

Applications

Files

Tables

CRDTs

Application Interface

Directories

Key/Value Pairs

Graphs

Wide-area SDS System

Manifest

Manifest

Block

Block

Block

Block

chunks

Service Interface

create(), read(), delete()

Object storage

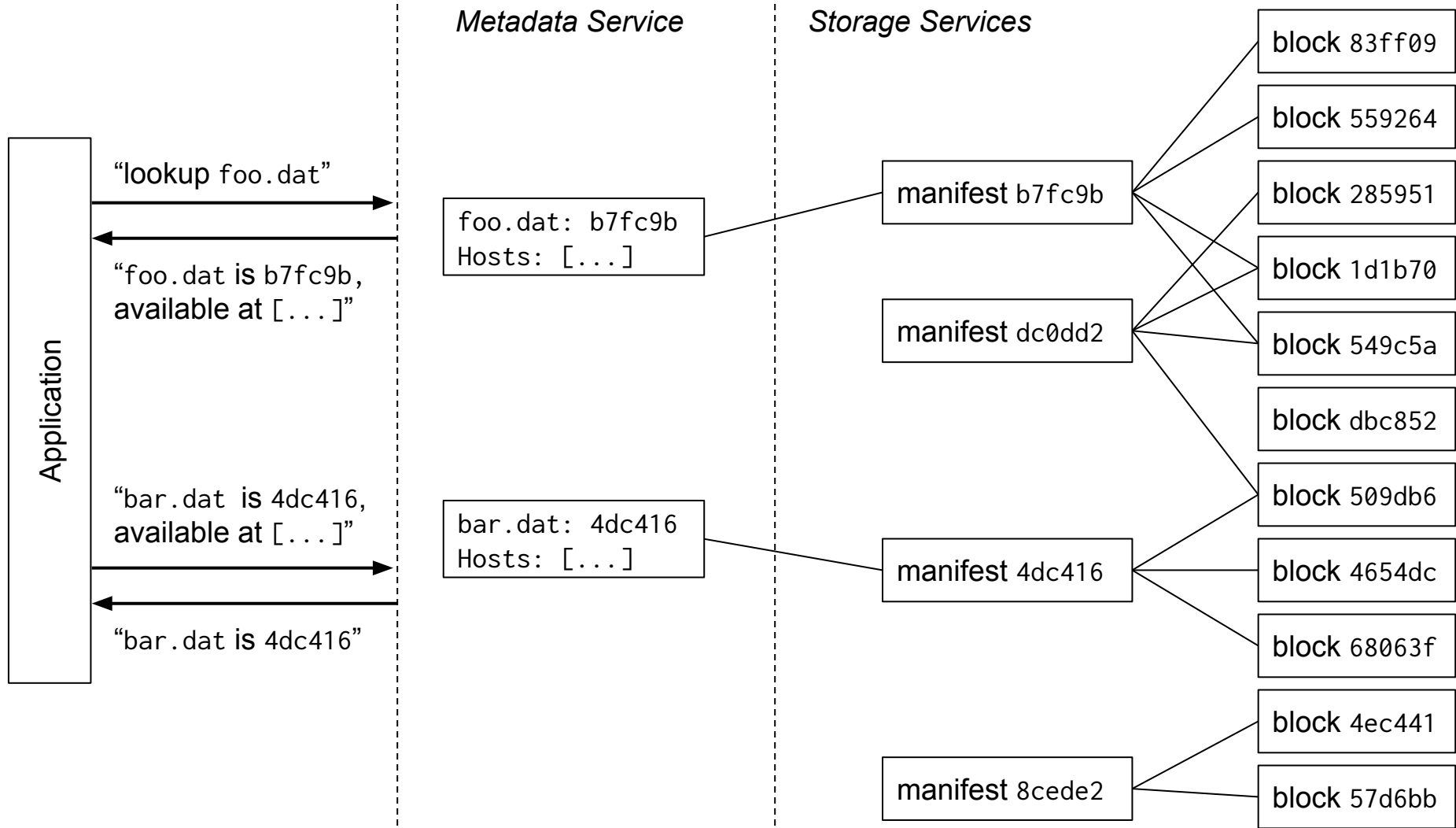
HTTP proxies

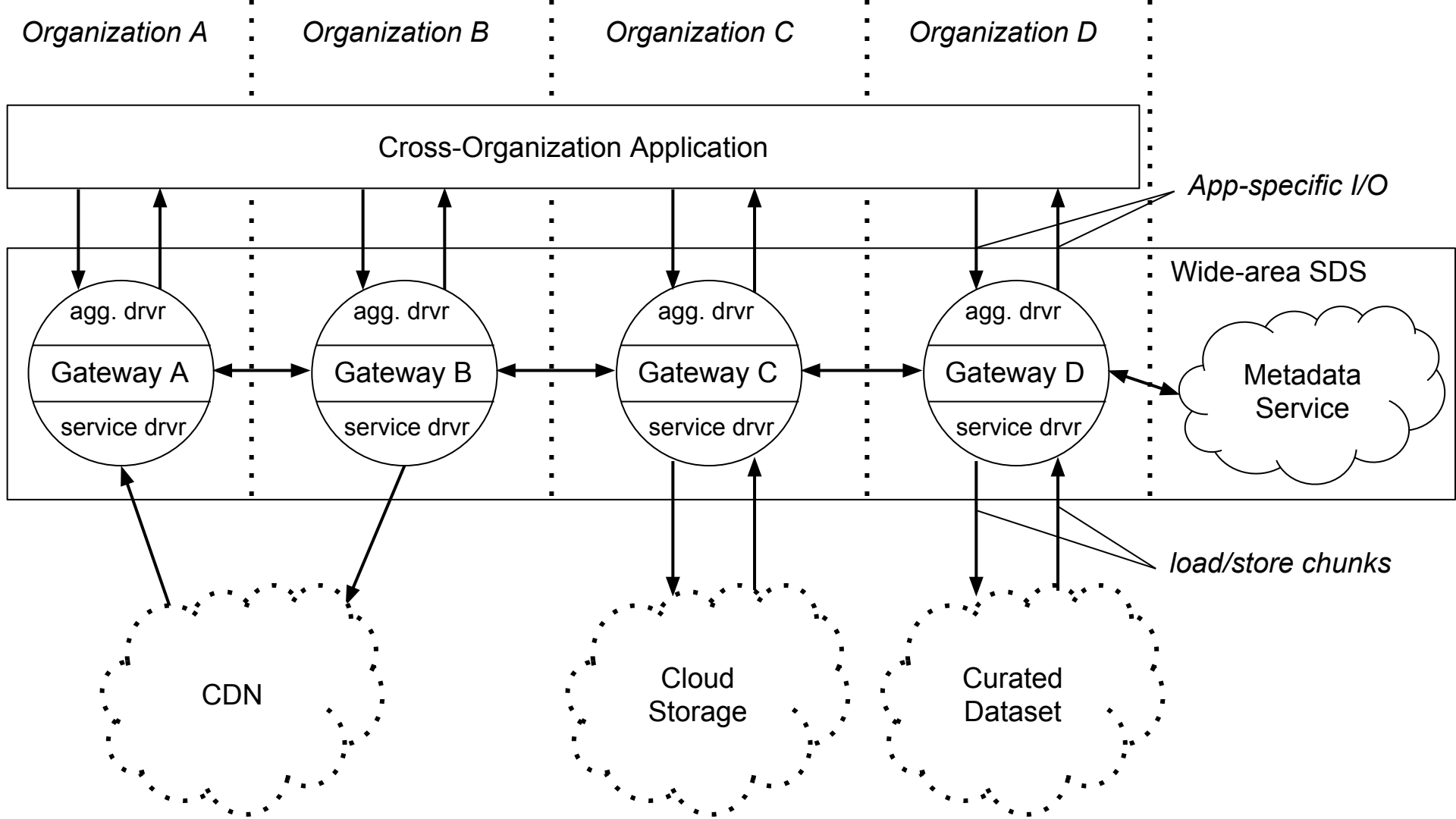
WebDAV

iSCSI

P2P networks

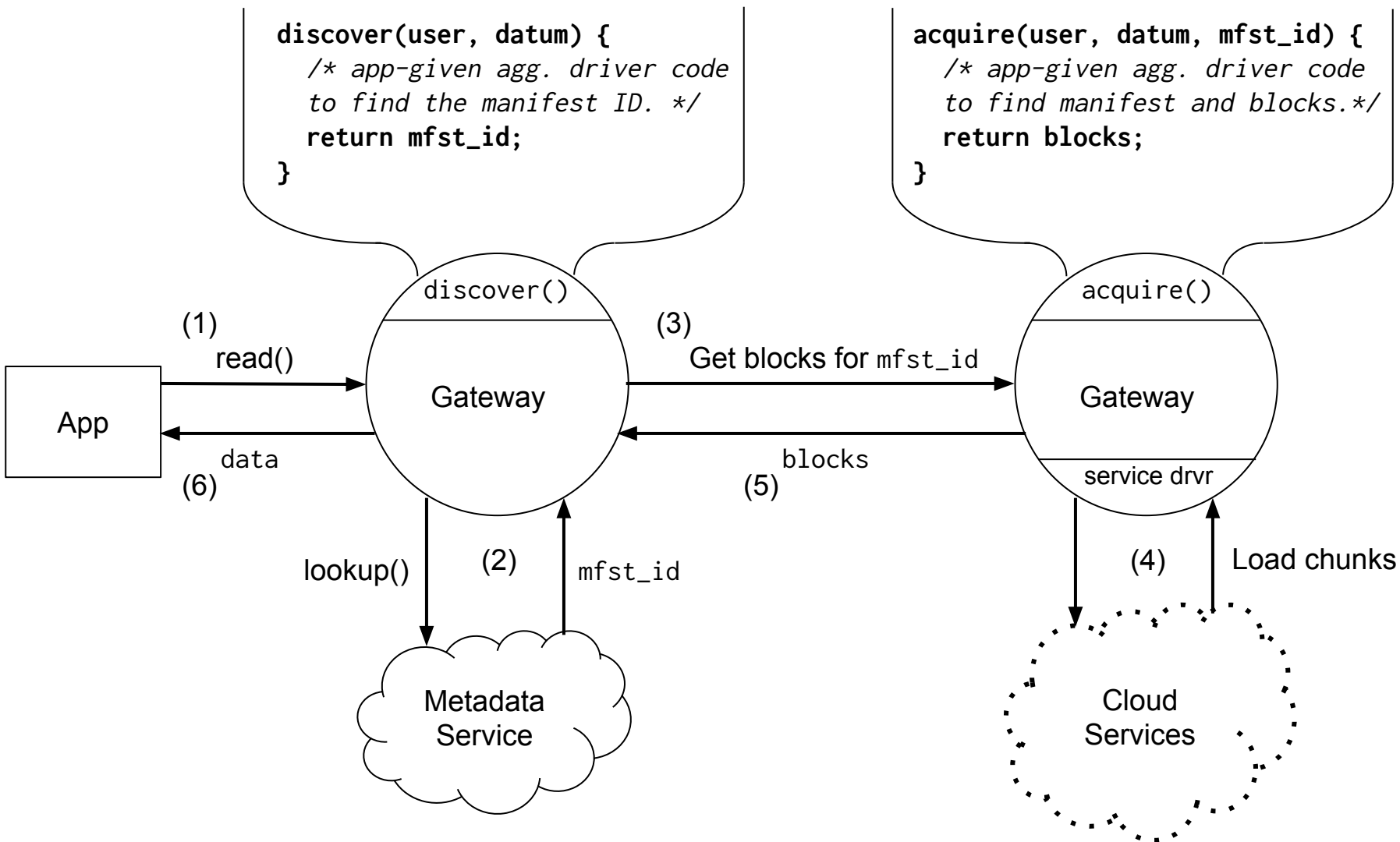
Services





```
discover(user, datum) {  
  /* app-given agg. driver code  
  to find the manifest ID. */  
  return mfst_id;  
}
```

