Unistore: A Unified Storage Architecture for Cloud Computing

Yong Chen

Assistant Professor, Computer Science Department

Director, Data-Intensive Scalable Computing Laboratory
Associate Director, Cloud and Autonomic Computing Site

Texas Tech University

(in collaboration with Nimboxx, Inc.)

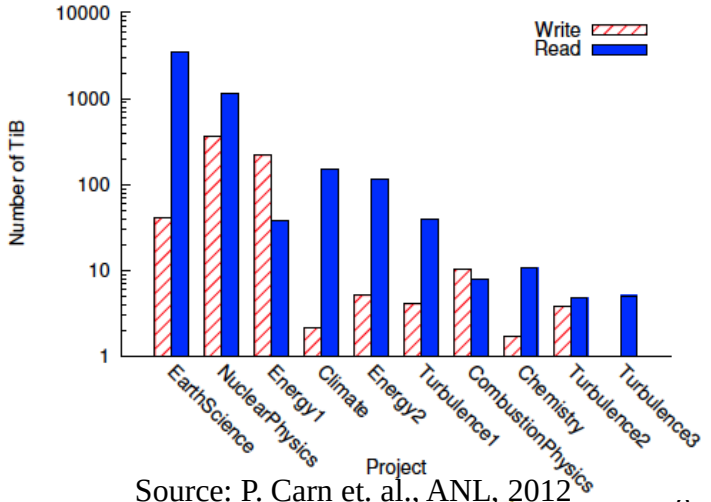
Background

- Enterprise computing apps are highly data intensive
  - Facebook: O(100PB) of storage, with 220M+ photos uploaded/25+TB growth per week years ago
  - Google: 1 trillion pages by 2008, 1 billion pages growth per day
  - Info. retrieval, data mining, online business, social network, etc.
- Scientific computing apps follow the similar trend
  - Scientific computing in Cloud
  - "Research as a service"
- Pressure on the storage system capability increases

Note: data not the latest

Project	On-Line Data	Off-Line
FLASH: Buoyancy-Driven Turbulent Nuclear Burning	75TB	300TB
Reactor Core Hydrodynamics	2TB	5TB
Computational Nuclear Structure	4TB	40TB
Computational Protein Structure	1TB	2TB
Performance Evaluation and Analysis	1TB	1TB
Kinetics and Thermodynamics of Metal and	5TB	100TB
Climate Science	10TB	345TB
Parkinson's Disease	2.5TB	50TB
Plasma Microturbulence	2TB	10TB
Lattice QCD	1TB	44TB

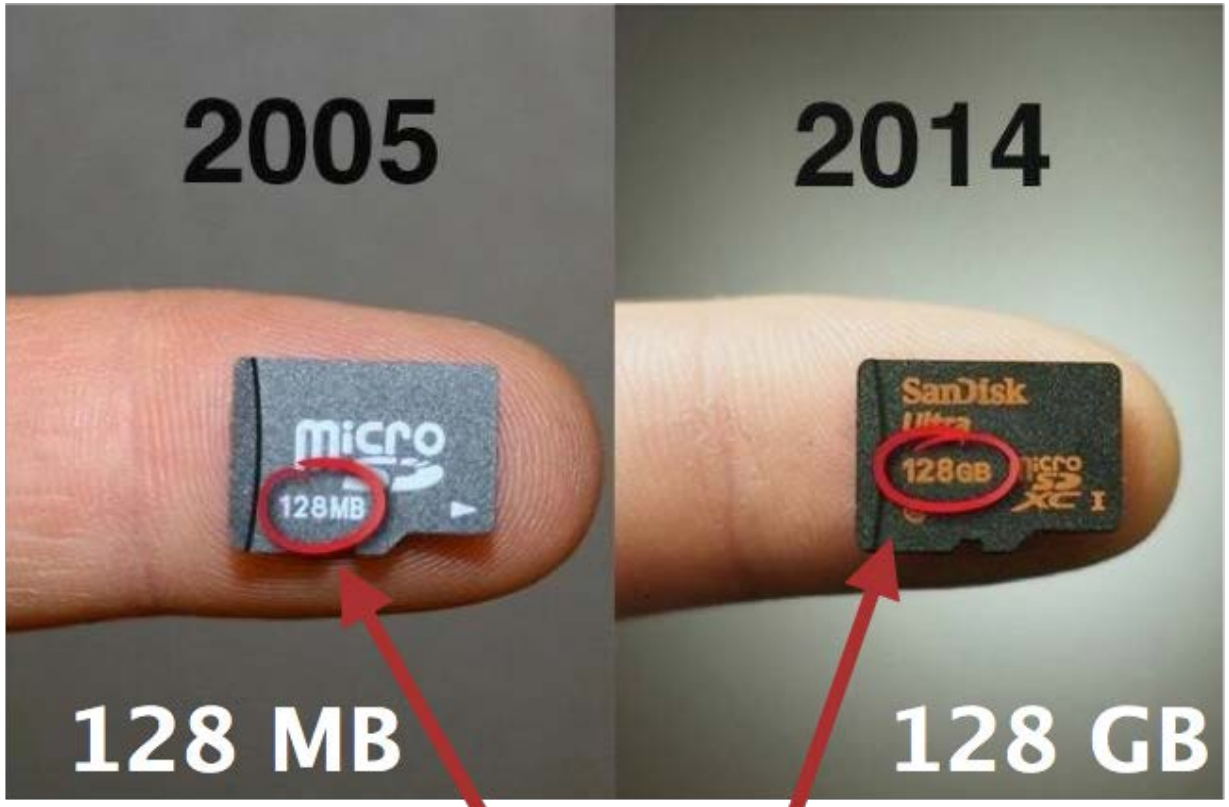
Source: R. Ross et. al., Argonne National Laboratory 2008



