S4: A Simple Storage Service for Sciences

Matei Ripeanu
University of British Columbia

Adriana Iamnitchi
University of South Florida

The Situation

- Services as utilities have gained traction
  - Economy of scale → lower costs
  - One of the present drivers for Grid computing
- Success story: Amazon Simple Storage Service (S3)
  - S3 growth is capacity constrained
  - Direct access to storage: open protocols, APIs
  - Performance claims:
    - Infinite data durability, 99.99% availability, fast access
  - Billing: pay-as-you go
    - $0.15/month/GB stored; $0.13-0.18/GB transferred
- Science communities are huge storage users

The Motivating Questions

The immediate question: Is offloading data storage to a storage utility feasible and cost-effective for science Grids?

The long-term question: How should a storage utility that targets scientific applications look like?

The Approach

- Characterize S3
  - Does it live up to its own objectives?
- Toy scenario: consider a representative scientific application (DZero)
  - Is the functionality provided adequate?
  - Estimate performance and costs

Q: Is offloading data storage from an in-house storage system to S3 feasible and cost-effective for science Grids?

The Answer: Risky.

- New risk: direct monetary loss
  - Magnified as there is no built-in solution to limit loss
  - In addition to well-known risk in distributed systems
- Security mechanisms -- too simple to be useful for large collaborations
  - Access control using ACLs,
    - hard to use in large systems, needs at least groups
  - No support for delegation
  - Implicit trust between users and the S3 service
    - No transaction 'receipts', no support for un-repudiabiliy
- But … standard techniques to deal with these problems

The Answer: Costly.

- Scenario: S3 used by a high-energy physics collaboration
  The DØ Experiment
  - Traces from January '03 to March '05 (27 months)
  - 375TB stored, 5.2 PB processed, 561 users, 13 countries

Data Processing	S3 DØ	S3 EC2	S3 DØ
Storage	$675K	$675K	$200K…$400K
Access	$462K	$44K	$66K
	(per year)	$66K	

Add caching:
4TB/site cooperative cache

Move processing to EC2

Realize that data gets 'cold':
- archive cold raw-data
- throw away cold derived data (keep definitions)

Guidelines for a Simple Storage Service for Sciences (S4)

- Unbundle performance characteristics
  - S3: high-availability, high-durability, high-access performance, bundled at a single pricing point
    - Applications often do not need all three
    - Each characteristic requires different resources and generates different costs
  - Solution: classes of service that allow applications to specify their requirements and chose pricing point
- Exploit usage patterns
  - e.g., data gets cold
- Facilitate the use of application-level information to reduce costs
  - E.g., raw vs. derived data

Questions?

To access the S3 evaluation technical report: http://www.ece.ubc.ca/~matei

–

Simple Storage Service (S3) Architecture

- Two level namespace
  - Buckets (think directories)
    - Unique names
    - Two goals: data organization and charging
  - Data objects
    - Opaque object (max 5GB)
    - Metadata (attribute-value, up to 4K)

- Functionality
  - Simple put/get functionality
  - Limited search functionality
  - Objects are immutable, cannot be renamed

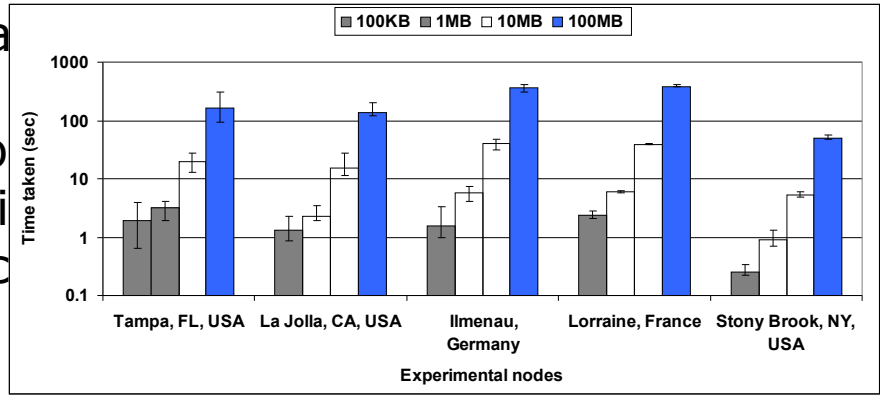
- Data access protocols
  - SOAP
  - REST
  - BitTorrent

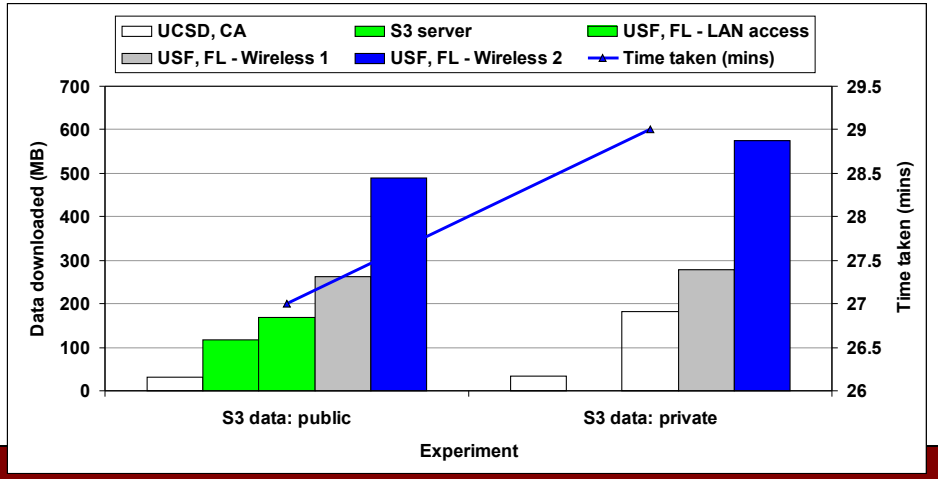
S3 Architecture (…cont)

- Security
  - Identities
    - Assigned by S3 when initial contract is ‘signed’
  - Authentication
    - Public/private key scheme
    - But private key is generated by Amazon!
  - Access control
    - Access control lists (limited to 100 principals)
    - ACL attributes
      - FullControl,
      - Read & Write (for buckets only for writes)
      - ReadACL & WriteACL (for buckets or objects)
  - Auditing (pseudo)
    - S3 can provide a log record

S3 Evaluation

- Durability
  - Perfect (but based on limited scale experiment)
- Availability
  - Four weeks of traces, about 3000 access requests from 5 PlanetLab nodes
  - Retry protocol, exponential ba
  - 'Cleaned' data
    - 99.03% availability after o
    - 99.55% availability after fi
    - 100% availability after sec
- Access performance





Characteristics	***Resources and techniques to provide them***
High-performance data access	Geographical replication to improve access locality, high-speed storage, fat networks.
Durability	Replication at various scales: RAID, erasure codes, multiple locations, multiple media;.
Availability	service replication, hot-swap technologies, multi-hosting, increase availability for auxiliary services (e.g., authentication, access control)

Application class	***Durability***	***Availability***	***High access speed***
Cache	No	Depends	Yes
Long-term archival	Yes	No	No
Online production	No	Yes	Yes
Batch production	No	No	Yes