GOETHE
UNIVERSITÄT
FRANKFURT AM MAIN

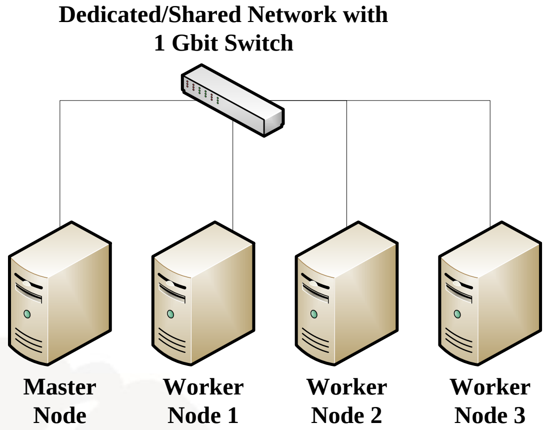
TPCx-HS Experiments

Todor Ivanov (todor@dbis.cs.uni-frankfurt.de)
Sead Izberovic

Frankfurt Big Data Lab
-understanding and applying technologies for Big Data-

Goethe University Frankfurt am Main, Germany
http://www.bigdata.uni-frankfurt.de/

New Cluster Setup



Setup Description	Summary
Total Nodes:	4 x Dell PowerEdge T420
Total Processors/ Cores/Threads:	5 CPUs/ 30 Cores/ 60 Threads
Total Memory:	4x 32GB = 128 GB
Total Number of Disks:	13 x 1TB,SATA, 3.5 in, 7.2K RPM, 64MB Cache
Total Storage Capacity:	13 TB
Network:	1GBit Ethernet

- Operating System: Ubuntu Server 14.04.1. LTS
- Cloudera’s Hadoop Distribution - CDH 5.2
- Replication Factor of 2 (only 3 worker nodes)

Goal → Run end-to-end, analytical Big Data benchmark (BigBench) to evaluate the platform!

Initial results → BigBench performance was very slow! → **Shared 1Gbit Network**

Solution → Upgrade to Dedicated 1Gbit Switch (around 30 €)

Questions

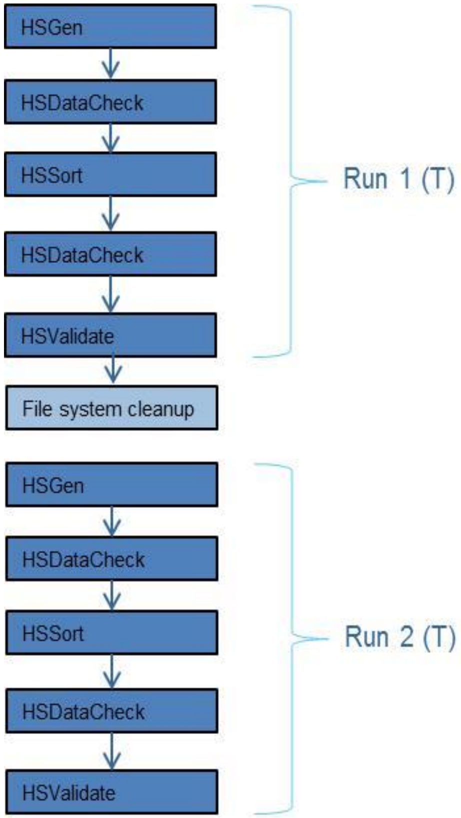
- What about the performance differences between the two setups (shared vs. dedicated networks)?

- How can we measure the network improvement in the new setup (dedicated network)?

→ Use **network intensive** workload/benchmark like **TPCx-HS.**

TPCx-HS: TPC Express for Hadoop Systems

- X: Express, H: Hadoop, S: Sort
- TPCx-HS [1],[2] is the ***first industry standard Big Data Benchmark*** released in July 2014
- Based on ***TeraSort*** and consists of 4 modules: HSGen, HSDataCkeck, HSSort & HSValidate
- Scale Factors following stepped size model: 100GB, 300GB, 1TB, 3TB,10TB ….
- The TPCx-HS specification defines three major metrics:
  - Performance metric (HSph@SF)
  - Price-performance metric ($/HSph@SF)
  - Power per performance metric (Watts/HSph@SF)
- Use **TPCx-HS Kit 1.1.2**