Source: P. Carn et. al., ANL, 2012

More Real World Examples

Factor of 1000x increase in less than a decade!



Present day *real world*:

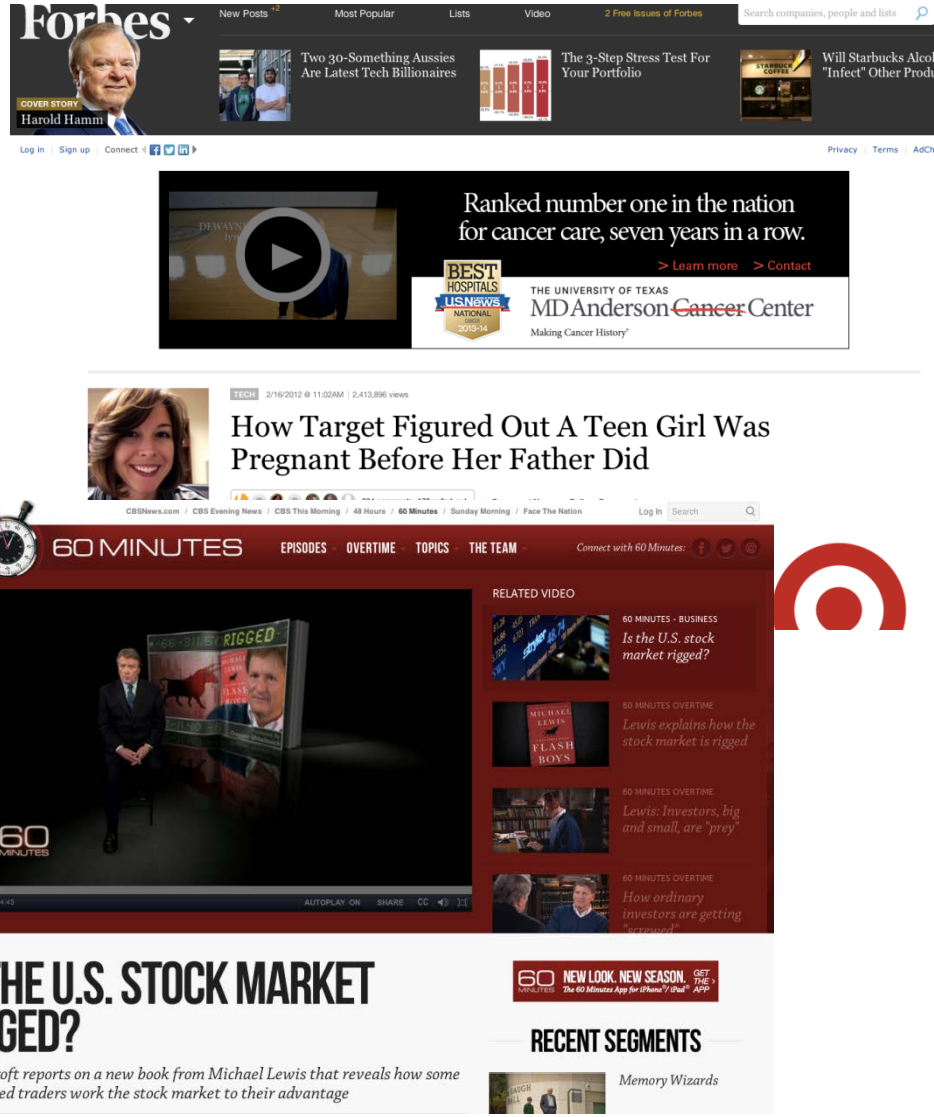
Phones: 100+ *Gigabytes*

Science and Business: 100s to 10,000s of *Petabytes*

Adapted from: A. Sill

Data-driven Discovery Impact

- How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did
- High Frequency Trading 'Flash Boys: A Wall Street Revolt,' by Michael Lewis



