Unistore: A Unified Storage Architecture for Cloud Computing

Progress and Proposal

Presenter: Wei Xie

Project Members: Wei Xie, Jiang Zhou, and Yong Chen

Data-Intensive Scalable Computing Laboratory (DISCL)

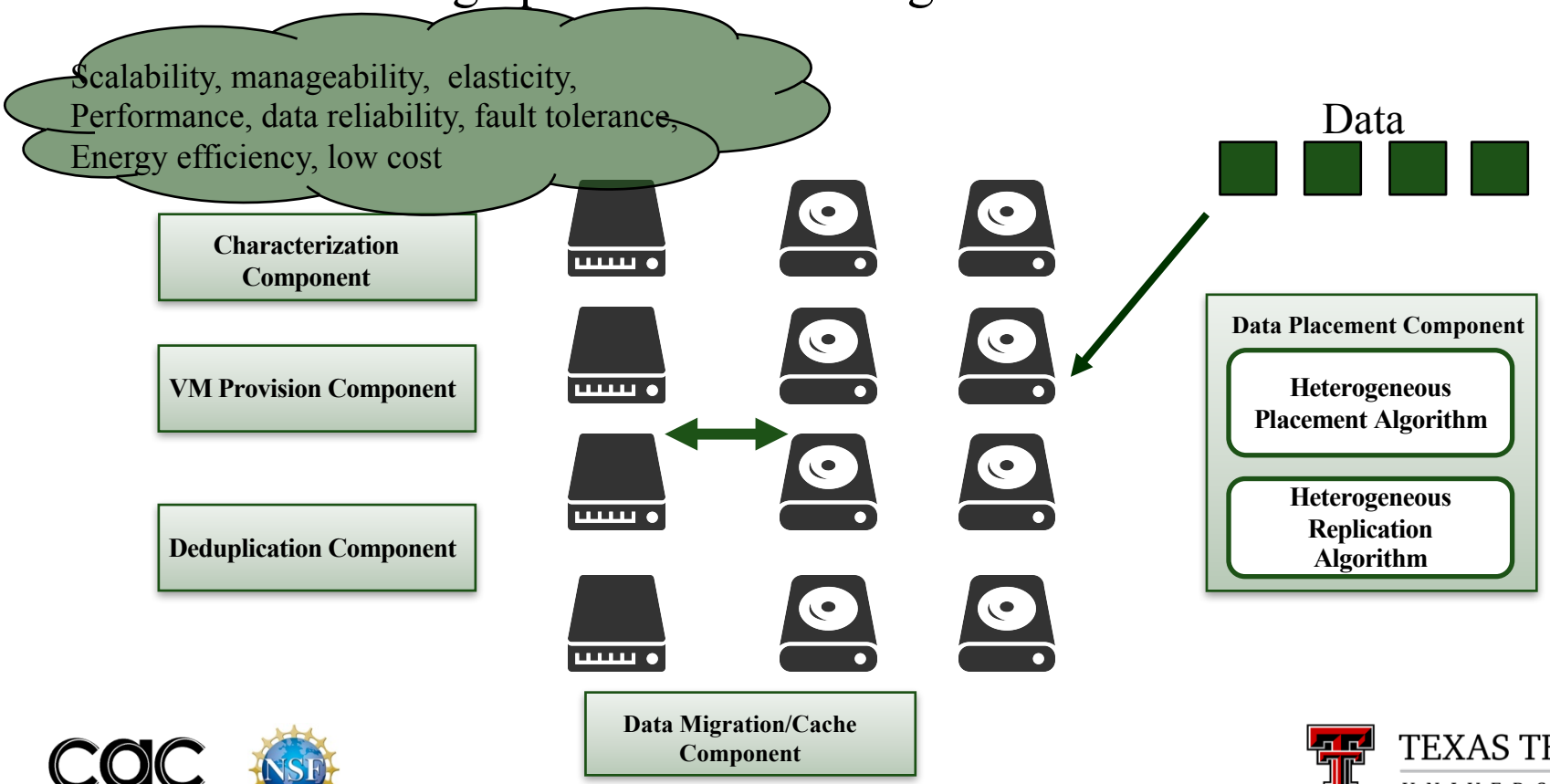
Computer Science Department

Texas Tech University

We are grateful to the Nimboxx and the Cloud and Autonomic Computing site at Texas Tech University for the valuable support for this project.