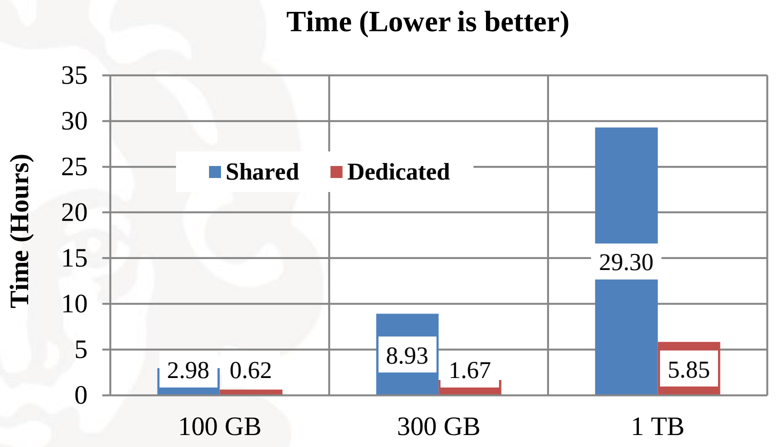
[1] TPC website - http://www.tpc.org/tpcx-hs

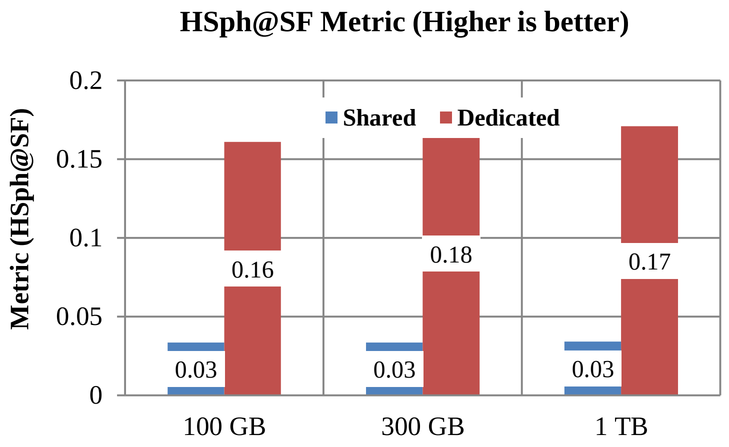
[2] Nambiar et al.,“Introducing TPCx-HS: The First Industry Standard for Benchmarking Big Data Systems,” in Performance Characterization and Benchmarking. Traditional to Big Data, Eds. Springer International Publishing, 2014.

Performance (Shared vs. Dedicated)

- The presented results are obtained by executing the TPCx-HS kit provided on the official TPC website (www.tpc.org). However, the reported times and metrics are experimental, not audited by any authorized organization and therefore not directly comparable with other officially published full disclosure reports.

- Tested with 3 scale factors: 100GB, 300GB, 1TB





- The shared setup is ***5 times slower*** compared to the dedicated one.

