Thank you

for more information

DuraSpace organization
- http://duraspace.org

DuraCloud project
- http://duracloud.org

awoods@duraspace.org

Preservation Infrastructure in the Cloud

CNI, April 2010

Andrew Woods
awoods@duraspace.org

Outline

- what is DuraCloud?
- pilot partners
- architecture
- quick look
- roadmap

Organizational Mission

- promoting durable, persistent access to digital data
- creating practical solutions in collaboration with academic, scientific, cultural, and technological communities
- development/deployment of open technologies

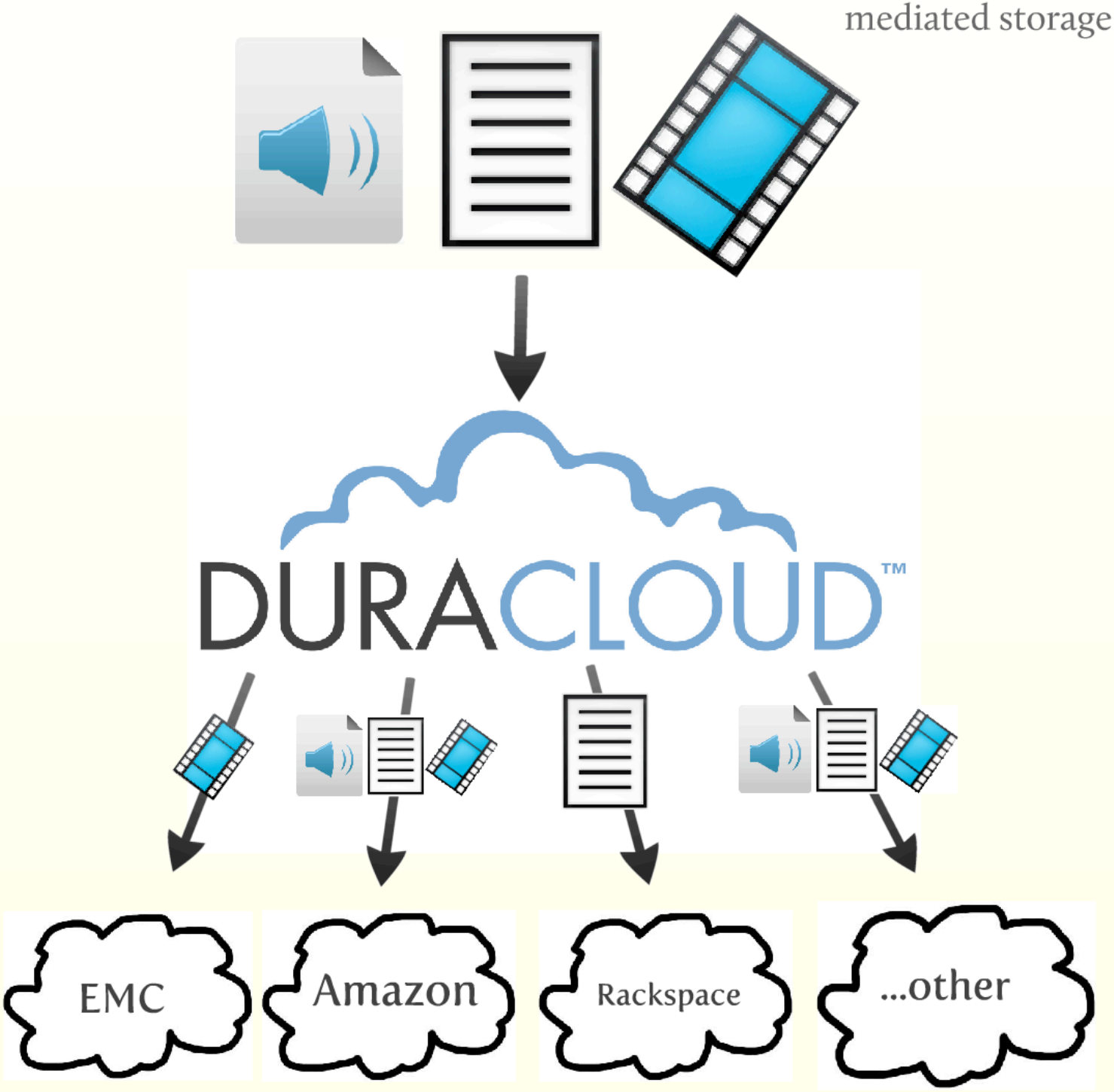
What is it?

- cloud-based service offered by the not for profit organization, DuraSpace
- cloud storage across multiple commercial & non-commercial providers
- an open source, cloud storage/compute application
  ...focused on preservation support and
  ...data access for reuse and sharing
- an open canvas for cloud-based services

Key notions

- mediated cloud storage
- ready to use, value-added services
- dead simple

mediated storage



DURACLOUD™

Initial Pilot Partners

Use Cases

- geographically distributed online backup
- bulk image conversion
- jpeg2000 serving/viewing
- international replication
- streaming video
- access to corpus for processing
- ingest staging area

