTracker executes individual map( ) and reduce( ) tasks on each DataNodes



Hadoop* MapReduce Architecture

- Distributed processing of large datasets
- Good for batch processing and data-parallel processing
- Bulk Synchronous Parallel

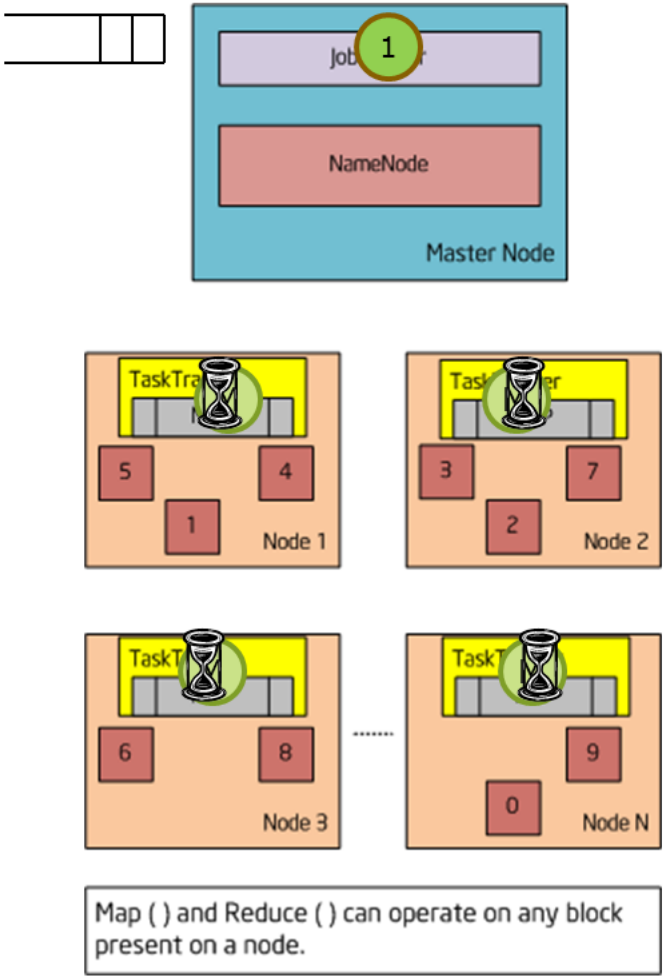
- JobTracker schedules and manages jobs on the NameNode
- TaskTracker executes individual map( ) and reduce( ) tasks on each DataNodes



Hadoop* MapReduce Architecture

- Distributed processing of large datasets
- Good for batch processing and data-parallel processing
- Bulk Synchronous Parallel

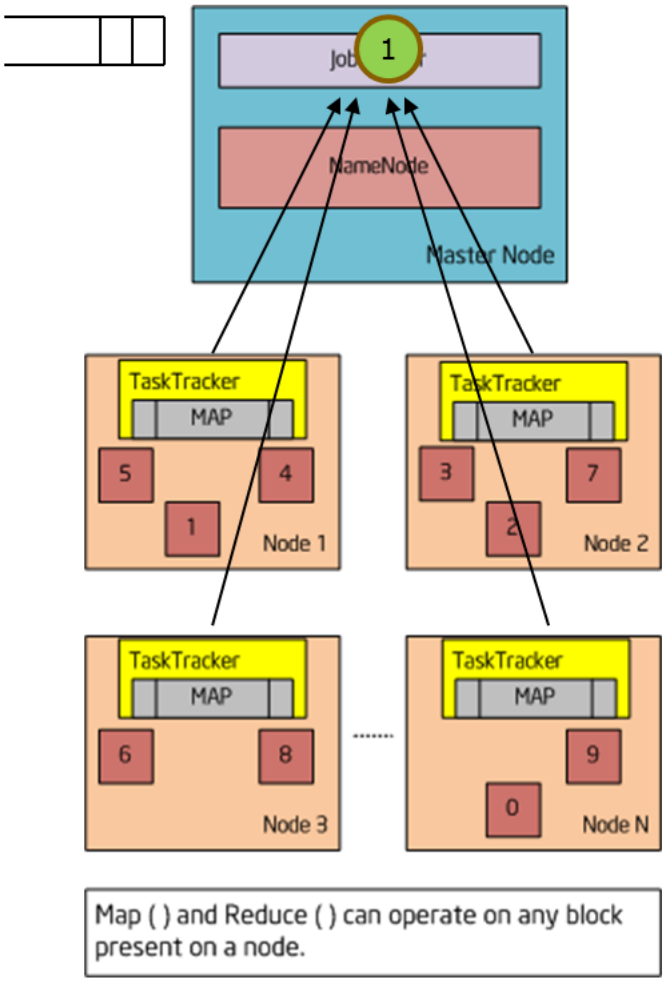
- JobTracker schedules and manages jobs on the NameNode
- TaskTracker executes individual map( ) and reduce( ) tasks on each DataNodes



Hadoop* MapReduce Architecture

- Distributed processing of large datasets
- Good for batch processing and data-parallel processing
- Bulk Synchronous Parallel

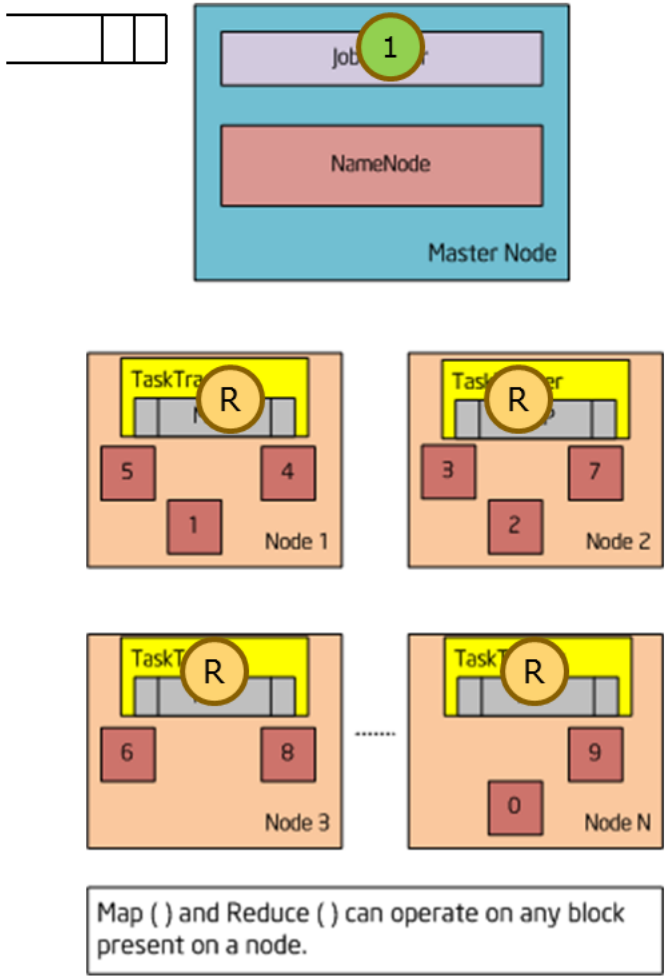
- JobTracker schedules and manages jobs on the NameNode
- TaskTracker executes individual map( ) and reduce( ) tasks on each DataNodes



Hadoop* MapReduce Architecture

- Distributed processing of large datasets
- Good for batch processing and data-parallel processing
- Bulk Synchronous Parallel

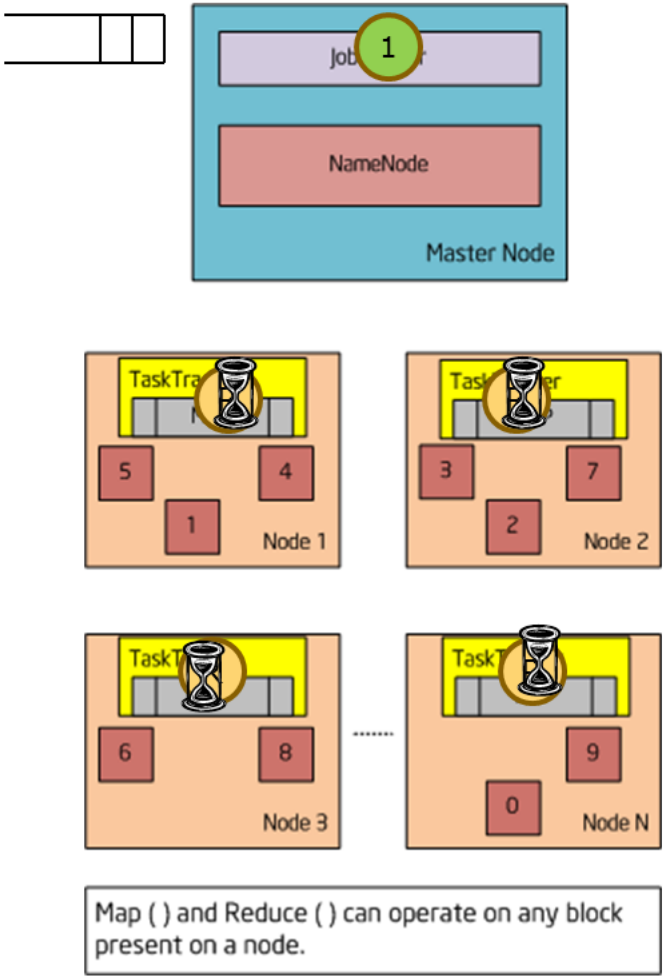
- JobTracker schedules and manages jobs on the NameNode
- TaskTracker executes individual map( ) and reduce( ) tasks on each DataNodes



Hadoop* MapReduce Architecture

- Distributed processing of large datasets
- Good for batch processing and data-parallel processing
- Bulk Synchronous Parallel

- JobTracker schedules and manages jobs on the NameNode
- TaskTracker executes individual map( ) and reduce( ) tasks on each DataNodes