Unistore Overview

- To build a unified storage architecture (Unistore) for Cloud storage systems with the co-existence and efficient integration of heterogeneous HDDs and SCM (Storage Class Memory) devices
- Scalable and high performance storage for virtual machine

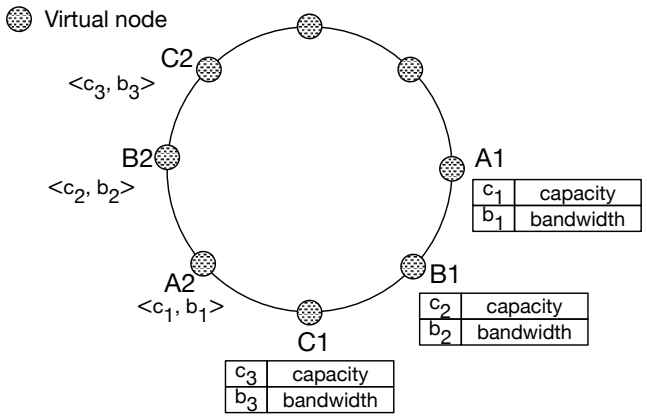


Project Highlights

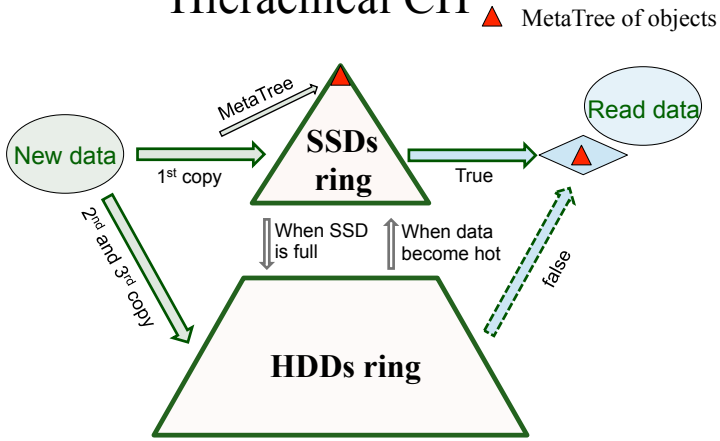
- Two papers published:
  - SUORA paper published at IEEE NAS'16
  - Hierarchical Consistent Hashing at IEEE ISPA'16
- One paper submitted:
  - Strategy Consistent Hashing submitted to PPoPP'17
- Ongoing work and New proposal
  - Elasticity in decentralized storage
  - Deduplication of SSD cache in cloud storage