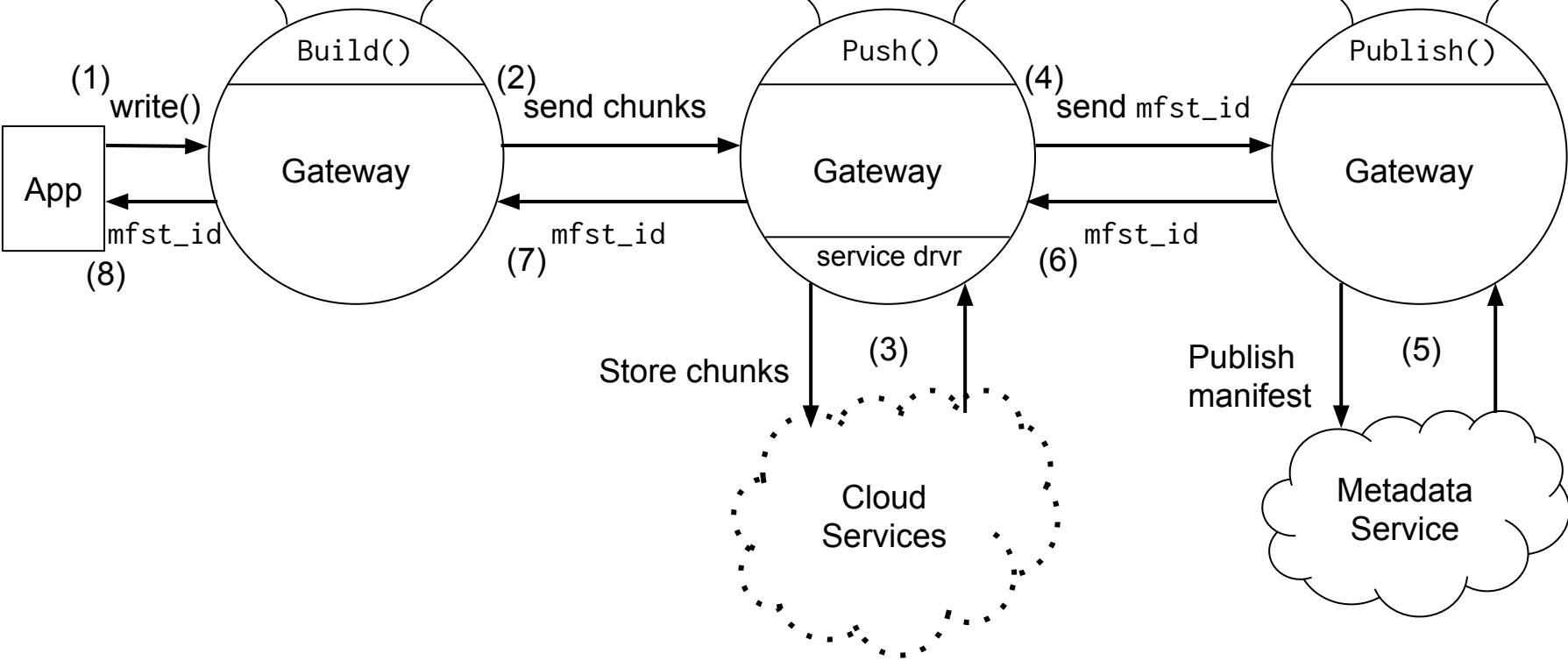
```
acquire(user, datum, mfst_id) {  
  /* app-given agg. driver code  
  to find manifest and blocks.*/  
  return blocks;  
}
```

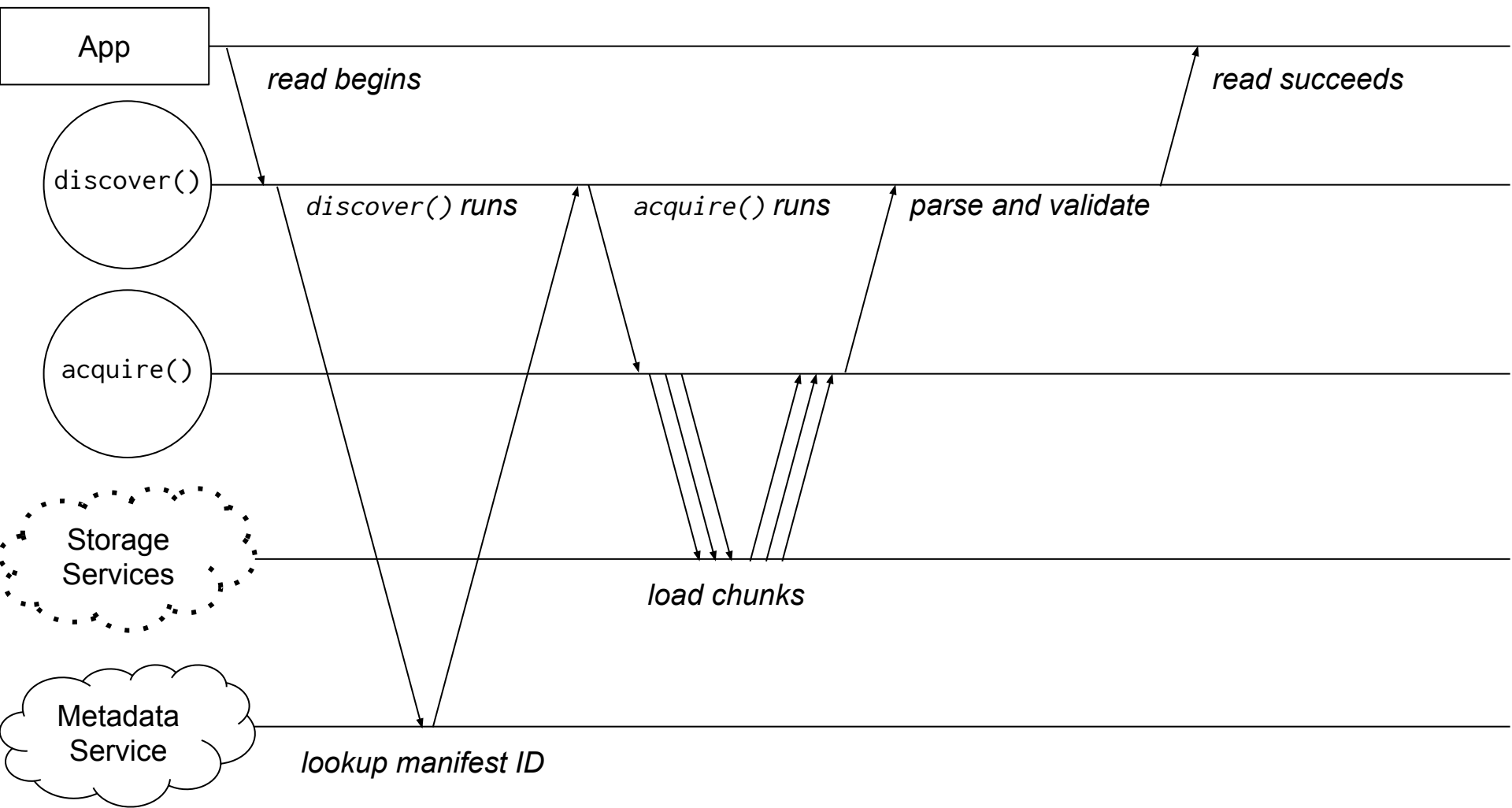


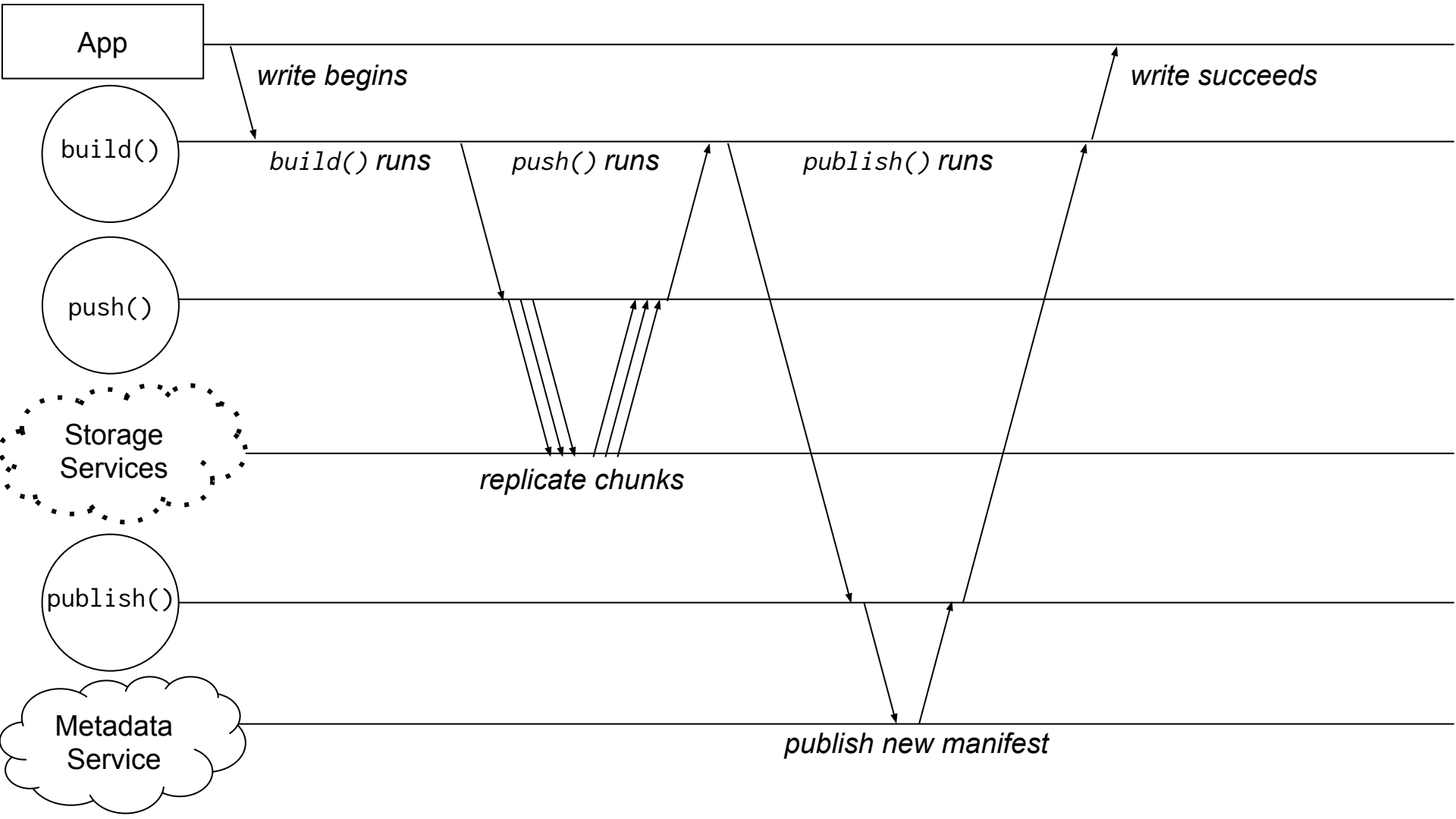
```
Build(user, datum, data) {  
  /* app-given agg. driver code  
  to generate new data. */  
  return (manifest, blocks);  
}
```