Lessons learned

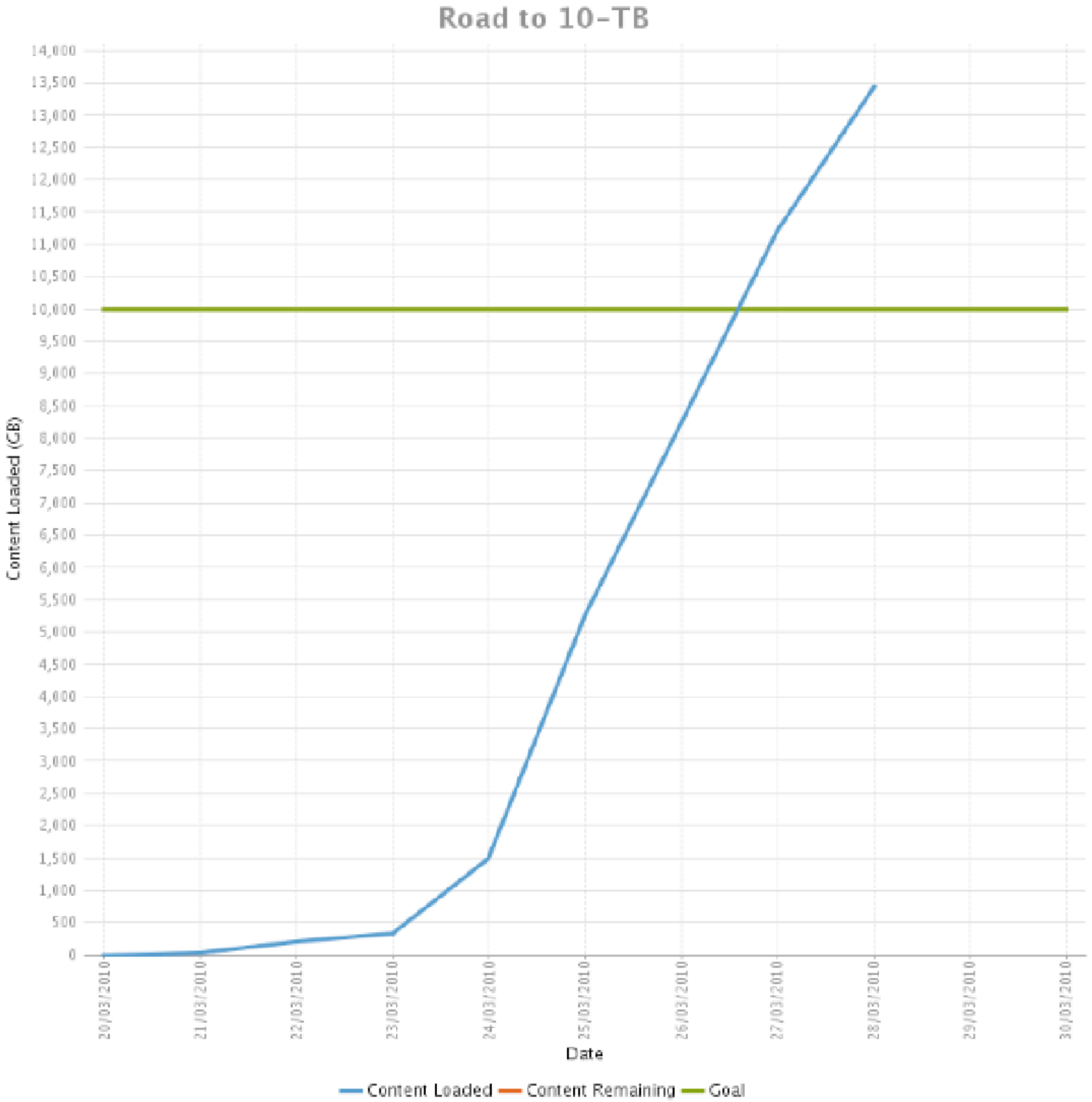
- bulk ingest: chunking / stitching
- bulk ingest: parallel upload
- bulk ingest: hard disk process