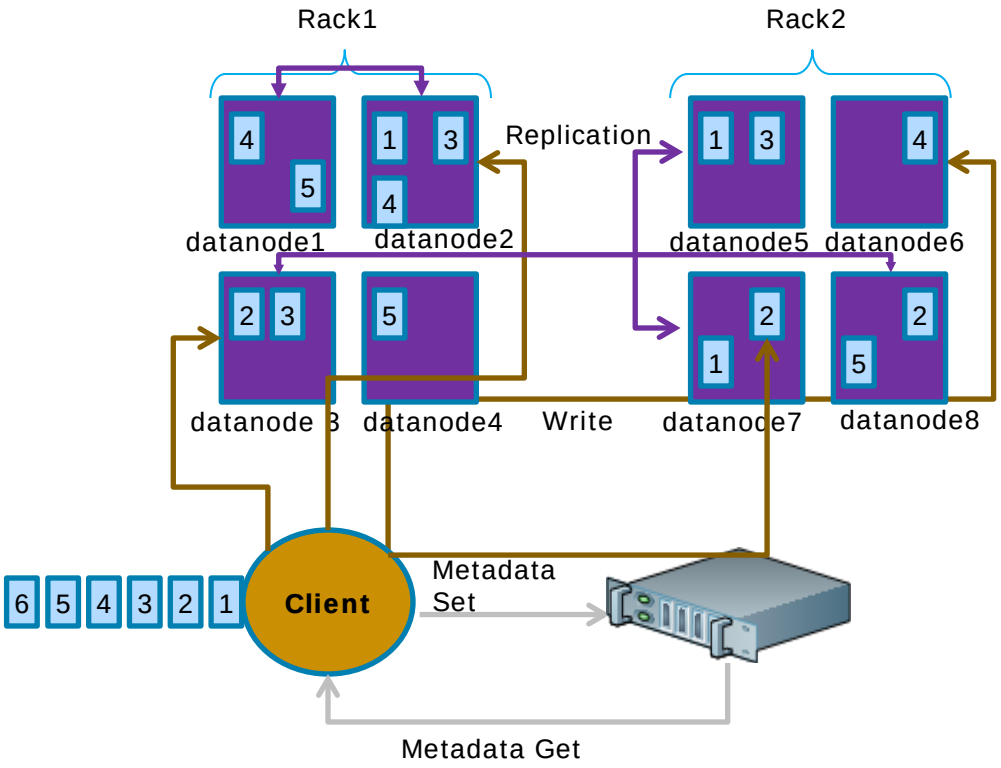


Hadoop* Distributed File System

- A scalable, fault tolerant distributed file system
- Data replication over clusters to address machine failures
- Scales to thousands of nodes

Hadoop* Distributed File System

- A scalable, fault tolerant distributed file system
- Data replication over clusters to address machine failures
- Scales to thousands of nodes
- Files are broken and spread over DataNodes
- Dedicated NameNode to store system metadata



Data-parallel

▼

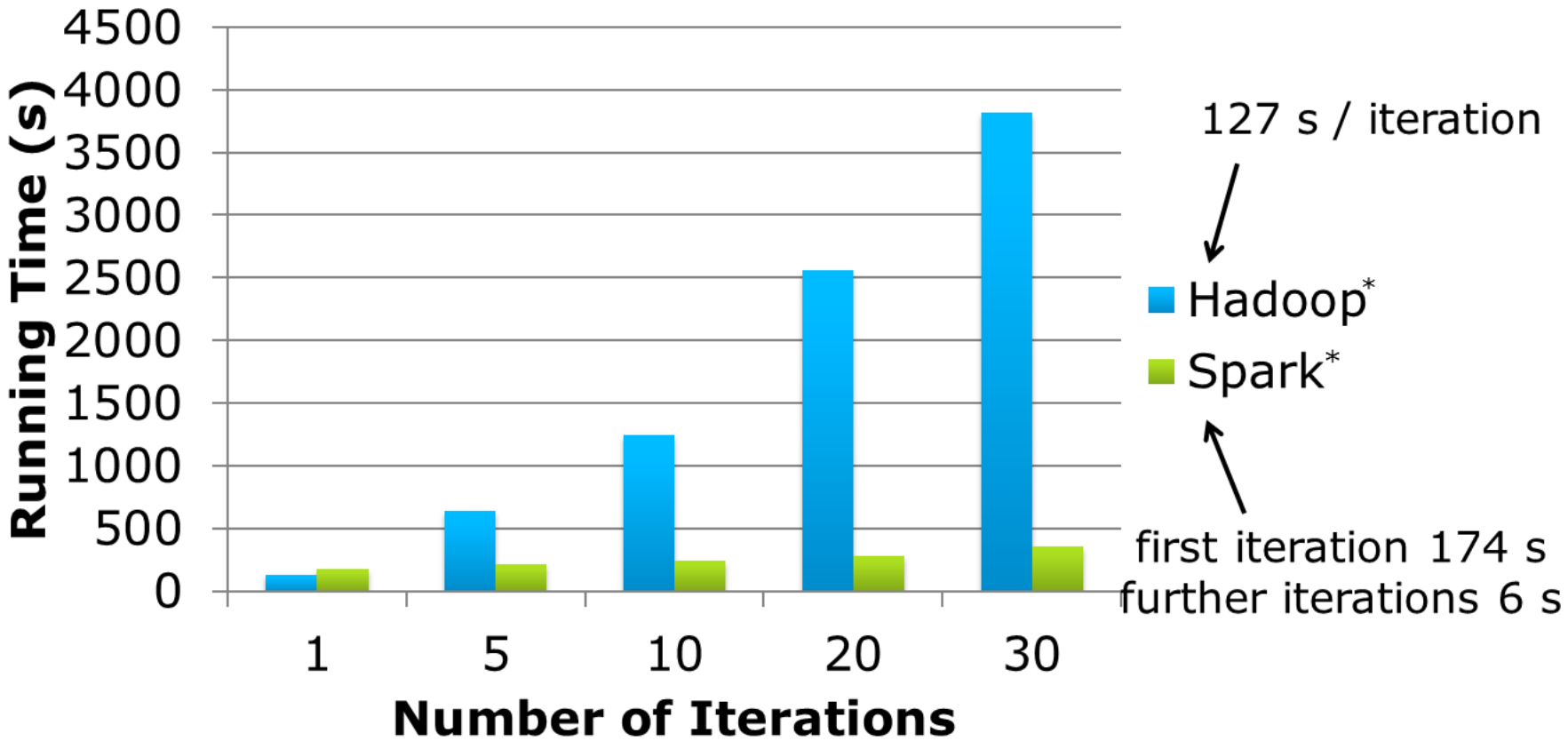
Data-parallel + Iterative

▼

Graph-parallel + Iterative + …

Spark*

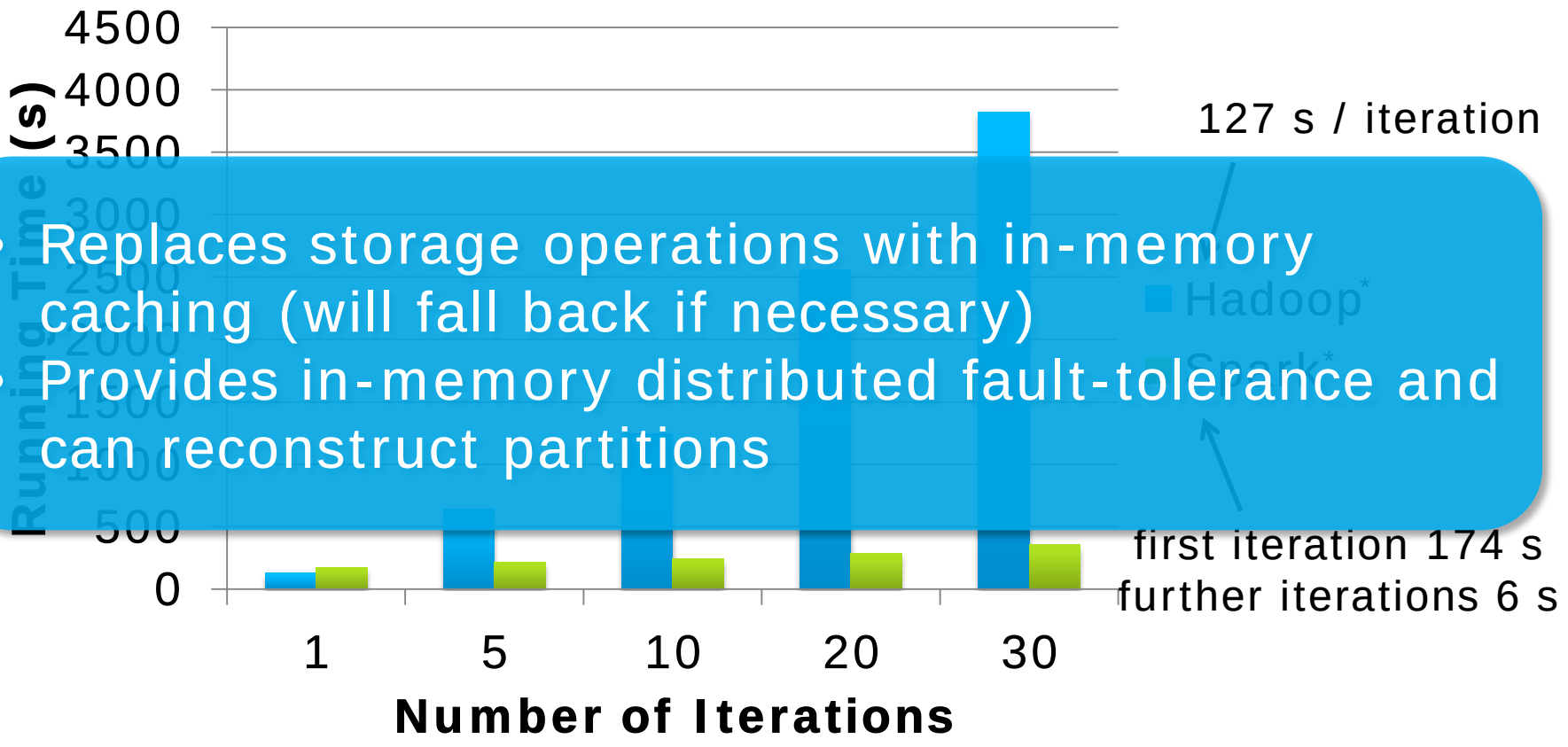
Fast, Interactive, Language-Integrated Cluster Computing



Zaharia et al. UC Berkeley. Retrieved from www.spark-project.org.

Spark*

Fast, Interactive, Language-Integrated Cluster Computing



- Replaces storage operations with in-memory caching (will fall back if necessary)
- Provides in-memory distributed fault-tolerance and can reconstruct partitions

Zaharia et al. UC Berkeley. Retrieved from www.spark-project.org.