Consistent Hashing for Heterogeneous Storage

Strategy CH



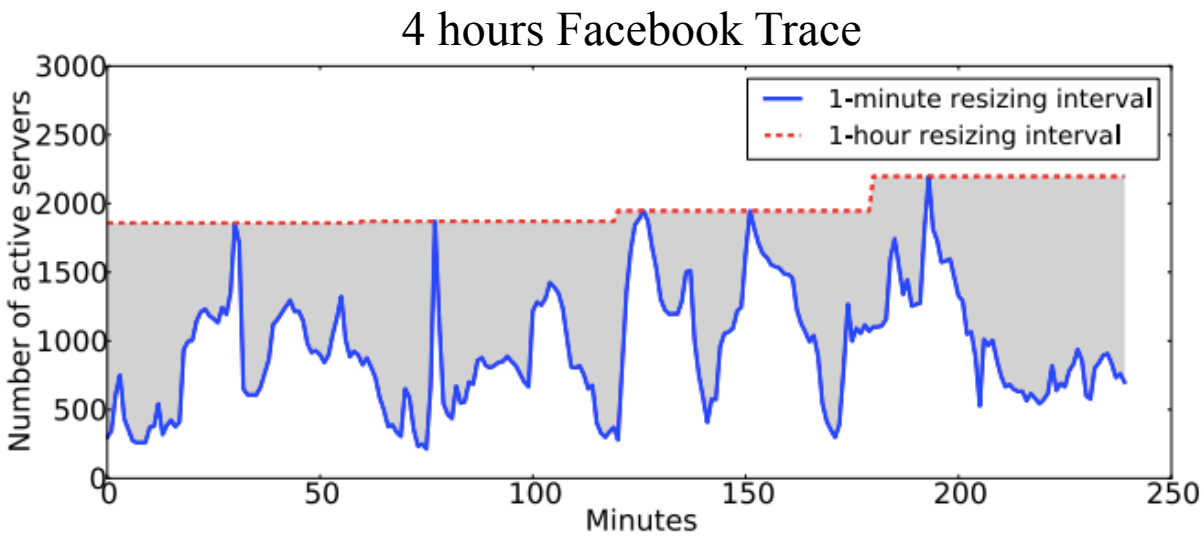
Hierachical CH



- Adopt a unified hashing ring to manage heterogeneous nodes
- Maintain attributes of each node
- Use a selection strategy for mapping nodes
  - Location strategy
  - Uniform strategy
  - Performance strategy
- Data placement strategy
  - When new data objects come, the first choice is to place the data on SSD ring
  - Data movement between SSD and HDD rings

Elasticity in Scale-out Storage

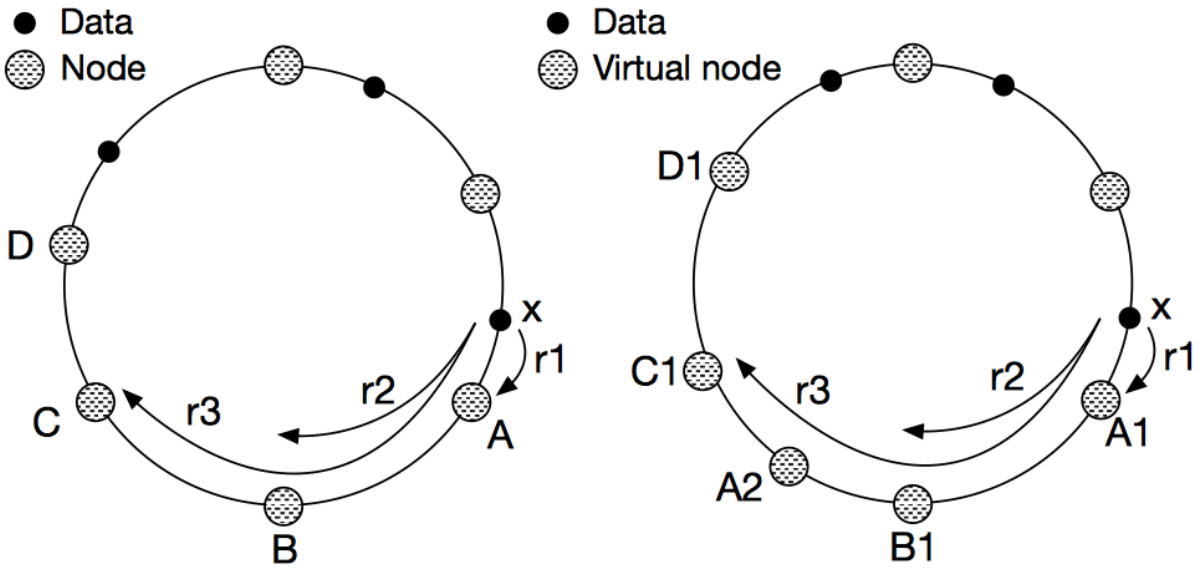
- Size-up and size-down in elastic storage
  - Size-up to meet I/O demand
  - Size-down to save energy or free up resource for other use
- Agility to scale is critical to better satisfy the goals



[1]L. Xu et al, "SpringFS: Bridging Agility and Performance in Elastic Distributed Storage.", FAST'14

Consistent Hashing based Storage

- Consistent hashing
  - For decentralized distributed system
  - Sizing up and down without completely changing data layout
  - Sizing down multiple nodes will need to migrate data first



