



Google
Developers



Power your Application's Data using Google Cloud Storage

Navneet Joneja, Product Manager
Nathan Herring, Senior Software Engineer

Overview

- Why Cloud Storage?
- Introduction to the Service
- Advanced Uses / Cloud Storage in Your Application



What could you do with
Storage
as a
Service?



SLA

99.9%

High Availability & Performance





60+ hrs/min

Transparent scaling





Cost-Effective



<http://goo.gl/gxJJ1>

"We're able to have our site **up and running 24 hours a day, seven days a week**. We don't have to manage and maintain servers."



– Brigitte Ganter, Director of Product

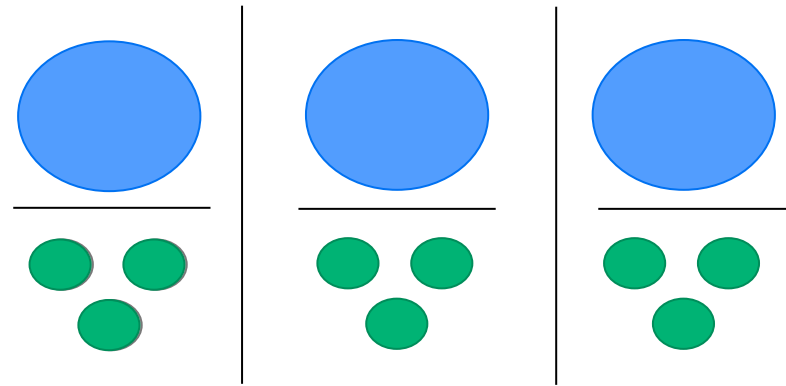




The Foundations

<http://developers.google.com/storage>

Storage



Multiple Layers of Redundancy; Multiple Data Centers.
Regions: US & EU



Network



Powered by the Google Network



Scale



Unlimited data; 5 TB per Object





Cloud Storage in Action

Cloud Storage in Action

- Sign up using the Google APIs Console
 - <https://developers.google.com/console>
- Managing your data using gsutil or the browser interface
- Build new applications using the RESTful API





Service Overview

The Basics



RESTful API; Bucket/Object model



Easy Access



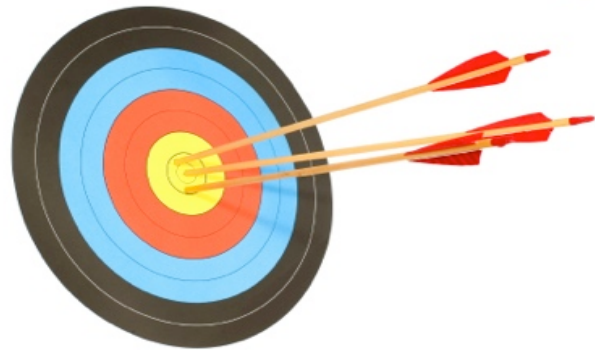
Streaming &
Resumable
Transfers

Optimized browser
Uploads &
Downloads

Transparent Auth from
App Engine, Compute
Engine



Strong Data Consistency

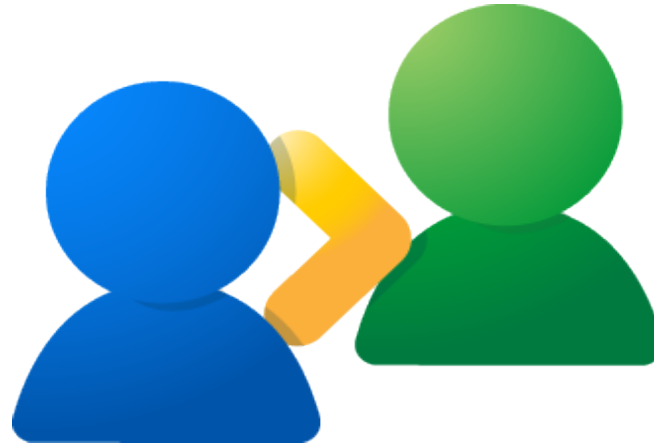


Read your
Writes

Update
Ordering



Control & Share Your Data



Group, User,
Default ACLs

Application-Driven
Sharing

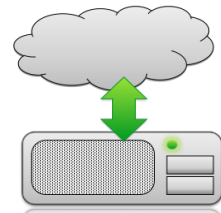
OAuth2;
Browser Access



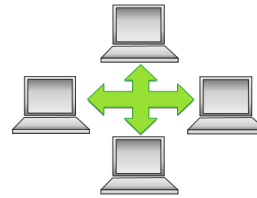
Real-World Use Cases



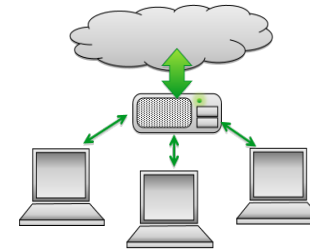
Content
Serving



Archives,
Backups



Data
Sharing



Storage for
applications

01001010101
10101010010
00101010101
42!
Computation



What's new since IO 2011?

