Jfokus 2011

Google Cloud for Java Developers: Platform and Monetization

Patrick Chanezon, Developer Advocate, Cloud
@chanezon, chanezon@google.com

Agenda

- Google Apps Marketplace
- Google App Engine
- Google App Engine for Business
- Google Storage for Developers
- Prediction API
- BigQuery
- Google Fusion Tables
- Google Visualization API
- Google Refine

The benefits of Cloud Computing

Economics

- Pay for only what you use
- TCO
- OPEX vs CAPEX

Operations

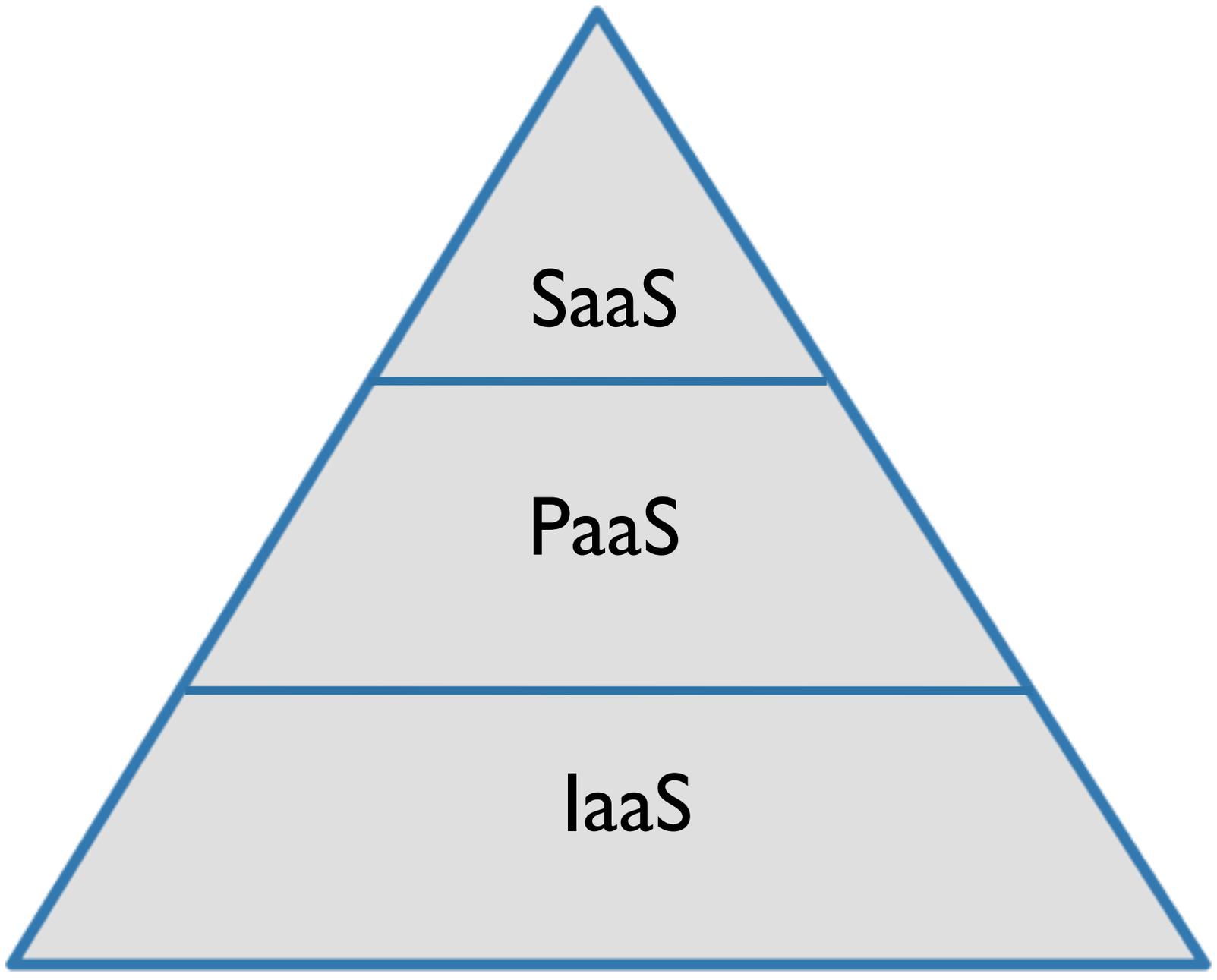
- Day to day: no maintenance
- Fighting fires: no Pagers