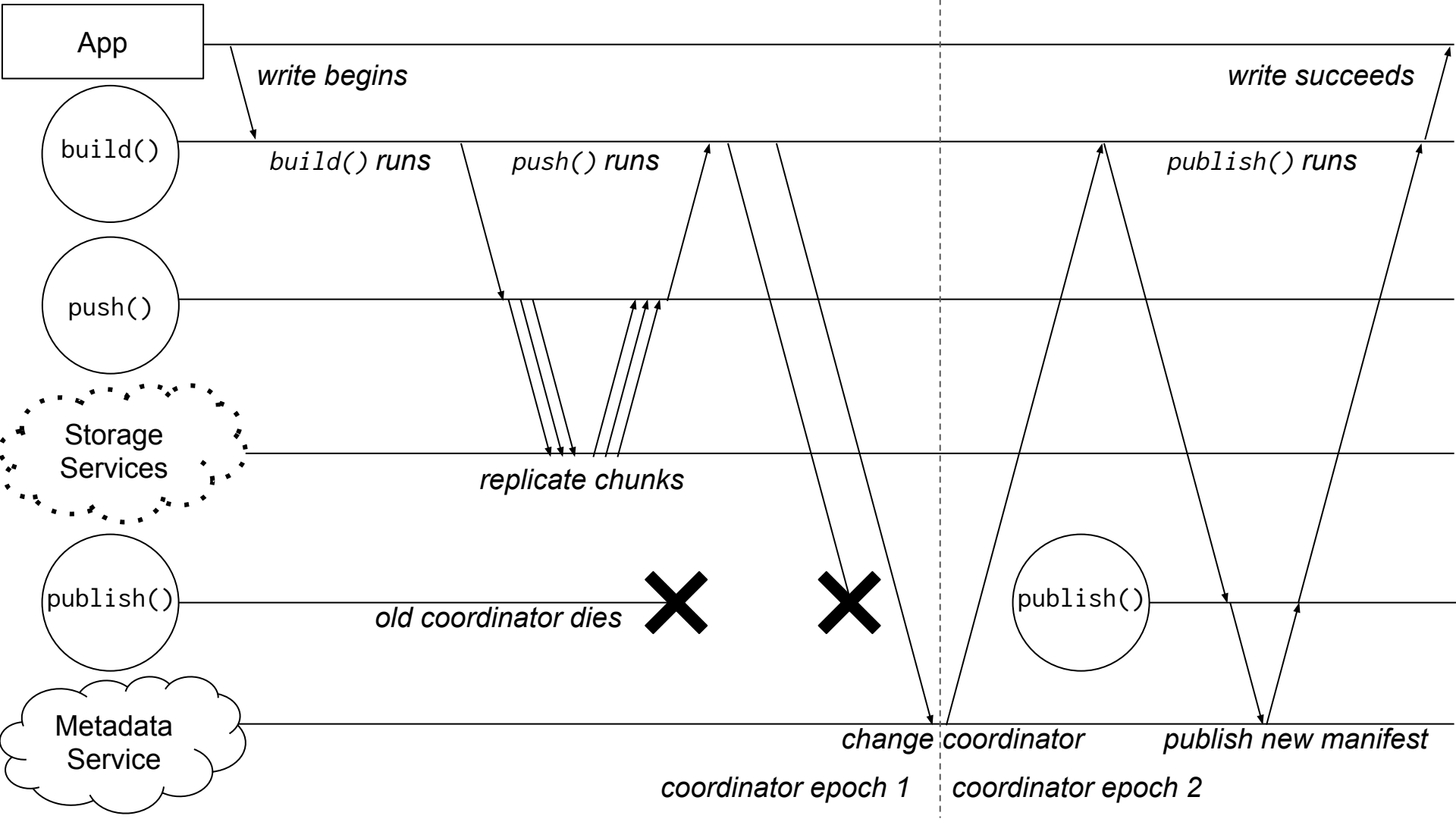
```
Push(user, datum, chunks) {  
  /* app-given agg. driver code  
  to replicate new chunks */  
  return manifest_id;  
}
```

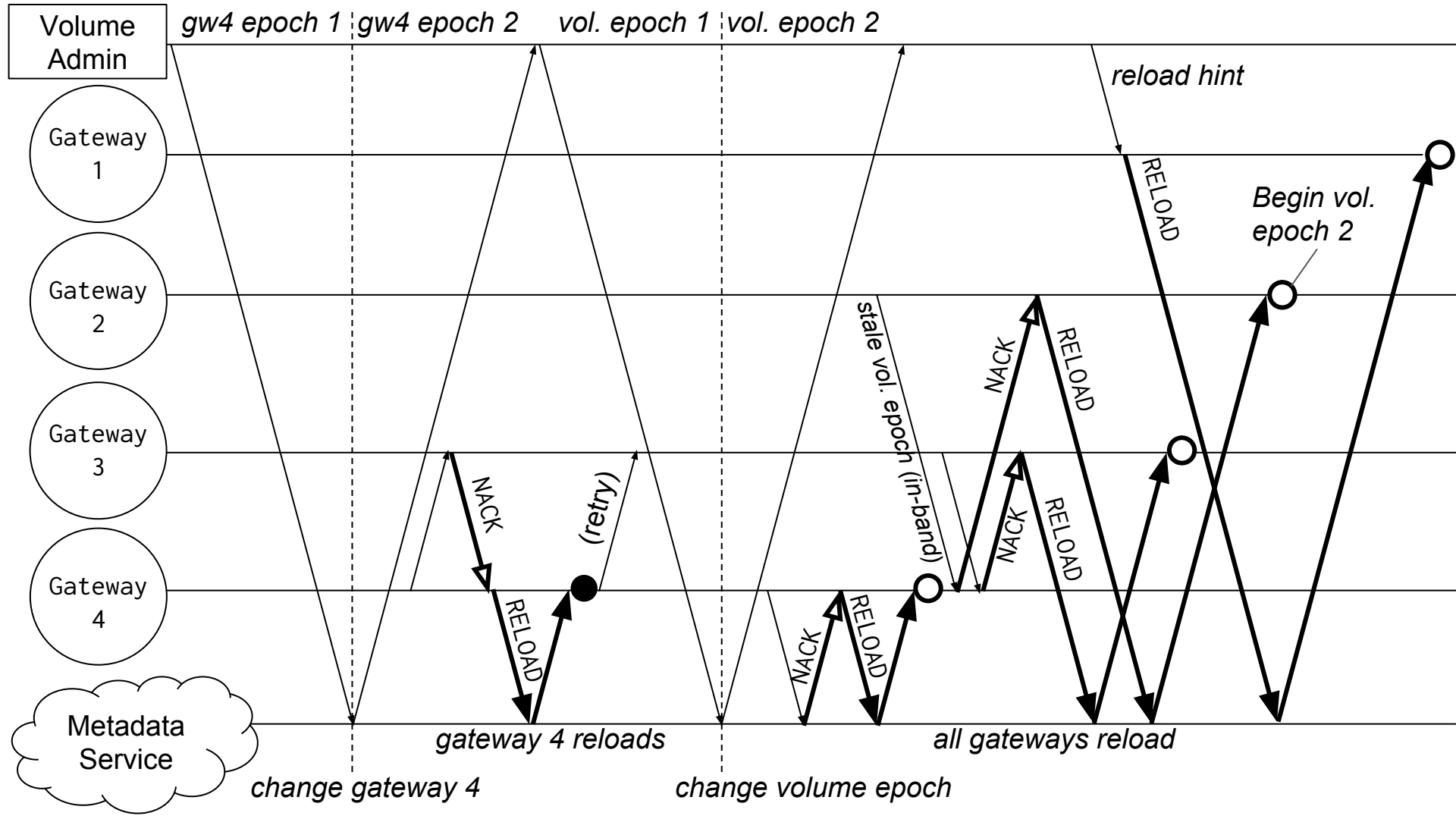
```
Publish(user, datum, mfst_id) {  
  /* app-given agg. driver code  
  to expose new data globally */  
  return mfst_id;  
}
```