- Files API integration with Google App Engine
- Write ordering / concurrency control
- Access Logs
- Short-Lived Pre-Authenticated URLs
- Direct Browser Uploads
- Cross-Origin Resource Sharing (CORS)
- Per-bucket default object access controls
- Image Transformations in GAE
- Easily serve static sites (index & error page semantics)
- Faster access log delivery
- Technology Preview: JSON API
- Technology Preview: Notifications
- Technology Preview: Multiple Object Versions

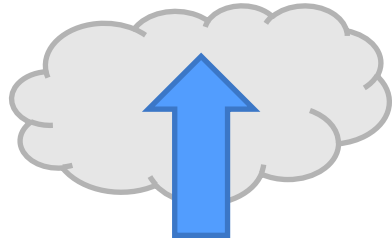




Advanced Uses

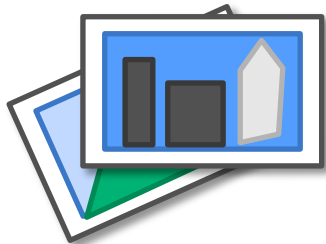
...or how Cloud Storage fits into *your* next app

Building your data-driven application



Handle end-user uploads

Take actions on uploaded objects



Serve objects and other results



Analyze usage





Demo



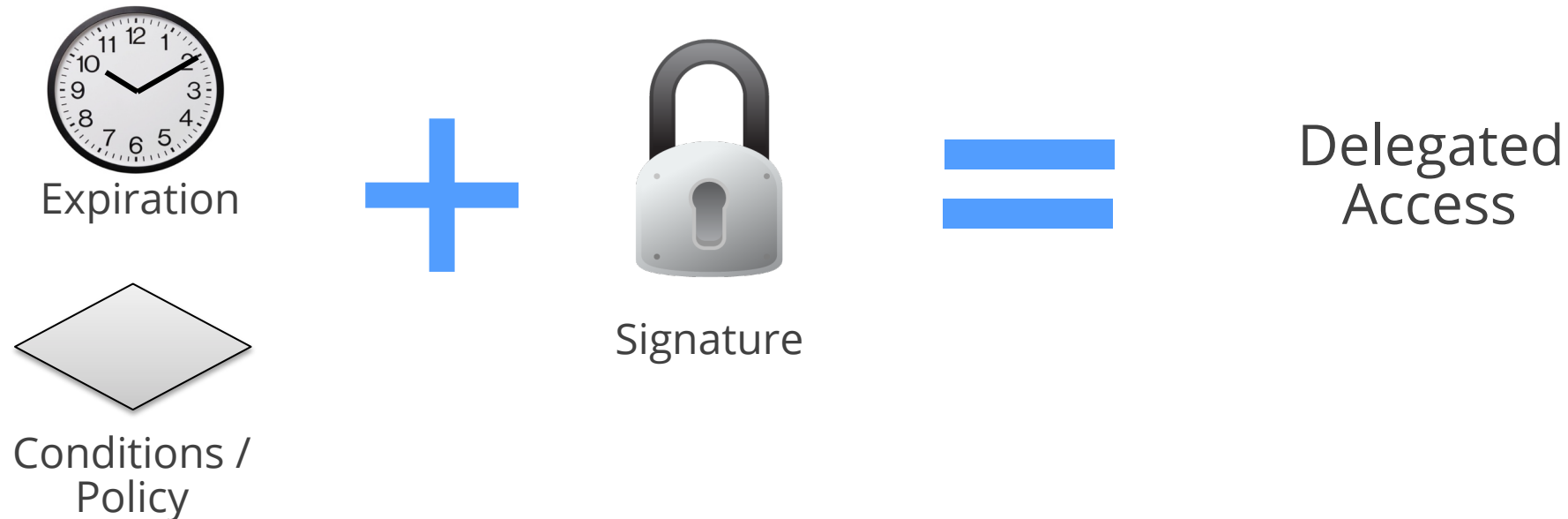
Storing End User Data

Signed URLs and POST Policies

Why?

