DataCenters & Supercomputers II

Master Data Science

Noviembre 2017

Ramón Beivide

Universidad de Cantabria

Tema 2 Outline

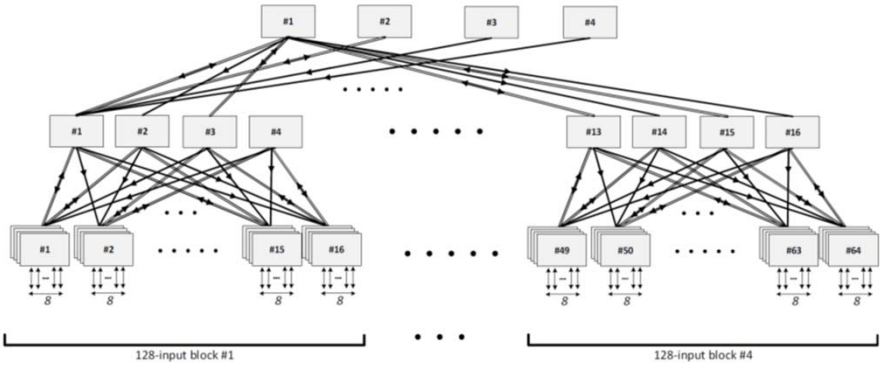
- 1. Paper on Clos for NNs
- 2. Warehouse-Scale Computers (WSCs)
- 3. Architectural Overview of WSCs
  - Networks, Storage & Processing power
- 4. Benchmarks: Graph500 (& Top500)

Clos networks for NoCs

IEEE TRANSACTIONS ON COMPUTERS

Customizing Clos Network-on-Chip for Neural Networks

Reza Hojabr, Mehdi Modarressi, Masoud Daneshtalab, Ali Yasoubi, and Ahmad Khonsari



Tema 2 Outline

- 1. Paper on Clos for NNs
- **2. Warehouse-Scale Computers (WSCs)**
- 3. Architectural Overview of WSCs
- 4. Benchmarks: Graph500 (& Top500)

2. Warehouse-scale computing

The ARPANET is about to turn forty, and the WWW is twenty

- Early Internet services were mostly informational, today many Web applications offer services that previously resided in the client, including email, photo and video storage and office applications.
- User experience improvements
  - no configuration or backups needed
  - a browser is all you need
- Advantages it offers to vendors
  - faster application development
  - servers may be shared among thousands of active users
  - easier to manage than the desktop or laptop equivalent

2. WSCs vs Enterprise Traditional DCs

- Traditional DCs host a number of small/medium applications, each running on a dedicated hardware, decoupled and protected from other systems in the same facility.
- Traditional DCs host hardware and software for multiple organizational units or even different companies.
- WSCs power the services offered by companies such as Google, Amazon, Yahoo, and Microsoft's online services division.
  - Use a relatively homogeneous hardware and system software platform, and share a common system management layer
  - Much of the application, middleware, and system software is built in-house compared to the predominance of third-party software running in conventional DCs.
  - Run a smaller number of very large applications (or Internet services)
  - Internet services must achieve high availability, typically aiming for at least 99.99% uptime (about an hour of downtime per year)

2. Some new WSC Networks

VL2: A Scalable and FlexibleData Center Network
By Albert Greenberg, James R. Hamilton, Navendu Jain, Srikanth Kandula, Changhoon Kim,Parantap Lahiri, David A. Maltz, Parveen Patel, and S. Sengupta

Figure 1. A conventional network architecture for data centers (adapted from figure by Cisco[5]).

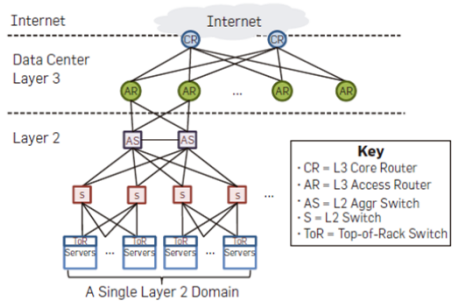
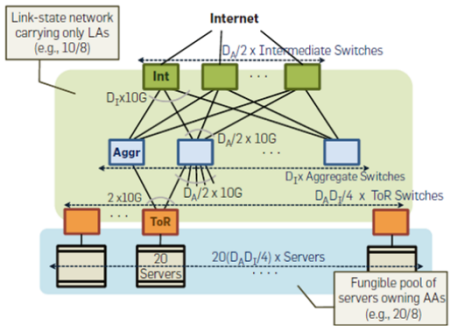


Figure 4. An example Clos network between aggregation and intermediate switches provides a richly connected backbone well suited for VLB. The network is built with two separate address families—topologically significant locator-specific addresses (LAs) and flat application-specific addresses (AAs).



CANTABRIA CAMPUS INTERNACIONAL

2. NOT JUST A COLLECTION OF SERVERS

- The software for Gmail or Web search services, execute at a scale far beyond a single machine or a single rack. Hundreds to thousands servers
- The ***computer*** is this big collection of hardware.
- Its size makes it difficult to experiment with or simulate efficiently
- Fault behavior and power/energy considerations
- New challenge to programmer productivity, perhaps greater than programming multicore systems
- High complexity:
  - larger scale of the application domain
  - deeper and less homogeneous storage hierarchy
  - higher fault rates
  - higher performance variability

CANTABRIA CAMPUS INTERNACIONAL

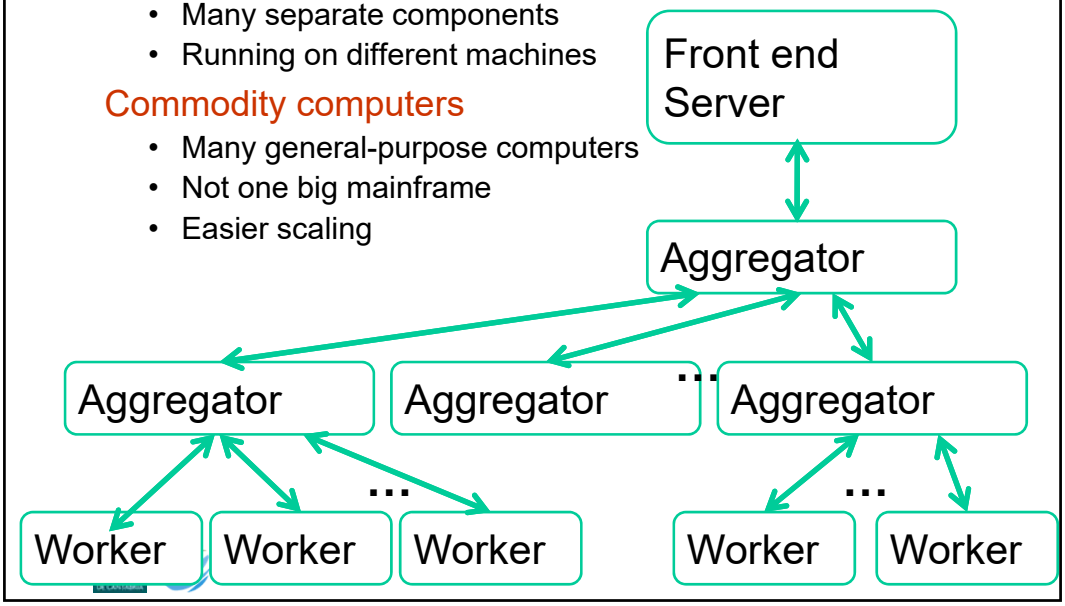
2. Warehouse-scale computing

- Some workloads require a massive computing infrastructure not possible for client-side computing
  - Search services (Web, images, maps, etc.)
  - language translation
- In WH scale computing, the program is an Internet service,
  - tens individual programs that interact to implement complex user services: email, search, or maps
- These programs implemented and maintained by different teams of engineers, even across organizational, geographic, and company boundaries (as is the case with mashups)
- The hardware consists of thousands of servers with networking and storage subsystems, power distribution and extensive cooling.
- The enclosure for these systems is a building structure and often indistinguishable from a large warehouse