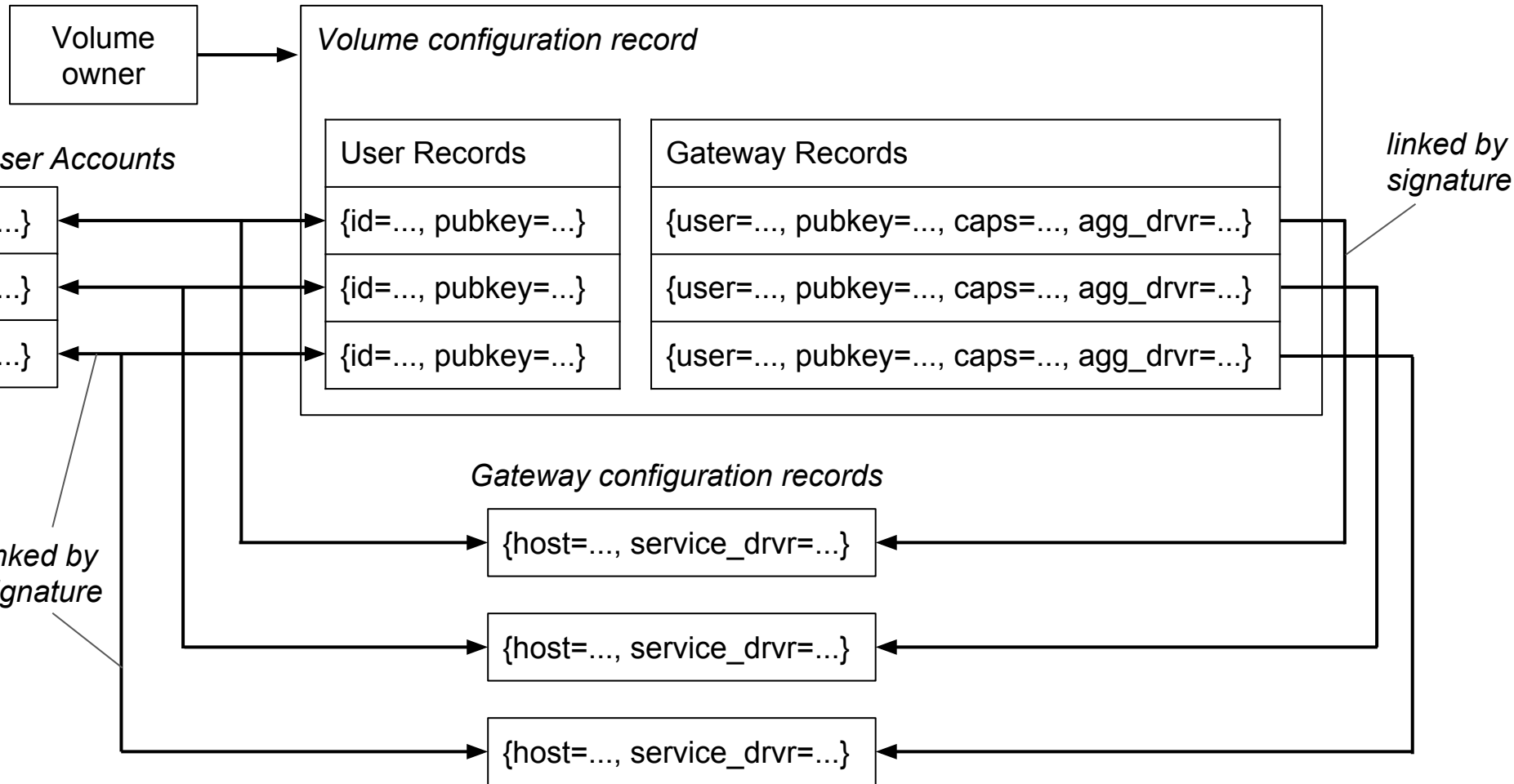


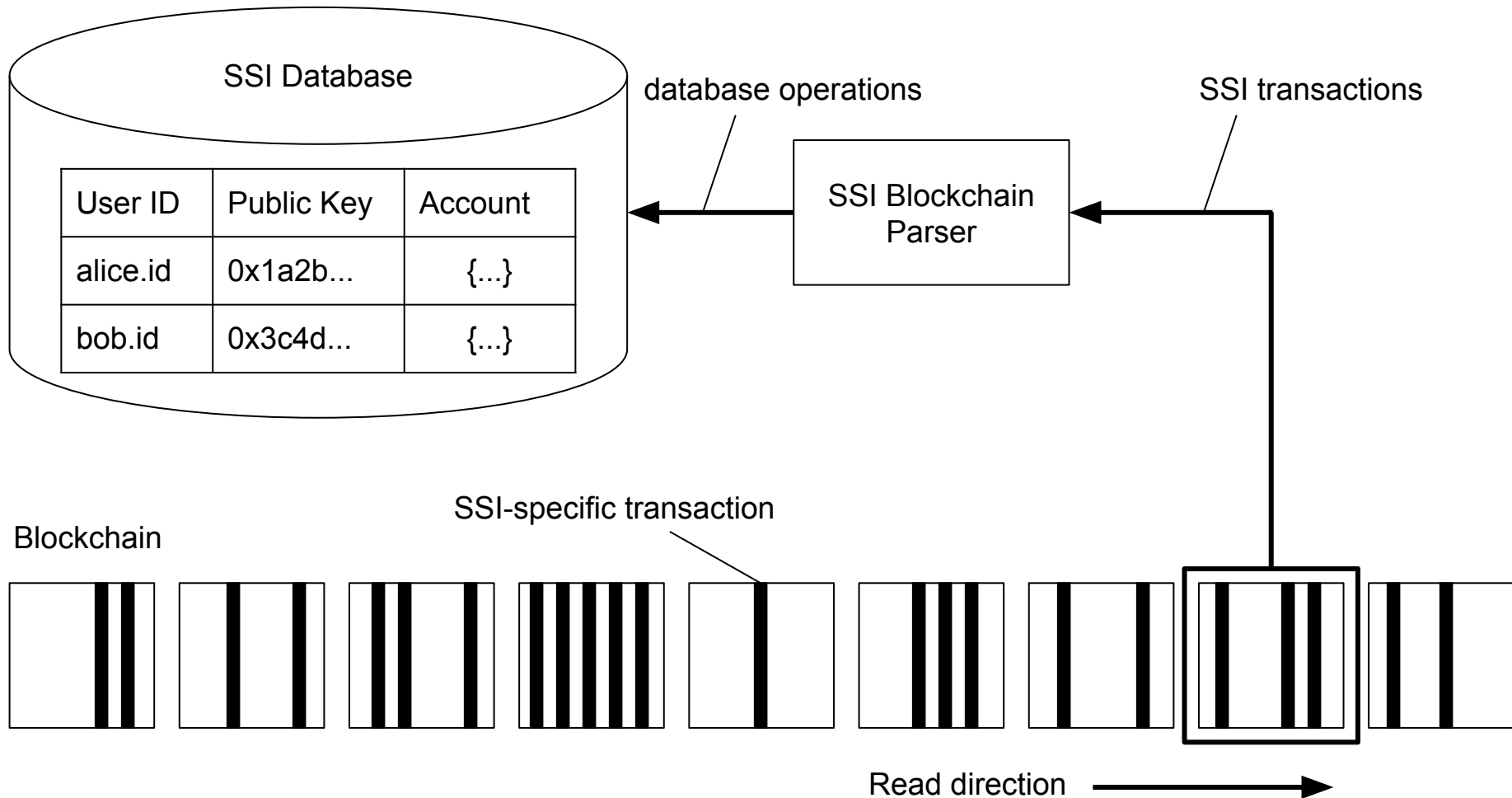










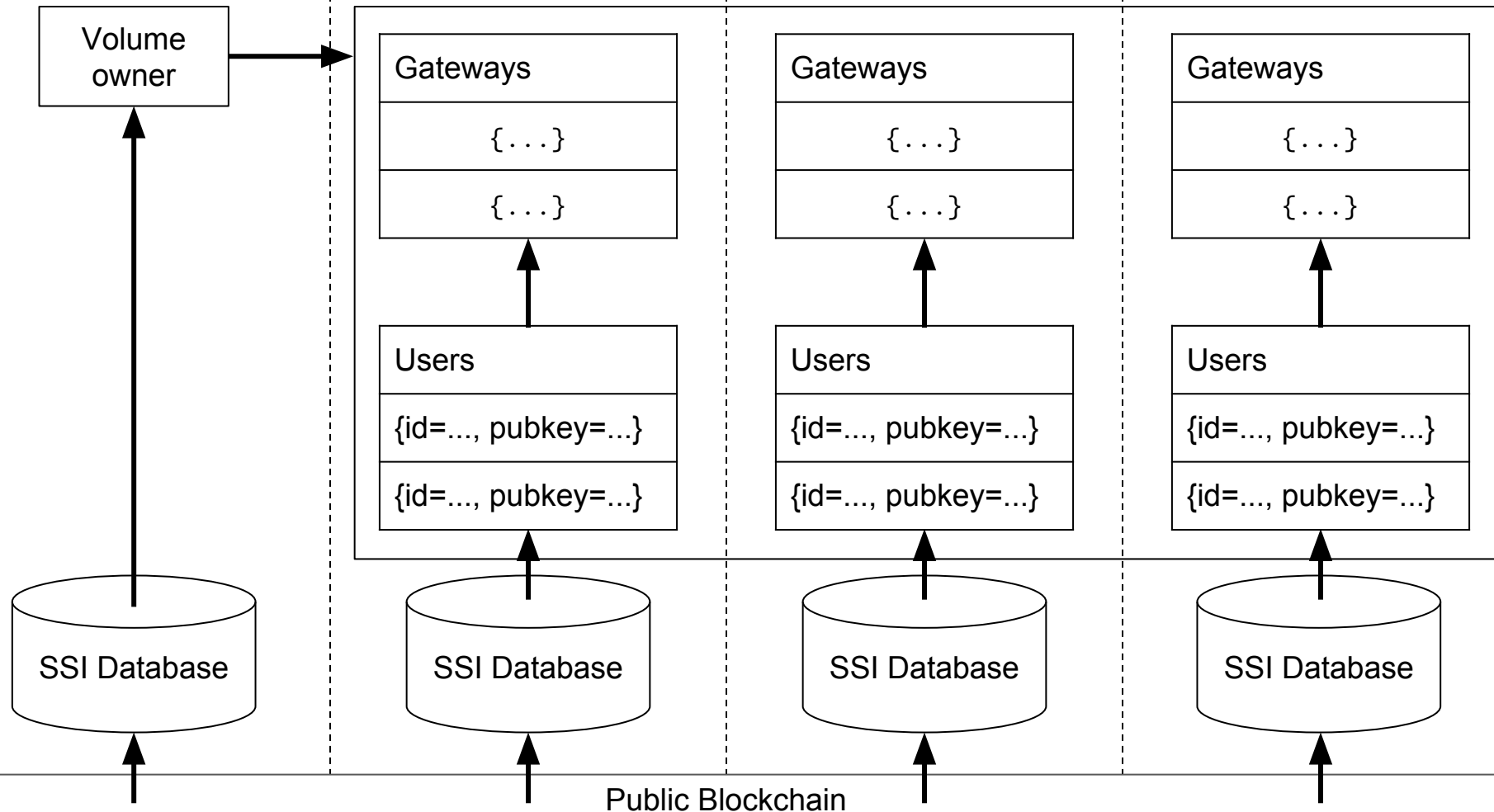


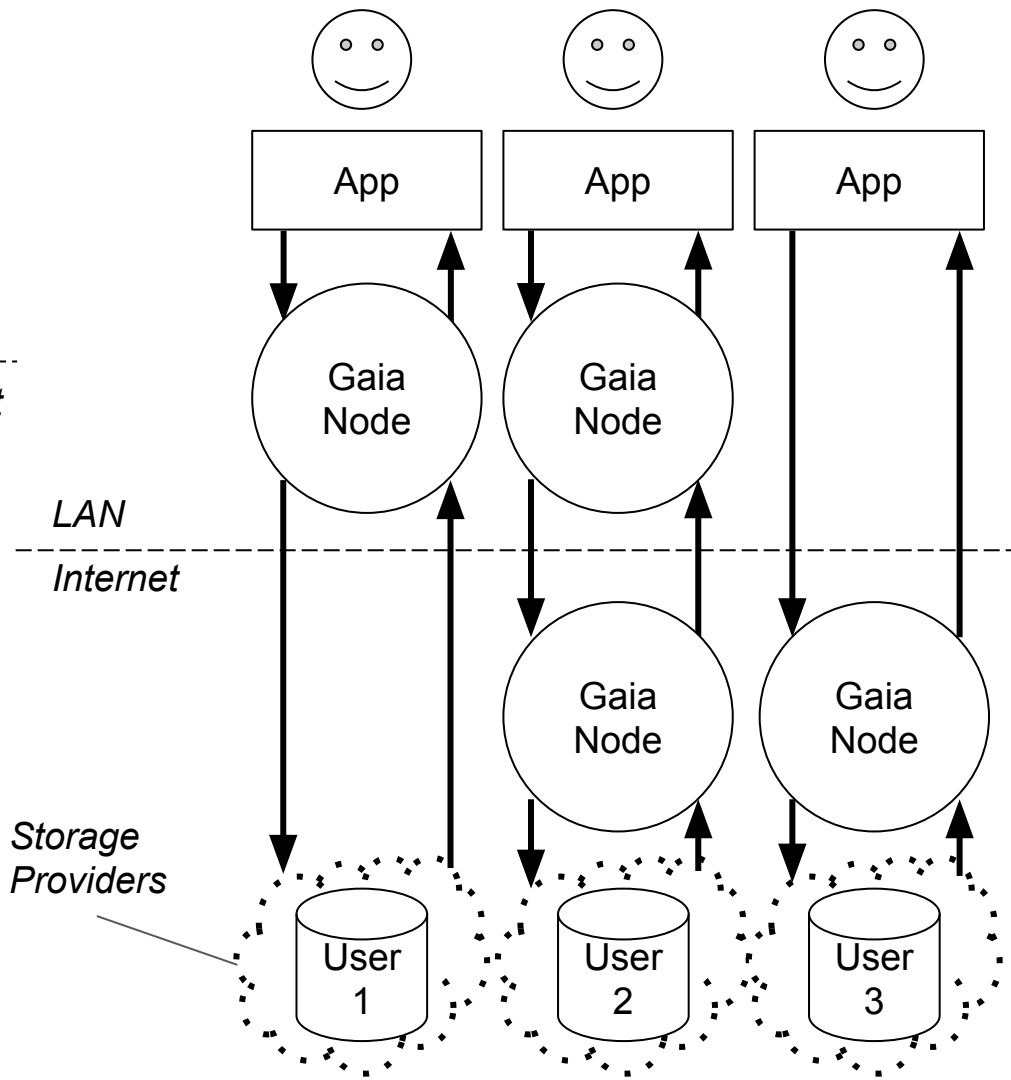
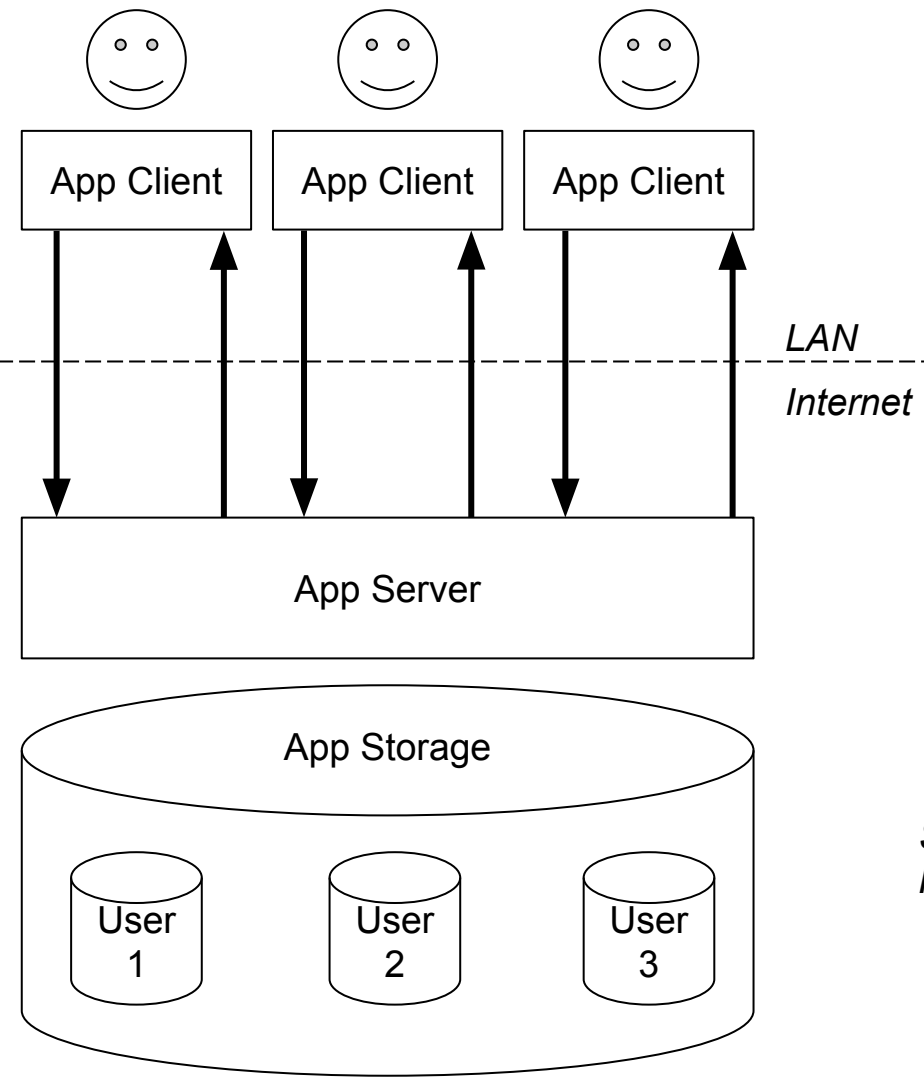
Organization A

Organization B

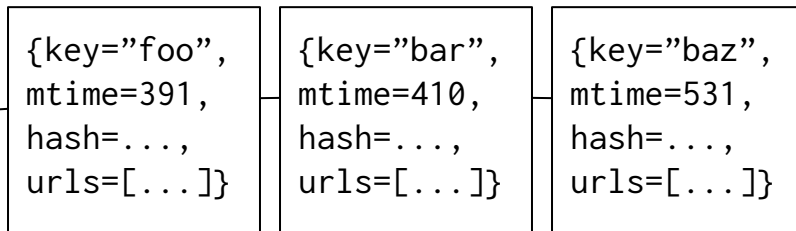