- Application-Driven access control
- Premium Content / Software Distribution
- Allow End-Users to directly upload data to the cloud

How does it work?



Cross-Origin Resource Sharing (CORS)

Letting the browser play Good Cop

Allows:

- Form post
 - JavaScript XMLHttpRequest (XHR)
- from your application's pages to Cloud Storage.

The browser sends:

Origin: <http://www.example.appspot.com>

Cloud Storage returns:

Access-Control-Allow-Origin: <http://www.example.appspot.com>



Access Control on Customer Uploads

- Optionally configure groups:
 - [API project teams](#) (owner, editor, viewer) are on ACLs by default.
 - Use a [Google Group](#) to manage bucket- and object-specific ACLs
- Specify ACLs to use for uploaded objects:
 - Put predefined ACL ([x-goog-acl](#)) request header in the signed URL, or
 - Set a [default object ACL](#) on the bucket.





Serving Static Sites

Serving the Gallery

Gallery served as a static website

JavaScript

```
PHOTO_URL = 'http://commondatastorage.googleapis.com/marc-photos/'
THUMB_URL = 'http://commondatastorage.googleapis.com/marc-thumbs/'
$.each(thumbs, function (index, photo) {
    photo_url = PHOTO_URL + photo;
    thumb_url = THUMB_URL + photo;
    $('<a rel="gallery"/>').append($('<img>').prop('src', thumb_url))
        .prop('href', photo_url).prop('title', photo).appendTo(gallery);
});
// Yields 
```



Buckets as Static Sites

New

Who needs a separate web server?

- Custom “directory” pages – e.g., example.com loads from example.com/index.html
- Custom 404 page

XML

```
PUT /example.com?websiteConfig HTTP 1.1
```

```
...
```

```
<WebsiteConfiguration>
```

```
  <MainPageSuffiix>index.html</MainPageSuffix>
```

```
  <NotFoundPage>404.html</NotFoundPage>
```

```
</WebsiteConfiguration>
```





Acting on Uploaded Data

Triggering Application Code via Object Notifications

Managing Uploaded Images

1. Customer uploads a photo
2. Service creates a thumbnail
3. Customer's galleries are updated



Wiring Up Notifications

Part 0: Authorization

- The subscriber needs to have FULL_CONTROL permission on the bucket.
- Presuming the subscriber is a [service account](#), the ACL will need to contain:

XML

```
<Entry>
  <Scope type="UserByEmail">
    <EmailAddress>appid@appspot.gserviceaccount.com</EmailAddress>
  </Scope>
  <Permission>FULL_CONTROL</Permission>
</Entry>
```



Wiring Up Notifications

Part 1: Subscription

Python

```
from google.appengine.api import app_identity
from google.appengine.api import urlfetch
...
scope = "https://www.googleapis.com/auth/devstorage.full_control"
subscribe_url = "https://www.googleapis.com/storage/v1beta1/b/" + yourbucket + "/o"
callback_url = "https://" + app_identity.get_application_id() + ".appspot.com/
&appEngine=true"
authorization_token, _ = app_identity.get_access_token(scope)
response = urlfetch.fetch(subscribe_url, method=urlfetch.GET, headers = {
    "Authorization": "Bearer " + authorization_token, "X-Subscribe": "web_hook?url=" +
callback_url})
```



Wiring Up Notifications

Interlude: What's in a notification?

JSON

```
{ "kind": "storage#objects", "items": [  
  { "kind": "storage#object", "id": "myBucket/myObj",  
    "name": "myObj", "bucket": "myBucket",  
    "selfLink": "https://www.googleapis.com/storage/v1beta1/b/myBucket/myObj?alt=json",  
    "media": { "contentType": "text/plain", "timeCreated": "2011-04-25T19:11:46.890Z",  
              "length": "1234567", "hash": "01234abcdef...", "algorithm": "MD5",  
              "link": "https://www.googleapis.com/storage/v1beta1/b/myBucket/myObj" },  
    "owner": { "entity": "user-00b4903a97...", "entityId": "00b4903a97..." } },  
  { ... }  
]
```



Wiring Up Notifications

Part 2: Responding to Notifications

Python

```
import webapp2
import json
...
class MainPage(webapp2.RequestHandler)
    def post(self):
        notification = json.loads(self.request.body)
        bucket = notification.items[0]['bucket']
        object = notification.items[0]['object']
        // Profit!
```





Navel Gazing

Usage Logs & Analytics

Analyzing Usage

- Configure bucket to get access (and storage) logs
- Use Google BigQuery to analyze your data



Analyzing Usage

Configuring your bucket to get Access Logs

Need a second bucket...