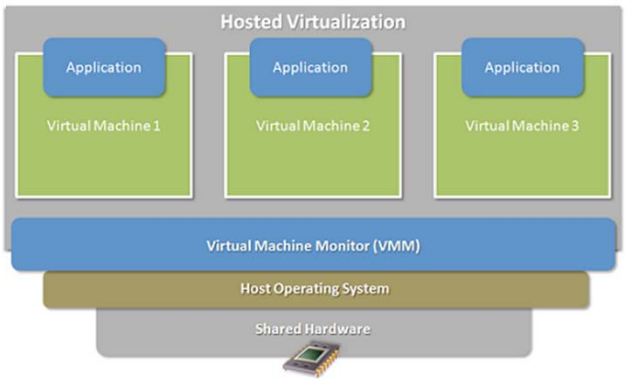
2. Multi-Tier Applications

Applications consist of tasks
- Many separate components
- Running on different machines

Commodity computers
- Many general-purpose computers
- Not one big mainframe
- Easier scaling



2. Enabling Technology: Virtualization



Multiple virtual machines on one physical machine

Applications run unmodified as on real machine

VM can migrate from one computer to another

2. Cloud Service Models

Software as a Service

- Provider licenses applications to users as a service
- E.g., customer relationship management, e-mail, ..
- Avoid costs of installation, maintenance, patches, …

Platform as a Service

- Provider offers software platform for building applications
- E.g., Google's App-Engine
- Avoid worrying about scalability of platform

Infrastructure as a Service

- Provider offers raw computing, storage, and network
- E.g., Amazon's Elastic Computing Cloud (EC2)
- Avoid buying servers and estimating resource needs

2. Cloud Computing Basis

Elastic resources

- Expand and contract resources
- Pay-per-use
- Infrastructure on demand

Multi-tenancy

- Multiple independent users
- Security and resource isolation
- Amortize the cost of the (shared) infrastructure

Flexibility service management

- Resiliency: isolate failure of servers and storage
- Workload movement: move work to other locations

2. Emphasis on COST/EFFICIENCY

A large computing platform is expensive. Quality of service depends on the aggregate processing, storage and networking capacity available

Web search

- Increased service popularity translates into higher request loads.
- The size of the problem keeps growing. Web grows by millions of pages per day, increasing costs of building and serving a Web index
- Even if throughput and data repository could be held constant, the competitive nature of this market continuously drives innovations to improve the quality of results and the frequency with which the index is updated, (synonyms,…)

The relentless demand for more computing capabilities makes cost/efficiency a primary metric for WSCs. Accounting for all the components of cost, including hosting-facility capital and operational expenses (including power provisioning and energy costs), hardware, software, management personnel, and repairs

2. WHY WSCs MIGHT MATTER TO YOU

- The attractive economics of low-end server class computing platforms puts clusters of hundreds of nodes within the reach of a relatively broad range of corporations and research institutions
- A rack with 40 servers, each with four 8-core dual-threaded CPUs, would contain more than two thousand hardware threads
- Such systems will be affordable to a very large number of organizations within just a few years, while exhibiting some of the scale, architectural organization, and fault behavior of today's WSCs (Google, Amazon,…).

Tema 2 Outline

- 1. Basic Terms
- 2. Warehouse-Scale Computers (WSCs)
- **3. Architectural Overview of WSCs**
  - **Networks, Storage & Processing power**
- 4. Benchmarks: Graph500 (& Top500)

3. WSCs Array: Enrackable boards or blades + rack router

Figure 1.1: Sketch of the typical elements in warehouse-scale systems: 1U server (left), 7' rack with Ethernet switch (middle), and diagram of a small cluster with a cluster-level Ethernet switch/router (right).

Top-of-Rack Architecture

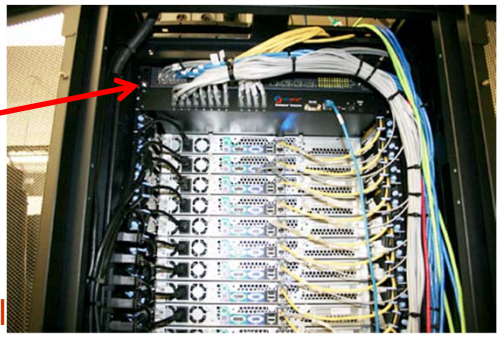
Rack of servers

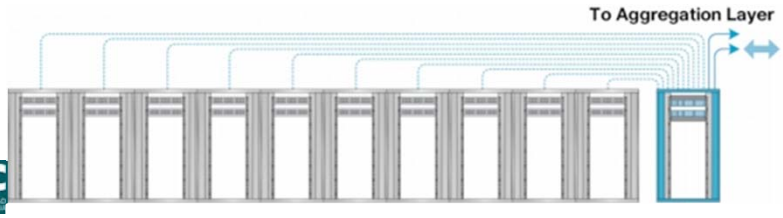
- Commodity servers
- And top-of-rack switch

Modular design

- Preconfigured racks
- Power, network, and storage cabling

Aggregate to the next level





Modularity, Modularity, Modularity

Containers

3. Storage I

Disk drives or Flash devices are connected directly to each individual server and managed by a global distributed file system (such as Google's GFS)

Attaching disks directly to compute nodes reduces hardware costs (the disks leverage the existing server enclosure) and improve networking fabric utilization (each server network port is dynamically shared between the computing and the file system)

Trading off among write overheads, high availability, and increased read bandwidth seems the right solution.

Another advantage of having disks co-located with servers is that it enables distributed system software to exploit data locality.

As networking performance has outpaced disk performance for the last decades such locality advantages are less useful for disks but may remain beneficial to faster modern storage devices such as those using Flash.

3. Storage II

Some WSCs deploy desktop-class disk drives instead of enterprise-grade disks because of the substantial cost differential.

Typically, consumer (or “desktop”) drives offer the lowest price but they may not be designed for continuous operation.