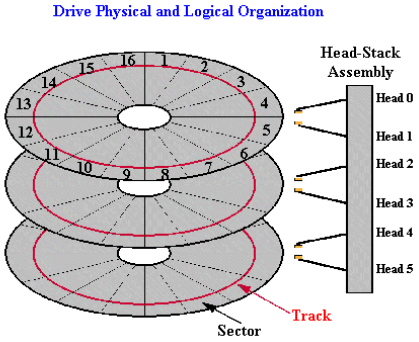
Big Data, Cloud, and Storage Systems

- Big data problem needs highly efficient storage system support
- Data-intensive scientific discovery – the fourth paradigm
  - Theory, experiments, computer simulation
- "Big data" refers to this fourth paradigm of data-driven scientific discovery and innovation
- Regardless of definition/tried definition, scope, HW/SW stack
  - 3-V definitions
    - **Volume**: too many bytes
    - **Velocity**: too high a rate
    - **Variety**: too many sources (structured and unstructured)
  - 5-V definitions
    - Volume, Velocity, Variety, **Veracity/validity, Value**
  - Scope: Infrastructure, Management, Search and mining, Security and privacy, Applications, etc.

Traditional Storage Media

- Over 90% of all data in the world is being stored on magnetic media (hard disk drives, HDDs)
- IBM invented in 1956
- Mechanism remains the same since then
- Various mechanical moving parts
- High latency, slow random access performance, unreliable, power hungry
- Large capacity, low cost (USD 0.10/GB), impressive sequential access performance

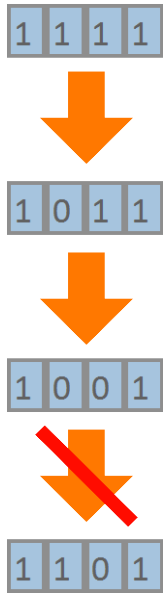
Source: online



Emerged Storage Media

- Non-volatile storage-class memory (SCM)
- Flash-memory based Solid State Drives (SSDs), PCRAM, NRAM, …
- Use microchips which retain data in non-volatile memory (array of floating gate transistors isolated by an insulating layer)
- High throughput, low latency (esp. random accesses), less susceptible to physical shock, power efficient
- Low capacity, high cost (USD 0.90-2/GB), block erasure, memory wear out (10K-100K P/E cycles)

Intel® X25-E SSD

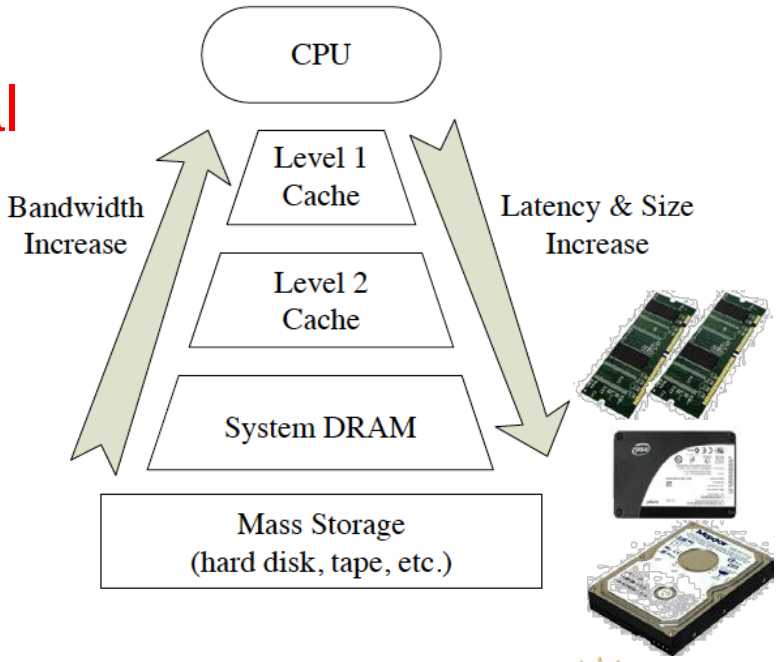


Storage System Needs and Challenges

- Conventional HDDs and emerged SCM complement each other well
- Traditional parallel/distributed file systems designed for HDDs not managing SCM well
- Straightforward solution not optimal
  - Placing SCM as additional memory hierarchy in the current storage tier
  - Intensive writes go through SCM
  - Inclusive setting limits capacity
  - Do not distinguish criticalness of data

Media	Access Time (μs)	Endurance (Write Times)	Norm. Cost[1]
DRAM	<0.01	>1E+16	200
PCM	(<0.055) Read, (>0.15) Write	1E+9	24
SSD	(<45) Read, (>200) Write	1E+5	6
HDD	<5000	>1E+16	1

