www.bsc.es

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

ALOJA: Cost-effectiveness study of Hadoop deployments

Nicolas Poggi, Senior Researcher

EXCELENCIA
SEVERO
OCHOA

BSC~Microsoft Research Centre

December 2014

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

INSTITUTIONAL
ABOUT THE PROJECT

Barcelona Supercomputing Center (BSC)

- 22 year history in Computer Architecture research
  - European Center for Parallelism of Barcelona (CEPBA)
    - Based at the Technical University of Catalonia (UPC) in 1991
  - Long track record with IBM in chip Architecture & Parallelism
- Led by Mateo Valero
  - ACM fellow, Eckert-Mauchly award 2007, Goode award 2009
  - Active research staff with more than 1000 publications
  - Large ongoing life science computational projects
    - Computational Genomics
    - Molecular modeling & Bioinformatics
    - Protein Interactions & Docking
  - In place computational capabilities
    - Mare Nostrum Super Computer

- Prominent body of research activity around Hadoop since 2008
  - Previous to ALOJA
    - SLA-driven scheduling (Adaptive Scheduler), in memory caching, etc.
  - Group page: http://www.bsc.es/autonomic

BSC-MSRS Centre and ALOJA

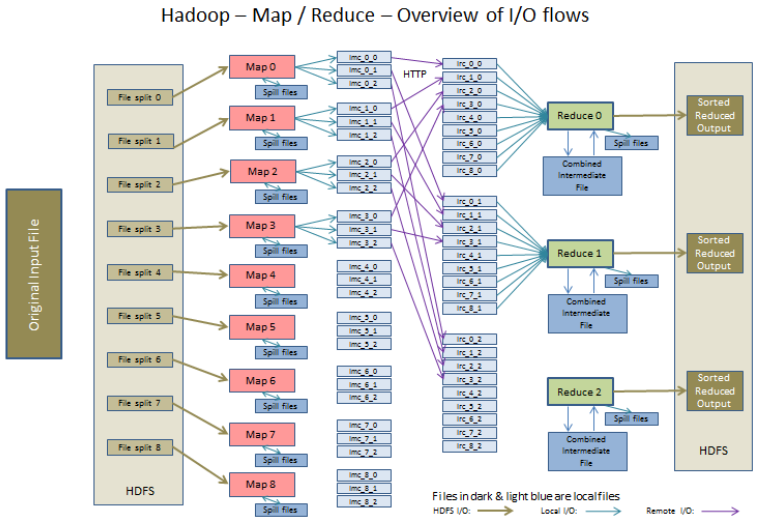
BSC Microsoft Research Centre

- Long-term relationship between
  - BSC and Microsoft Research and Microsoft product teams
- Previous research at the intersection of computer architecture, language implementation, and systems software, and performance profiling
- Open model:
  - **No patents, public IP, publications and open source main focus**
  - 87 publications, 4 Best paper awards

- **ALOJA** is the latest phase of the engagement
- With intent to explore:
  - upcoming hardware architectures
  - and building automated mechanism
- for deploying cost-effective Hadoop clusters.

