

Benchmarking Data Centers, Big Data, and Cloud Computer Systems

Jianfeng Zhan

Institute of Computing Technology, Chinese Academy of Sciences

HVC tutorial
in conjunction with The 19th IEEE International Symposium
on High Performance Computer Architecture (HPCA 2013)



中科院计算所
INSTITUTE OF COMPUTING TECHNOLOGY

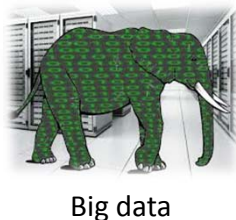
Acknowledgement

- This work is supported by the Chinese 973 project (Grant No.2011CB302500), the NSFC project (Grant No.60933003), the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDA06010401), the Huawei, Baidu, and Intel Funding.



Question one

- Too many concepts!

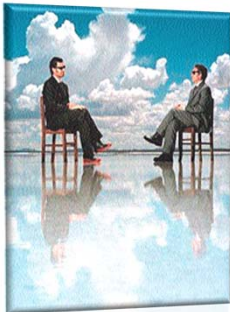


- Life is so tough!



Question two

- Gap between Industry and Academia
 - Longer and longer distance
 - Code
 - Data sets



Question three

- Different benchmark requirements
 - Architecture communities
 - Simulation is very slow
 - Small data and code sets
 - System communities
 - Large-scale deployment is valuable.
 - Real-world users
 - There are three kind of lies: **lies, damn lies, and benchmarks**
 - Real-world applications

ICTBench

- Benchmarking
 - foundation of computer system and architecture researches.
 - clarify the concepts
- ICTBench: three benchmark suites
 - HVCBench: typical data center workloads
 - BigDataBench: big data applications
 - CloudRank: cloud benchmarks
- Appear soon on <http://prof.ict.ac.cn/ICTBench>

HVCBench

- High Volume Computing (HVC)
 - Different from scientific computing: FLOPS
 - Datacenter based distributed computing: OPS
- HVCBench: typical data center workloads
 - Investigate popular applications in important domains
 - Search engine, electronic commerce etc.
 - Each benchmark = a single application.
- Purposes
 - Architecture, system (small-to-medium) researches

BigDataBench

- Characterizing big data applications
 - Not including data-intensive super computing
 - Synthetic data sets varying from 10G~ PB
- Purposes
 - large-scale system and architecture researches
- An incremental approach
 - Release a start-up benchmark suite
 - Workloads in the search engine system
 - Will investigate other important domains

CloudRank

- Cloud computing
 - Elastic resource management
 - Consolidating different workloads
- Cloud benchmarks
 - *Each benchmark = **a group of consolidated data center workloads.***
 - *Three benchmarks: services/ data processing/ desktop*
- Purposes
 - Capacity planning, system evaluation and researches
 - User can customize their benchmarks.

Schedule

- Morning
 - HVCBench: Zhen Jia, 09:00am to 10:00am
 - Coffee break
 - BigDataBench: Wanling Gao, 10:30am to 11:40pm
- Afternoon
 - CloudRank part one: Lin Cai, 02:00pm to 02:45pm
 - CloudRank part two: Jing Quan, 02:45pm to 03:30pm
 - Coffee break
 - GPGPU benchmarks: Zhibin Yu, 04:00pm to 05:00pm

Work in near future

- Establish a community
 - Focus on data center, big data, and cloud benchmarks.
- Hope more organizations and researchers can join us.

- Thanks
- ICTBench homepage:
<http://prof.ict.ac.cn/ICTBench>
- Email: zhanjianfeng@ict.ac.cn