Organization C

Organization D





Application-level Key/Value Storage View



Gaia Volume

Per-device manifest

Manifest
7ea4da

Manifest
b86a62

Manifest
135f0e

`{key="foo", mtime=100, hash=..., urls=[...]}`

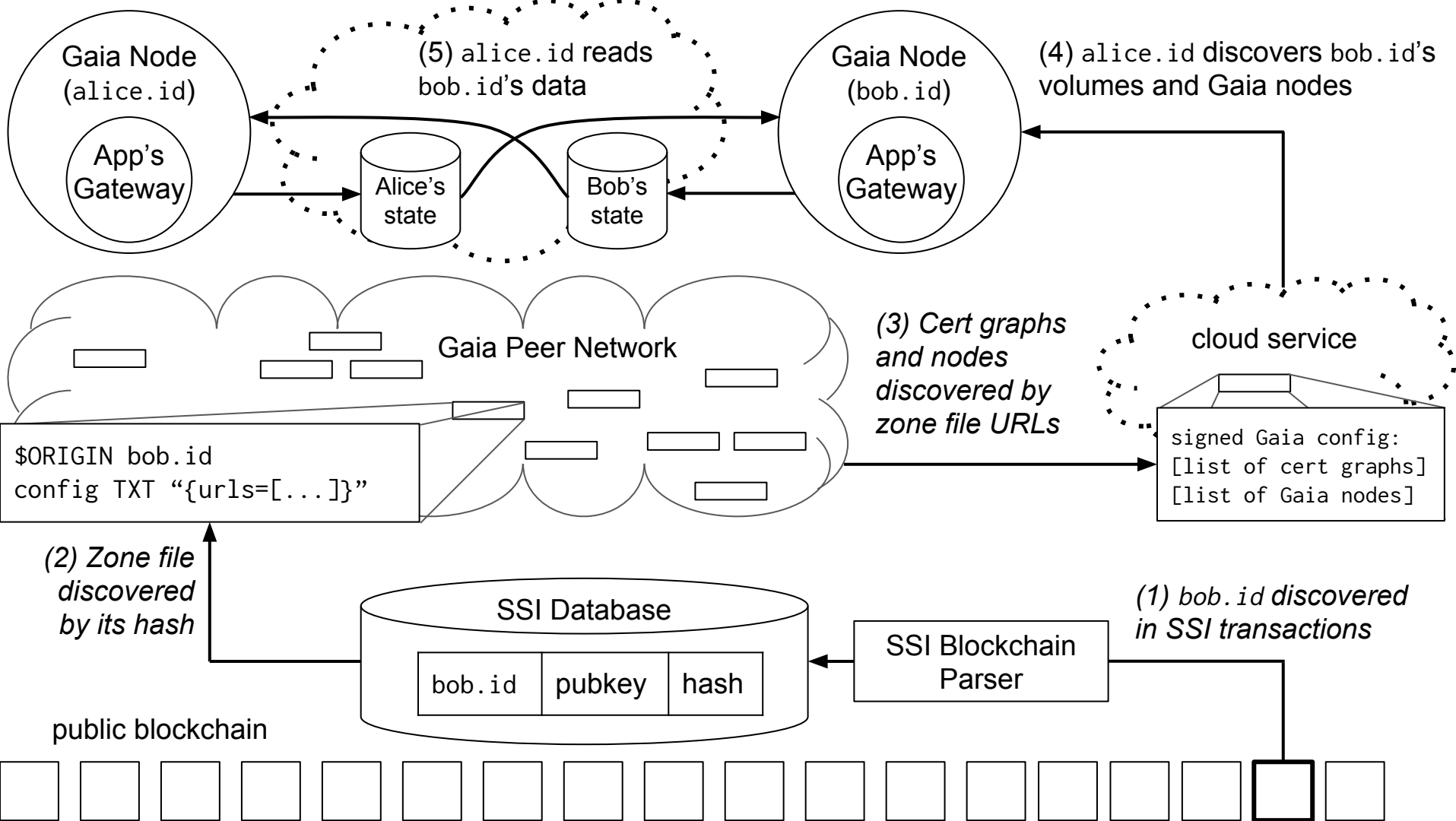
`{key="foo", mtime=210, hash=..., urls=[...]}`

