Karlsruhe Institute of Technology

Research Group Cloud Computing
http://www.scc.kit.edu/personen/rg-cc.php

Steinbuch Centre for Computing (SCC)
http://www.scc.kit.edu

A Redundant Array of Independent Services (RAIS)

Christian Baun, Marcel Kunze, Denis Schwab, Tobias Kurze

When using Cloud Storage Services the Risk for Data loss always exists

- Risk in case of an unplanned outage
  - Independent from guarantees stated in SLAs of public clouds
- Examples for unplanned outages:
  - provider goes out of business
  - storage service is not accessible due to network problems
- Private cloud deployments often do not guarantee SLAs the same way public cloud offerings do

State of the Art

- Several storage services implement the interface of the Amazon Simple Storage Service (S3)

Public Cloud Services	Private Cloud Services
Amazon S3	Eucalyptus Walrus
Google CS	OpenStack Swift
Connectria CS	Nimbus Cumulus
Host Europe CS	Riaks CS

Challenges

- Integrate different cloud storage services in a single service
- Check whether objects in storage services are in sync or not

Challenge: Integrate different cloud storage services to improve availability of users' data

Approach

Concept of a Redundant Array of Independent Services

- Stores users' data at federated object-based storage services
- Ensures data access
- Based on the simultaneous usage of multiple storage services in a RAID (Redundant Array of Independent Disks)-like manner to increase availability and to decrease provider dependency

Prototype called Octopus exists

- Implements two modes of operation
- RAID-1: mirroring without parity or striping
- RAID-5: block-level striping with distributed parity data

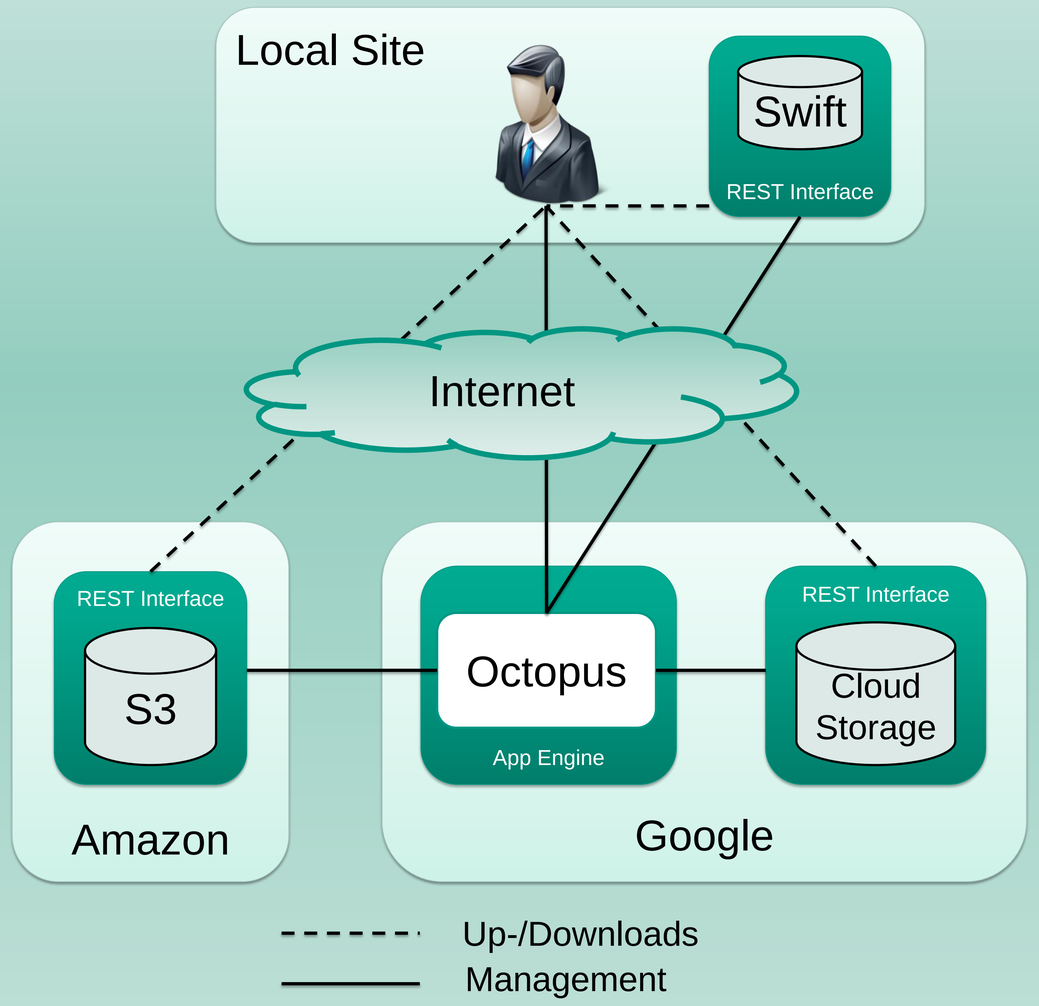
Check if Objects are in sync

- RAIS strongly relies on MD5 checksums
- Checksums and other meta-data are transferred automatically whenever an object or a list of objects is requested
- Important meta-data is `<Key>...</Key>`, holding the display name of the object, and `<ETag>...</ETag>`, holding the MD5 checksum
- Through object lists from registered storage services, RAIS can check if the data is still synchronized across the registered storage services by comparing MD5 checksums

Szenario

Octopus running inside the Public Cloud PaaS

- Up-/Downloads go directly to the storage services
- Management of services and objects is done via Octopus



Object-related meta-data as returned by Amazon S3

```
<Key>testobject.txt</Key>
<LastModified >2012-01-
29T21:18:24.000Z</LastModified>
<ETag>
  &quot;71388ba9a76ddb7ecd43a14f2a9ae216&quot;
</ETag>
<Size>2163</Size>
<Owner>
  <ID>af0af9137ff6...97272818796</ID>
  <DisplayName>username</DisplayName>
  </Owner>
<StorageClass>STANDARD</StorageClass>
```

Details of the Prototype

- Runs either inside the Google App Engine Public Cloud PaaS or in a Private Cloud PaaS based on AppScale or typhoonAE
- No need of local installation (runs in the cloud)
- Interoperability with all browsers (pure HTML) and devices

Octopus is an Open Source product

- http://code.google.com/p/octopuscloud/
- Apache License 2.0