(a) Without virtual nodes

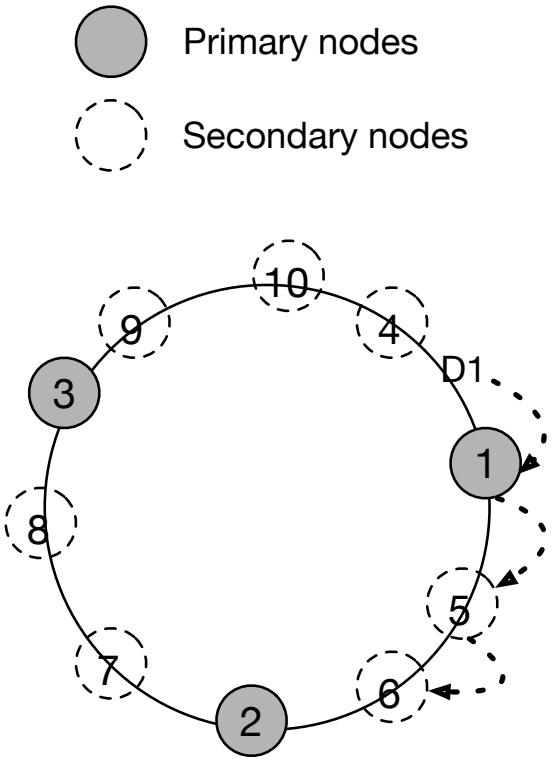
(b) Using virtual nodes

Elasticity in CH based Storage

- Challenges to improve agility
  - Scaling-up:
    - data movement to redistribute data
    - If node is turned down but not failed, there is no need to move data when it is turned on again
  - Scaling-down:
    - keep data available and reduce data movement
  - Data consistency and failure handling
    - The redundancy level may decrease when scaling down
    - Keep data replicas consistent across multiple nodes

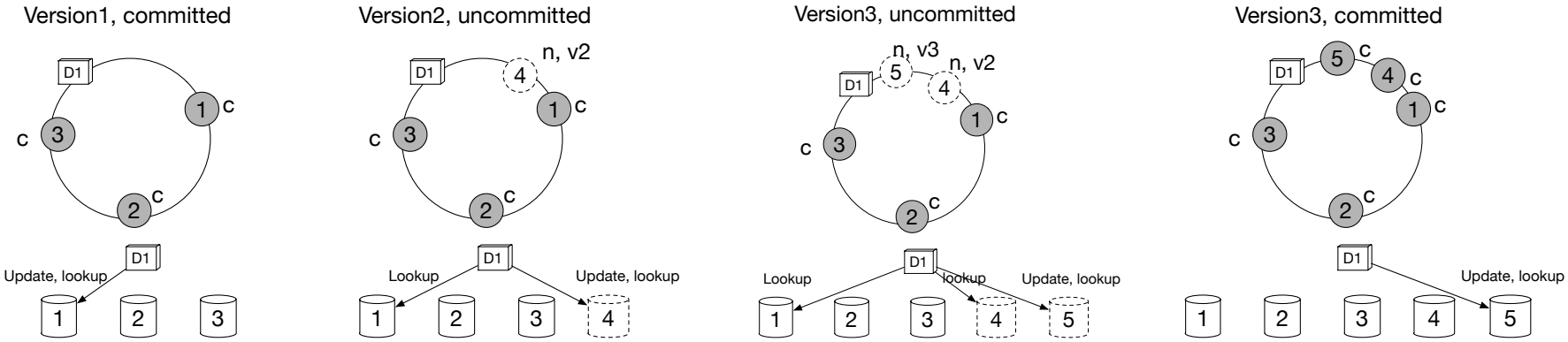
Elasticity in CH based Storage

- Our solution
  - Primary and secondary nodes: primary keeps one copy of whole data set, while secondary keep replicas
  - Size-down
    - Re-replicate to ensure data availability
    - Track modified data so that the data in the shutdown nodes can be updated when turned on
  - Size-up
    - CH automatically size-up
    - Migrate updated data to new nodes but no need to transfer existing data
  - Data consistency: need to keep data consistent when migrating modified data
  - Version consistent hashing

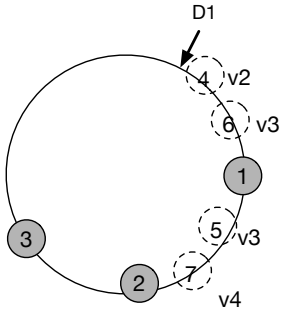


Version Consistent Hashing Scheme

- Build versions into the virtual nodes
- Avoid data migration when adding nodes or node fails
- Maintain efficient data lookup