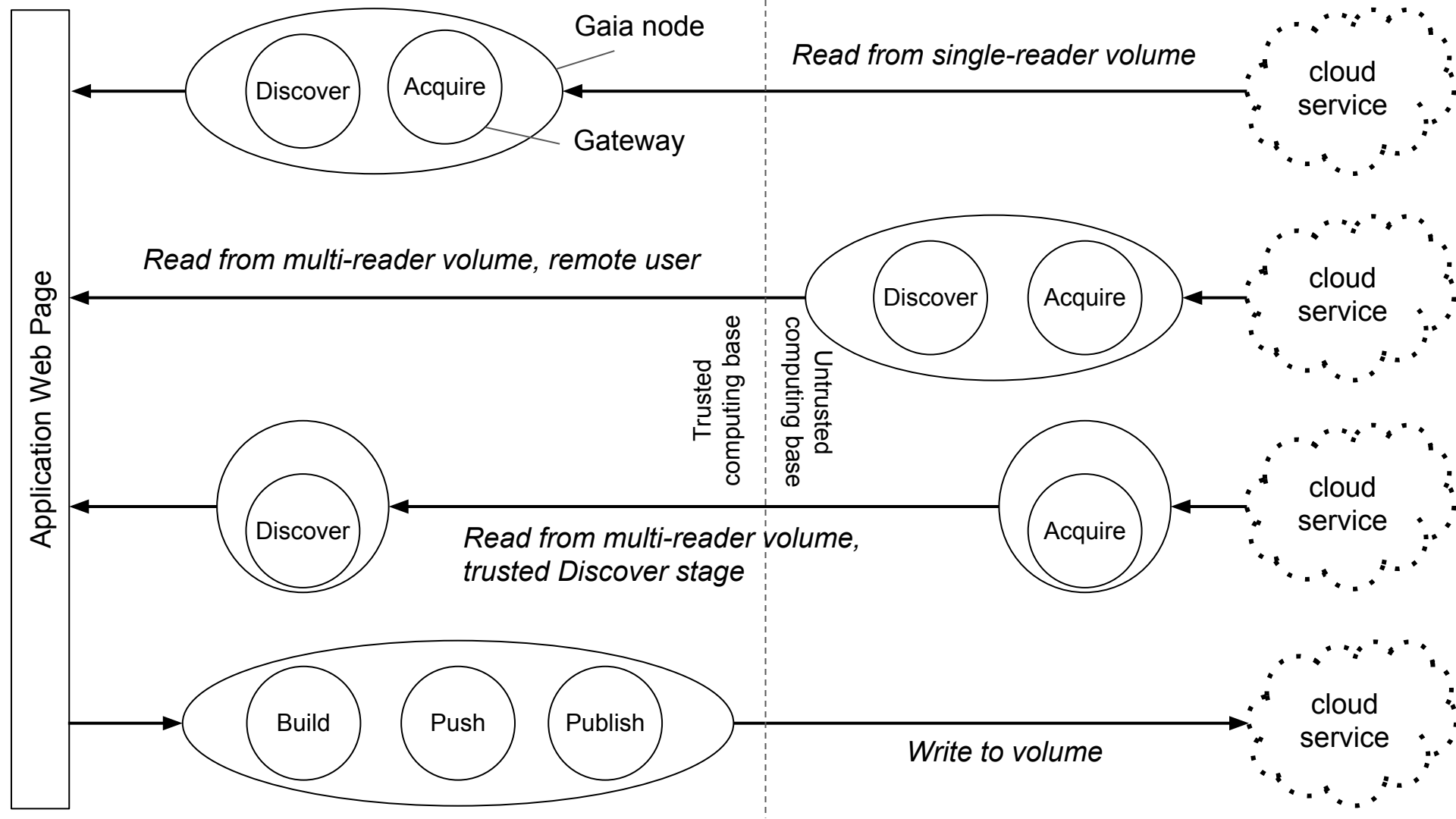
`{key="foo", mtime=391, hash=..., urls=[...]}`