Elasticity

Focus on your Business

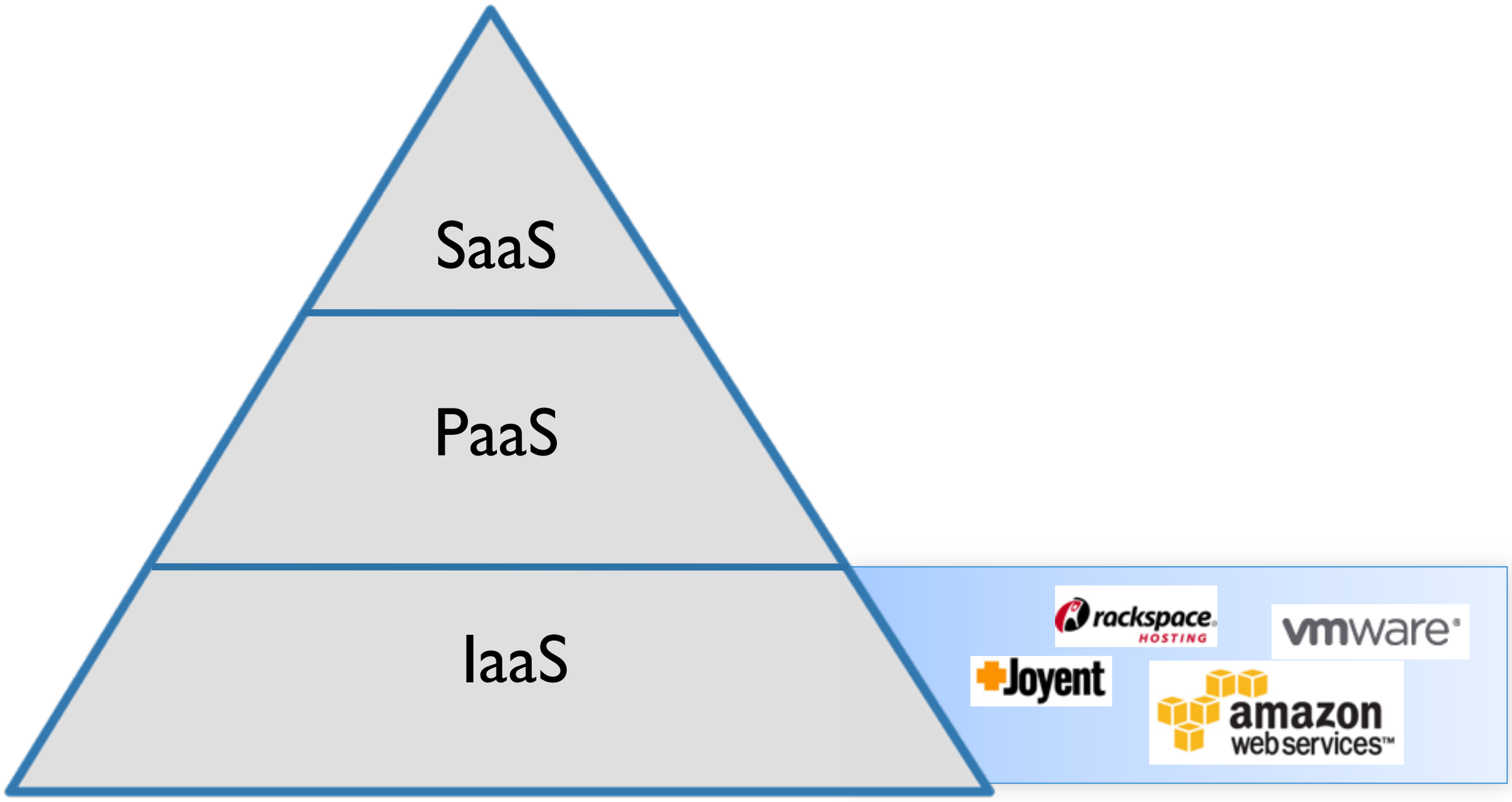
What is cloud computing?