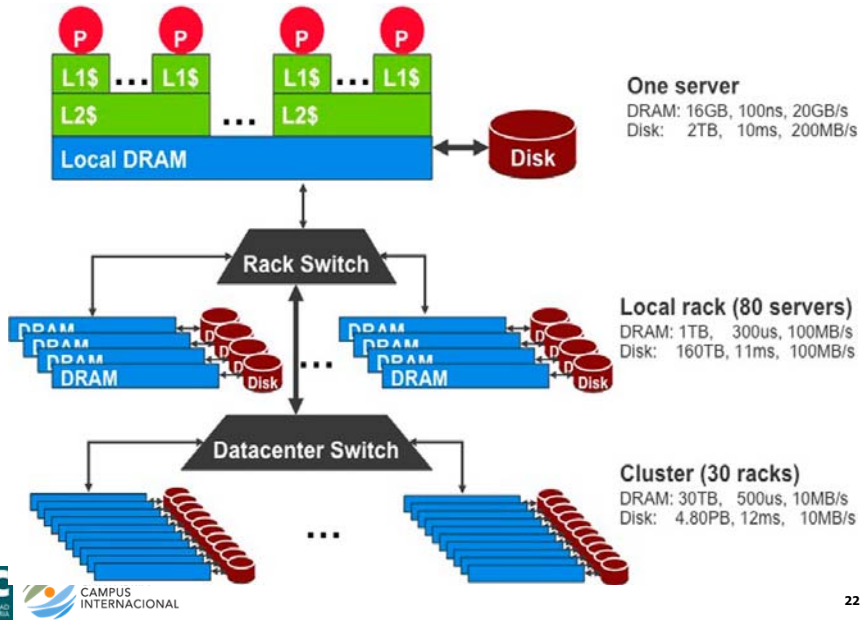
Nearline (higher desktop-class) drives, originally created for disk-based backup servers, add enterprise features such as increased vibration tolerance and are suitable for continuous operation.

Enterprise grade disks offer the highest performance at the highest cost.

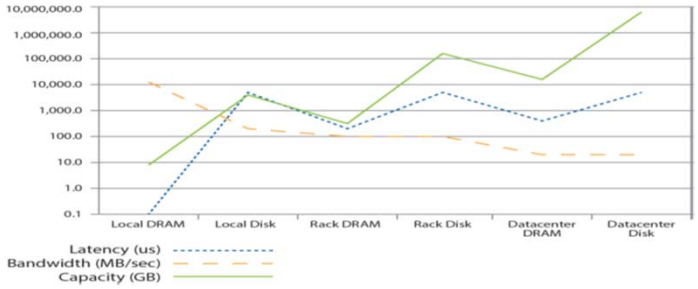
Since data is nearly always replicated in some distributed fashion (as in GFS), higher fault rates of non-enterprise disk models can often be tolerated.

NAND Flash has made Solid State Drives (SSDs) affordable. While the cost per byte in SSDs will remain much higher than in disks for the foreseeable future, many Web services have I/O rates that cannot be easily achieved with disk based systems. Since SSDs can deliver IO rates orders of magnitude higher than disks, they are displacing disk drives as the repository of choice for databases in Web services.

3. WSC Memory Hierarchy

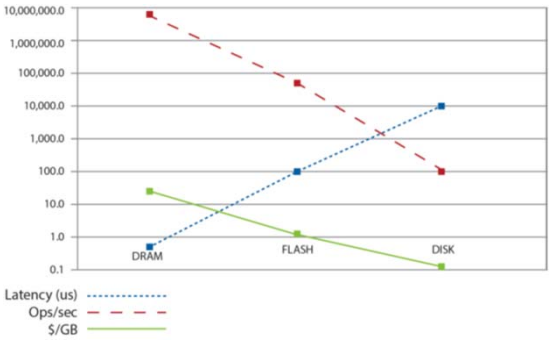


3. Latency, BW and capacity



2,400 servers, each with 16 GB of DRAM and four 2 TB disk drives. Each group of 80 servers is connected through a 1-Gbps link to a rack-level switch that has additional eight 1-Gbps ports used for connecting the rack to the cluster-level switch (oversubscription of 5). Network latency assume a TCP-IP transport, and networking bandwidth values assume that each server behind an oversubscribed set of uplinks is using its fair share of the available cluster-level bandwidth. For disks, we show typical commodity disk drive (SATA) latencies and transfer rates.
The bandwidth available from local disks is 200 MB/s, whereas the bandwidth from offrack disks is just 25 MB/s via the shared rack uplinks. On the other hand, total disk storage in the cluster is almost ten million times larger than local DRAM

3. Flash



NAND Flash originally developed for portable electronics has found target use cases in WSC systems.
Today Flash is a viable option for bridging the cost and performance gap between DRAM and disks
Flash's most appealing characteristic with respect to disks is its performance under random read operations, which is nearly three orders of magnitude better.
Flash's performance is so high that it becomes a challenge to use it effectively in distributed storage systems since it demands much higher bandwidth from the WSC fabric

3. Handling failures

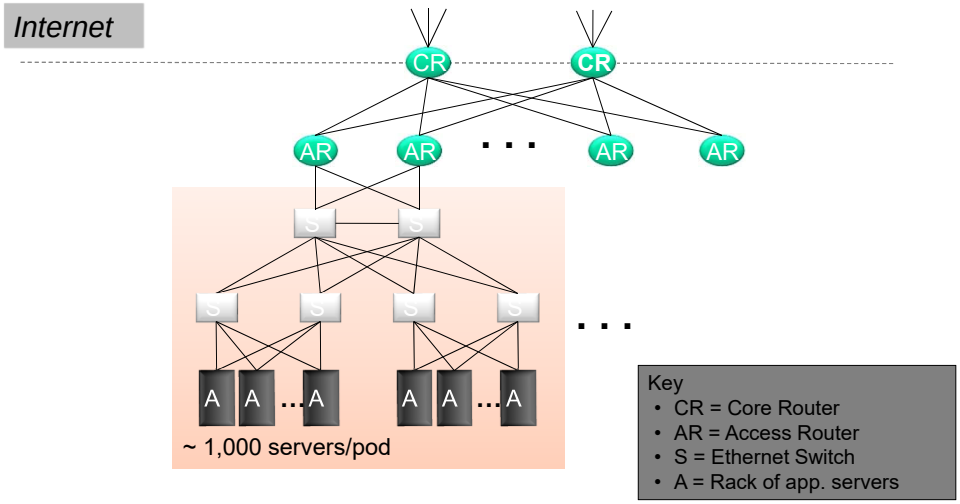
The scale of WSCs requires that Internet services software tolerate relatively high component fault rates.

Disk drives, for example, can exhibit annualized failure rates higher than 4%.

Different deployments have reported between 1.2 and 16 average server-level restarts per year.

With such high component failure rates, an application running across thousands of machines may need to react to failure conditions on an hourly basis.

3. Traditional Data Center Network Topology



3. Networking Fabric I

1-Gbps Ethernet switches with up to 48 ports are commodity components, costing less than $30/Gbps per server to connect a single rack (including switch port, cable, and server NIC)

Bandwidth within a rack of servers tends to be homogeneous.

Network switches with high port counts, needed to tie together WSC clusters, have a much different price structure and are more than ten times more expensive (per port) than commodity rack switches.