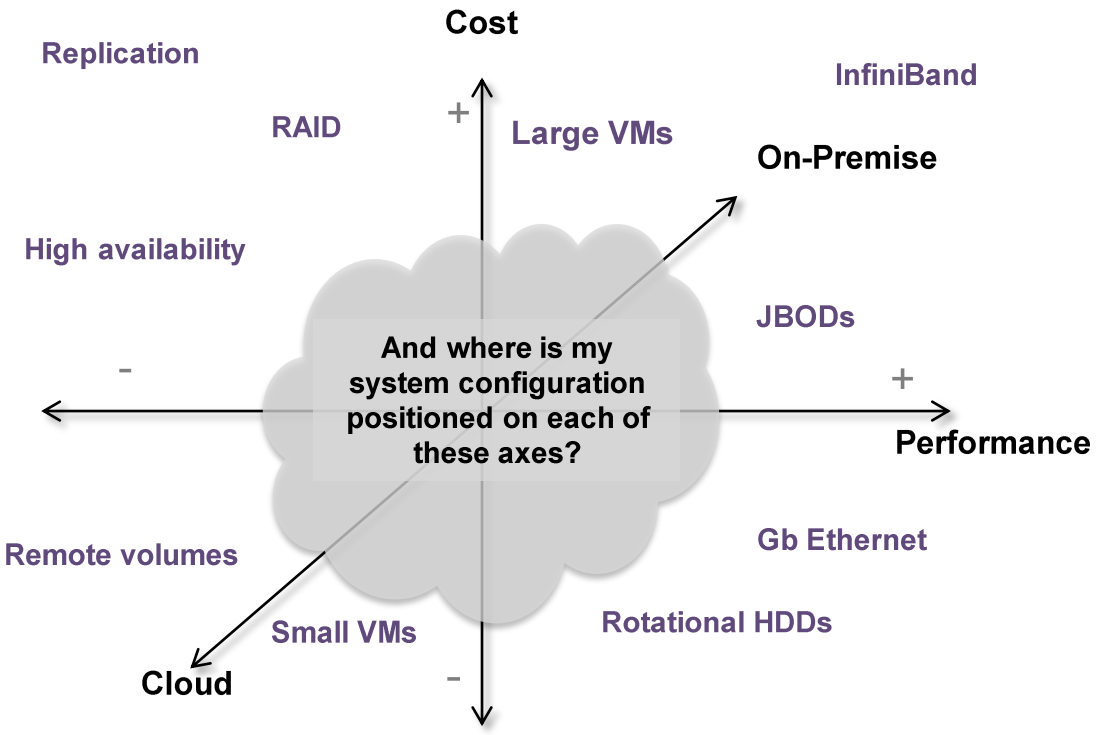
Motivation

- The Hadoop framework implements a complex distributed execution model
  - Over 100 interrelated config parameters
  - Requires manual iterative benchmarking and tuning
- Early results show that Hadoop's price/performance
  - are affected by relatively simple SW **>3x**
  - and HW configuration choices **> 3x**
- Commodity HW no longer low-end
  - new affordable hardware from original design (ie., SSDs)
  - Hadoop performs poorly on scale-up
  - or low power HW
- New Cloud services for Hadoop
  - IaaS and PaaS
  - Direct vs. remote attached volumes
- Spread Hadoop ecosystem
  - Dominated by vendors
  - Lack of verifiable benchmarks

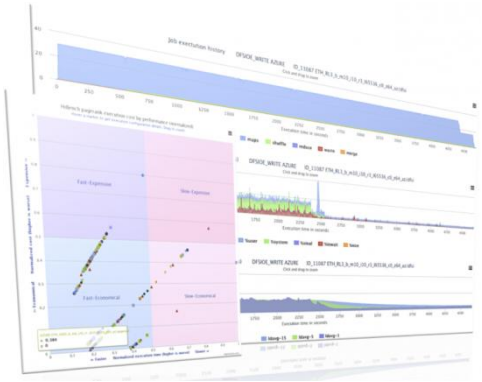


Current scenario

- What is the most cost-effective configuration for my needs?



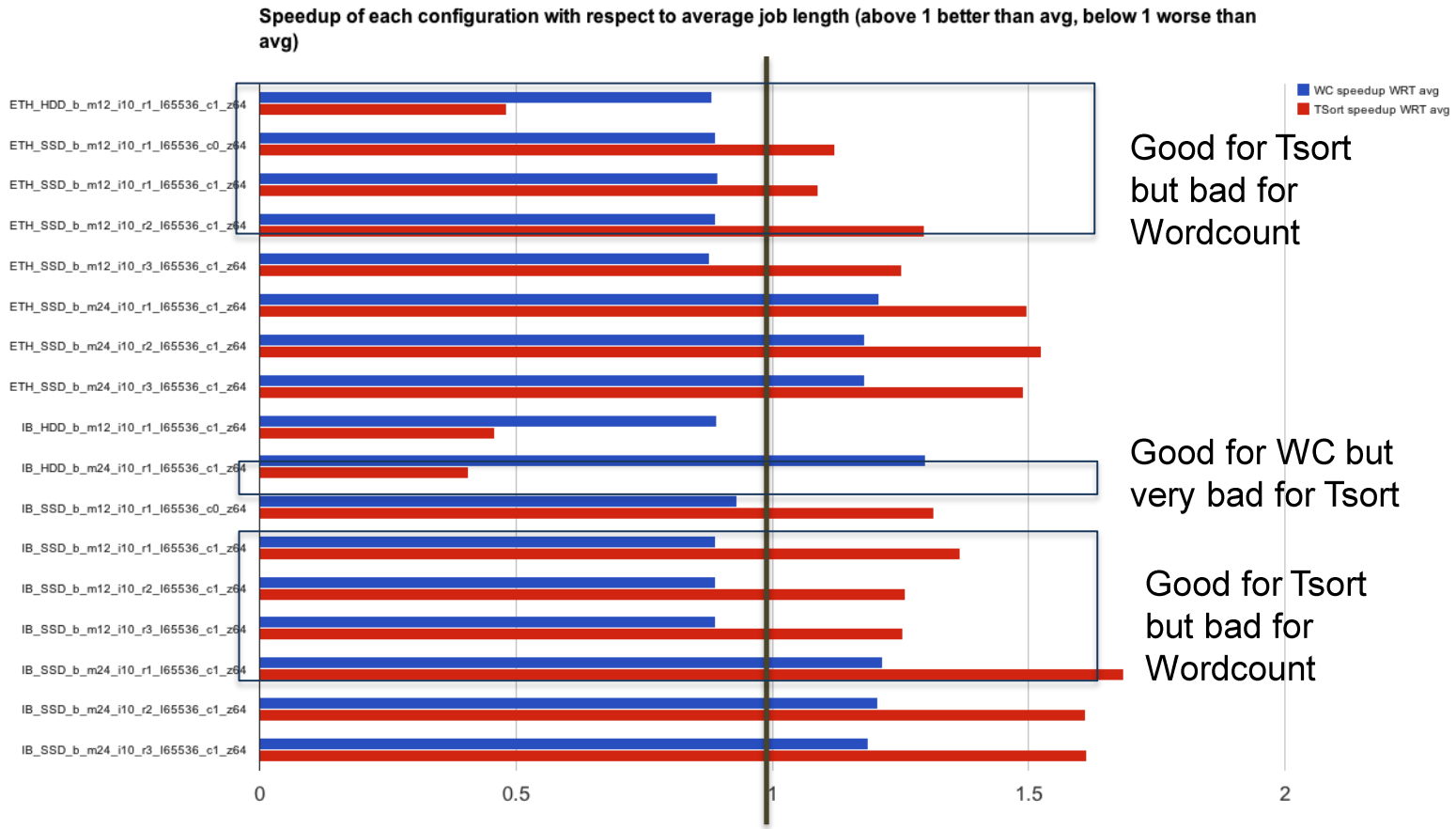
Project ALOJA



- Initiative to produce mechanisms for an **automated characterization of cost-effectiveness** of Big Data deployments with focus on Hadoop
- Results from of a growing need of the community to understand job execution
- Explore different configuration deployment options and their tradeoffs
  - Both Software and Hardware
  - Cloud and on-premise
  - High-end, mid range, low-end
- Seeks to provide both **knowledge**, tools, and an online service
  - to with which users make better informed decisions
  - and reduce the TCO for their Big Data infrastructures