Data-parallel

▼

Data-parallel + Iterative

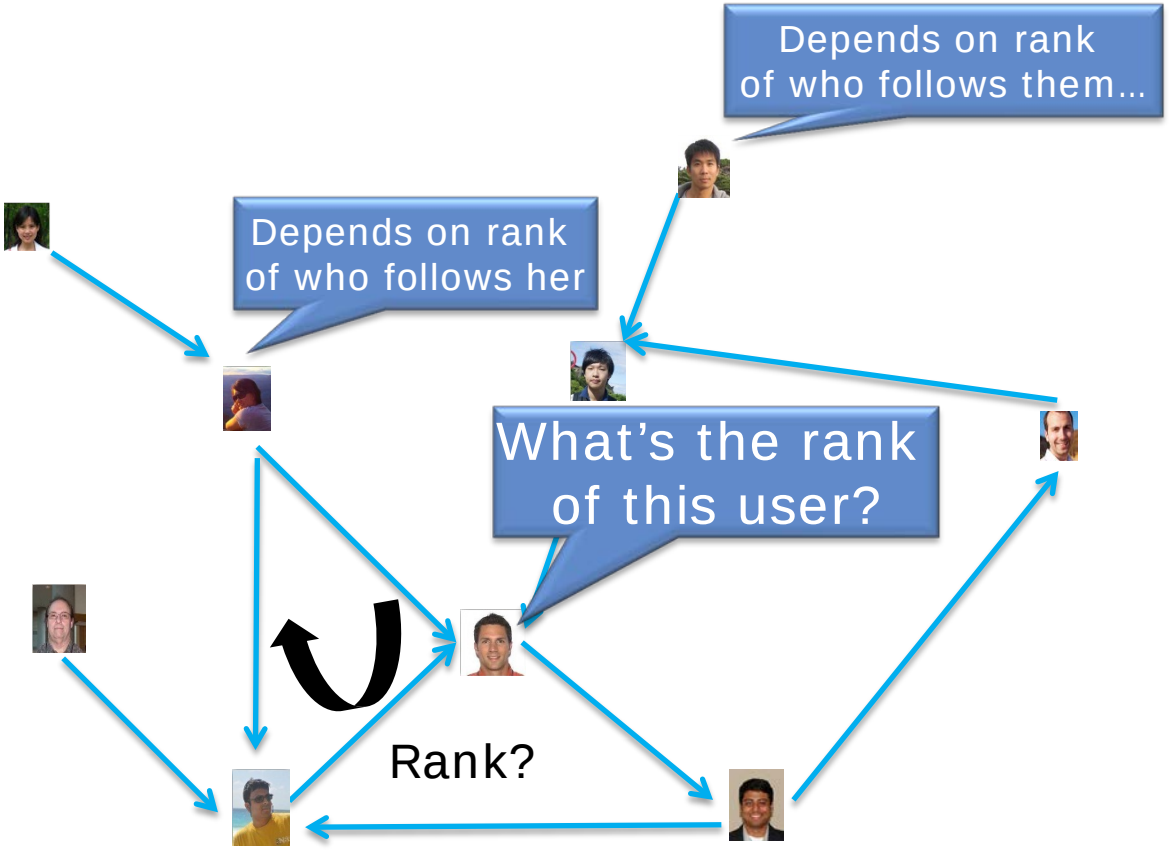
▼

Graph-parallel + Iterative + …

Graph-Based Processing:
Joint Work With the Intel Science & Technology Center for Cloud Computing

A Simple Large-Scale Graph Problem

How many people are pointing to you and what's their relative importance?

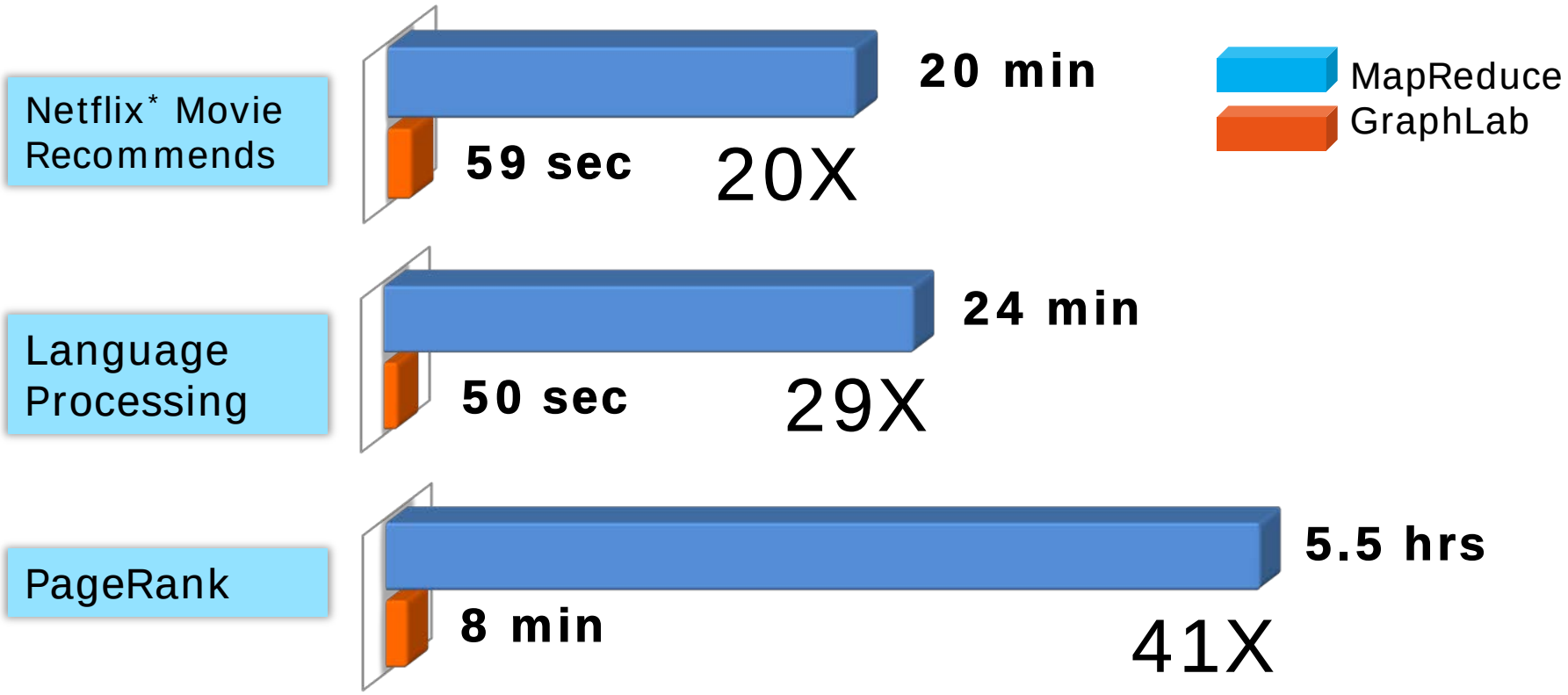


Loops in graph – Iterate!

Graphics source: [Joseph Gonzalez (CMU)]

Graph Engine Performance on Compute Clusters

Fast Predictions will enable new usages

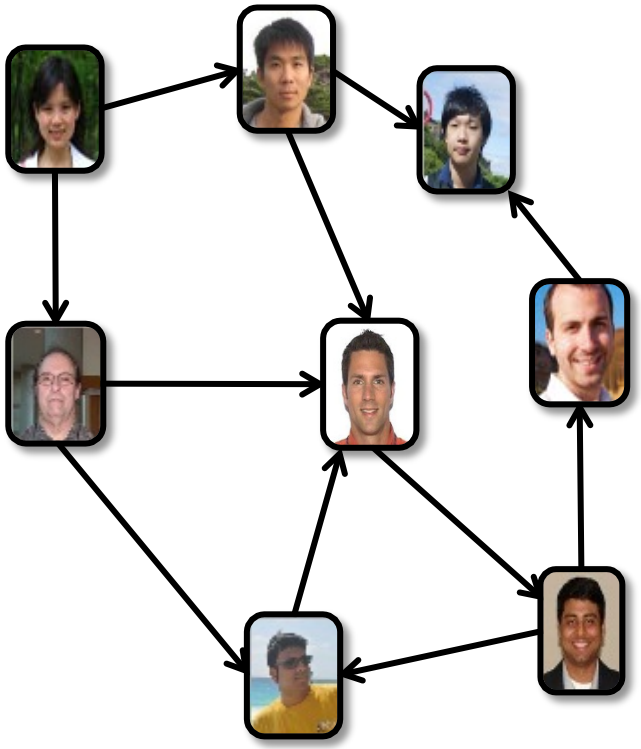


Graph Engines deliver significant speedup for a wide range of applications

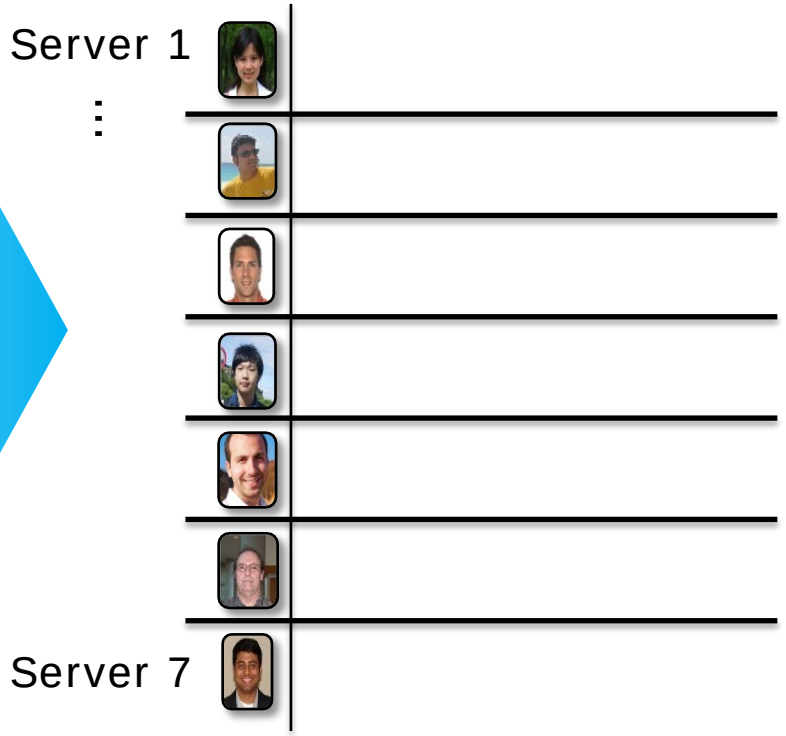
System configuration for Netflix results in backup.
Language Processing (NER) results from Low et al., "Distributed GraphLab: A Framework for Machine Learning and Data Mining in the Cloud," VLDB 2012.
PageRank results from Guestrin et al., "GraphLab 2: A Distributed Abstraction for Large-Scale Machine Learning," presented at GraphLab Workshop 2012.