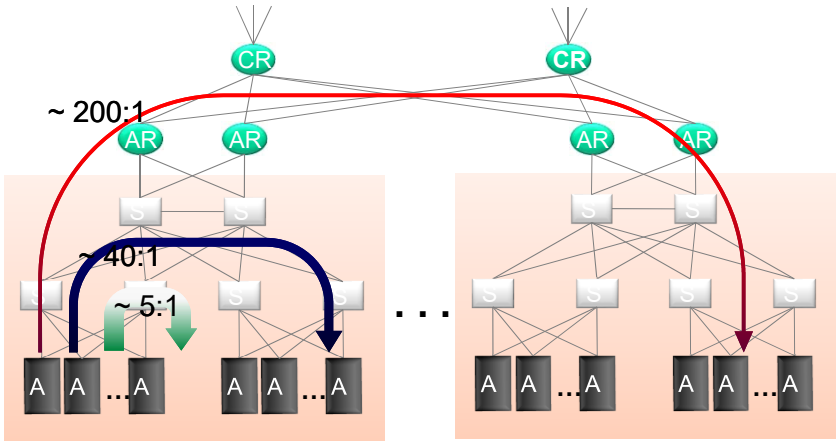
Commodity rack switches provide a fraction of their bandwidth for inter-rack communication through a handful of uplinks to the more costly cluster-level switches.

A rack with 40 servers, each with a 1-Gbps port, might have between four and eight 1-Gbps uplinks to the cluster-level switch

Oversubscription factor between 10 and 5 for communication across racks

Programmers must be aware of the scarce cluster-level bandwidth and try to exploit rack-level networking locality, complicating software development and impacting resource utilization

3. Capacity Mismatch (oversubscription)



3. Networking Fabric II

Alternatively, one can remove some of the cluster-level networking bottlenecks by spending more money on the interconnect fabric.

For example, Infiniband interconnects typically scale to a few thousand ports but can cost $500–$2,000 per port.

Similarly, some networking vendors are starting to provide larger-scale Ethernet fabrics, but again at a cost of at least hundreds of dollars per server.

How much to spend on networking vs. spending the equivalent amount on buying more servers or storage is an application-specific question that has no single correct answer.

There are other hardware platforms proposed for both HPC and BigData applications that choose more sophisticated networks

3. Some new WSC Networks

VL2: A Scalable and FlexibleData Center Network
By Albert Greenberg, James R. Hamilton, Navendu Jain, Srikanth Kandula, Changhoon Kim,Parantap Lahiri, David A. Maltz, Parveen Patel, and S. Sengupta

Figure 1. A conventional network architecture for data centers (adapted from figure by Cisco[5]).

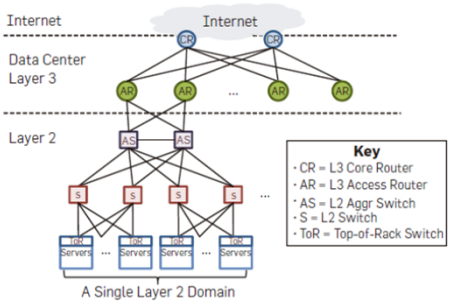
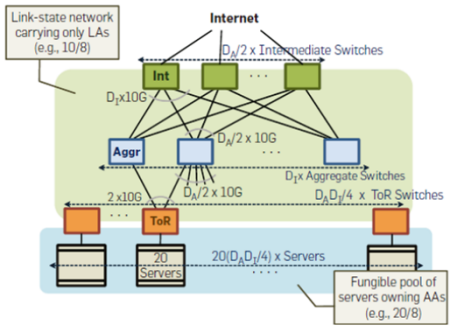
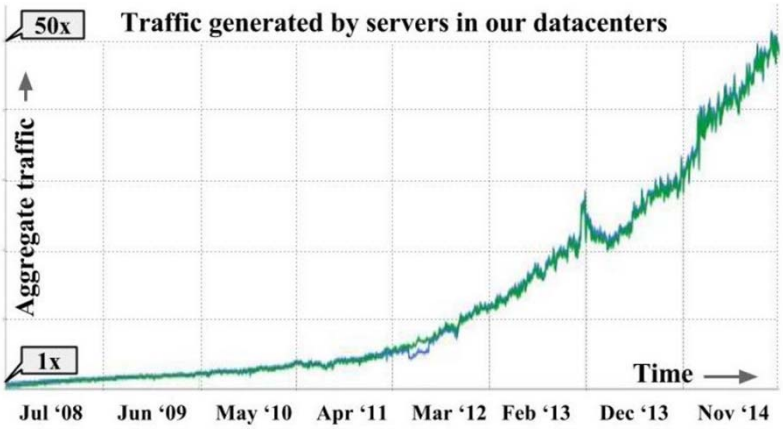


Figure 4. An example Clos network between aggregation and intermediate switches provides a richly connected backbone well suited for VLB. The network is built with two separate address families—topologically significant locator-specific addresses (LAs) and flat application-specific addresses (AAs).

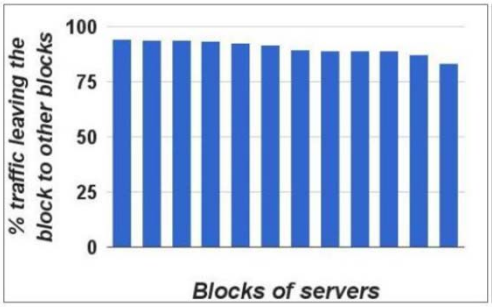


3. Traffic increasing demands (Google)



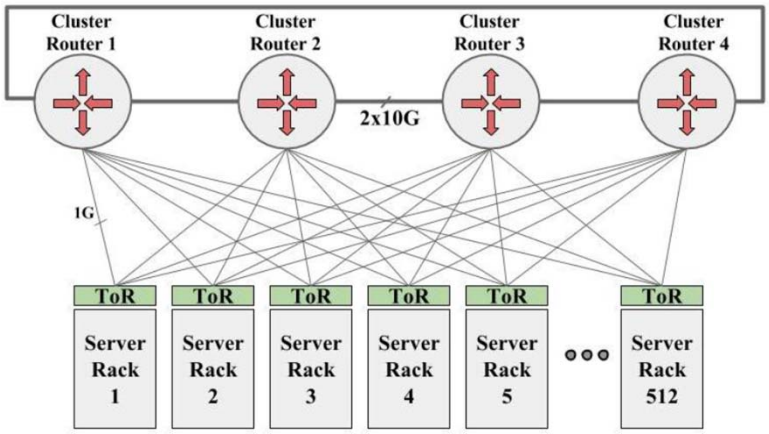
3. Most DC traffic tends to be Uniform

Job Category	B/w (%)
Storage	49.3
Search Serving	26.2
Mail	7.4
Ad Stats	3.8
Rest of traffic	13.3



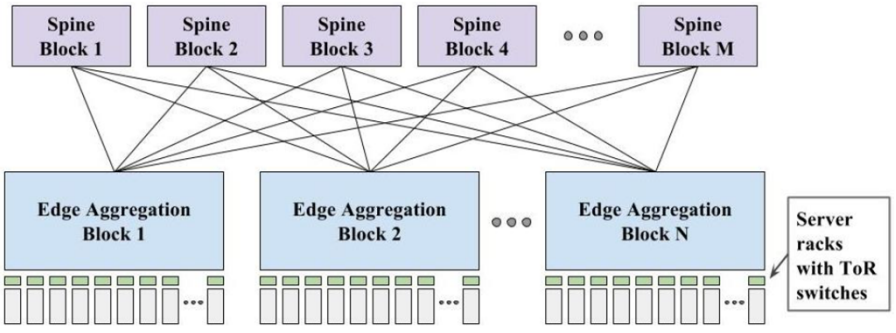
Mix of jobs in an example cluster with 12 blocks of servers (left). Fraction of traffic in each block destined for remote blocks (right). Mostly UNIFORM