- bulk image conversion

- de-mystifying third party hosted content

- asynchronous nature of cloud storage

Initial 10-TB Progress Meter



Initial Pilot Partners

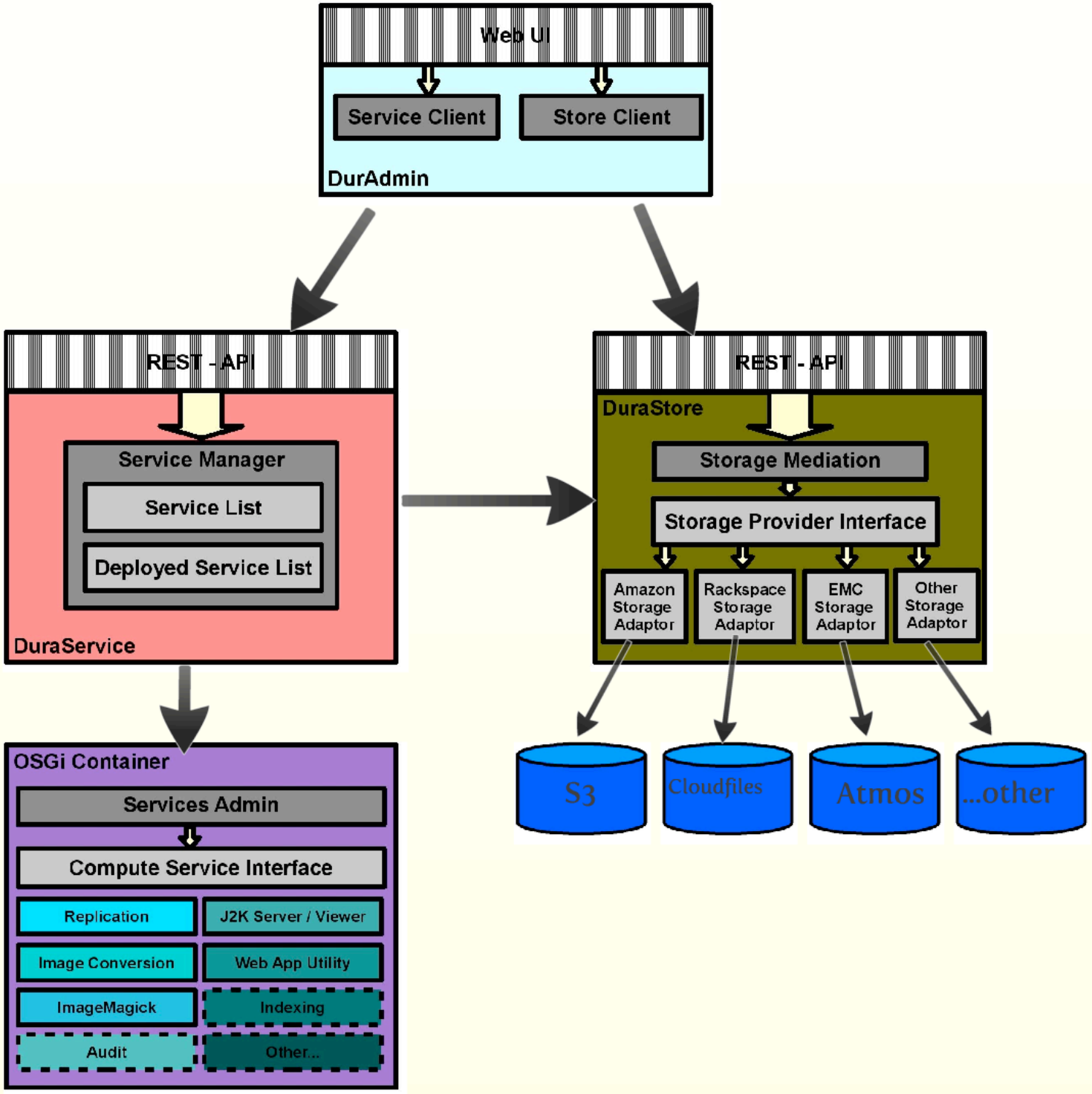
lau

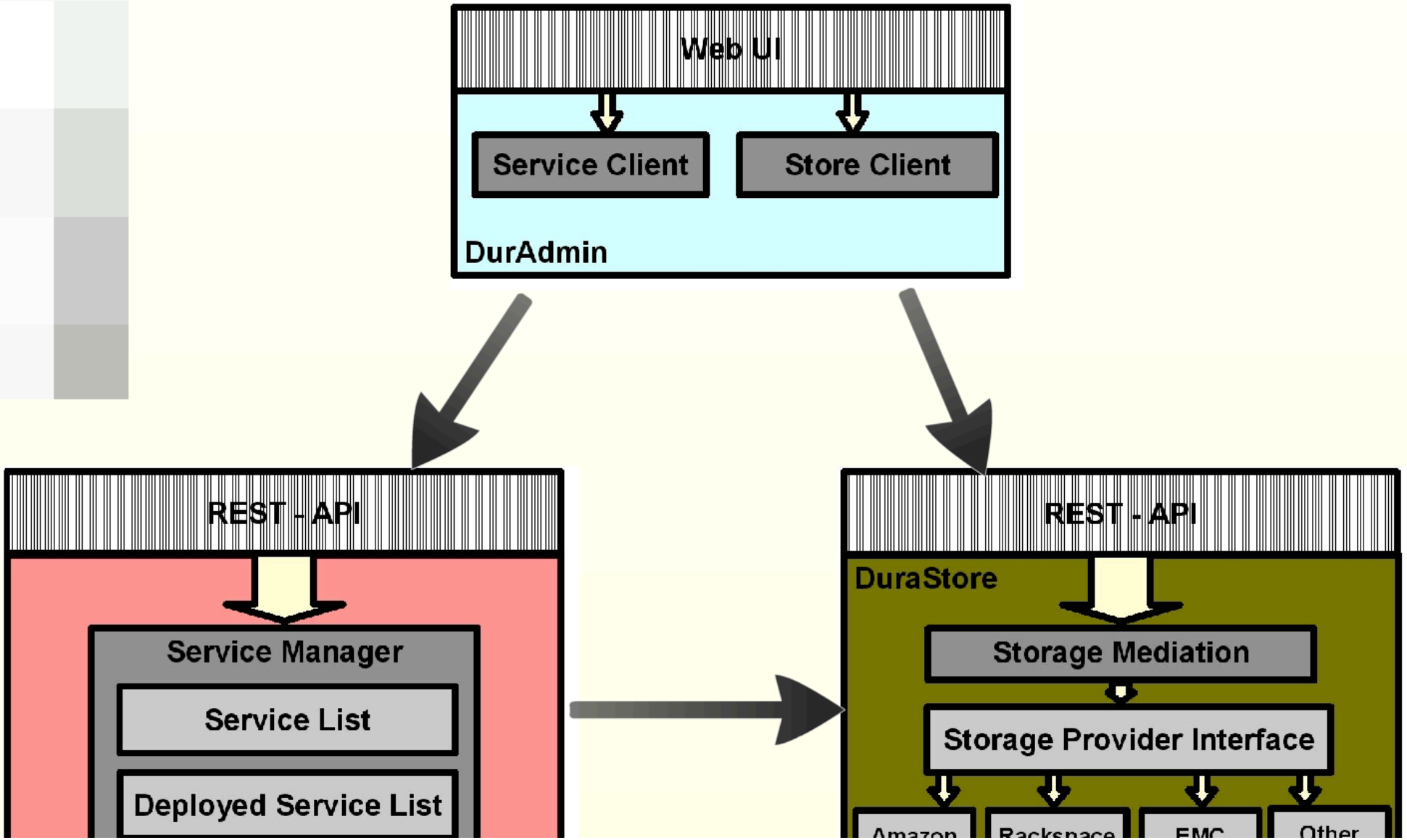
launch

primary
instance