MapReduce Challenged

Example Graph

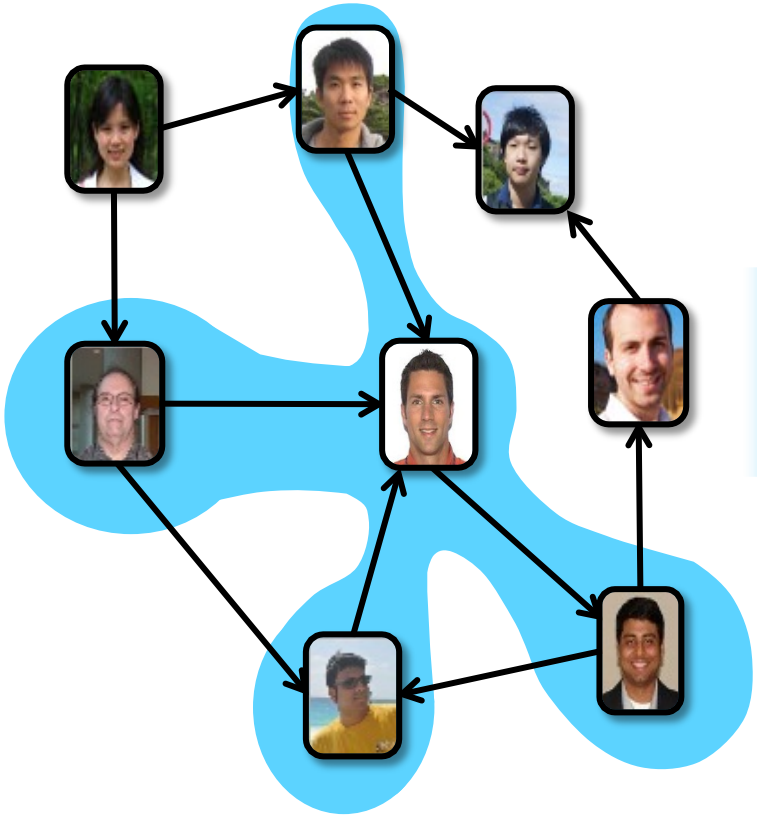


MapReduce Data
Partitioning & Replication

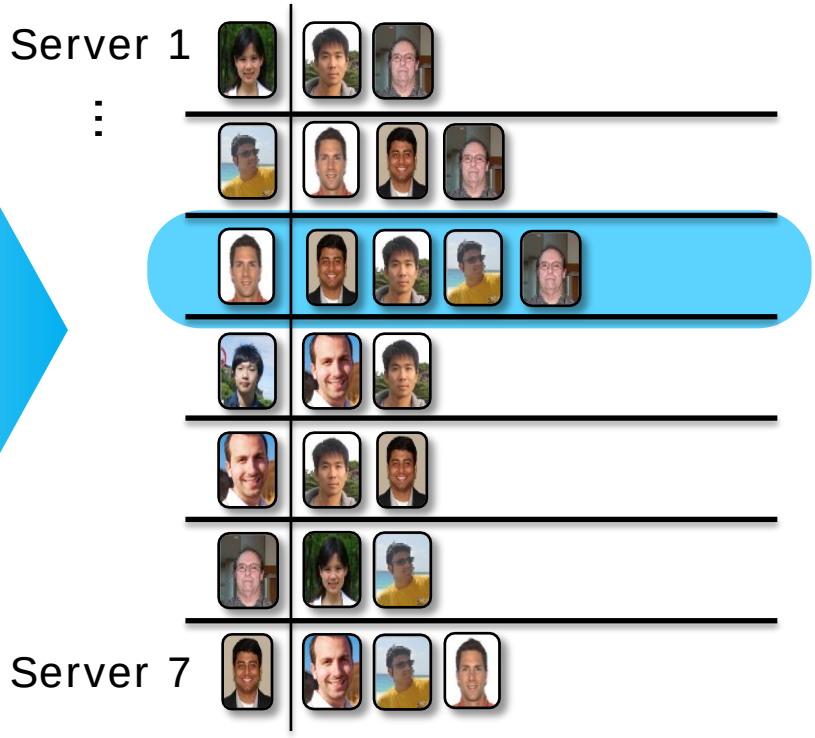


MapReduce Challenged

Example Graph

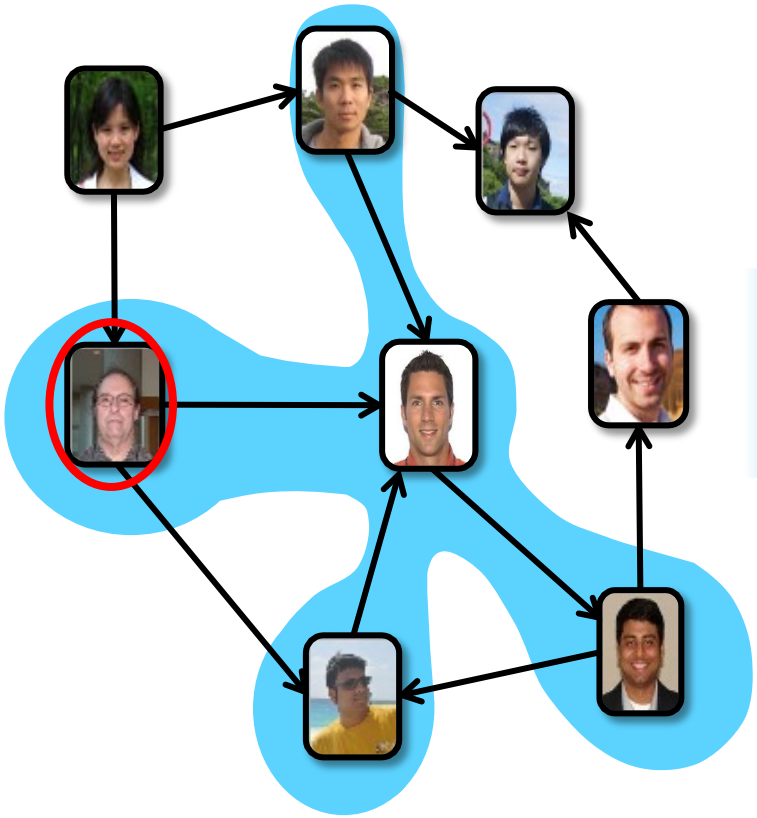


MapReduce Data
Partitioning & Replication

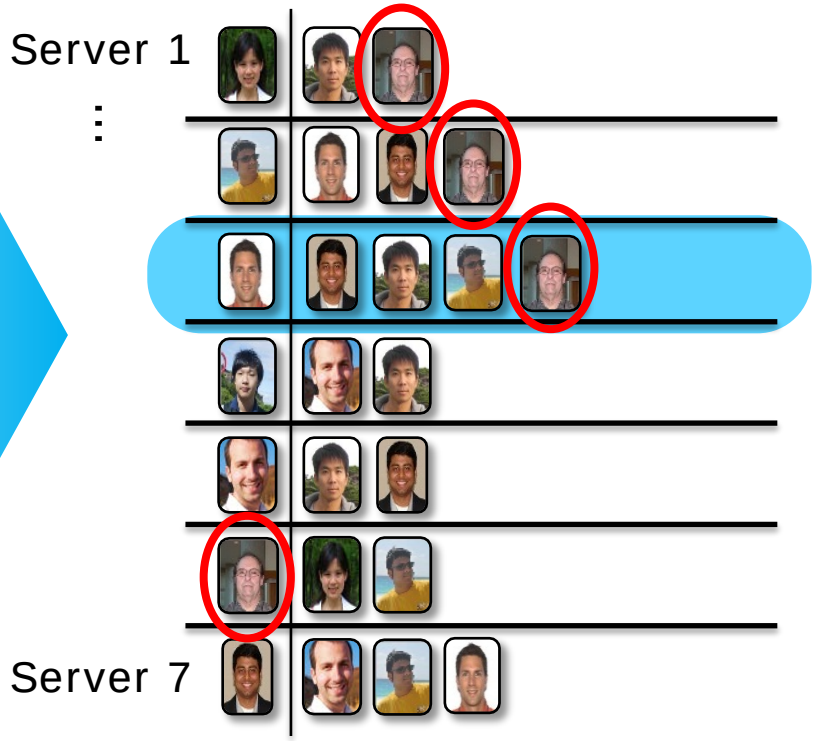


MapReduce Challenged

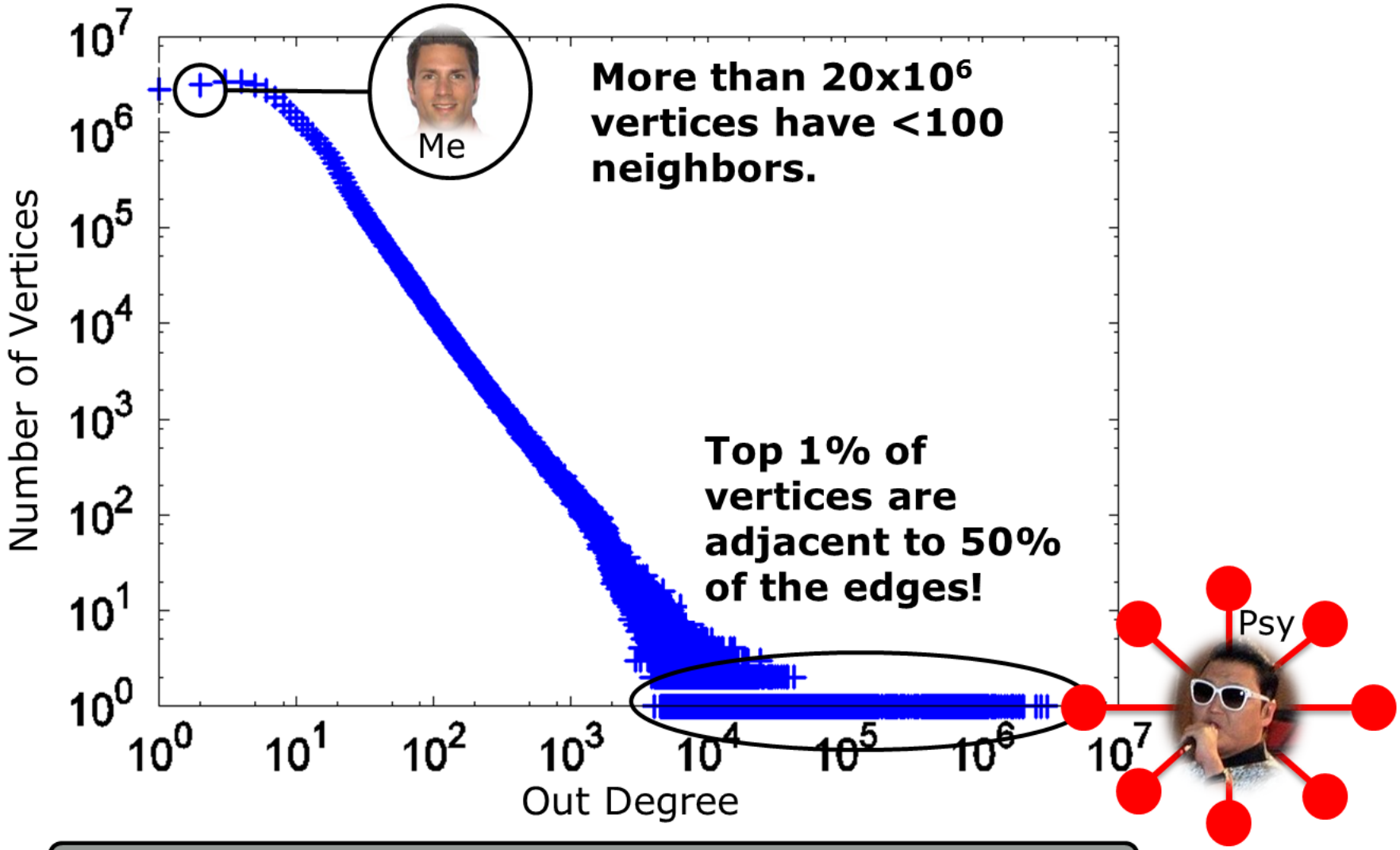
Example Graph



MapReduce Data
Partitioning & Replication



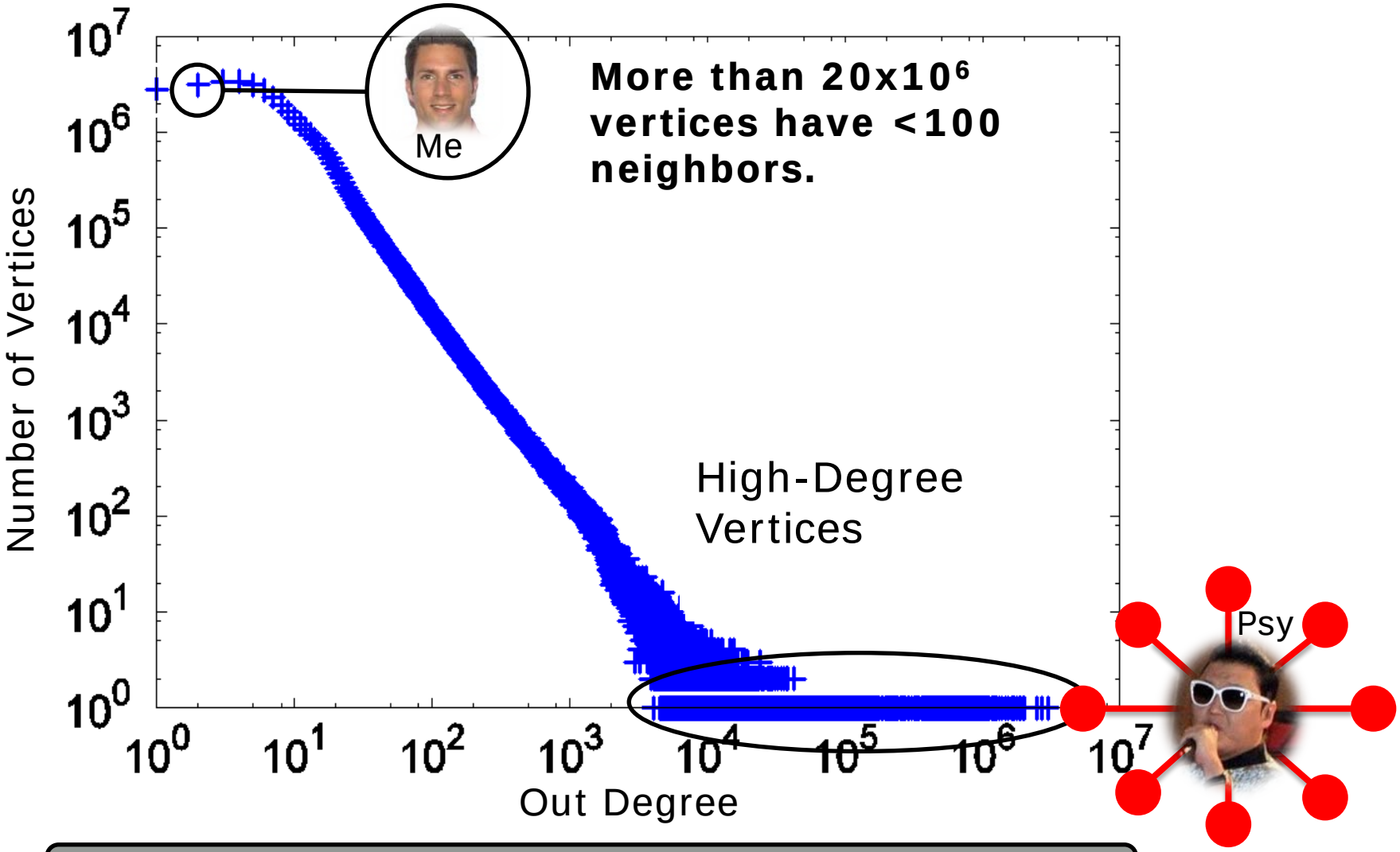
Complicating Things Further…



Power-law graphs = highly uneven processing!

Image source: [Wikipedia] [cmu.edu/~pegasus]

Complicating Things Further…



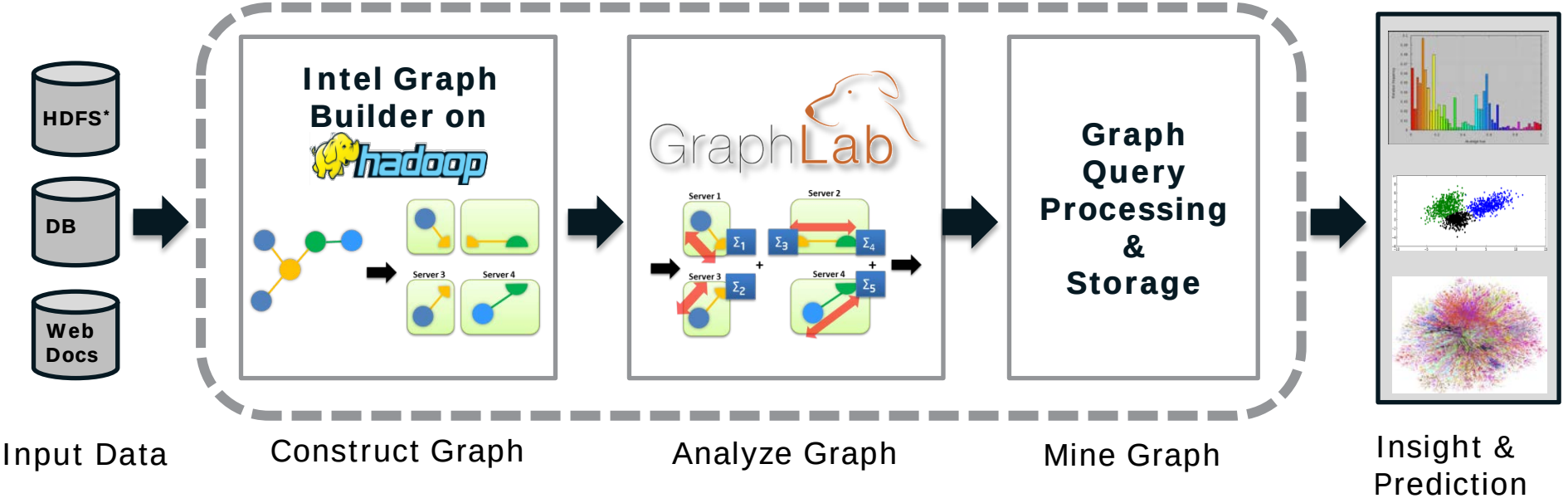
Power-law graphs = highly uneven processing!