3. Old Google DC Network deployment (2004)



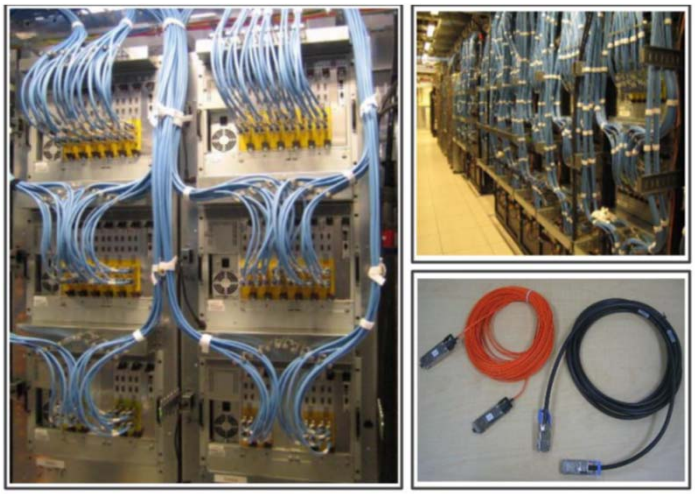
A traditional 2Tbps four-post cluster (2004). Top of Rack (ToR) switches serving 40 1G-connected servers were connected via 1G links to four 512 1G port Cluster Routers (CRs) connected with 10G sidelinks.

3. Google Folded Clos Networks



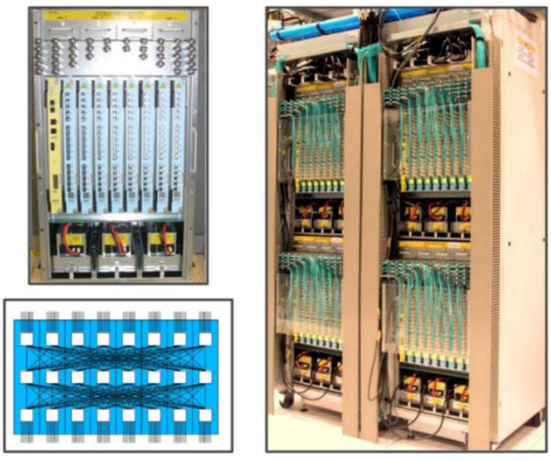
A generic 3 tier Clos architecture with edge switches (ToRs), aggregation blocks and spine blocks. All generations of Clos fabrics deployed in Google datacenters follow variants of this architecture.

3. Google Firehose



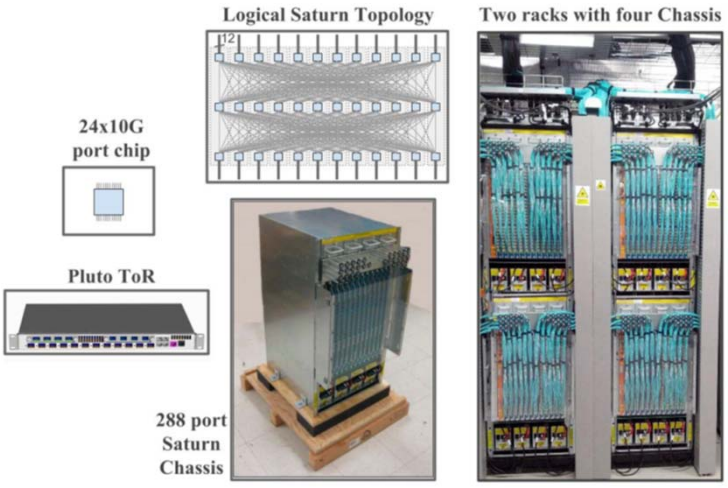
Two Firehose racks (left), each with 3 chassis with bulky CX4 cables from remote racks. The top right figure shows an aisle of cabled racks.

3. Google Watchtower



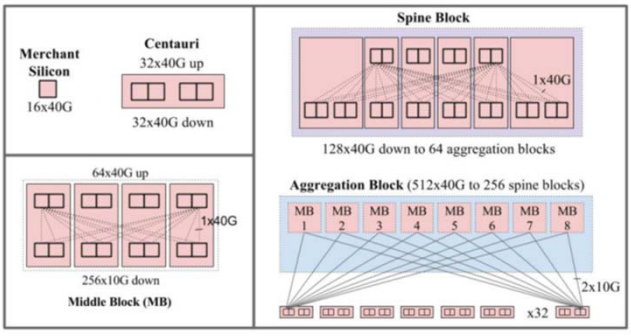
A 128x10G port Watchtower chassis (top left).
The internal non-blocking topology over eight linecards (bottom left). Four chassis housed in two racks cabled with ber (right).

3. Google Saturn



Components of a Saturn fabric. A 24x10G Pluto ToR Switch and a 12-linecard 288x10G Saturn chassis (including logical topology) built from the same switch chip.
Four Saturn chassis housed in two racks cabled with fiber (right).

3. Goggle Jupiter



Building blocks used in the Jupiter topology.

Jupiter Middle blocks housed in racks.

Tema 2 Outline

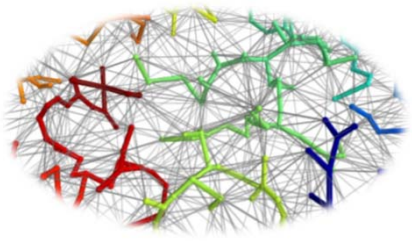
- 1. Basic Terms
- 2. Warehouse-Scale Computers (WSCs)
- 3. Architectural Overview of WSCs
  - Networks, Storage & Processing power
- **4. Benchmarks: Graph500 (& Top500)**

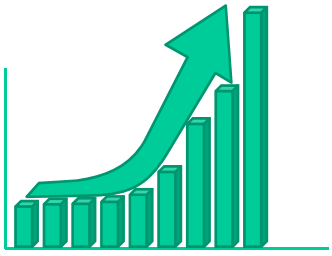
4. More and more data

There is an exponential rise in the amount of data and its complexity.

This growth introduces several algorithmic challenges to extract information out of the data.

Many data-intensive applications are search-based over graph-stored data.





4. Analytics, Big Data, Data Intensive

BigData es un término que hace referencia a una cantidad de datos tal que supera la capacidad del software "tradicional" para ser capturados, administrados y procesados en un tiempo razonable

Existen muchas herramientas para tratar con Big Data. Hadoop, MapReduce, Cassandra son algunas de los más conocidos.

Otras: Dryad, Sawzall, BigTable, Dynamo, Dremel, Spanner, Chubby.

4. Almacenamiento NoSQL (Not Only SQL)

Sistemas de almacenamiento que no cumplen con el esquema entidad-relación. Más flexibles y concurrentes; permiten manipular grandes cantidades de información de manera mucho más rápida que las bases de datos relacionales.

- Almacenamiento Clave-Valor (Cassandra)
- Almacenamiento Documental (MapReduce, Hadoop, MongoDB)
- Almacenamiento en Grafo (Neo4J, GraphDB)
- Almacenamiento Orientado a Columnas (BigTable, Hbase, HyperTable)

4. Accepted Benchmarks

Linpack reproduces average HPC applications behavior, but does not suit data-intensive workloads.

