CSAL: A CLOUD STORAGE ABSTRACTION LAYER TO ENABLE PORTABLE CLOUD APPLICATIONS

Zach Hill & Marty Humphrey
Dept. of Computer Science, University of Virginia

zjh5f@cs.virginia.edu
A Cloud Application

User Requests → Front-End → Worker

- Queue Service
- Table Service
- Object/Blob Service
Many clouds, many code versions

- Front-End
- Worker
- Tables
- Blobs
- Queues
- Storage Services

Z. Hill and M. Humphrey, Dept. of Computer Science, University of Virginia  Dec. 2, 2010
Single code version, many clouds
CSAL Overview

Application

CSAL

BlobStore
- createContainer
- listContainers
- deleteContainer
- getBlob
- putBlob
- deleteBlob

TableStore
- createTable
- deleteTable
- Insert
- Update
- Delete
- Query
- getItem

QueueStore
- createQueue
- deleteQueue
- getMessage
- putMessage
- peekMessage
- deleteMessage

Metadata Store
- table service(s)

Blob Namespace
Table Namespace
Q Namespace

S3 Plugin
SimpleDB Plugin
Azure Queue Plugin
...
CSAL Namespaces

- One namespace for each abstraction type
- Metadata only for containers
  - Service endpoint, identifier, user credentials
- Each abstraction has an independent metadata store
- Metadata caching
  - Container ops are not very common
  - If data is stale, simply re-fetch and retry
CSAL Namespaces

1. Call getBlob("X", "foo");

2. Lookup "X" metadata – cache first

3. Retrieve from table if not in cache

4. Use metadata to determine plugin to use

5. Access "foo" in "X" in Service1

("X", http://service1.com, testusr, testKey,...)
CSAL Implementation

- Client-side java library
  - i.e. `BlobStore.getBlob("Container", "foo");`
- Metadata backing store in the cloud
- Currently supports Azure & AWS storage
  - Both SOAP and REST
Performance of CSAL

- Adding software layers isn’t free
- Compare CSAL to Azure’s and Amazon’s SDK APIs
- Set of micro-benchmarks to test operation latency
- Container Ops and Data Ops
  - Expect a slowdown for container ops due to metadata
Performance - Container Ops

CSAL in AWS

Note: Error Bars indicate 1 Standard Deviation

CSAL in Windows Azure

Z. Hill and M. Humphrey, Dept. of Computer Science, University of Virginia  Dec. 2, 2010
Performance – Data Ops

CSAL in AWS

CSAL in Windows Azure

Note: Error Bars indicate 1 Standard Deviation

Z. Hill and M. Humphrey, Dept. of Computer Science, University of Virginia Dec. 2, 2010
What about Standards?

- Standards Efforts
  - OCCI
  - OVF
- Standards take time to develop and are resisted by vendors
- Multi-cloud APIs
  - SimpleCloud, jClouds, DeltaCloud, LibCloud
  - SAGA
Future work

- What if Cloud X doesn’t have tables/blobs/queues?
  - Map one abstraction to other (i.e. filesystem)
  - 3rd party services: Hbase, HyperTable, Cassandra...

- Placing, replicating, and migrating data in real-time for performance and/or cost

- Real-world applications such as multi-cloud MR
Summary

- Application lock-in and portability are problems in clouds
- Standards are great, but don’t hold your breath just yet
- CSAL provides storage abstractions to make the application code itself portable with little performance impact for common data operations
Questions?

zjh5f@cs.virginia.edu