Image source: [Wikipedia] [cmu.edu/~pegasus]

MapReduce in Summary

- Difficult to load balance graph problems
- Lots of data replication for independence
- Programmers must reimagine problems – not a natural abstraction
- And, it was not designed for iterative computation and stores everything away at each step

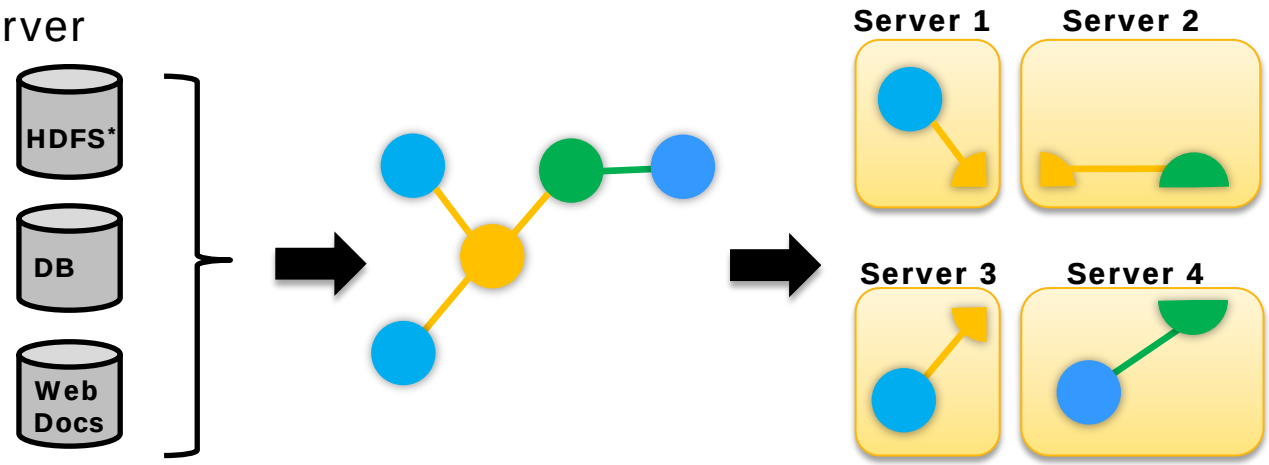
Graph Analytics: A System Perspective



Graph Technologies for Speed and Scaling

Intel Graph Builder constructs Big Data Graphs

- Leverages Intel Hadoop* to extract key features
- Minimizes data replication and communications by making smart cuts in the graph
- Balances compute effort by placing the same number of *relationships* on each server