- granting write access to the group cloud-storage-analytics@google.com
- with an appropriate default object ACL

XML

```
PUT /example.com?logging HTTP 1.1
```

```
...
```

```
<Logging>
```

```
  <LogBucket>logs.example.com</LogBucket>
```

```
  <LogObjectPrefix>example.com</LogObjectPrefix>
```

```
</Logging>
```



Analyzing Usage

Using BigQuery (<http://goo.gl/1MjqV>)

Compose Query



```
1 SELECT cs_object, sum(sc_bytes) out_bytes FROM
2     my_logs_dataset.usage_2012_06_21_v0,
3     my_logs_dataset.usage_2012_06_22_v0,
4     my_logs_dataset.usage_2012_06_23_v0
5 GROUP BY cs_object order by out_bytes desc limit 40;
```

RUN QUERY

[Show previous query results](#)

Query Results 6:56am, 23 Jun 2012

Download as CSV

Save as Table

Row	cs_object	out_bytes
1	lulzcats.jpg	7436489
2	cheezburger.jpg	5245774
3	n00b.jpg	5164660





Programmatic Access

Discovery-based API access to Cloud Storage

Discovery-based API

JSON Metadata Sample

Limited Availability

JSON

```
{  
  "kind": "storage#bucket",  
  "id": "myBucket",  
  "selfLink": "https://www.googleapis.com/storage/v1beta1/b/myBucket",  
  "projectId": "myProjectId",  
  "timeCreated": "2011-04-25T19:11:46.890Z",  
  "owner": {  
    "entity": "group-00b4903a97...",  
    "entityId": "00b4903a97..."  
  },  
  "location": "US"  
}
```



Discovery-based API

Improvements

- Metadata Patch
- Metadata Batch
- Line-item Access Control editing

Limited Availability



Discovery-based API

Libraries

Available Now

- JavaScript
- PHP
- Python
- .NET
- Go
- Java

... more support coming!

Limited Availability





Miscellany

A couple other interesting things

Concurrency Control

Traditional HTTP precondition headers:

- If-Match / If-None-Match – Use with ETag.
- If-Modified-Since / If-Unmodified-Since – Use with Last-Modified.

Extension HTTP precondition headers for Copy operation:

- x-goog-copy-source-if-match / x-goog-copy-source-if-non-match
- x-goog-copy-source-if-modified-since / x-goog-copy-source-if-unmodified-since



Object Versions

Data retention in the cloud.

- Per-bucket configuration
- Keeps deleted and overwritten objects in history
- Historical objects have separate, mutable metadata, including
 - Access controls
 - Content-Type, Content-Encoding, Cache-Control
 - User-defined metadata





Wrap up

All that and a bag of chips...

<https://developers.google.com/storage/>

Use it now:

- Access and Storage Logs (<http://goo.gl/J2IWQ>)
- Signed URLs (<http://goo.gl/qGi4r>)
- POST policies (<http://goo.gl/thXLS>)
- Cross Origin Resource Sharing (<http://goo.gl/o0zR9>)
- Buckets as Websites (<http://goo.gl/z88o1>)
- Default Object ACLs (<http://goo.gl/zYokl>)
- Concurrency Control (<http://goo.gl/upGKZ>)

Discovery-based (JSON) API access

- Limited Preview – Request access (<http://goo.gl/h5OYc>).
- Read the documentation (<http://goo.gl/j9wEN>) and try it out in the API Explorer (<http://goo.gl/cJtba>).

Technology Preview:

- Notifications
- Object Versions

Tune to gs-announce@googlegroups.com and give us feedback at gs-discussion@googlegroups.com!



<https://developers.google.com/storage/docs/signup>

Related Sessions

- Building Data Pipelines at Google Scale (<http://goo.gl/1bz05>)
 - Running MapReduce on data stored in Cloud Storage.
 - Workflows to move data through Cloud Storage to BigQuery.
- OAuth 2.0 for Identity and Data Access (<http://goo.gl/T0ilU>)
 - More about using Service Accounts to interact with Cloud Storage and other APIs.



Thank you!

<https://developers.google.com/storage>

gs-discussion@googlegroups.com

#GCS / Stack Overflow Tag "google-cloud-storage"

+Navneet Joneja and +Nathan Herring





Google
Developers