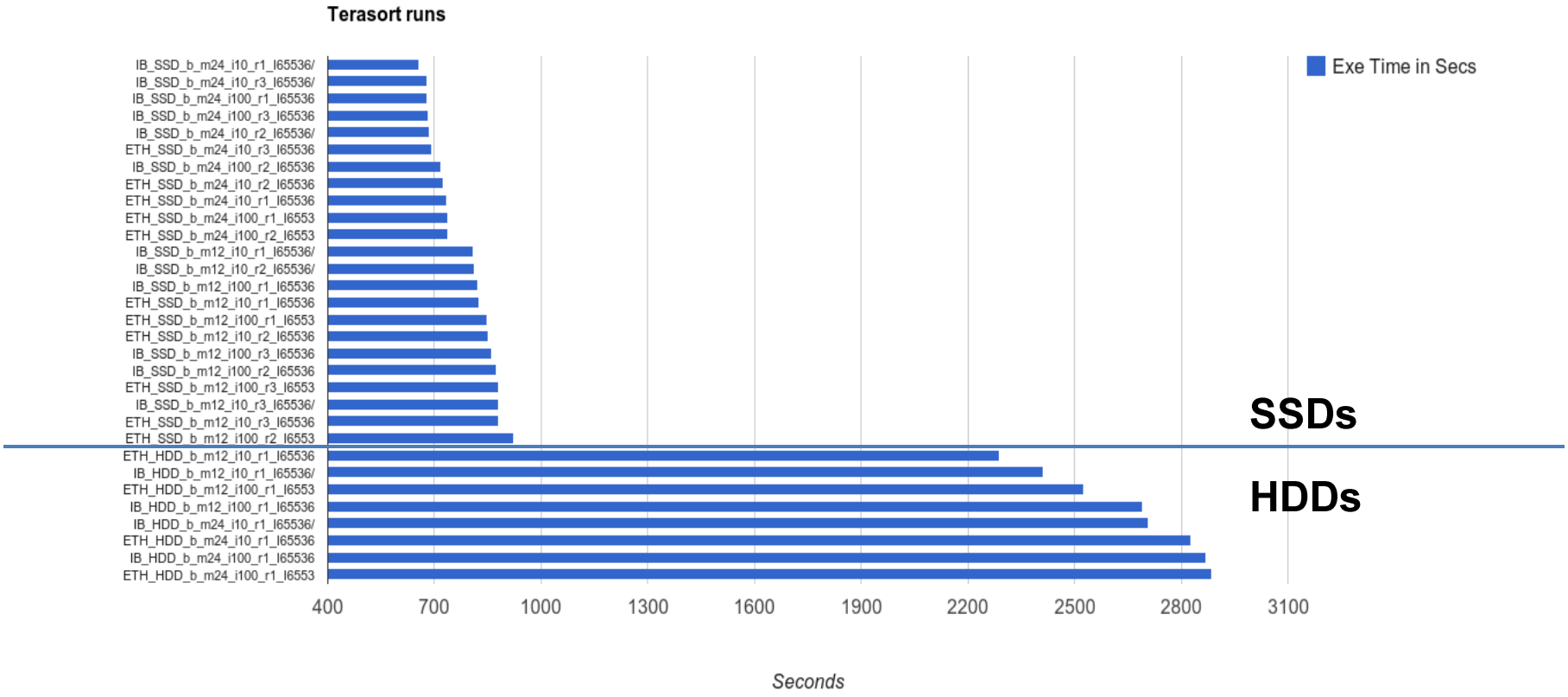
Early exploration: Job resource configs

Is there one software configuration iteration that fits everybody?



Vertical line: Average performance for this workload across configurations
Values to the right: above average
Values to the left: below average