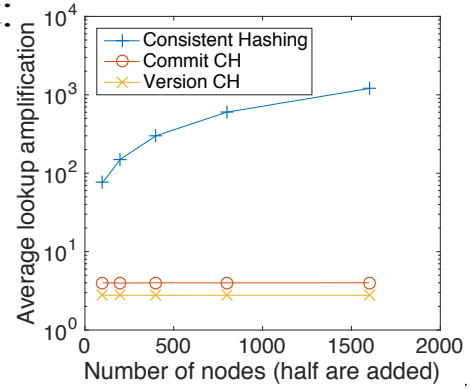


Data lookup algorithm



- v1: 1,2
- v2: 4,1
- v3: 4,6
- v4: 4,6
- Lookup locations: {4, 6, 1, 2}

Performance improvement:



Cloud and Autonomic Computing Center

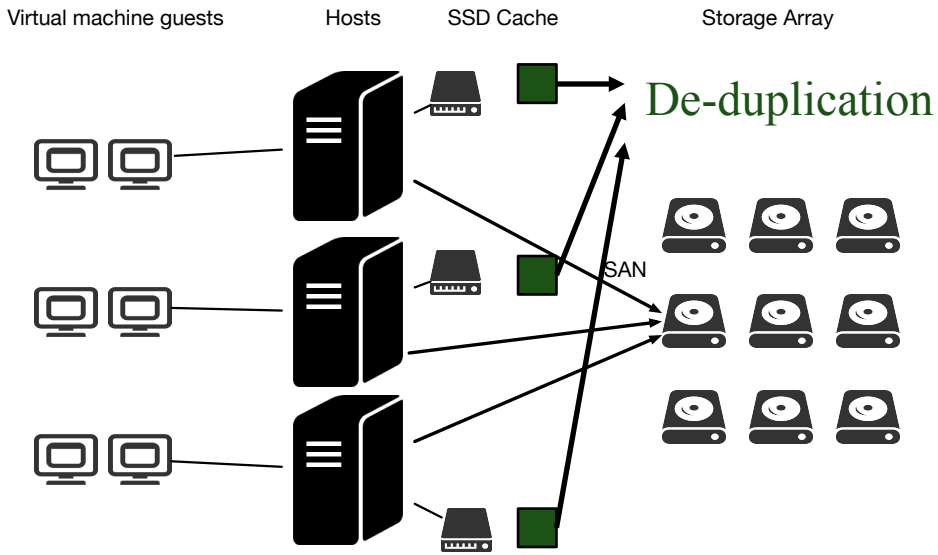
TEXAS TECH UNIVERSITY

On-going/Future Work

- Implement/evaluate the elastic consistent hashing in Sheepdog that is deployed on the DISCI cluster at TTU
- Use Microsoft Enterprise I/O trace and compare the agility of resizing with existing techniques
- Prepare a submission for IPDPS17 conference

Proposal of New Project

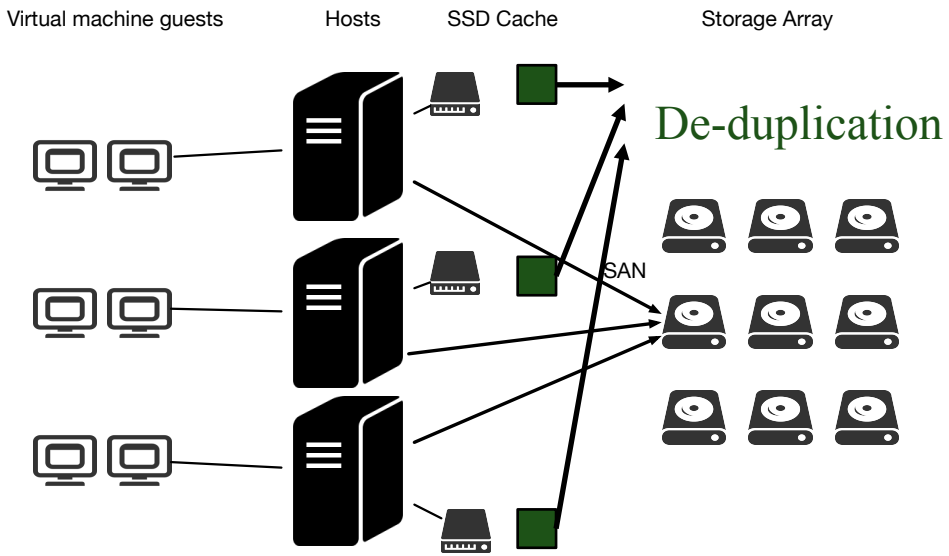
- Deduplication is critical especially in storage systems with SSD cache due to SSDs' limited endurance and capacity
- Current study focuses on single machine deduplication, but redundancy could occur across machines
- We propose distributed deduplication in SSD caches in the cloud storage environment