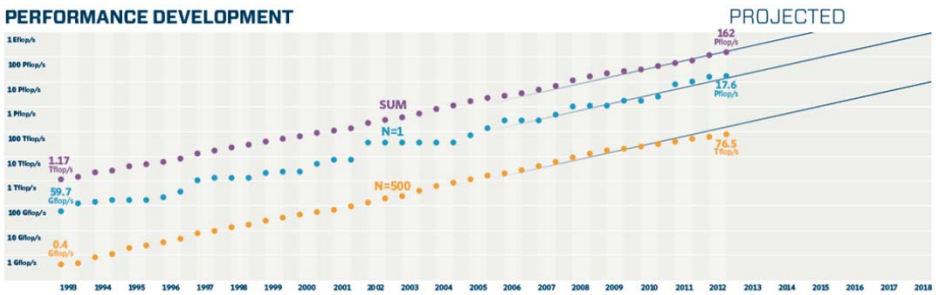
Graph500 benchmark appears to evaluate system performance and adequacy to a given case of Big Data applications.

A thorough understatement of a benchmark behavior can help us to improve its performance in current systems.

UC UNIVERSIDAD DE CANTABRIA
CANTABRIA CAMPUS INTERNACIONAL

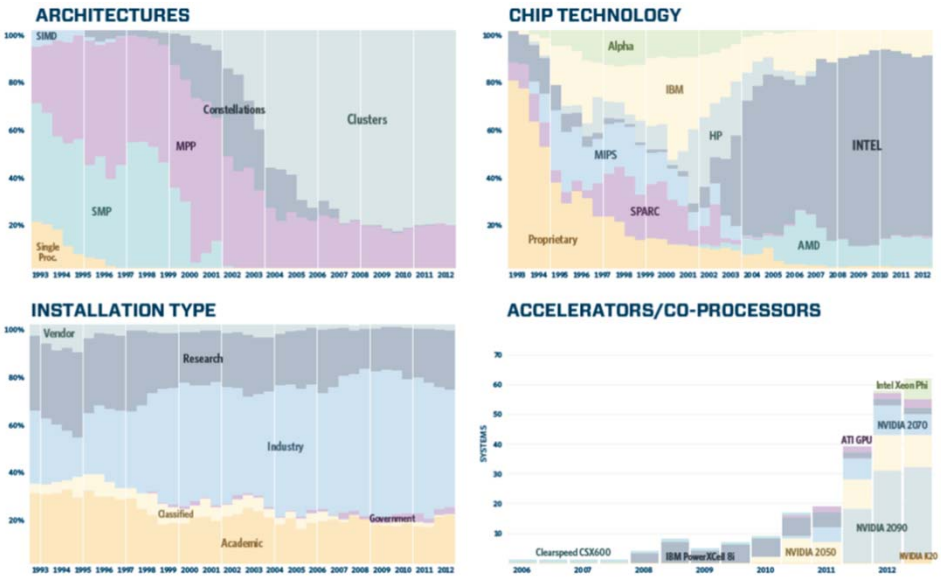
4. Top500 I

	NAME	SPECS	SITE	COUNTRY	CORES	Rmax Pflop/s	POWER MW
1	TITAN	Cray XK7, Operon 6274 16C 2.2 GHz + Nvidia Kepler GPU, Custom Interconnect	DOE/OS/ORNL	USA	560,640	**17.6**	8.3
2	SEQUOIA	IBM BlueGene/Q, Power BQC 16C 1.60 GHz, Custom Interconnect	DOE/NNSA/LLNL	USA	1,572,864	**16.3**	7.9
3	K COMPUTER	Fujitsu SPARC64 VIIIfx 2.0GHz, Custom Interconnect	RIKEN AICS	Japan	705,024	**10.5**	12.7
4	MIRA	IBM BlueGene/Q, Power BQC 16C 1.60 GHz, Custom Interconnect	DOE/OS/ANL	USA	786,432	**8.16**	3.95
5	JUQUEEN	IBM BlueGene/Q, Power BQC 16C 1.60 GHz, Custom Interconnect	Forschungszentrum Jülich	Germany	393,216	**4.14**	1.97



UC UNIVERSIDAD DE CANTABRIA
CANTABRIA CAMPUS INTERNACIONAL

4. Top500 II



MontBlanc 3

As part of the Phase 3 of Mont-Blanc, a new prototype is built by Atos. It is named Dibona, after the Dibona peak in the French Alps, and it will start operation in Fall 2017. It is based on 64 bit ThunderX2 processors from Cavium®, relying on the ARM® v8 instruction set.

MontBlanc 1

8 nodes, each equipped with:

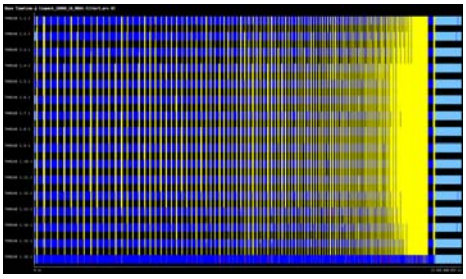
2 racks, 8 standard BullX chassis, 72 compute blades fitting 1080 compute cards, for a total of 2160 CPUs and 1080 GPUs.
SoC Samsung Exynos 5 Dual.
CPU Cortex-A15@1.7GHz dual core.
GPU ARM Mali T-604 (OpenCL 1.1 capable).

MontBlanc 2

70 nodes, each equiped with:
CPU Cortex-A9 Quad @1.4 GHz
GPU Nvidia Tesla K20
4 GB DDR3 RAM
1 Gb Ethernet interconnection network
QDR Infiniband interconnection*

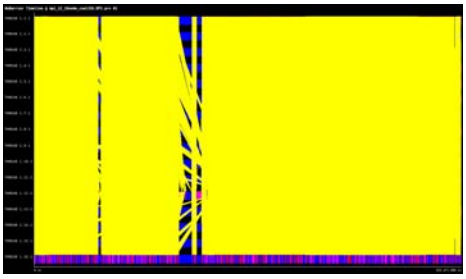
4. Trace Comparison

Linpack



- Bulk-Synchronous-Parallel model
- Computation time is vastly higher than communications time

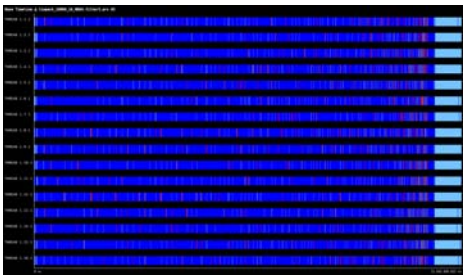
Graph500



- Asynchronous uniform communication
- Communication represents a significant amount of total execution time

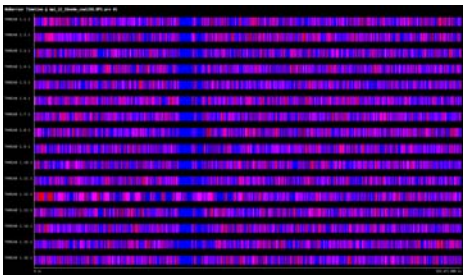
4. Trace Comparison

Linpack



- Bulk-Synchronous-Parallel model
- Computation time is vastly higher than communications time