Early exploration: HW technology impact



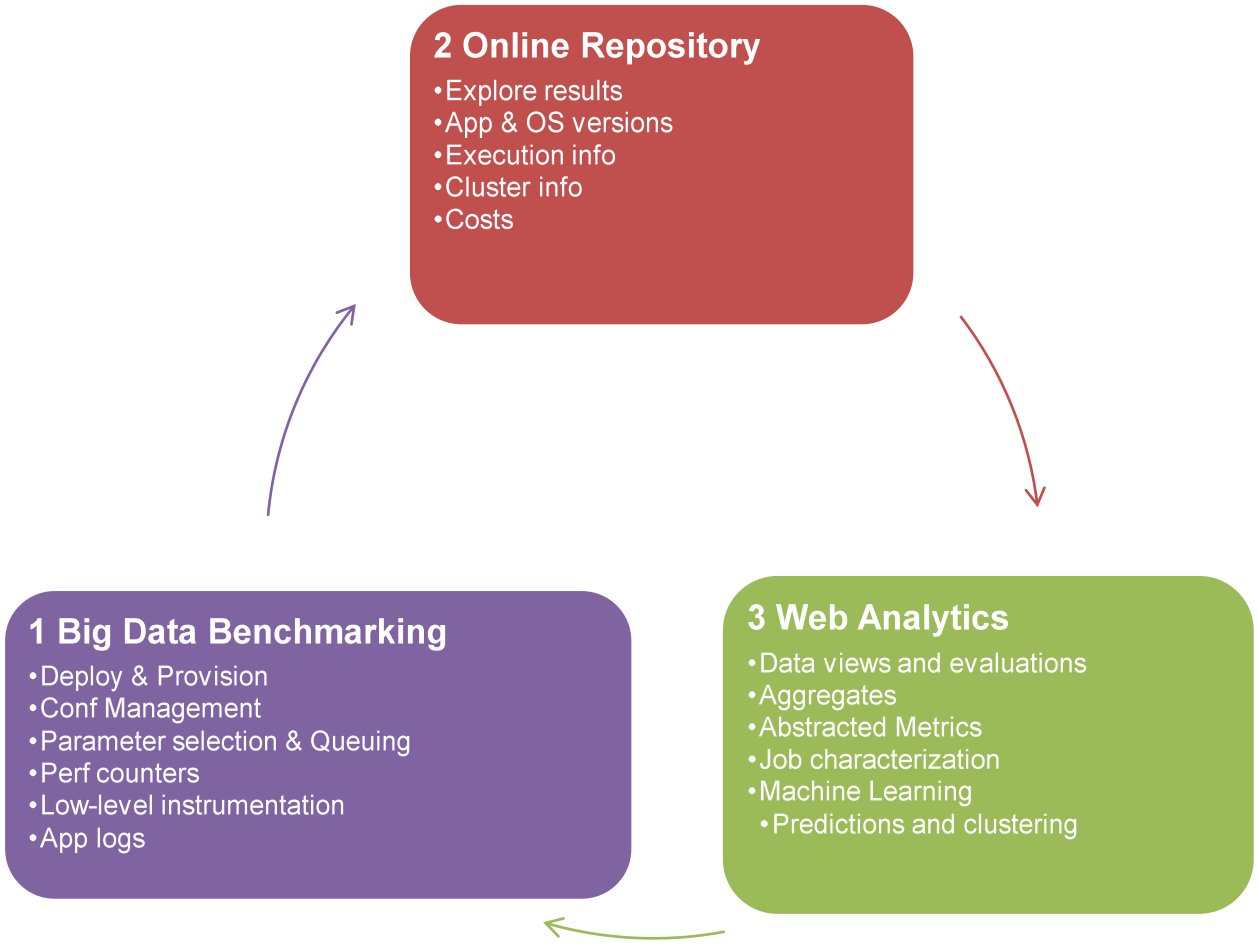

BSC
Barcelona
Supercomputing
Center
Centro Nacional de Supercomputación
PLATFORM

ALOJA Platform

- Benchmarking, Repository, and Analytics tools for Big Data

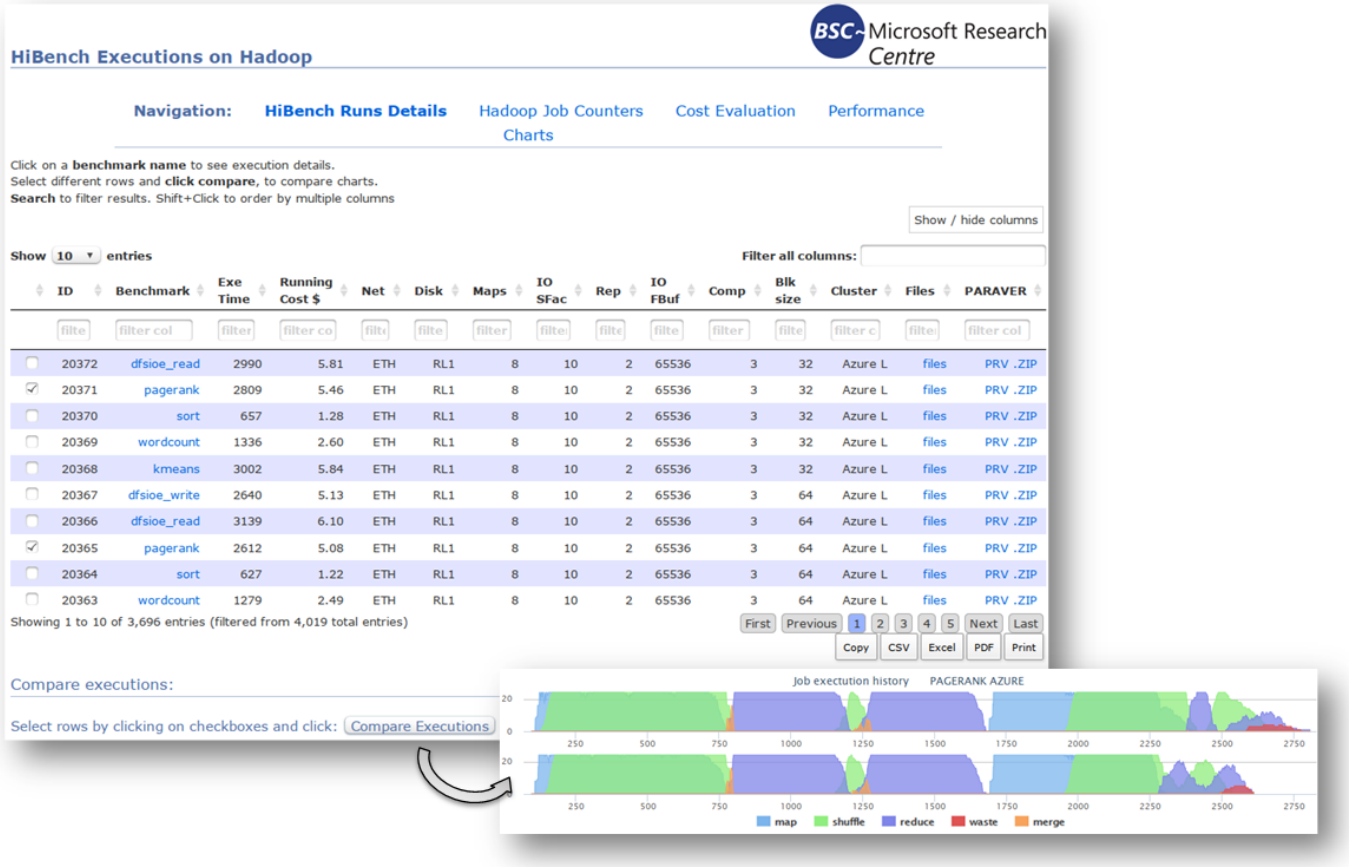
- Composed of open-source
  - benchmarking and configuration management tools,
  - high-level system performance metric collection,
  - low-level Hadoop instrumentation based on **BSC Tools**
  - and Web based data analytics tools
- Online Big Data Benchmark repository of:
  - + 5000 runs (from HiBench)
  - Sharable, comparable, repeatable, verifiable executions
- Not reinventing the wheel, but
  - most current tools designed for production, not for benchmarking
  - leverages current compatible tools and projects