`{key="bar", mtime=240, hash=..., urls=[...]}`

`{key="bar", mtime=410, hash=..., urls=[...]}`

`{key="baz", mtime=531, hash=..., urls=[...]}`



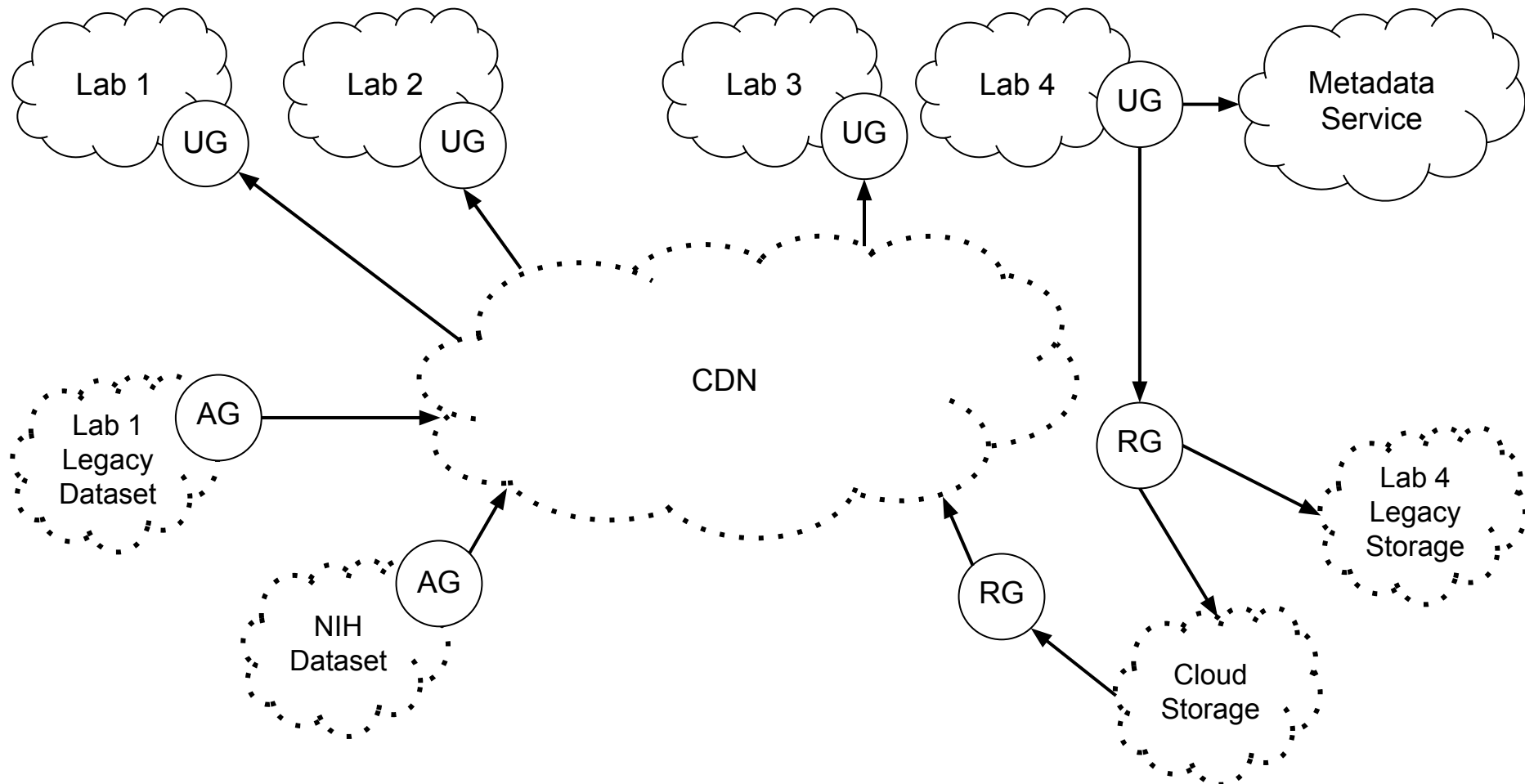


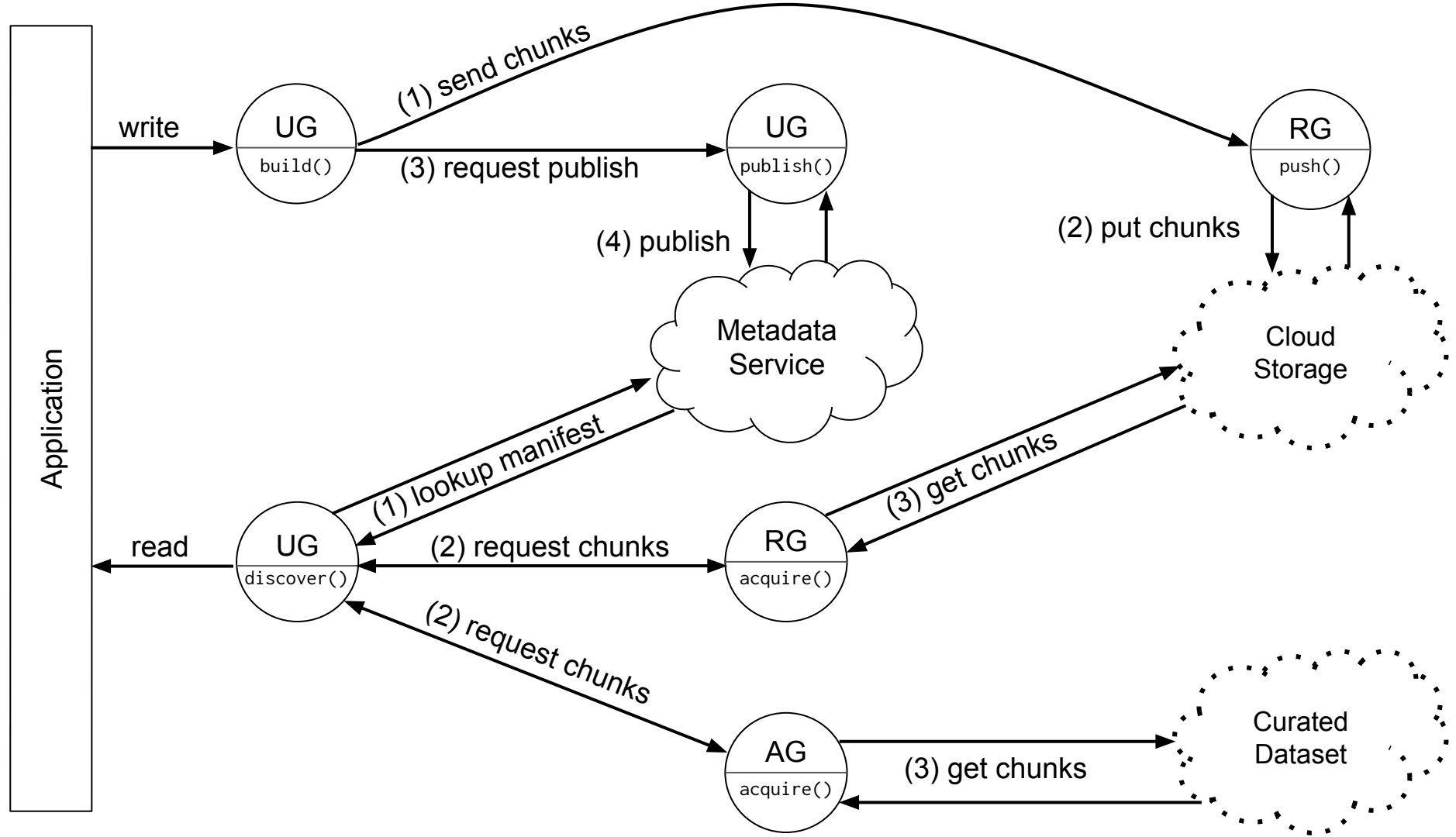
*Lab 1: Reads
NIH data*

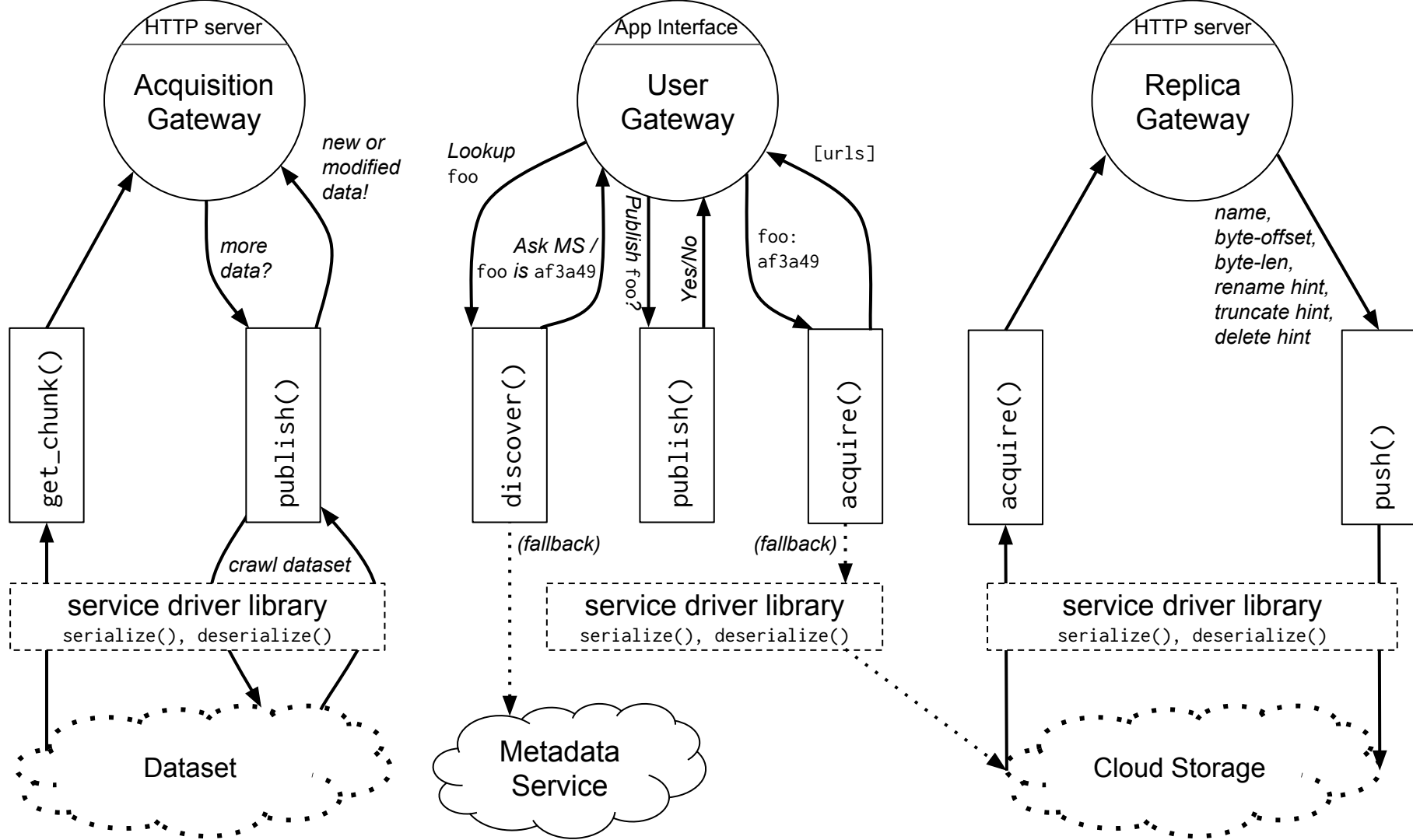
*Lab 2: Reads
Lab 1's dataset*

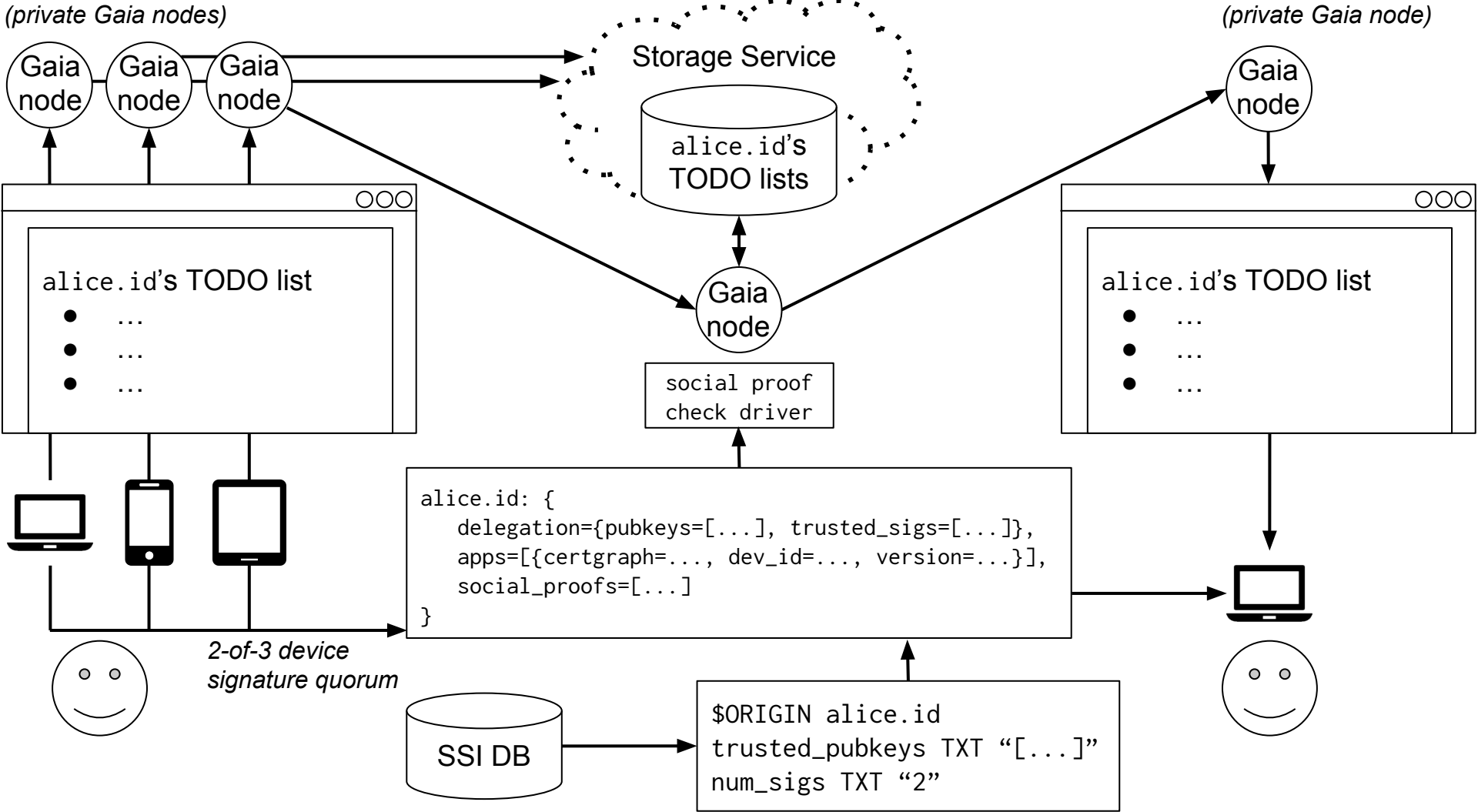
*Lab 3: Reads Lab 4's
data from cloud storage*

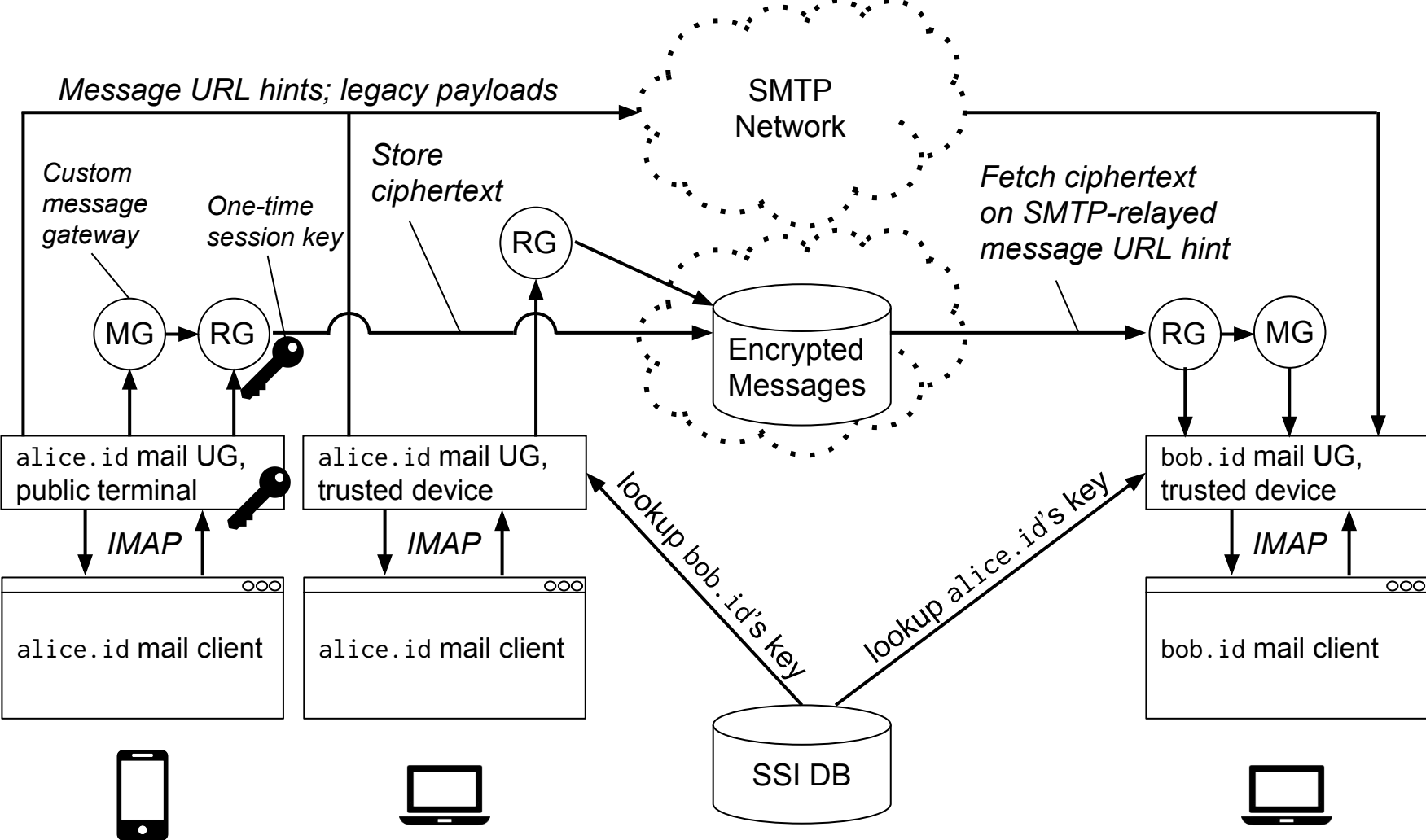
*Lab 4: Mirrors data to legacy
and cloud storage*