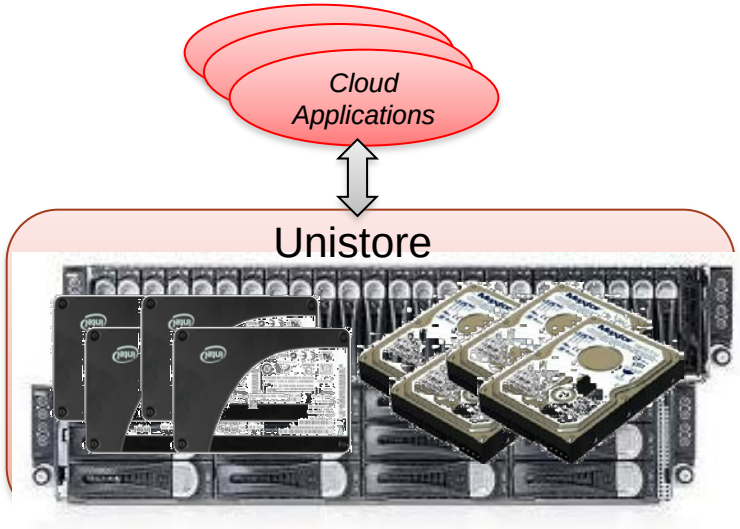
[1] Normalized average cost per GB based on HDD.



Our Solution - Unistore: A Unified Storage Architecture

- A distributed file system store
- Distinguish workload and access patterns
- Consider SCM and HDD features
  - Block erasure, internal concurrency, wear-leveling
- High performance, high cost SCM keeps
  - Semantically critical data, e.g. metadata blocks
  - Performance critical data, e.g. frequently accessed
- High capacity, low cost HDD keeps
  - Low-priority large data sets
- In-situ processing runtime optimizations

- Goals
  - Performance close to SCM devices
  - Capacity close to the SCM+HDDs
  - Combines merits and avoids drawbacks