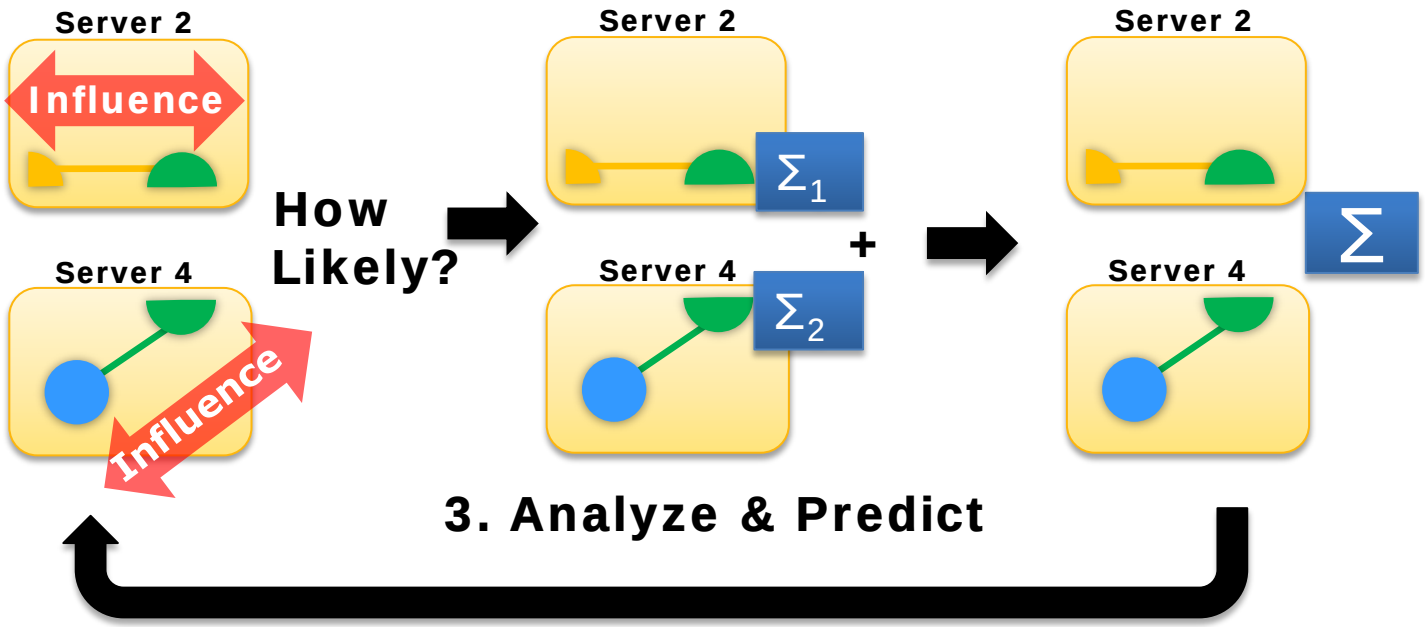
1. Input Data **2. Distribute & Organize**

Extracts timely insights that are invisible to current solutions

Graph Technologies for Speed and Scaling

Graph-parallel engines and algorithms provide significant speed advantage

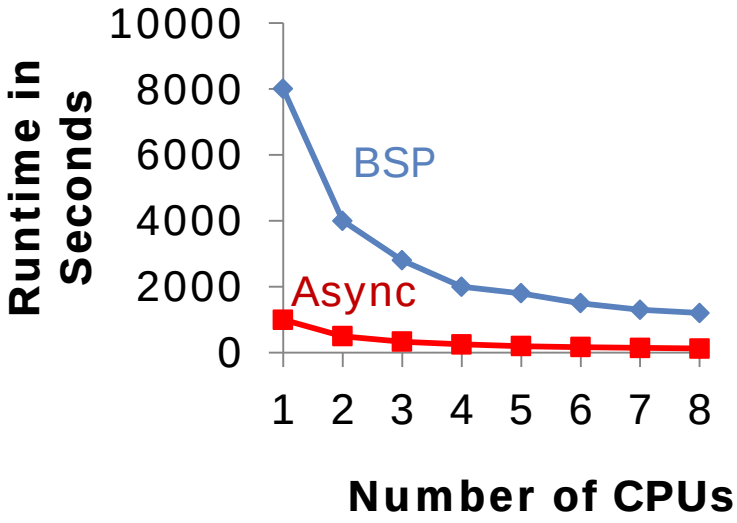
- Algorithms iterate asynchronously over relationships on each machine
- Subtotals are calculated and shared asynchronously



Asynchronous Graph Engines for disruptive speed

Other Graph Engines

- Bulk Synchronous Processing (BSP) Graph-Parallel
  - Giraph on Hadoop* (Inspired by Google* Pregel)
  - Dryad (Microsoft* Research)
  - Apache* Hama* on Hadoop (Twitter*)



- Asynchronous Graph-Parallel
  - Galois (UT Austin) → Edge partitioning
  - GraphLab (CMU) → Vertex partitioning

Asynchronous frameworks have an edge but are difficult to program

Graphics source: [Joseph Gonzalez (CMU)]

This is just the beginning…

For graph analytics to be practical, the open source ecosystem must fully embrace the challenge.

- Better glue between graph processing components
- Complete solutions for graph-based OLTP + OLAP
- More productive programming languages and data workflow tools

Agenda

- Introduction
- The Map Reduce Framework and Beyond Hadoop*
- Big Data & Solid State Storage
- Addressing Gaps through University Research

Big Data & Solid State Storage

- Christian Black

Intel® SSD Data Center Family

Marketing Pitch

Intel® 910 Series

Workload Acceleration
Highest Performance
High Endurance
PCI Express* (PCIe) X4
[1]Up to 180K/75K IOPs 4K Rand R/W
800GB: Up to 10PB with HET

Intel® DC S3700 Series

Enterprise Performance
High Write Performance
High Endurance
SATA 6Gb
[1]Up to 75K/36K IOPs 4K Rand R/W
800GB: 10DW/day with HET

Intel® DC S3500 Series

Enterprise Mainstream
High Performance
Standard Endurance
[1]Up to 72K/10K IOPs 4K Rand R/W
800GB: 0.3 DW/day

1 Data based on Intel® SSD DC S3700 and DC 3500 Series data sheets, see http://www.intel.com/go/ssd for detailed products specifications *HET = High Endurance Technology

Intel® SSDs and Hadoop* - One Example

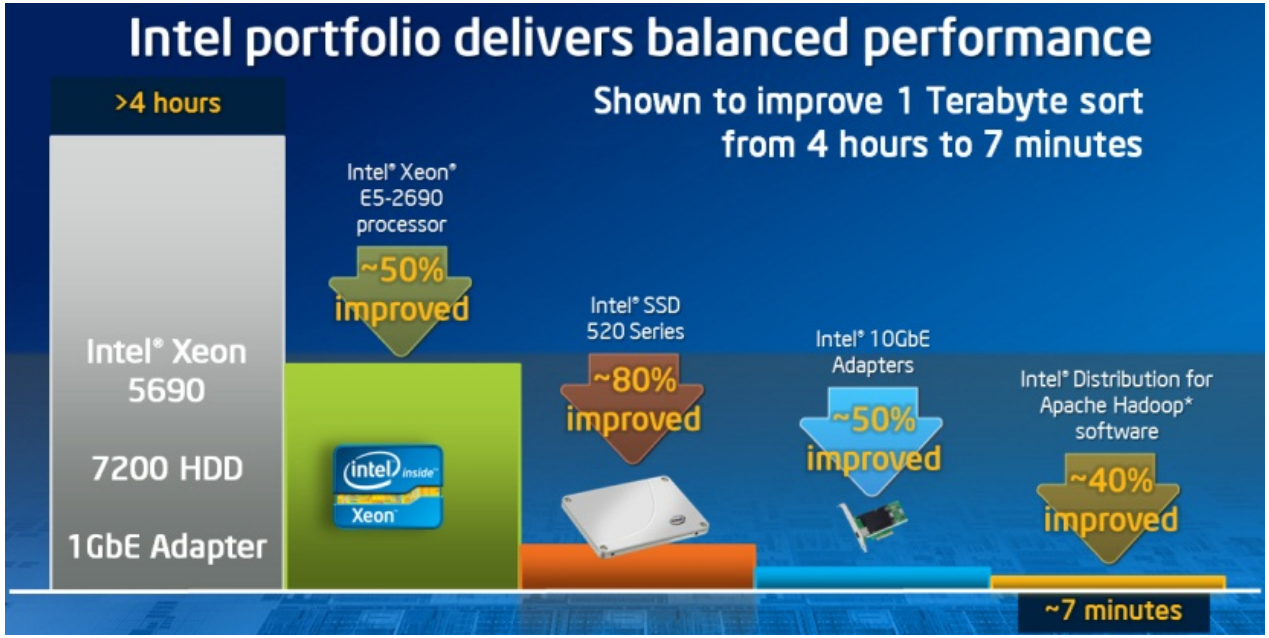
- Hadoop* from the disk perspective
  - 128MB-256MB sequential IO operations
  - Write once, read many, occasional rebalance
  - Perfect for $.04/GB 7k RMP spinning rust @ 130-150MB/Sec
- Temp/Intermediate data creates disk contention
- SSDs provide SSD 450-500MB/Sec Sustained
  - Intel internal testing shows 'pure SSDs' provide up to 80%[1] performance increase for 1TB Terasort* in Hadoop*
- $1-$2.35/GB for SSD…
  - Tough sell with typical enterprise IT financial constraints

Complete replacement of Big Data spinning rust with SSDs is both impractical and improbable today…

1 Results have been estimated based on internal Intel analysis and are provided for informational purposes only. Any difference in system hardware or software design or configuration may affect actual performance. Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as Terasort or HiBench, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. Source: Internal Testing Configuration: Intel ® Xeon 5600 & E5, 7200 RPM HDD & Intel ® 520 Series SSD, Intel ® 1GbE and 10Gb Ethernet, and open source Apache Hadoop * & Intel ® Distribution for Apache Hadoop*

Intel® SSDs and Hadoop*

- Results from Intel® research white paper
  - http://www.intel.com/content/dam/www/public/us/en/documents/white-papers/big-data-apache-hadoop-technologies-for-results-whitepaper.pdf
- Match SSD to cluster write rate or cluster refresh rate – the right SSD for your workload!