ALOJA main components



ALOJA-WEB

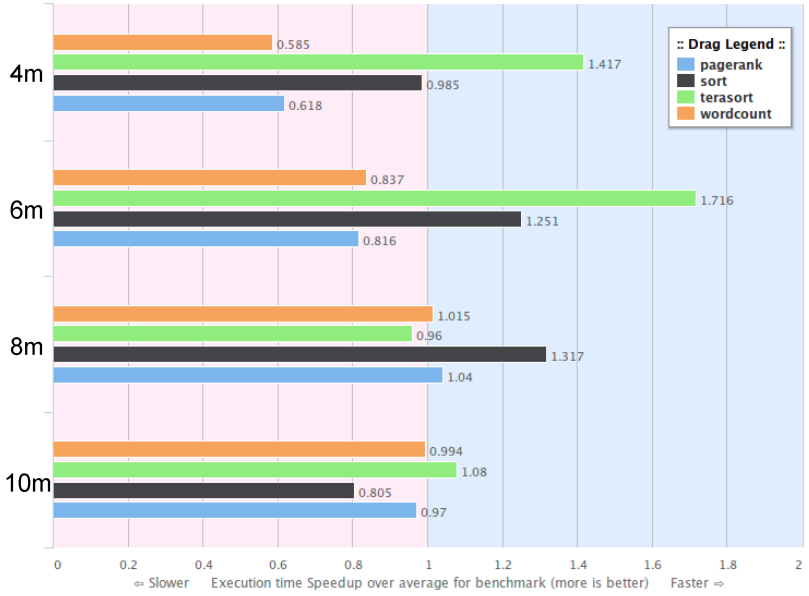
- http://hadoop.bsc.es
- Entry point for explore the results collected from the executions, providing insights on the obtained results through continuously evolving data views.



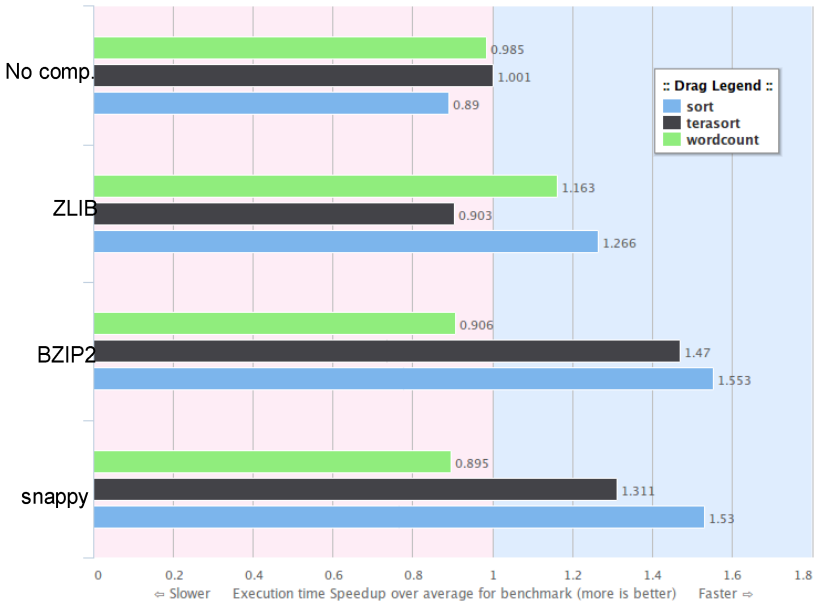

BSC
Barcelona
Supercomputing
Center
Centro Nacional de Supercomputación
ONLINE DEMO