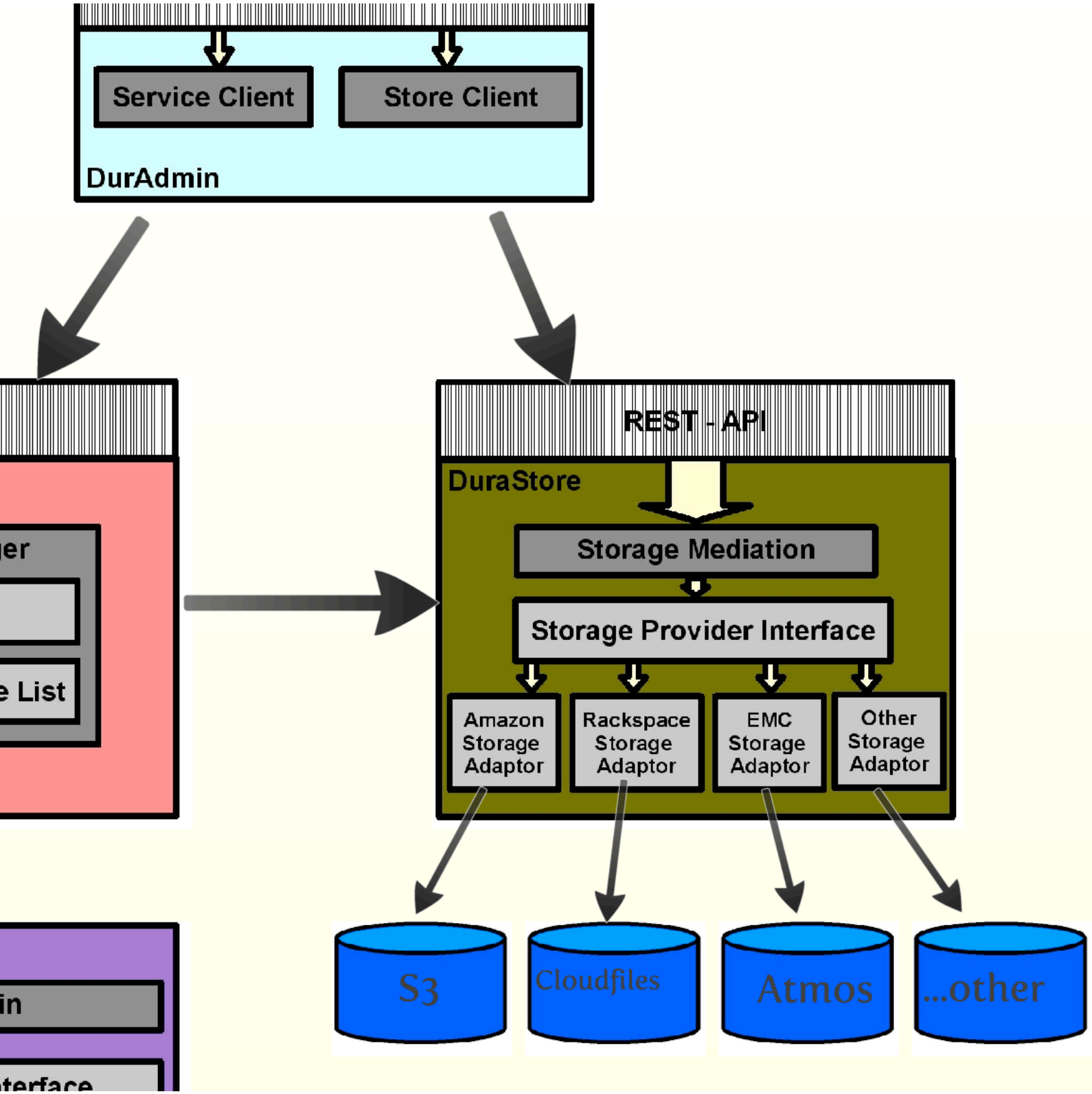
nch

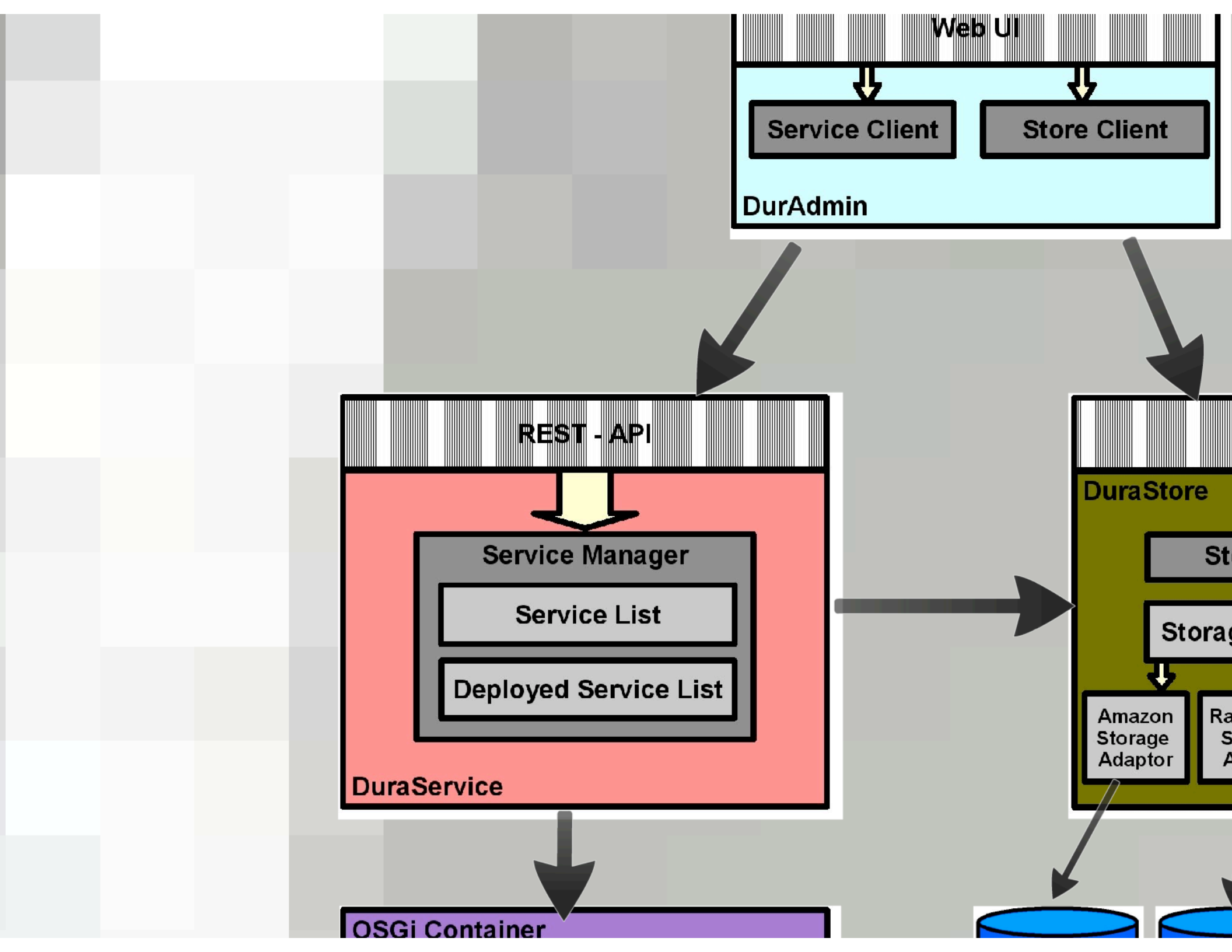
primary
instance

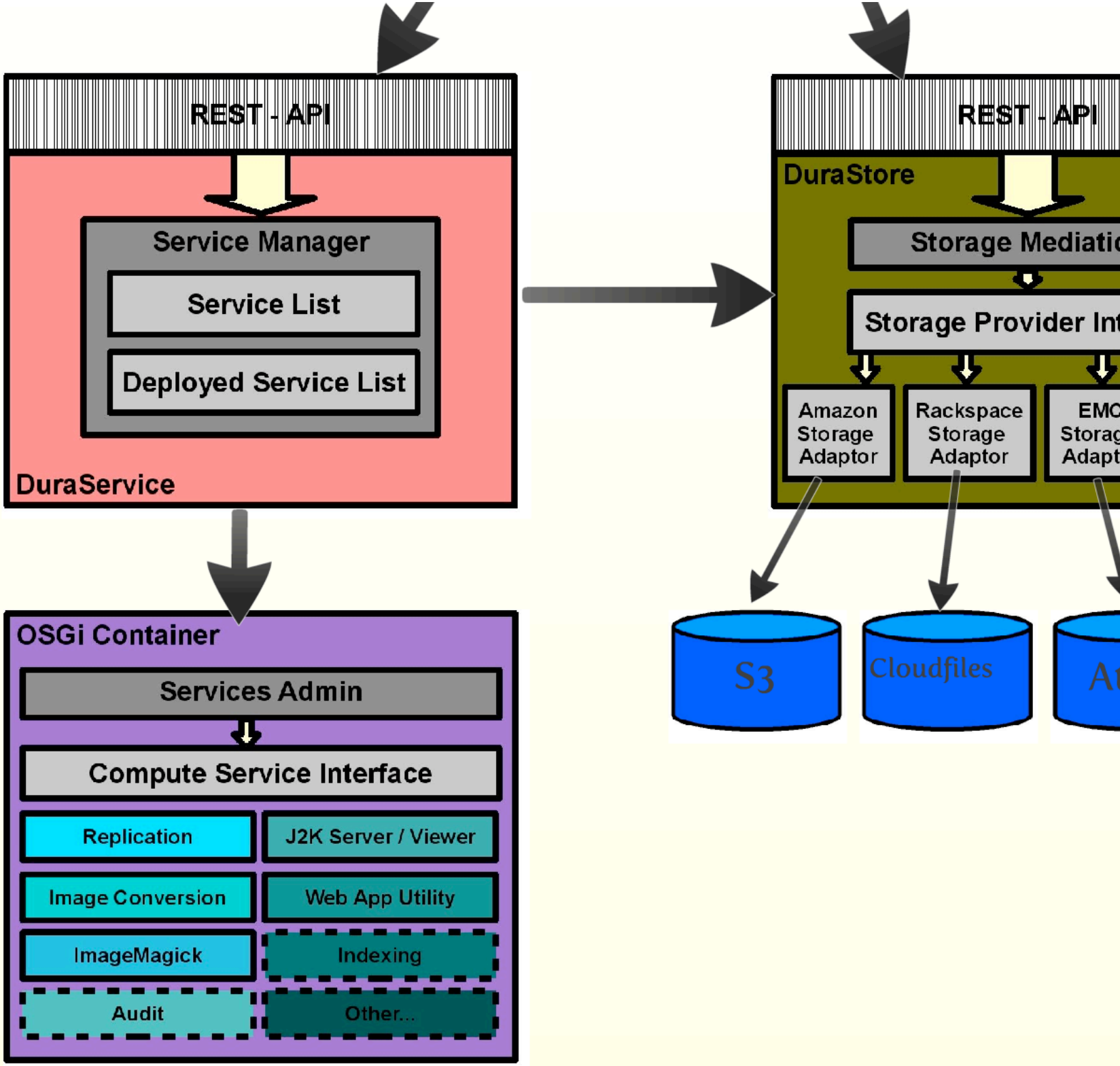

Web UI
Service Client
Store Client
DurAdmin
REST - API
Service Manager
Service List
Deployed Service List
DuraService
REST - API
DuraStore
Storage Mediation
Storage Provider Interface
Amazon Storage Adaptor
Rackspace Storage Adaptor
EMC Storage Adaptor
Other Storage Adaptor
S3
Cloudfiles
Atmos
...other
OSGi Container
Services Admin
Compute Service Interface
Replication
J2K Server / Viewer
Image Conversion
Web App Utility
ImageMagick
Indexing
Audit
Other...