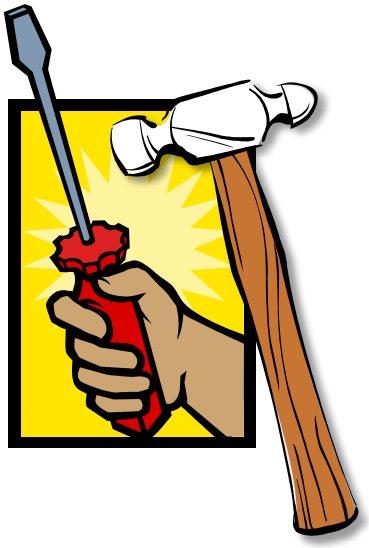


Intel research shows potential benefits in Dev, Real-Time Query, and Temp data use cases…

1 Results have been estimated based on internal Intel analysis and are provided for informational purposes only. Any difference in system hardware or software design or configuration may affect actual performance. Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as Terasort or HiBench, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. Source: Internal Testing Configuration: Intel ® Xeon 5600 & E5, 7200 RPM HDD & Intel ® 520 Series SSD, Intel ® 1GbE and 10Gb Ethernet, and open source Apache Hadoop * & Intel ® Distribution for Apache Hadoop*

SSDs are Tools…

- Specialized
  - $/GB – Traditional HDDs lowest cost/GB
  - $/IOP – SSDs lowest cost/IOP
- Both devices have a place in the DC!
  - Writing applications around HDDs for decades
  - SSD open up new possibilities for applications
    - High Speed, Low Latency, & Random IO
    - Caching, Heat-Based ***Tiering***, & Segmentation
- The Changing Data Center
  - New capabilities - IO closer to CPU
  - Storage tiering & resource balance imperative
  - Knowing ***your*** workload ***absolutely*** matters!
  - Leverage DC tools where appropriate



How can your Big Data application benefit from Solid State Storage, New Platforms, 40GbE?

Stepping into Tomorrow…

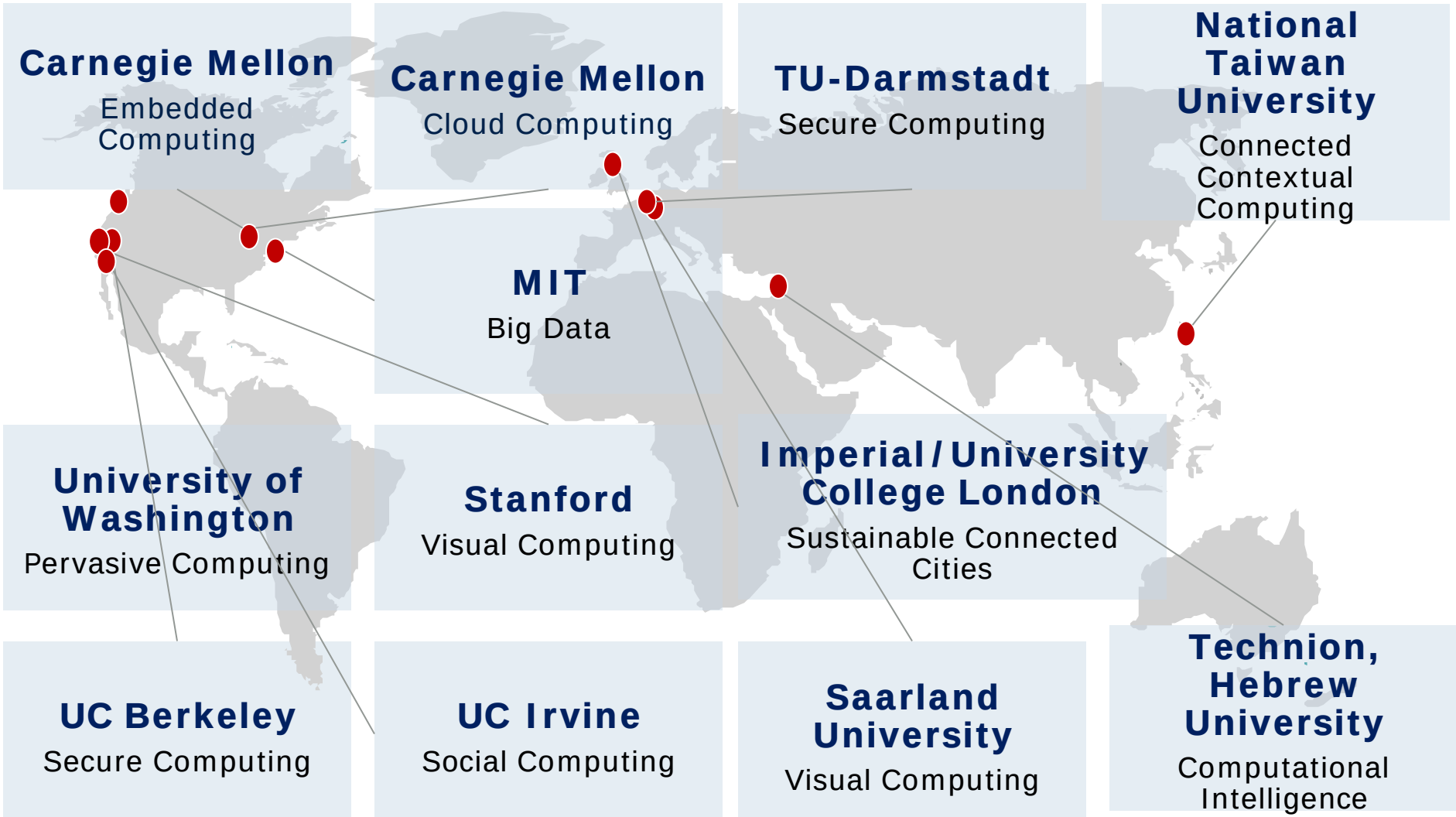
- We've designed applications around spinning rust for decades…
- We've designed applications around volatile memory for decades…
- What if…
  - Applications knew how to leverage solid state specifically?
  - Hadoop* and other Big Data platforms had heat based data tiering?
  - There were multiple levels of volatile and non-volatile RAM?

Application programmers need to start thinking about the "What Ifs" of NV storage/RAM

Agenda

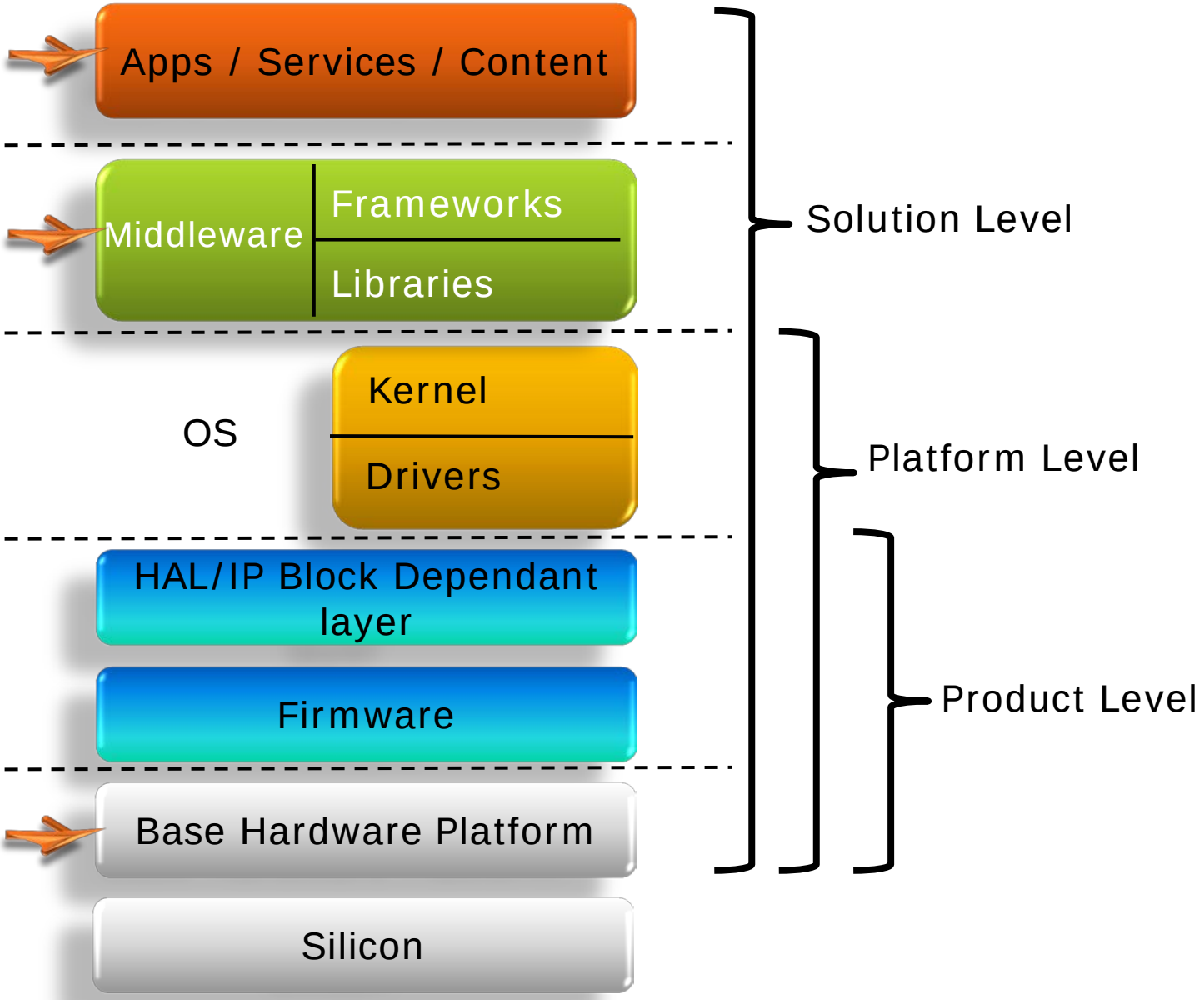
- Introduction
- The Map Reduce Framework and Beyond Hadoop*
- Big Data & Solid State Storage
- Addressing Gaps through University Research

UCO Research Communities Worldwide



>$140M commitment over 5 years. >40 universities. >200 professors.