Graph500



- Asynchronous uniform communication
- Communication represents a significant amount of total execution time

4. Trace Comparison

Linpack

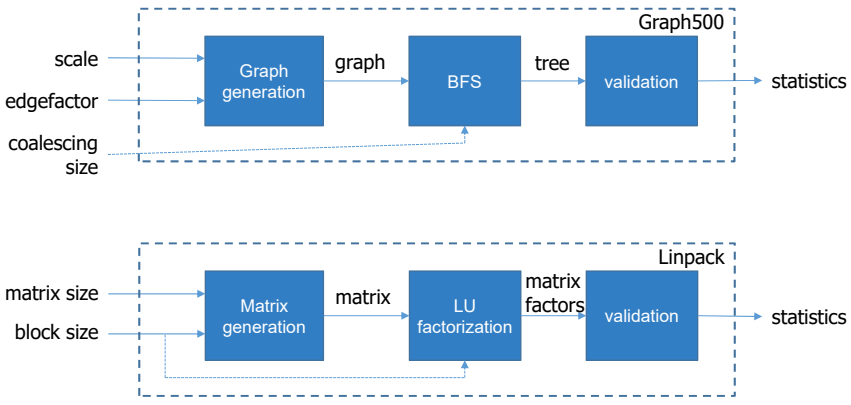
	Running	Wait/WaitAll	Point-to-point communication
THREAD 1.1.1	90,47 %	0,51 %	2,52 %
THREAD 1.2.1	89,70 %	1,17 %	2,40 %
THREAD 1.3.1	89,56 %	1,40 %	2,50 %
THREAD 1.4.1	88,04 %	2,80 %	2,73 %
THREAD 1.5.1	89,79 %	1,08 %	2,73 %
THREAD 1.6.1	90,33 %	0,41 %	2,50 %
THREAD 1.7.1	89,46 %	1,44 %	2,50 %
THREAD 1.8.1	87,86 %	3,05 %	2,72 %
THREAD 1.9.1	89,55 %	1,47 %	2,41 %
THREAD 1.10.1	90,65 %	0,54 %	2,30 %
THREAD 1.11.1	90,42 %	0,64 %	2,50 %
THREAD 1.12.1	87,99 %	2,97 %	2,48 %
THREAD 1.13.1	89,60 %	1,37 %	2,66 %
THREAD 1.14.1	90,54 %	0,57 %	2,47 %
THREAD 1.15.1	90,57 %	0,38 %	2,56 %
THREAD 1.16.1	90,37 %	0,45 %	2,76 %
Total	1.434,91 %	20,24 %	40,76 %
Average	89,68 %	1,27 %	2,55 %
Maximum	90,65 %	3,05 %	2,76 %
Minimum	87,86 %	0,38 %	2,30 %
StDev	0,92 %	0,89 %	0,13 %
Avg/Max	0,99	0,41	0,92

Graph500

	Running	Wait/WaitAll	Point-to-point communication	Group Communication
THREAD 1.1.1	73,29 %	2,71 %	23,90 %	0,10 %
THREAD 1.2.1	73,96 %	2,36 %	23,58 %	0,09 %
THREAD 1.3.1	68,97 %	2,53 %	28,42 %	0,08 %
THREAD 1.4.1	66,35 %	3,06 %	30,51 %	0,07 %
THREAD 1.5.1	65,73 %	3,11 %	31,09 %	0,07 %
THREAD 1.6.1	66,85 %	3,02 %	30,08 %	0,05 %
THREAD 1.7.1	66,76 %	3,07 %	30,10 %	0,06 %
THREAD 1.8.1	74,45 %	2,44 %	23,07 %	0,05 %
THREAD 1.9.1	64,00 %	3,04 %	31,92 %	0,04 %
THREAD 1.10.1	65,41 %	3,09 %	31,43 %	0,07 %
THREAD 1.11.1	66,66 %	3,06 %	30,22 %	0,05 %
THREAD 1.12.1	48,98 %	6,87 %	44,09 %	0,07 %
THREAD 1.13.1	73,90 %	2,43 %	23,54 %	0,13 %
THREAD 1.14.1	71,78 %	2,81 %	25,30 %	0,12 %
THREAD 1.15.1	73,80 %	2,35 %	23,74 %	0,11 %
THREAD 1.16.1	73,05 %	2,56 %	24,28 %	0,11 %
Total	1.094,93 %	48,49 %	455,24 %	1,31 %
Average	68,43 %	3,03 %	28,45 %	0,08 %
Maximum	74,45 %	6,87 %	44,09 %	0,13 %
Minimum	48,98 %	2,35 %	23,07 %	0,04 %
StDev	6,13 %	1,03 %	5,20 %	0,03 %
Avg/Max	0,92	0,44	0,65	0,63

CANTABRIA CAMPUS INTERNACIONAL

4. Analysis of the code



CANTABRIA CAMPUS INTERNACIONAL

4. Empirical evaluation: Infrastructure

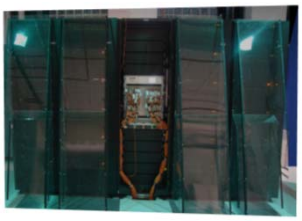
Altamira architecture:

IBM-iDataplex dx360m4 nodes

Intel Sandybridge Xeon E5-2670 processors per node, 8 cores per processor

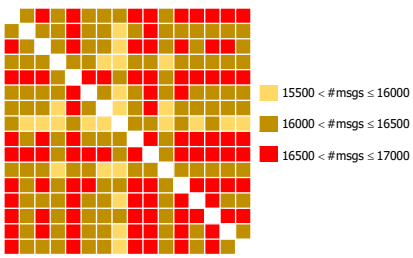
64 GB of RAM per node

Infiniband FDR10 network with a folded Clos topology

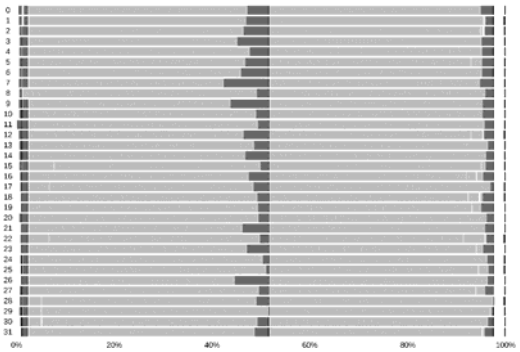


4. Graph500 benchmark

- Uniform communications over space
- Uniform communications over execution time



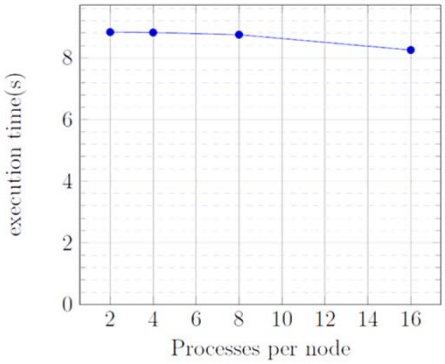
[scale 20, edgefactor 16, 16 processes]