Cloud Computing Defined



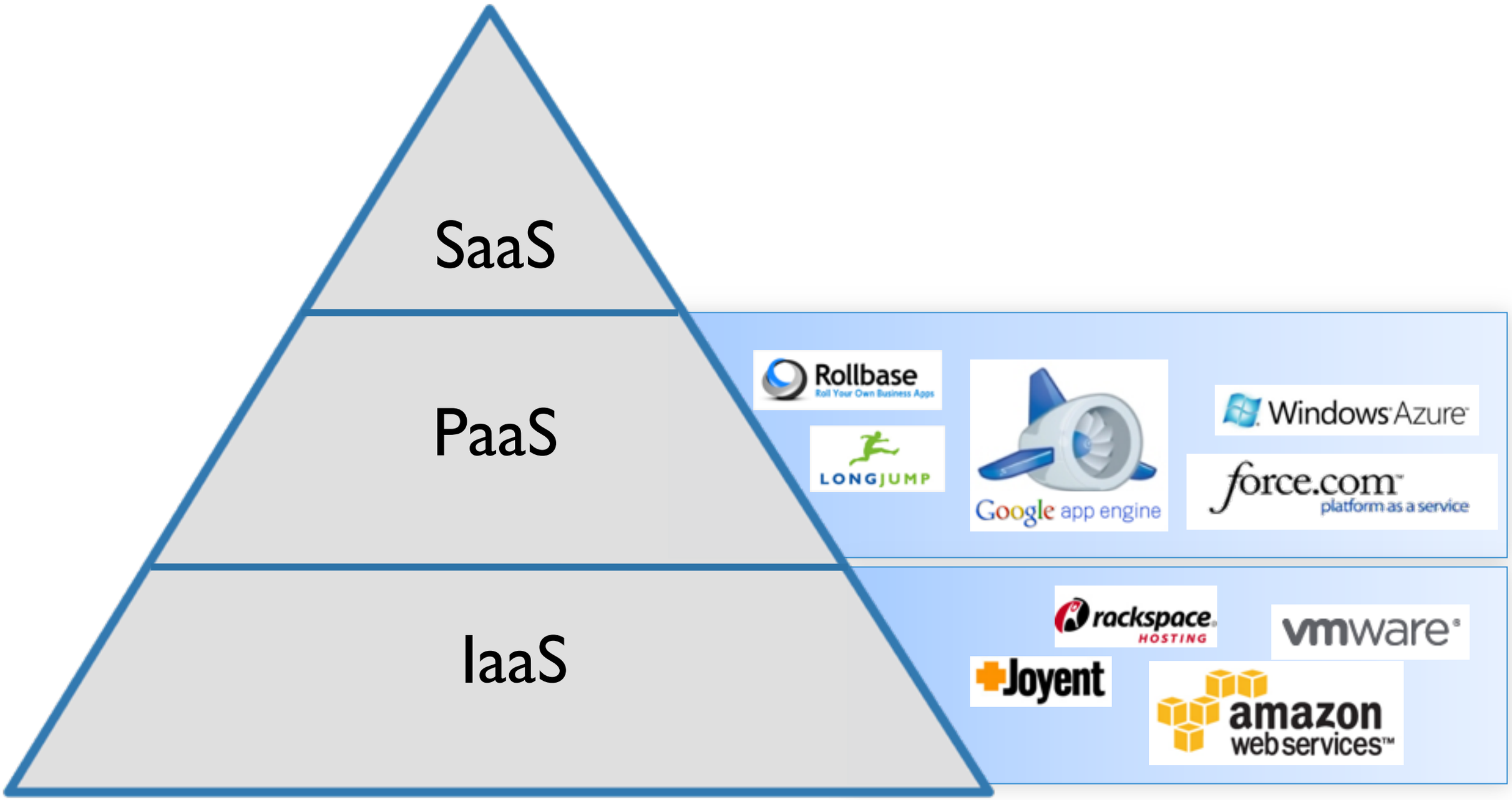
Source: Gartner AADI Summit Dec 2009