Impact of SW configurations in Speedup

Number of mappers

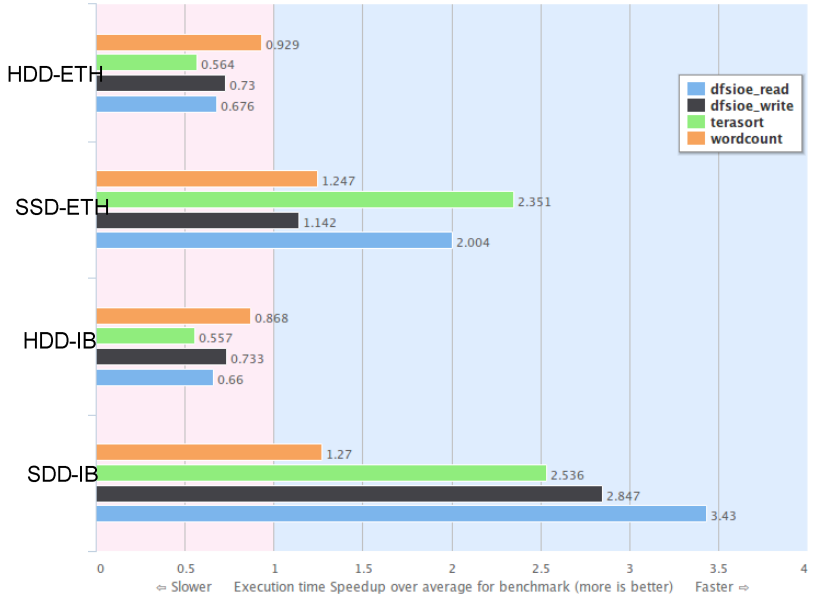


Compression algo.