A typical scenario in cloud-scale data centers

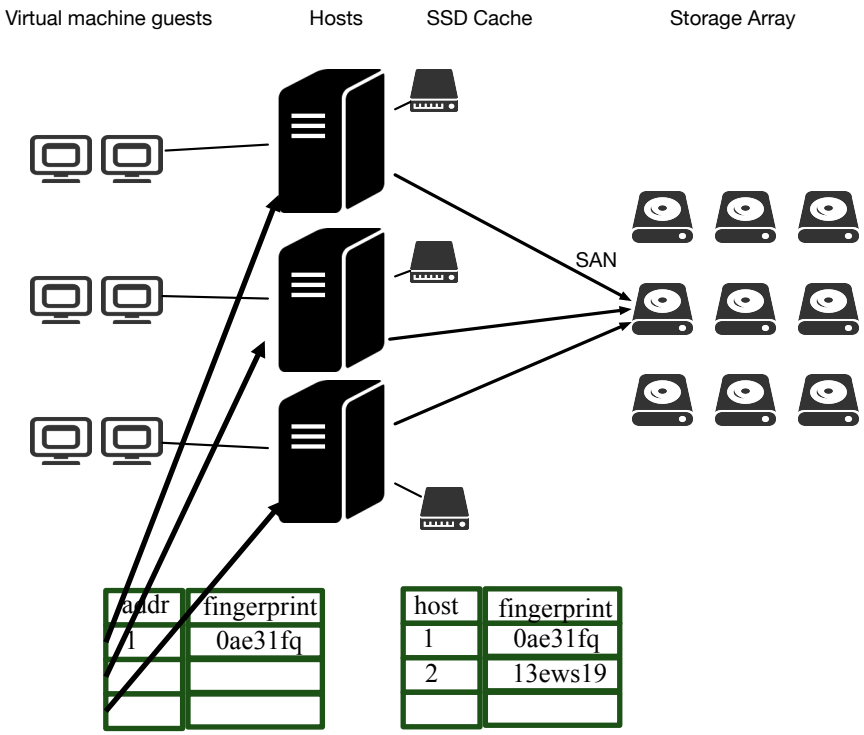
Motivation of the Study

- Motivation
  - VMs across hosts may share the same operating system and dataset
  - SSD cache is usually host-local, it is challenging to de-dup across multiple node local storage (how to reference data on remote host)



A typical scenario in cloud-scale data centers

Proposed Solution



Fingerprint table LRU table

- Metadata (fingerprints) is distributed using memcached
- Data not de-duplicated across hosts
- Consider a global LRU instead of local LRU
  - Recently used data on one host may be reused on another host
  - If duplicate data is written back to primary storage, only create a reference on primary storage
  - If duplicate data is inserted, a reference to the remote host is created

Summary

- Elasticity is an important feature in cloud computing and considered in scale-out storage systems
- We propose techniques to allow agile resizing and minimal performance degradation
- It is going to benefit data centers to optimize resource utilization and/or save power consumption
- We will continue investigating workload characterization based on statistical techniques

- We have proposed a de-duplication system for cloud storage as a new project and seeking sponsorship for either continuing the Unistore project or the new project

Thank You

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

Acknowledgement: 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.

Please take a moment to fill out your L.I.F.E. forms.

http://www.iucrc.com

Select “Cloud and Autonomic Computing Center”
then select “IAB” role.

What do you like about this project?
What would you change?
(Please include all relevant feedback.)