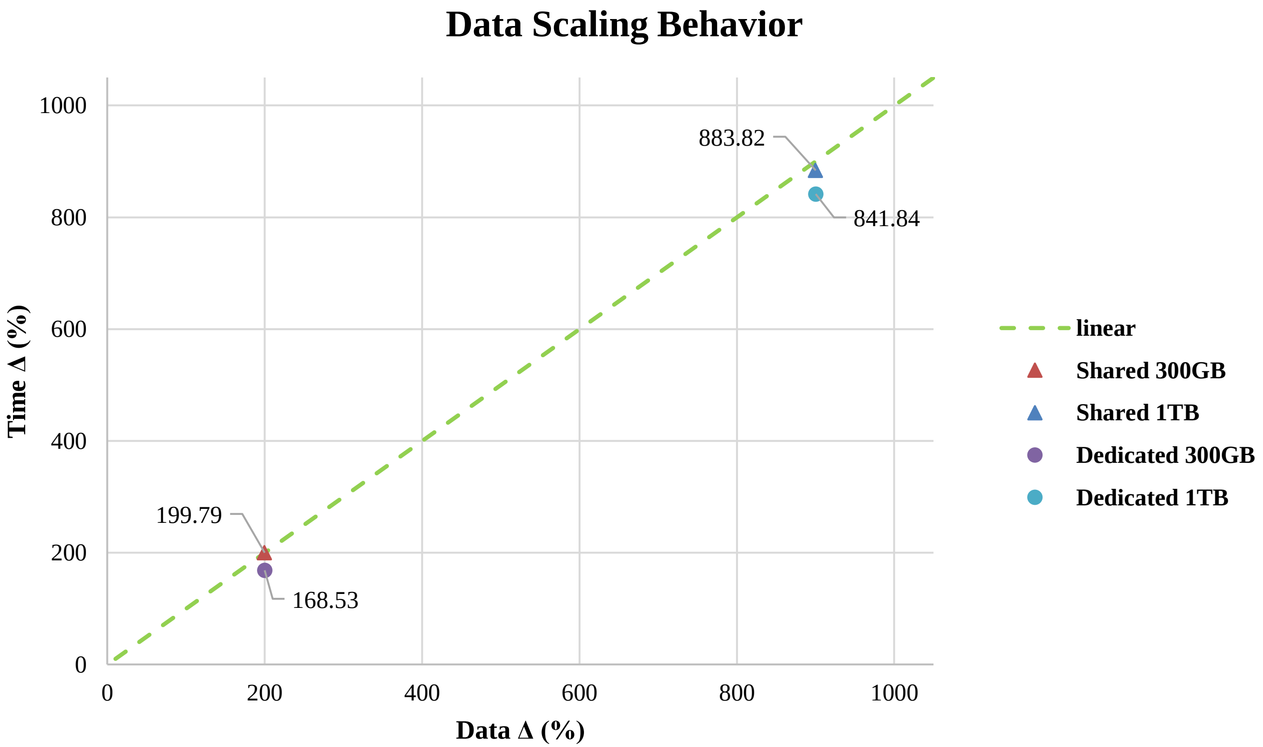
Results

- **Data Δ** column represents the difference in percent of the Data Size to the data baseline in our case 100GB.
- **Time (Sec)** shows the average time in seconds of **two** complete TPCx-HS runs.
- **Time Stdv (%)** shows the standard deviation of **Time (Sec)** in percent between the **two** runs.
- **Time Δ (%)** represents the difference in percent of **Time (Sec)** to the time **baseline** in our case scale factor 0.1.

Scale Factor	Data Size	Data Δ (%)	Network	Metric (HSph@SF)	Time (Sec)	Time Stdv (%)	Time Δ (%)
0.1	100 GB	baseline	shared	0.03	10721.75	0.59	baseline
0.3	300 GB	+200	shared	0.03	32142.75	0.50	+199.79
1	1 TB	+900	shared	0.03	105483.00	0.41	+883.82
0.1	100 GB	baseline	dedicated	0.16	2234.75	0.72	baseline
0.3	300 GB	+200	dedicated	0.18	6001.00	0.89	+168.53
1	1 TB	+900	dedicated	0.17	21047.75	1.53	+841.84