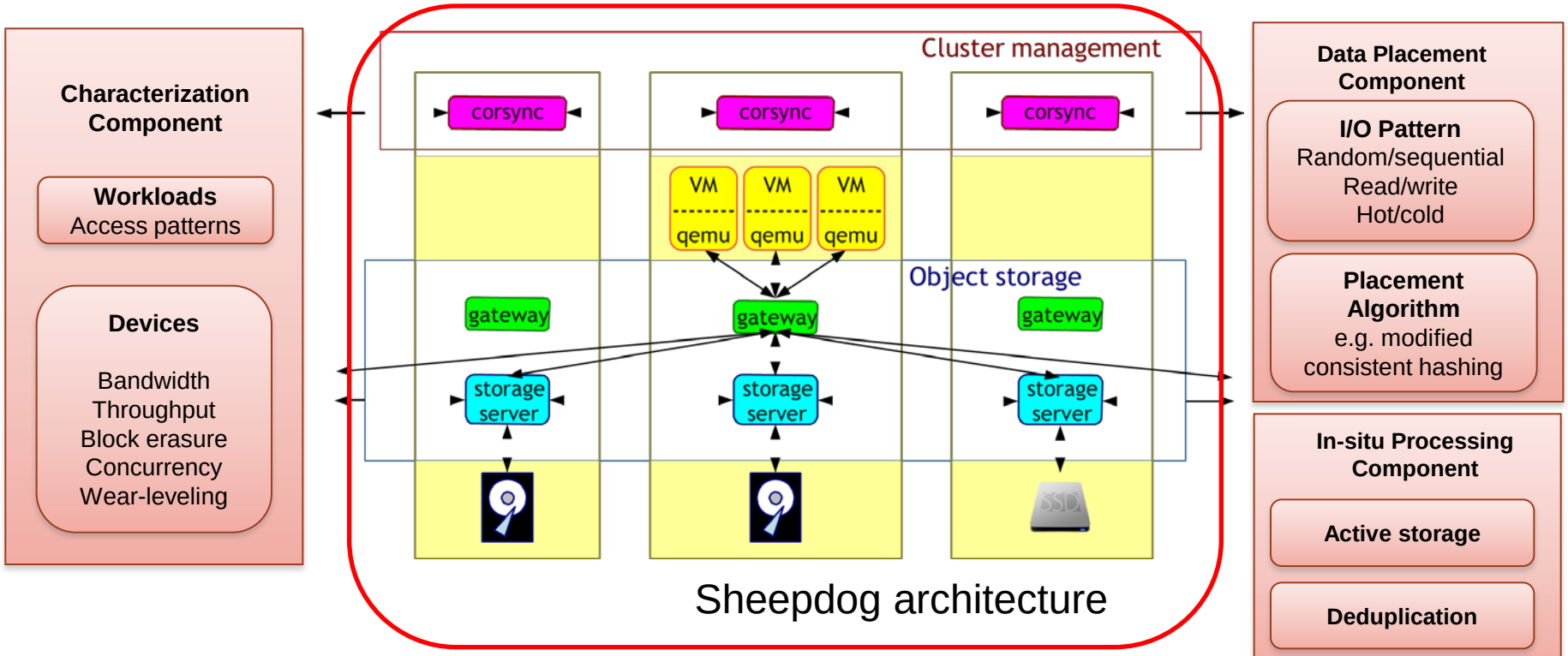
Cloud and Autonomic Computing Center

TEXAS TECH UNIVERSITY

NIMBOXX

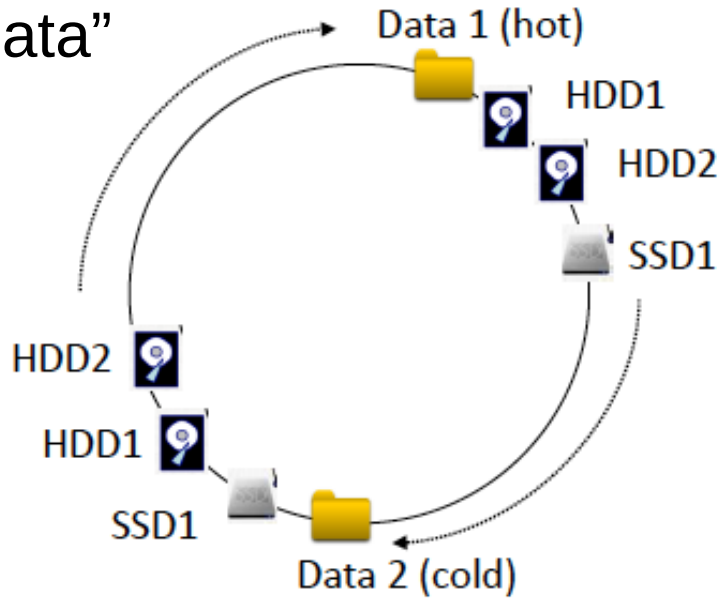
Unistore: High-level View

- We are developing a prototype based on Sheepdog, a distributed object storage system for volume and container services
- https://sheepdog.github.io/sheepdog/



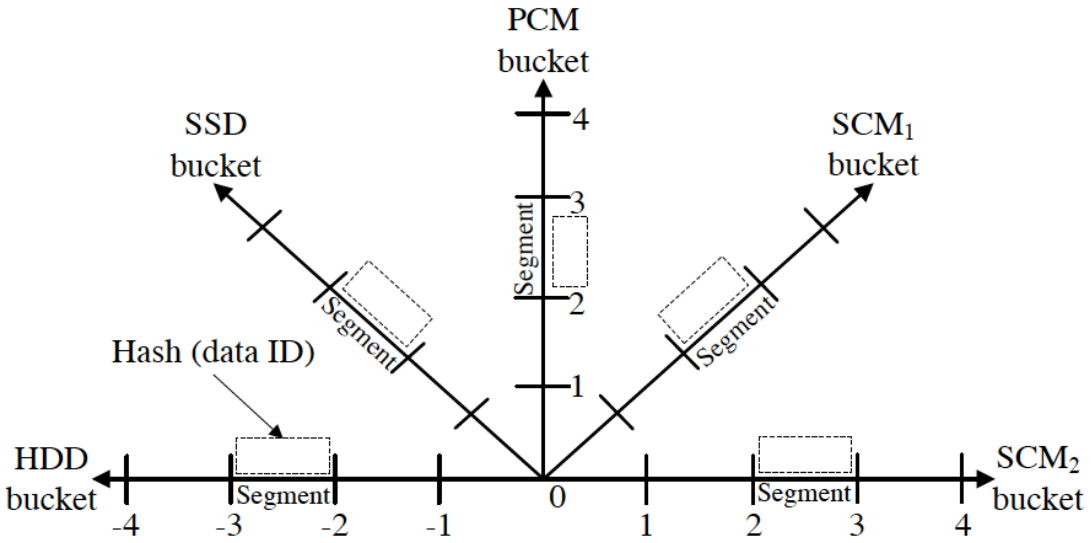
Enhanced Consistent Hashing

- No central namenode/metadata server or table-management server
- Only minimum information stored (node info)
- Handle node addition and removal naturally
- Data distributed uniformly
- Better scalability, handle “Big Data”
- However no consideration for distinct storage device features



SUORA Algorithm

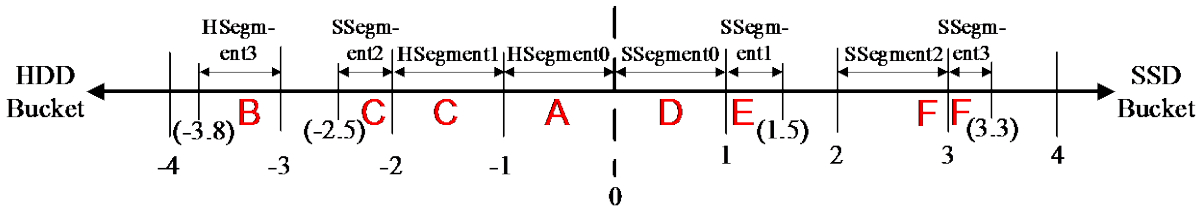
- SUORA: Scalable and Uniform storage via Optimally-adaptive and Random number Addressing
  - Motivated by ASURA algorithm by K. Ishikawa
  - Divide heterogeneous devices into multi-dimensional space
  - Assigns devices to different segments in each dimension
  - Map data among segments and dimensions
- Model