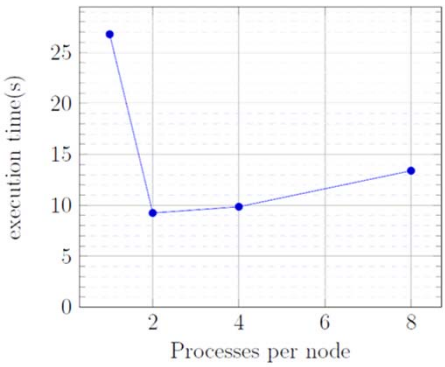
[scale 22, edgefactor 16, 32 processes]

4. Impact of process dispersion

Linpack — 32 procs — Graph500



- Limited impact

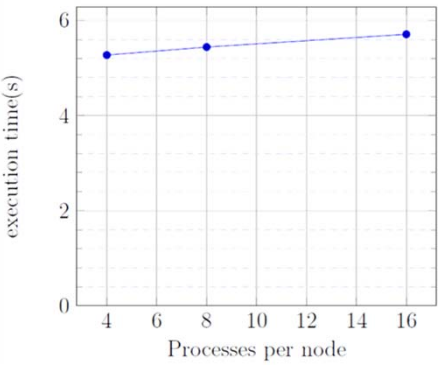


- Noticeable impact
- Tradeoff between network and memory impact

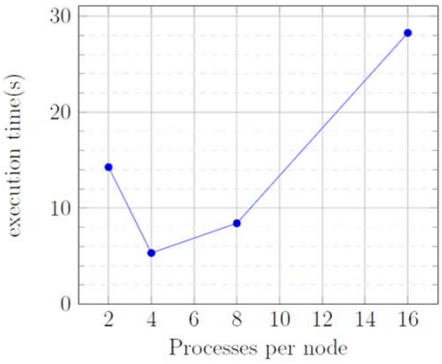
UC UNIVERSIDAD DE CANTABRIA — CANTABRIA CAMPUS INTERNACIONAL

4. Impact of process dispersion

Linpack — 64 procs — Graph500



- When the execution is divided into smaller pieces, concentration has a negative impact



- Noticeable impact
- Tradeoff between network and memory impact

UC UNIVERSIDAD DE CANTABRIA — CANTABRIA CAMPUS INTERNACIONAL

4. Graph500 coalescing size

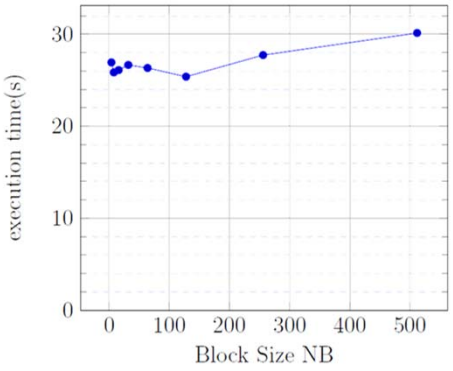
- Number of messages per process originated:

$$\#messages/_{process} = \frac{2^{Scale+1} \cdot edgefactor}{coalescing\_size} \cdot \frac{n-1}{n^2}$$

- A tradeoff can balance aggregation effects and increase performance
  - High aggregation reduces communication
  - Low aggregation diminishes length of active wait loops
  - Tradeoff is related both to network technology and algorithm behavior

4. Impact of parameter tuning

Linpack



- Better performance achieved with certain values that adjust better the workload distribution

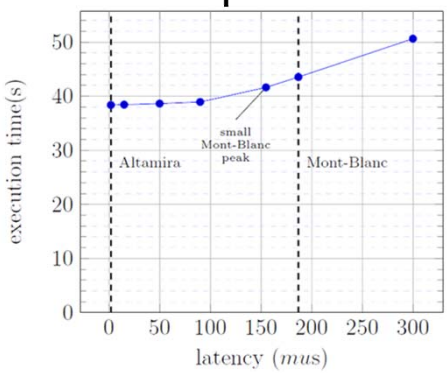
Graph500

- Tradeoff shows a minimum execution time around 12KB messages (3x default aggregation)

4. Network impact

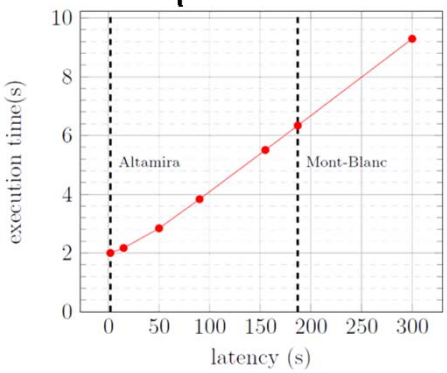
- *Latency*

Linpack



- Current latency value range isn't a bottleneck

Graph500

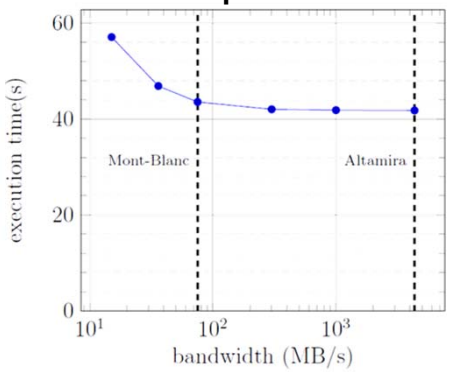


- Trend indicates room for improvement below current system parameters

4. Network impact

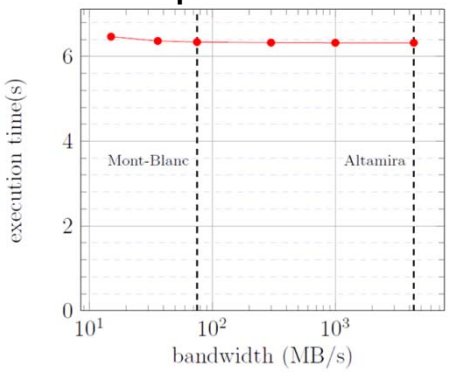
- *Bandwidth*

Linpack



- Impact from BW is higher in Linpack than in Graph500

Graph500



- Graph500 is more dependent on latency than BW

Tema 2 Outline

- 1. Basic Terms
- 2. Warehouse-Scale Computers (WSCs)
- 3. Architectural Overview of WSCs
  - Networks, Storage & Processing power
- 4. Benchmarks: Graph500 (& Top500)
- Dudas del día anterior

1. Response Time and Throughput

Response time
- How long it takes to do a task

Throughput
- Total work done per unit time
  - e.g., tasks/transactions/… per hour

How are response time and throughput affected by
- Replacing the processor with a faster version?
- Adding more processors?

1. Measuring Execution Time

Elapsed time

- Total response time, including all aspects
  - Processing, I/O, OS overhead, idle time
- Determines system performance

CPU time

- Time spent processing a given job
  - Discounts I/O time, other jobs' shares
- Comprises user CPU time and system CPU time
- Different programs are affected differently by CPU and system performance

Benchmarks & Workloads

1. Relative Performance

Performance = 1/Execution Time

"X is n time faster than Y"

$$\text{Performance}_X / \text{Performance}_Y = \text{Execution time}_Y / \text{Execution time}_X = n$$

- Example: time taken to run a program
  - 10s on A, 15s on B
  - Execution Time_B / Execution Time_A = 15s / 10s = 1.5
  - So A is 1.5 times faster than B