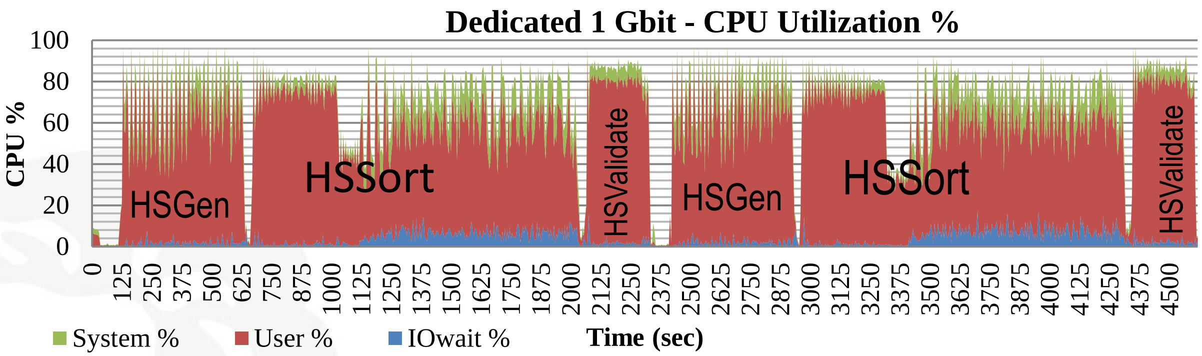
Data Scaling Behavior

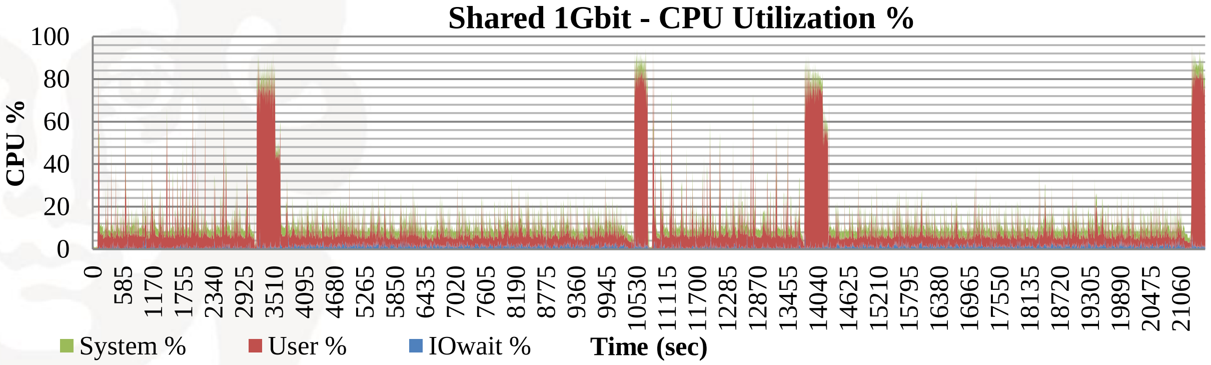
- TPCx-HS Scaling Behavior (0 on the *X* and *Y-axis* is equal to the baseline of SF 0.1/100GB)

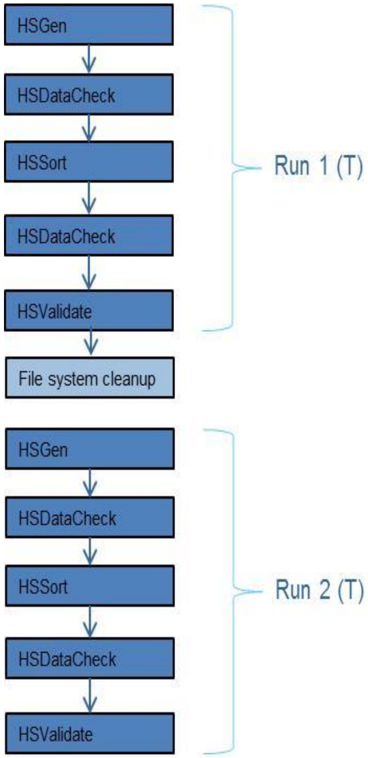


CPU (Shared vs. Dedicated)

- Performance Analysis Tool (PAT) (https://github.com/intel-hadoop/PAT)
- ***Worker Node*** average statistics - measured for 100GB scale factor.


Dedicated 1 Gbit - CPU Utilization %


Shared 1Gbit - CPU Utilization %


HSGen → HSDataCheck → HSSort → HSDataCheck → HSValidate → File system cleanup → HSGen → HSDataCheck → HSSort → HSDataCheck → HSValidate

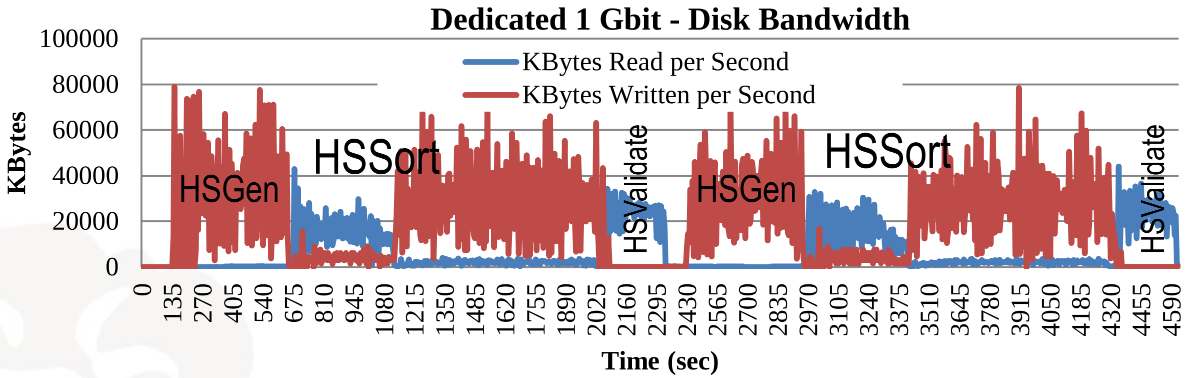
Run 1 (T)

Run 2 (T)