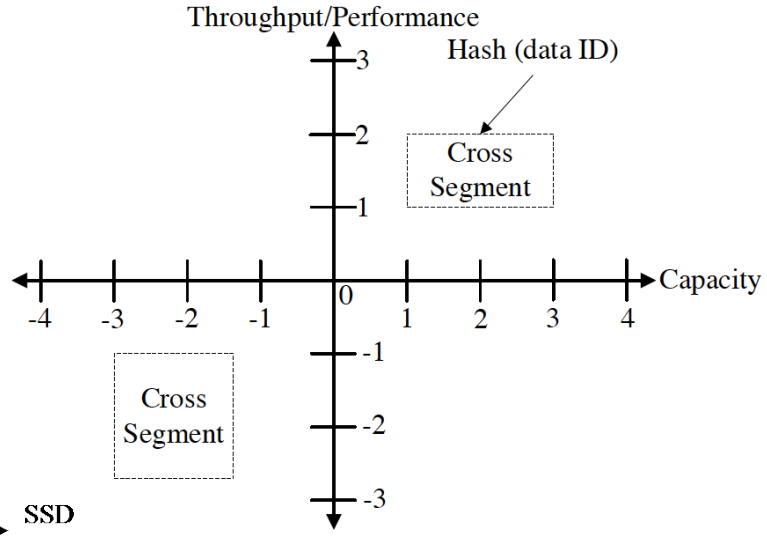


SUORA Algorithm (cont.)

- SUORA models devices as buckets and segments in numbered lines
- Different lines (dimensions) represent different factors considered (e.g. throughput, capacity, etc.)

Node	Bucket	Capacity	Assigned Segments
A	HDD	1TB	HSegment0 (-1, 0)
B	HDD	0.8TB	HSegment3 (-3.8, -3)
C	HDD	1.5TB	HSegment1 (-2, -1) HSegment2 (-2.5, -2)
D	SSD	0.6TB	SSegment0 (0, 1)
E	SSD	0.3TB	SSegment1 (1, 1.5)
F	SSD	0.8TB	SSegment2 (2, 3) SSegment3 (3, 3.3)

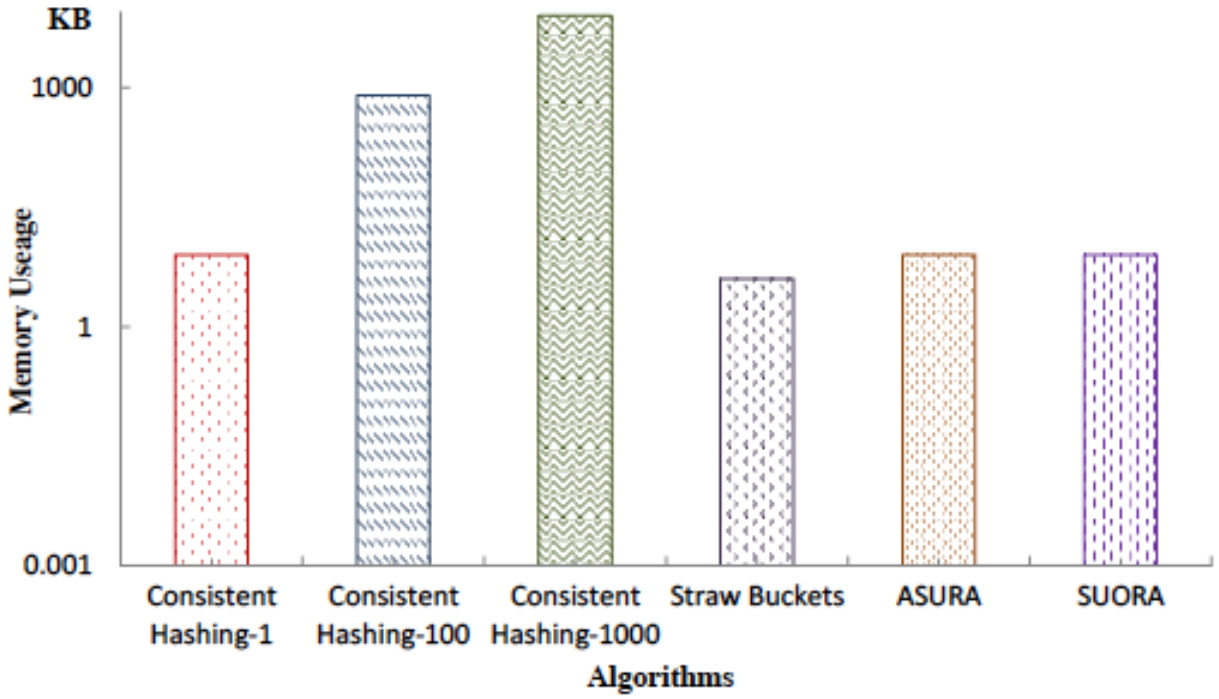




SUORA Algorithm (cont.)