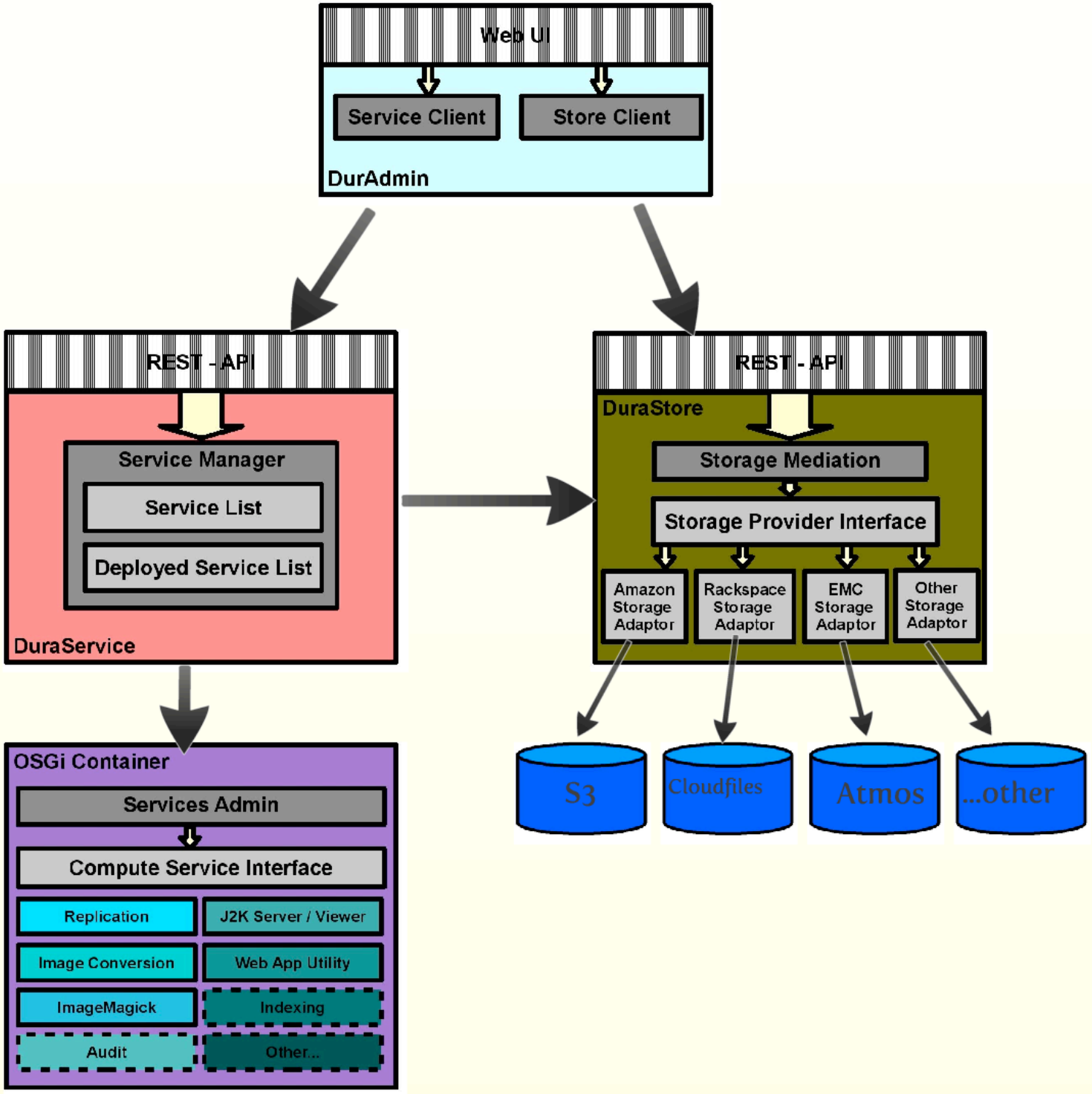

Web UI
Service Client
Store Client
DurAdmin
REST - API
Service Manager
Service List
Deployed Service List
REST - API
DuraStore
Storage Mediation
Storage Provider Interface
Amazon
Rackspace
EMC
Other

http://demo.duracloud.org:8080/duradmin


Service Client
Store Client
DurAdmin
REST - API
DuraStore
Storage Mediation
Storage Provider Interface
Amazon Storage Adaptor
Rackspace Storage Adaptor
EMC Storage Adaptor
Other Storage Adaptor
S3
Cloudfiles
Atmos
...other


Web UI
Service Client
Store Client
DurAdmin
REST - API
Service Manager
Service List
Deployed Service List
DuraService
DuraStore
Amazon Storage Adaptor
OSGi Container


REST - API
Service Manager
Service List
Deployed Service List
DuraService
OSGi Container
Services Admin
Compute Service Interface
Replication
J2K Server / Viewer
Image Conversion
Web App Utility
ImageMagick
Indexing
Audit
Other...
REST - API
DuraStore
Storage Mediatio
Storage Provider Int
Amazon Storage Adaptor
Rackspace Storage Adaptor
EMC Storag Adapt
S3
Cloudfiles


Web UI
Service Client
Store Client
DurAdmin
REST - API
Service Manager
Service List
Deployed Service List
DuraService
REST - API
DuraStore
Storage Mediation
Storage Provider Interface
Amazon Storage Adaptor
Rackspace Storage Adaptor
EMC Storage Adaptor
Other Storage Adaptor
S3
Cloudfiles
Atmos
...other
OSGi Container
Services Admin
Compute Service Interface
Replication
J2K Server / Viewer
Image Conversion
Web App Utility
ImageMagick
Indexing
Audit
Other...