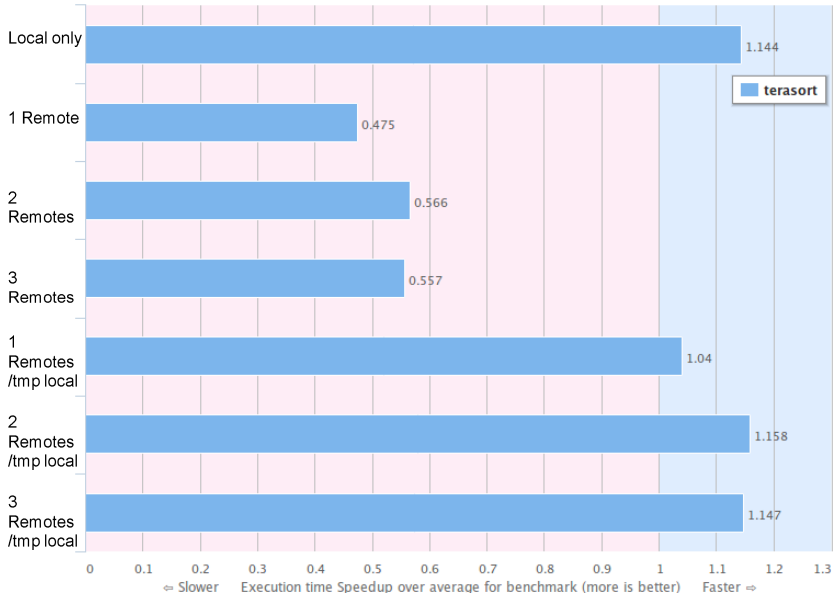


Impact of HW configurations in Speedup

Disks and Network



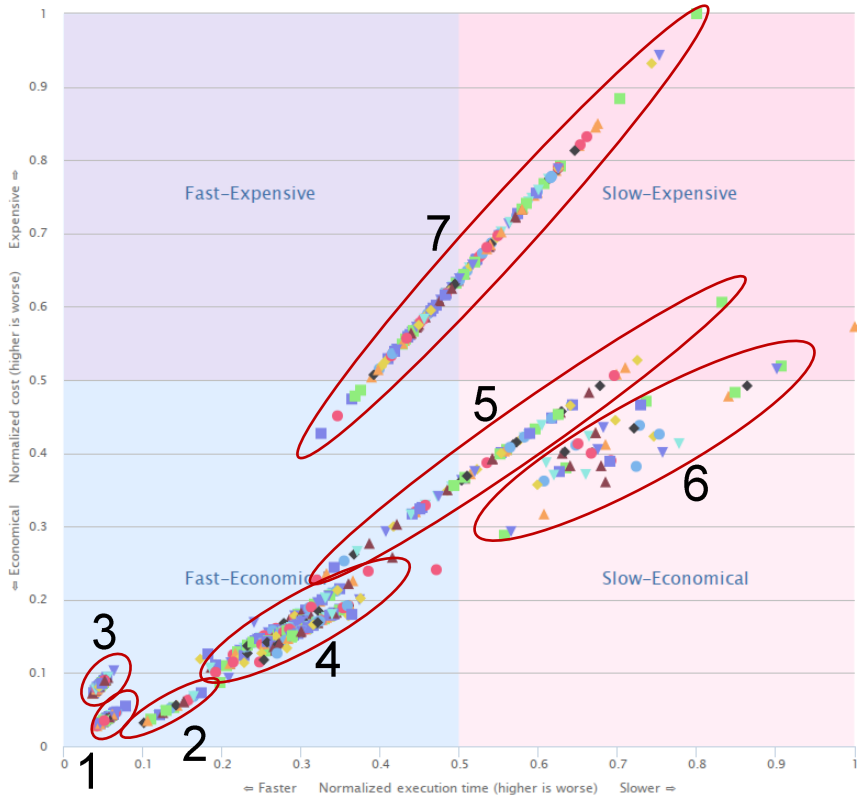
Cloud remote volumes



Cost-effectiveness of SW and HW

Point (0,0) represents most cost-effective execution

Terasort



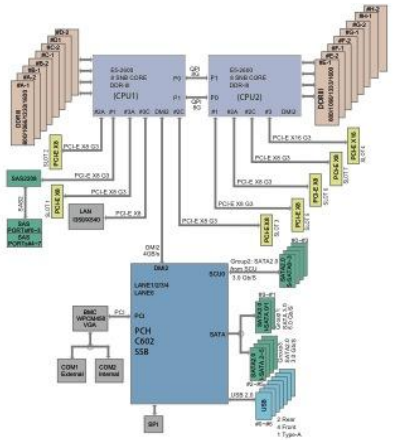
Wordcont

1) On-premise cluster: SSD disks + GbE.
2) Azure IaaS: Only local disk, virtualized SSD and GbE (baseline).
3) On-premise cluster: SSD disks + InfiniBand.

4) Azure IaaS: 1-3 remote vol. and Hadoop /tmp to local disk (SSD) and GbE
5) On-premise cluster: 1 SATA disk + GbE.
6) Azure IaaS: 1-3 remote volumes (Blob storage).
7) On-premise cluster: 1 SATA disk + InfiniBand.

Initial testing infrastructure

- High-End Cluster:
  - 4 nodes, 12 real cores, 64GB RAM, 6x SSDs , 56Gb InfiniBand, 4Gb GbE (bonding)
- Mid-end Cluster *:
  - 18 nodes, 8 real cores, 24GB RAM, 1x SSD, 2x HDDs, 1Gb GbE
- Cloud IaaS (Azure)
  - 4 node A7s, 9 node A7s*, 9 node A3s*
- Cloud PaaS (HDInsight)
  - 4, 8, 16, 32 data nodes*
- Low-powered cluster:
  - 10-node ARM based cluster*