- Initial comparison with CRUSH, consistent hashing, ASURA different algorithms
- In terms of computation time (of data distribution), memory footprint, uniform distribution, adaptation to device additions and removal
- The initial comparison has shown the promise of SUORA algorithm in Unistore architecture

SUORA Algorithm (cont.)



Memory consumption

Cloud and Autonomic Computing at Texas Tech University (CAC@TTU)

- Director: Dr. Alan Sill (alan.sill@ttu.edu)
- Associate Directors: Drs. Ravi Vadapalli and Yong Chen

Cloud and Autonomic Computing Center

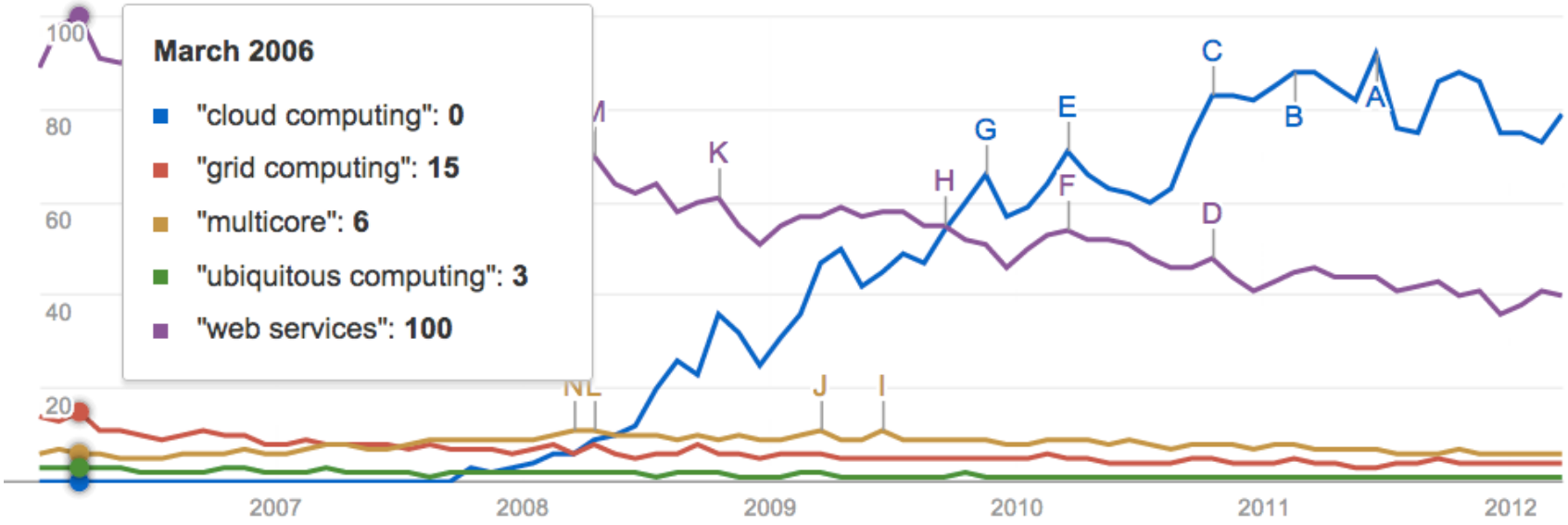
TEXAS TECH UNIVERSITY

NIMBOXX

Cloud Computing: An Emerged Paradigm

Cloud computing paradigm

Scientific computing in Cloud
"Research as a service"



Courtesy: Google Trends Google

Cloud and Autonomic Computing Center

TEXAS TECH UNIVERSITY

NIMBOXX

NSF

Concept and Motivations

- A computing paradigm that
  - Uses the Internet and central remote servers to offer data and application services for end-users/businesses (by utilizing service-oriented architecture, virtualization, and storage techniques)
- Driven by
  - Delivered on demand over the Internet, Wi-Fi, cell phone networks
  - No owned infrastructure, no maintenance, pay-as-you-go model
  - Ideal for personal and small- to medium-size businesses
- Clouds
  - Google App Engine, IBM SmartCloud, Amazon EC2, MS Azure, Apple iCloud, DOE Magellan, NSF Chameleon/CloudLab

Research and Development

- Challenges
  - Lack of consensus among Cloud providers and industry-wide standards for Cloud computing paradigm
  - Concerns of risks related to information security, privacy, compliance, and regulation
  - Cloud services in public, and business in private - interoperability issues
- Standards development and standards-based Cloud computing research and development
  - SOA (IaaS, PaaS, SaaS), security, storage, networking, virtualization, programming interface, applications, data centers.