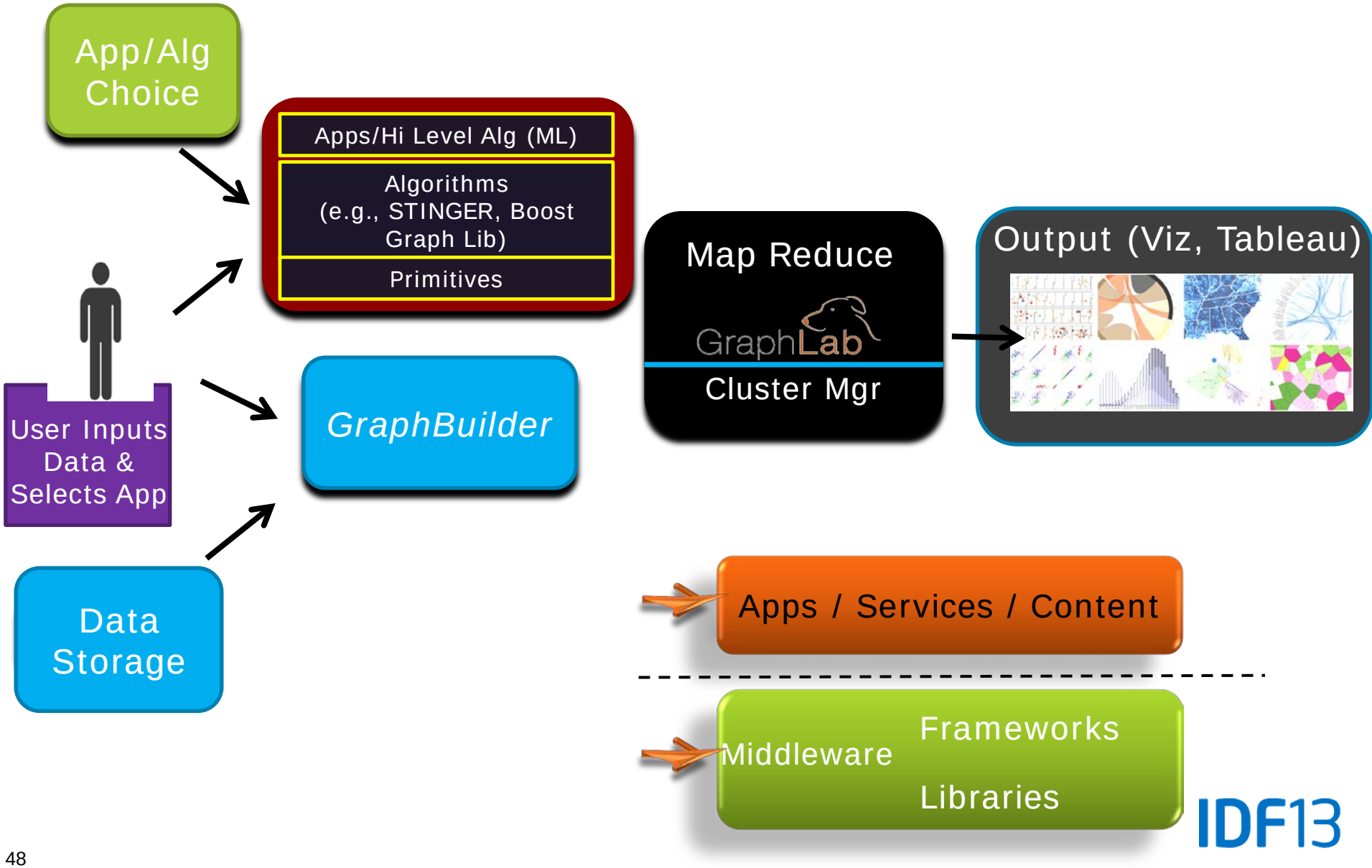
Research Focus: Exploring the Stack



Disruptive Nature of NVM

- Consider Database in main memory
  - Introduce temperature labeled data
    - Those accessed the most labeled as “highest temperature”
  - Introduce a tiered storage system (DRAM, NVM and SSDs)
    - Within memory, you could have compressed data as well
  - Consider that current database in memory techniques rely upon DRAM to flush out partially transferred data
    - With NVM, this can no longer be assumed
- Consider processing streamed data in real time
  - Indexing itself becomes extremely complex
  - Incrementally processing the new data is a hard problem
- Explore an automated algorithm to optimize where to store and process streamed data

Addressing Gaps in Processing Big Data



Research Focus: Processing at the Edge

- Motivation
  - Bandwidth limitations and latency
- Problem Statement
  - Many algorithms/apps for processing Big Data are computationally intensive
    - Image/Object Recognition
    - Scene reconstruction and identification
- Addressing this in ISTC for Embedded Computing
  - Modifying traditional algorithms to work on an embedded platform
  - Hardware/Software Co-Design
  - Application specific accelerators

Summary

- Hadoop* Map Reduce has its limitations
- Graph based analytics can fill these gaps
- New storage technologies can aid in the overall processing/handling of Big Data
- University research addressing gaps

Next Steps

- Consider Graph Based Analytics Algorithms when working with relation based data
- Consider SSDs as part of the overall frameworks storage system for better optimization
- Processing is moving to the edge

Additional Sources of Information

PDF of this presentation is available is available from our Technical Session Catalog: www.intel.com/idfsessionsSF. The URL is on top of Session Agenda Pages in Pocket Guide.

More web based info:

http://www.intel.com/content/www/us/en/research/intel-labs-istc.html

http://graphlab.org/

https://01.org/graphbuilder/

Legal Disclaimer

INFORMATION IN THIS DOCUMENT IS PROVIDED IN CONNECTION WITH INTEL PRODUCTS. NO LICENSE, EXPRESS OR IMPLIED, BY ESTOPPEL OR OTHERWISE, TO ANY INTELLECTUAL PROPERTY RIGHTS IS GRANTED BY THIS DOCUMENT. EXCEPT AS PROVIDED IN INTEL'S TERMS AND CONDITIONS OF SALE FOR SUCH PRODUCTS, INTEL ASSUMES NO LIABILITY WHATSOEVER AND INTEL DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY, RELATING TO SALE AND/OR USE OF INTEL PRODUCTS INCLUDING LIABILITY OR WARRANTIES RELATING TO FITNESS FOR A PARTICULAR PURPOSE, MERCHANTABILITY, OR INFRINGEMENT OF ANY PATENT, COPYRIGHT OR OTHER INTELLECTUAL PROPERTY RIGHT.

A "Mission Critical Application" is any application in which failure of the Intel Product could result, directly or indirectly, in personal injury or death. SHOULD YOU PURCHASE OR USE INTEL'S PRODUCTS FOR ANY SUCH MISSION CRITICAL APPLICATION, YOU SHALL INDEMNIFY AND HOLD INTEL AND ITS SUBSIDIARIES, SUBCONTRACTORS AND AFFILIATES, AND THE DIRECTORS, OFFICERS, AND EMPLOYEES OF EACH, HARMLESS AGAINST ALL CLAIMS COSTS, DAMAGES, AND EXPENSES AND REASONABLE ATTORNEYS' FEES ARISING OUT OF, DIRECTLY OR INDIRECTLY, ANY CLAIM OF PRODUCT LIABILITY, PERSONAL INJURY, OR DEATH ARISING IN ANY WAY OUT OF SUCH MISSION CRITICAL APPLICATION, WHETHER OR NOT INTEL OR ITS SUBCONTRACTOR WAS NEGLIGENT IN THE DESIGN, MANUFACTURE, OR WARNING OF THE INTEL PRODUCT OR ANY OF ITS PARTS.

Intel may make changes to specifications and product descriptions at any time, without notice. Designers must not rely on the absence or characteristics of any features or instructions marked "reserved" or "undefined". Intel reserves these for future definition and shall have no responsibility whatsoever for conflicts or incompatibilities arising from future changes to them. The information here is subject to change without notice. Do not finalize a design with this information.

The products described in this document may contain design defects or errors known as errata which may cause the product to deviate from published specifications. Current characterized errata are available on request.

Contact your local Intel sales office or your distributor to obtain the latest specifications and before placing your product order.

Copies of documents which have an order number and are referenced in this document, or other Intel literature, may be obtained by calling 1-800-548-4725, or go to: http://www.intel.com/design/literature.htm

Intel, Xeon, Look Inside and the Intel logo are trademarks of Intel Corporation in the United States and other countries.

*Other names and brands may be claimed as the property of others.

IDF13

Risk Factors

The above statements and any others in this document that refer to plans and expectations for the third quarter, the year and the future are forward-looking statements that involve a number of risks and uncertainties. Words such as “anticipates,” “expects,” “intends,” “plans,” “believes,” “seeks,” “estimates,” “may,” “will,” “should” and their variations identify forward-looking statements. Statements that refer to or are based on projections, uncertain events or assumptions also identify forward-looking statements. Many factors could affect Intel’s actual results, and variances from Intel’s current expectations regarding such factors could cause actual results to differ materially from those expressed in these forward-looking statements. Intel presently considers the following to be the important factors that could cause actual results to differ materially from the company’s expectations. Demand could be different from Intel's expectations due to factors including changes in business and economic conditions; customer acceptance of Intel’s and competitors’ products; supply constraints and other disruptions affecting customers; changes in customer order patterns including order cancellations; and changes in the level of inventory at customers. Uncertainty in global economic and financial conditions poses a risk that consumers and businesses may defer purchases in response to negative financial events, which could negatively affect product demand and other related matters. Intel operates in intensely competitive industries that are characterized by a high percentage of costs that are fixed or difficult to reduce in the short term and product demand that is highly variable and difficult to forecast. Revenue and the gross margin percentage are affected by the timing of Intel product introductions and the demand for and market acceptance of Intel's products; actions taken by Intel's competitors, including product offerings and introductions, marketing programs and pricing pressures and Intel’s response to such actions; and Intel’s ability to respond quickly to technological developments and to incorporate new features into its products. The gross margin percentage could vary significantly from expectations based on capacity utilization; variations in inventory valuation, including variations related to the timing of qualifying products for sale; changes in revenue levels; segment product mix; the timing and execution of the manufacturing ramp and associated costs; start-up costs; excess or obsolete inventory; changes in unit costs; defects or disruptions in the supply of materials or resources; product manufacturing quality/yields; and impairments of long-lived assets, including manufacturing, assembly/test and intangible assets. Intel's results could be affected by adverse economic, social, political and physical/infrastructure conditions in countries where Intel, its customers or its suppliers operate, including military conflict and other security risks, natural disasters, infrastructure disruptions, health concerns and fluctuations in currency exchange rates. Expenses, particularly certain marketing and compensation expenses, as well as restructuring and asset impairment charges, vary depending on the level of demand for Intel's products and the level of revenue and profits. Intel’s results could be affected by the timing of closing of acquisitions and divestitures. Intel's results could be affected by adverse effects associated with product defects and errata (deviations from published specifications), and by litigation or regulatory matters involving intellectual property, stockholder, consumer, antitrust, disclosure and other issues, such as the litigation and regulatory matters described in Intel's SEC reports. An unfavorable ruling could include monetary damages or an injunction prohibiting Intel from manufacturing or selling one or more products, precluding particular business practices, impacting Intel’s ability to design its products, or requiring other remedies such as compulsory licensing of intellectual property. A detailed discussion of these and other factors that could affect Intel’s results is included in Intel’s SEC filings, including the company’s most recent reports on Form 10-Q, Form 10-K and earnings release.

Rev. 7/17/13

IDF13

Backup

System Configuration for Netflix* Results

- 16 machines
- 64GB of memory
- 2 CPUs (Intel(R) Xeon(R) CPU E5-2670 @ 2.60GHz [8 cores each])
- 4 x 1 TB HDDs
- 10Gb Ethernet interconnect