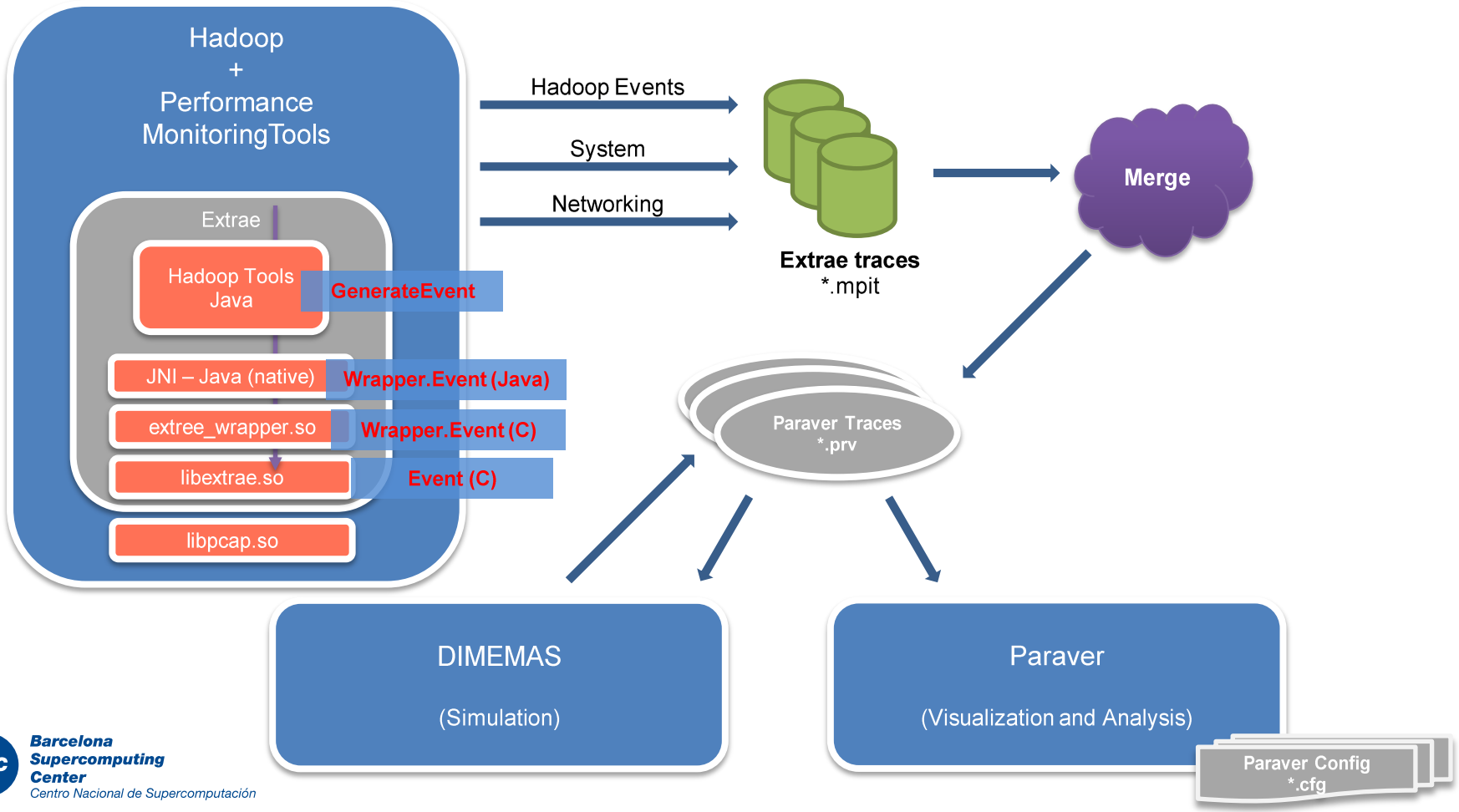


*Result numbers not online yet

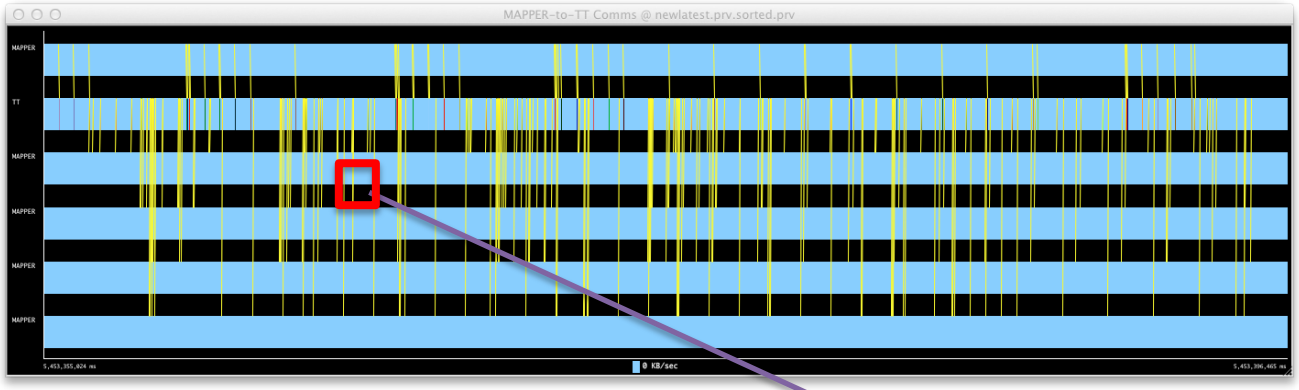

BSC
Barcelona
Supercomputing
Center
Centro Nacional de Supercomputación
DEEP INSTRUMENTATION

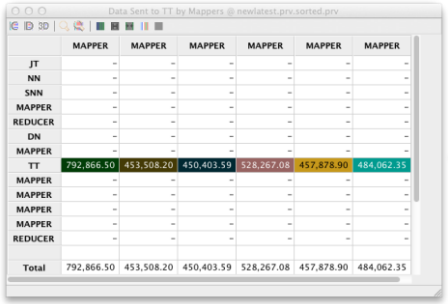
Overview

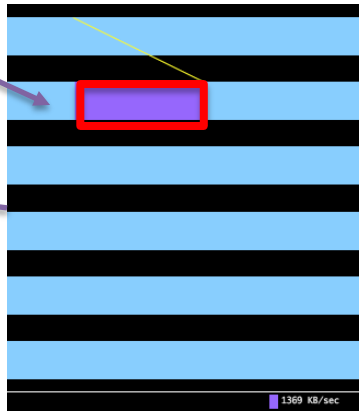
Hadoop Analysis Toolkit and BSC tools



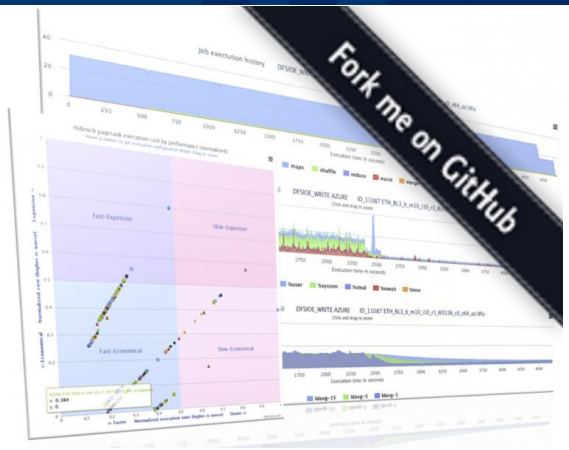
Example: Packet level communications







Extending ALOJA and Collaborating



1. Install prerequisites
2. git clone https://github.com/Aloja/aloja.git
3. cd aloja/vagrant
4. vagrant up
5. Go to: http://localhost:8080

Concluding remarks

- The early findings of the project already show significant value in understanding Hadoop's runtime
  - for optimizing executions times
  - understanding the cost-effectiveness of different configuration and deployment options

- Our intent is that researchers and organizations evaluating or deploying the Hadoop stack will benefit
  - from this growing database of performance results
  - and configuration guidance

www.bsc.es

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Thanks

Q&A

Contact: hadoop@bsc.es