Cloud and Autonomic Computing (CAC) Center

- Cloud and Autonomic Computing Industry/University Cooperative Research Center
- Goals
  - To leverage faculty expertise and industry interests to address pre-competitive R&D challenges that could improve business development and commercialization opportunities
  - To function as a multidisciplinary center fostering long-term collaborative partnerships among industry, academe, government
  - To train a diverse body of students, facilitate the creation of knowledge and technology, and to help accelerating their transfer into industry and commercial product development
  - Areas of R&D interest include all aspects of Cloud computing and advanced distributed and autonomic computing

Cloud and Autonomic Computing (CAC) Center

- Funded by NSF through the IUCRC program
  - Currently in second year of Phase-I
  - Multi-university cooperative research center
    - Texas Tech is becoming the lead site
    - Other center memberships are from Mississippi State and University of Arizona.
- Actively seeking industry partnerships
  - Need a minimum of $150 K across three memberships
  - Each member contributes a minimum of $35K/year (only 10% F&A cost; TTU has additional in-kind contributions)
  - Industry contributions play pivotal role in center activities
  - IAB governs the center

Cloud and Autonomic Computing Center

TEXAS TECH UNIVERSITY

NIMBOXX

NSF

Current Industry and Gov't Members

Covenant Health System

DISA DEFENSE INFORMATION SYSTEMS AGENCY DEPARTMENT OF DEFENSE

HAPPY STATE BANK

ERDC US Army Engineer Research and Development Center

NIMBOXX

SOLIEL

STACK VELOCITY

TTU

MSU

Technical Cooperation Agreements:

THE AEROSPACE CORPORATION

(Existing)

Cloud Standards Customer Council

OMG OBJECT MANAGEMENT GROUP

(In Progress)

...

Others to come!

CAC Cloud and Autonomic Computing Center

TEXAS TECH UNIVERSITY

NIMBOXX

NSF

Current Projects and Areas of Strengths

- Current Projects
  - Cloud testbeds deployment, performance benchmarking, interoperability testing, standards implementation platform deployment
  - Risk and financial analytics for population health
  - Design and development of advanced storage systems
  - Financial industry analytics for automated contract analysis
- Areas of Strengths
  - Cloud standards and interoperability
  - Cloud performance testing and application prototyping
  - Cloud and cyberinfrastructure security
  - Cloud storage system design and development
  - Healthcare and big data analytics
  - High performance computing including programming and code tuning
  - Cloud computing workforce training and recruitment assistance

Benefits to Industry Members

- Collaboration with faculty, graduate students, post-doctoral researchers, and other center partners; choice of topics funded
- Continuous interaction/networking with peer industry companies
- Timely access to reports, papers, patents, and intellectual property generated by the center
- Access to unique world-class equipment, facilities, and other CAC infrastructure
- Customized recommendations on standards, software, methods
- Recruitment opportunities among excellent graduate students
- Leveraging of investments, projects, and activities by the entire multi-university center and CAC members
- Spin-off initiatives leading to new partnerships, customers, or teaming for competitive proposals to funded programs

Benefits to Institutions

- Opportunities to work on practical problems picked by IAB with real impact and with business value
- Students get financially supported and develop pre-career relationship with interested industry companies
- Faculty can interact with industry, broaden their visions, address problems of industry interests, increase impact of their research
- Access to special NSF programs only available for IUCRC center
- Opportunities to create new higher educational programs and expertise development
- New grant opportunities and increased research expenditures
- Better chance of commercialization and technology transfer
- Enhanced visibility

Contacts

- Director: Dr. Alan Sill, alan.sill@ttu.edu, 806-834-5940
- Associate Directors:
  - Dr. Ravi Vadapalli, ravi.vadapalli@ttu.edu, 806-834-5941
  - Dr. Yong Chen, yong.chen@ttu.edu, 806-834-0284
- Website:
  - http://cac.ttu.edu
  - http://www.nsfcac.org/

Summary

- Unistore is an ongoing effort of attempting to build a unified storage architecture for Cloud storage system
- With the co-existence and efficient integration of heterogeneous HDDs and SCM devices
- An initial prototype being developed based on Sheepdog
- An enhanced consistent hashing algorithm and an SUORA algorithm are being developed
- An NSF funded Cloud and Autonomic Computing (CAC) center leverages faculty expertise and industry interests to address R&D challenges that could improve business development and commercialization opportunities

Thank You

Please visit:
http://cac.ttu.edu/, http://discl.cs.ttu.edu/

Acknowledgement: This project is funded by CAC@TTU through Nimboxx membership contribution and the CAC@TTU is funded by the National Science Foundation under grants IIP-1362134 and IIP-1238338.

National Science Foundation
WHERE DISCOVERIES BEGIN

cac
Cloud and Autonomic Computing Center

TEXAS TECH
UNIVERSITY

NIMBOXX

NSF