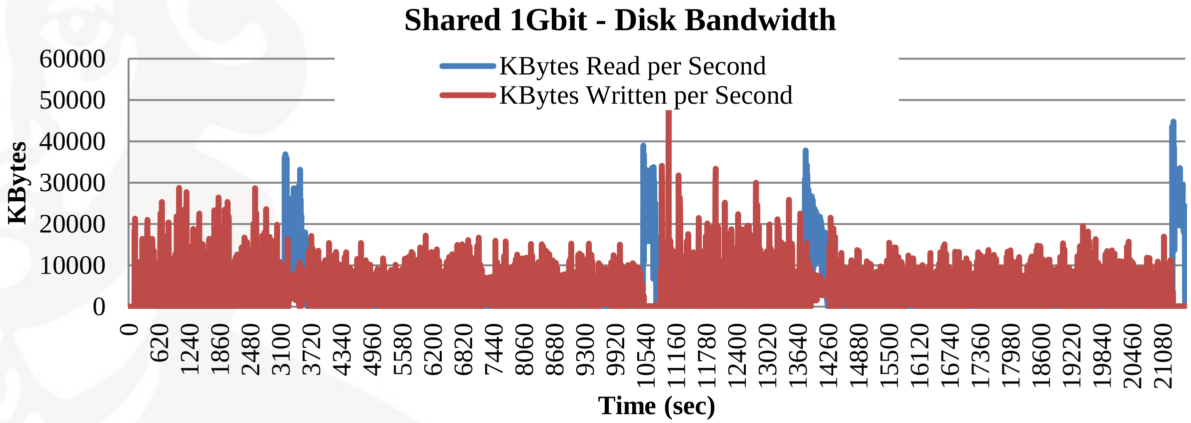
- The *shared* setup ***performs 5-6 times slower*** than the *dedicated* setup.

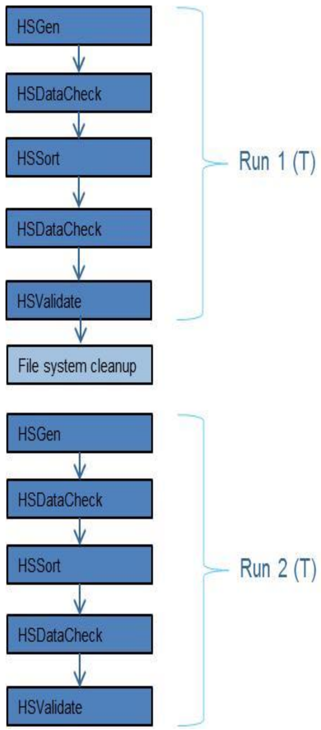
Disk Bandwidth (Shared vs. Dedicated)

- **Dedicated** setup: on average ***read throughput*** is around ***6.4 MB*** per second and ***write throughput*** is around ***18.6 MB*** per second.



- **Shared** setup: on average ***read throughput*** is around ***1.4 MB*** per second and ***write throughput*** is around ***4 MB*** per second.





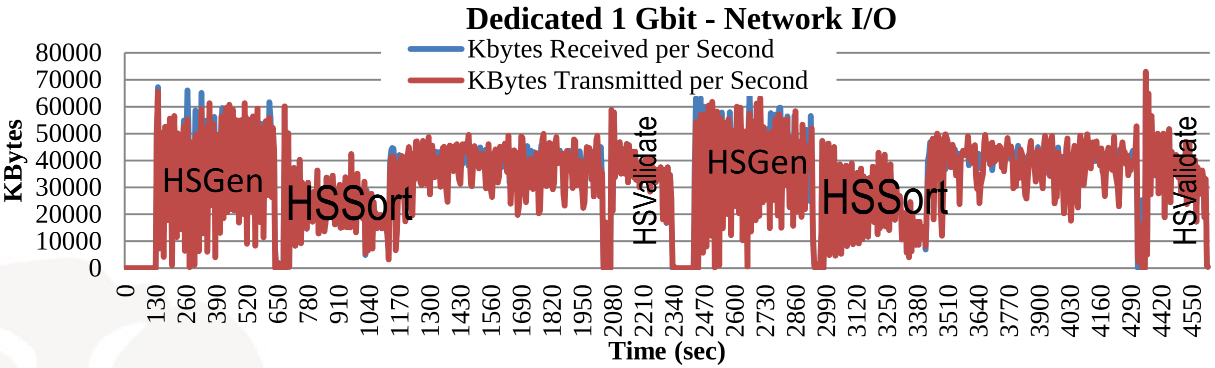
HSGen → HSDataCheck → HSSort → HSDataCheck → HSValidate — Run 1 (T)