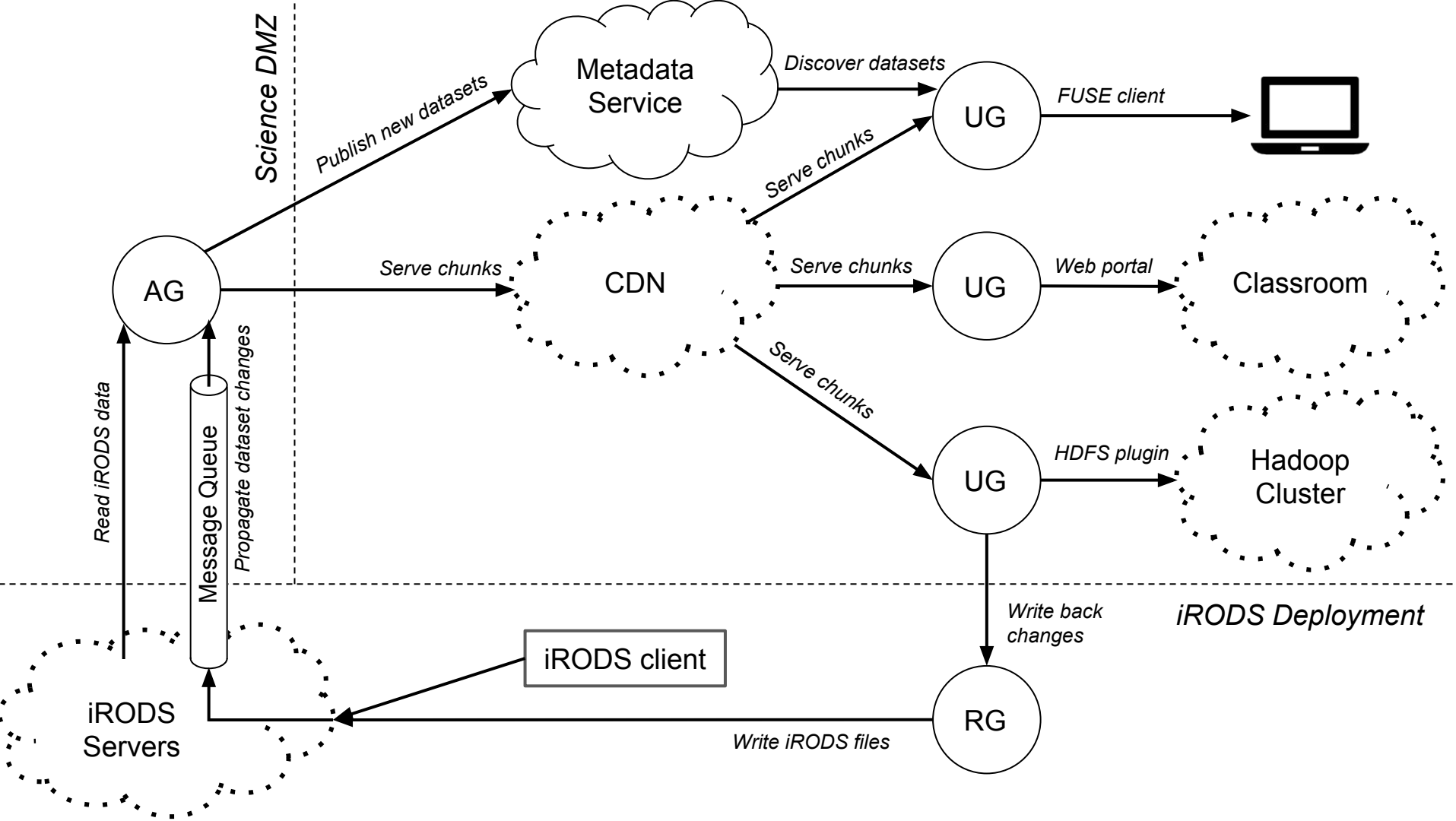


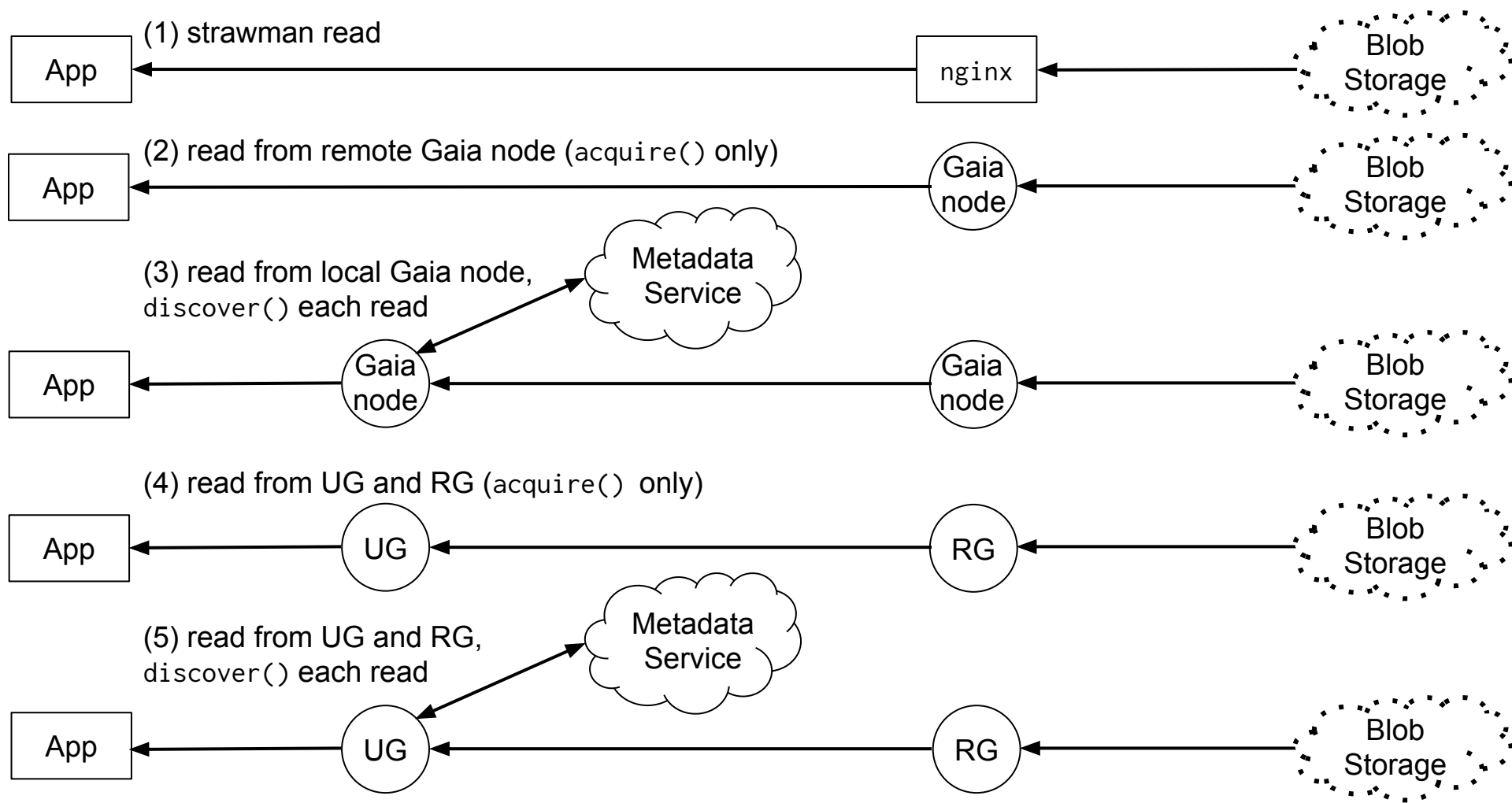


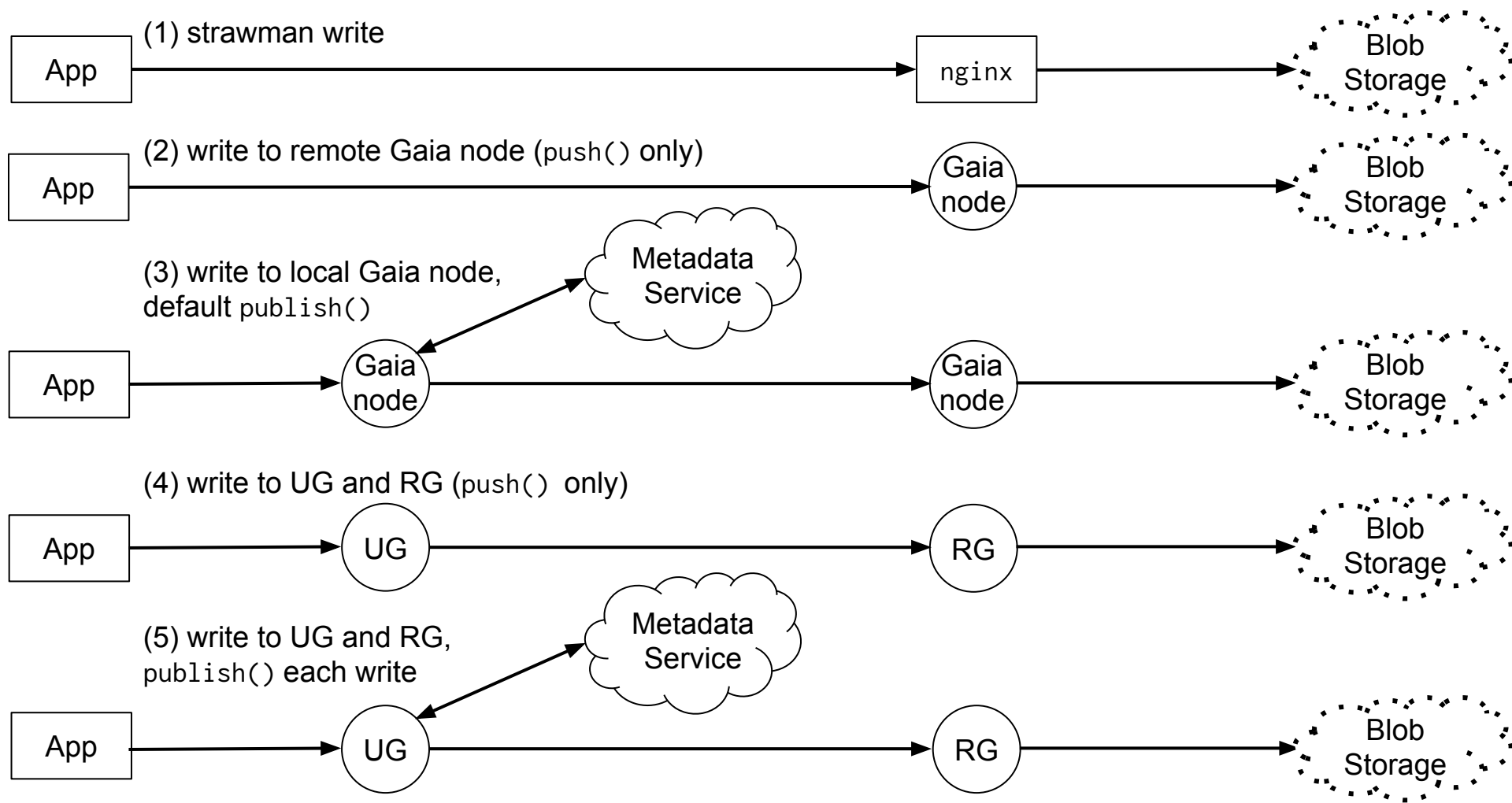












Applications

Scientific workflows

Encrypted email

Groupware

Online games

Blogging

Social media

Video-streaming

Aggregation drivers

Wide-area SDS System

Service drivers

Akamai CDN

Dropbox

Amazon Cloudfront

Amazon S3

NIH GenBank

iRODS catalogs

M-Lab dataset

Services