REST-API examples

store context

space id

- get contents of space

  http://your-domain.edu/durastore/images

content id

- get content

  http://your-domain.edu/durastore/images/rome.jpg

- deploy service

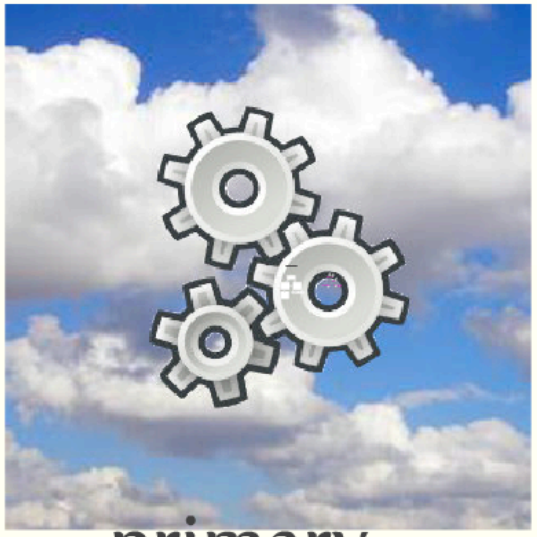
  http://your-domain.edu/duraservice/services/image-conversion-1.0.0

service context

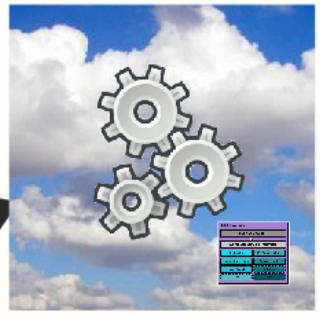
service id

nch

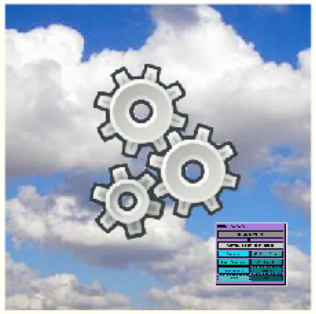
primary
instance



primary instance



managed service instance



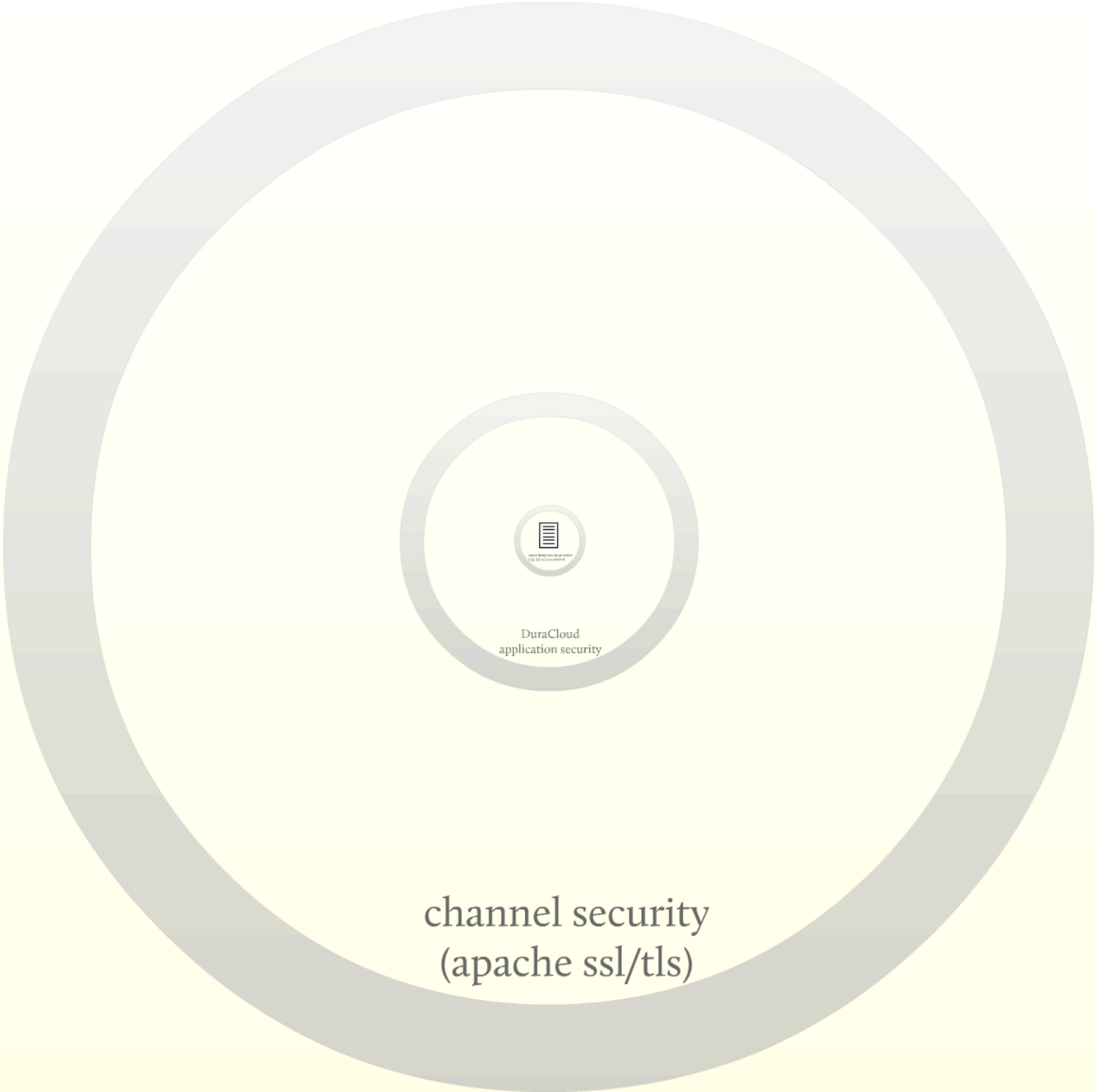
managed service instance

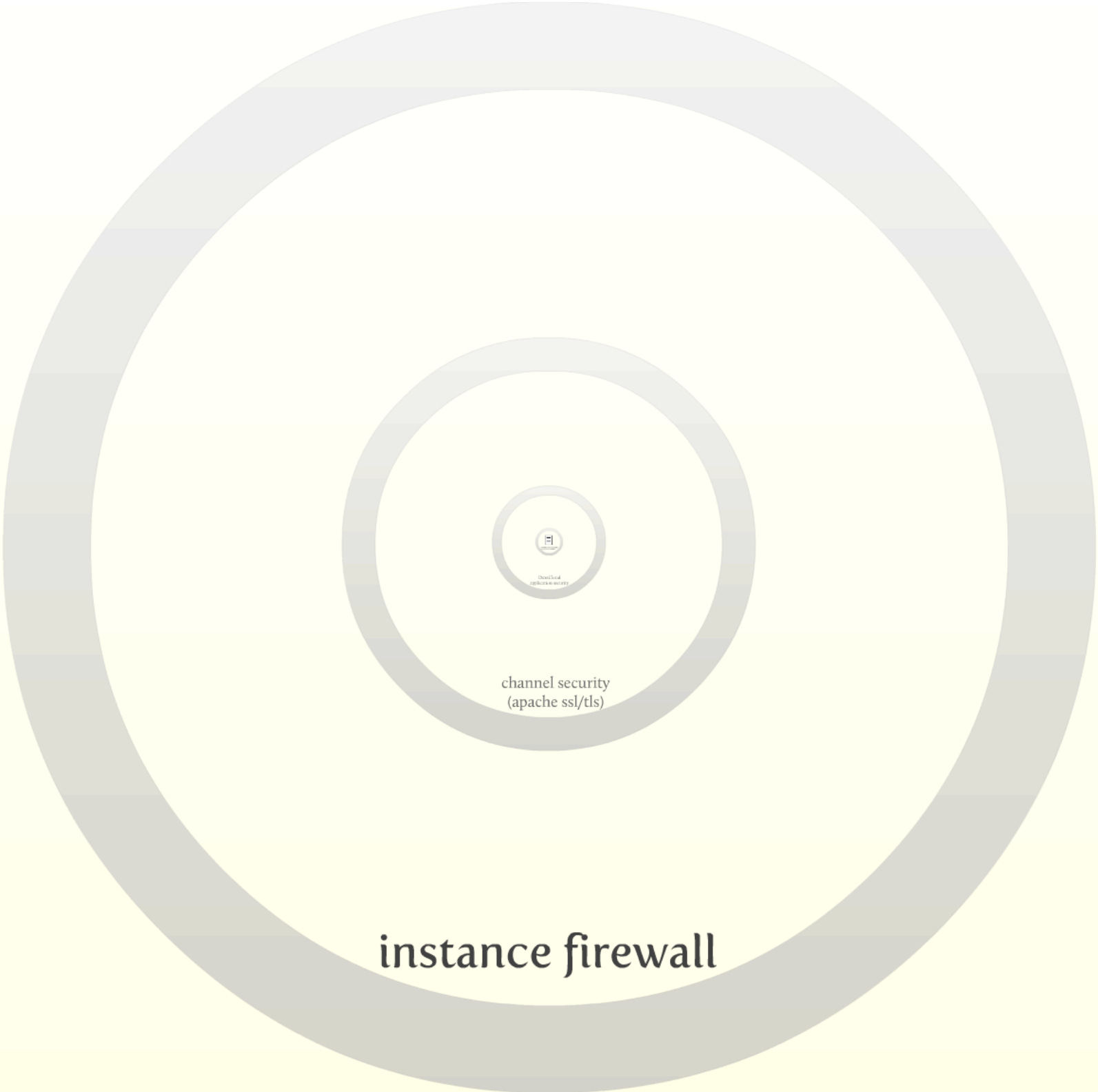
pre-configured service instance

Security

underlying storage provider
(e.g. S3) access control

DuraCloud
application security


DuraCloud
application security
channel security
(apache ssl/tls)


channel security
(apache ssl/tls)
instance firewall

Roadmap

- series of internal releases through Fall '10
  - repository integration (DSpace & Fedora)
  - expanded storage/compute provider support
  - expanded service selection
- open sourcing Summer '10
- second small set of pilot partners
- extended pilots Fall '10
- public beta release early '11

Roadmap

- series of internal releases through Fall '10
  - repository integration (DSpace & Fedora)
  - expanded storage/compute provider support
  - expanded service selection
- open sourcing Summer '10
- second small set of pilot partners
- extended pilots Fall '10
- public beta release early '11

Upcoming service developments focused on preservation support and access

- indexing / search
- bit integrity assurance
- expanded replication (on demand, filtered, etc)
- video streaming
- auditing

Roadmap

- series of internal releases through Fall '10
  - repository integration (DSpace & Fedora)
  - expanded storage/compute provider support
  - expanded service selection
- open sourcing Summer '10
- second small set of pilot partners
- extended pilots Fall '10
- public beta release early '11

Thank you

for more information

DuraSpace organization
- http://duraspace.org

DuraCloud project
- http://duracloud.org

awoods@duraspace.org

Roadmap

- series of internal releases through Fall '10
  - repository integration (DSpace & Fedora)
  - expanded storage/compute provider support
  - expanded service selection
- open sourcing Summer '10
- second small set of pilot partners
- extended pilots Fall '10
- public beta release early '11