File system cleanup

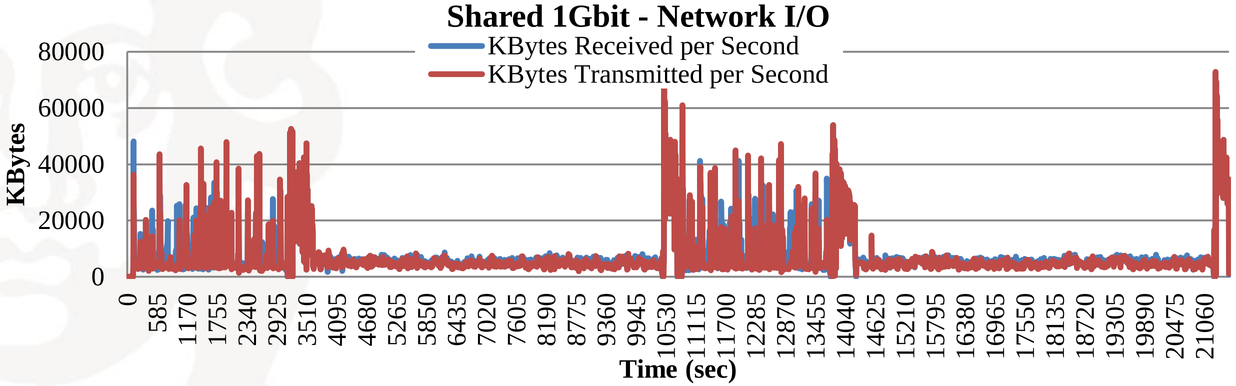
HSGen → HSDataCheck → HSSort → HSDataCheck → HSValidate — Run 2 (T)

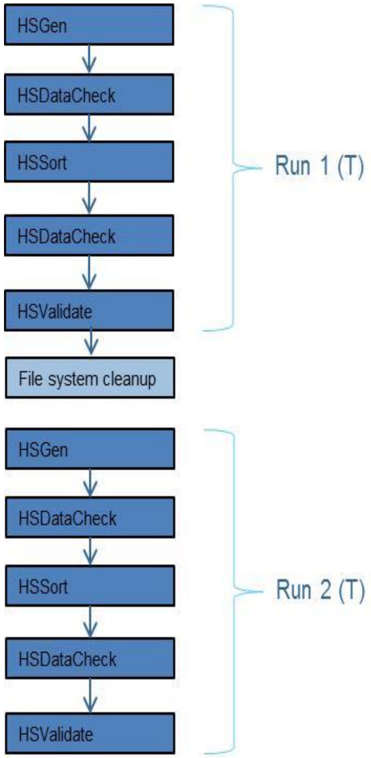
Network (Shared vs. Dedicated)

- ***Dedicated*** setup: on average received ***32.8MB per second*** and transmitted ***30.6MB per second.***



- ***Shared*** setup: on average are ***received 7.1MB per second*** and transmitted ***6.4MB per second***





- The dedicated setup achieves **almost 5 times better network utilization.**

Summary of **Resource Utilization**

Worker Nodes		
Network Type:	Dedicated 1Gbit	Shared 1Gbit
Scale Factor:	100GB	100GB
Avg. CPU Utilization - User %	56.04	12.44
Avg. CPU Utilization - System %	9.52	3.71
Avg. CPU Utilization - IOwait %	3.61	1.28
Avg. Context Switches per Second	20788.16	14233.08
Memory Utilization %	92.31	92.93
Avg. Kbytes Transmitted per Second	31363.57	6548.16
Avg. Kbytes Received per Second	33636.93	7297.27
Avg. Kbytes Read per Second	6532.24	1438.23
Avg. Kbytes Written per Second	19010.01	4087.33
Avg. Read Requests per Second	111.75	24.70
Avg. Write Requests per Second	39.50	9.71
Avg. I/O Latencies in Milliseconds	136.87	69.83

🟦 CPU

🟨 Memory

🟩 Network

⬜ Disk

Lessons Learned

- The dedicated network setup is around ***5 times faster*** than the shared network setup in terms of:
  - Execution time
  - HSph@SF metric
  - Average read and write throughput per second
  - Network utilization

- The TPCx-HS benchmark is a good choice for testing/comparing the network performance of clusters!

Technical Report

GOETHE UNIVERSITÄT FRANKFURT AM MAIN

Evaluating Hadoop Clusters with TPCx-HS

http://arxiv.org/abs/1509.03486

Todor Ivanov and Sead Izberovic

Contact

Todor Ivanov
todor@dbis.cs.uni-frankfurt.de

Frankfurt Big Data Lab
-understanding and applying technologies for Big Data-

Goethe University Frankfurt am Main, Germany
http://www.bigdata.uni-frankfurt.de/