Cloud Computing Defined



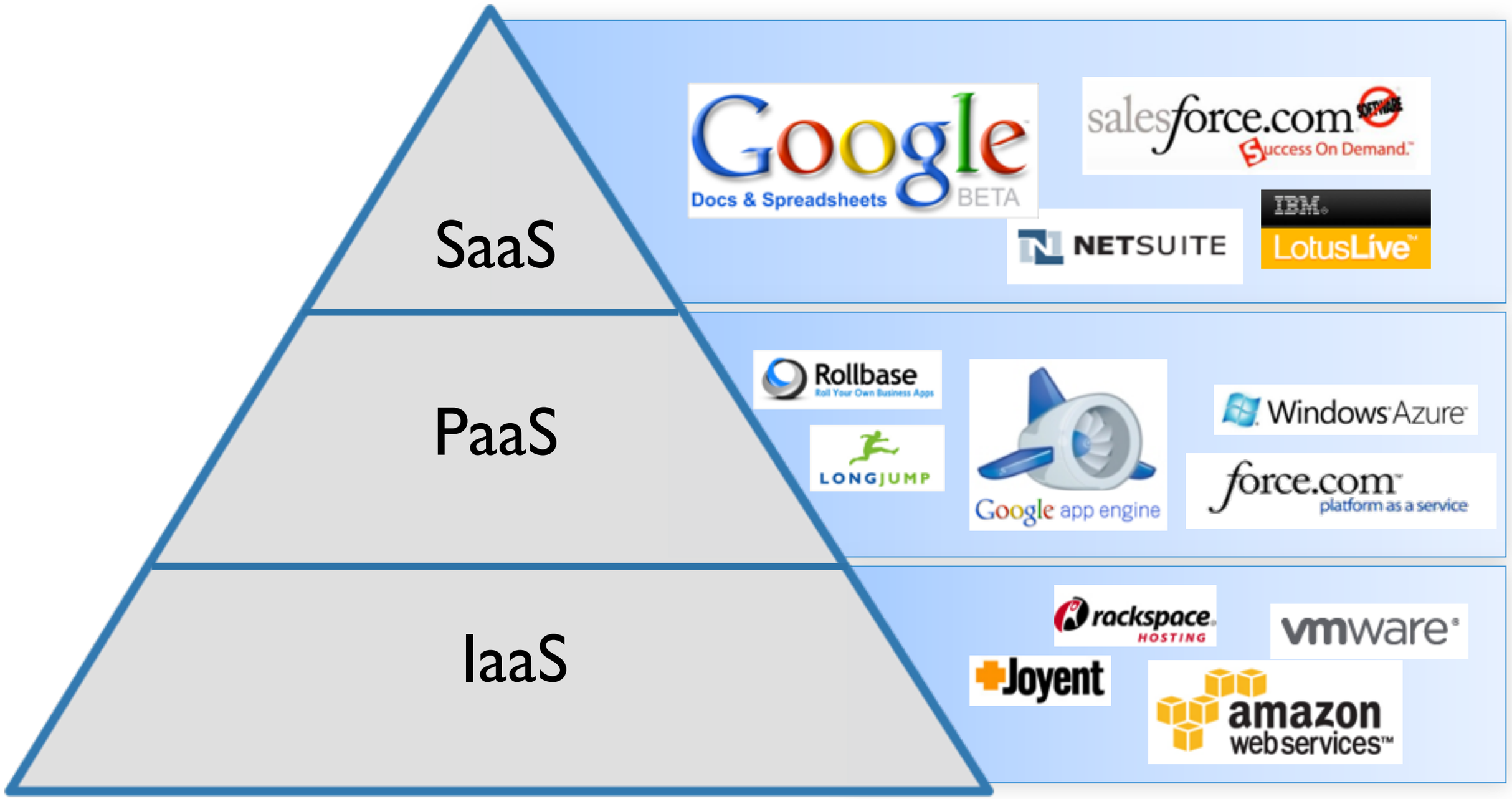
Source: Gartner AADI Summit Dec 2009

Cloud Computing Defined



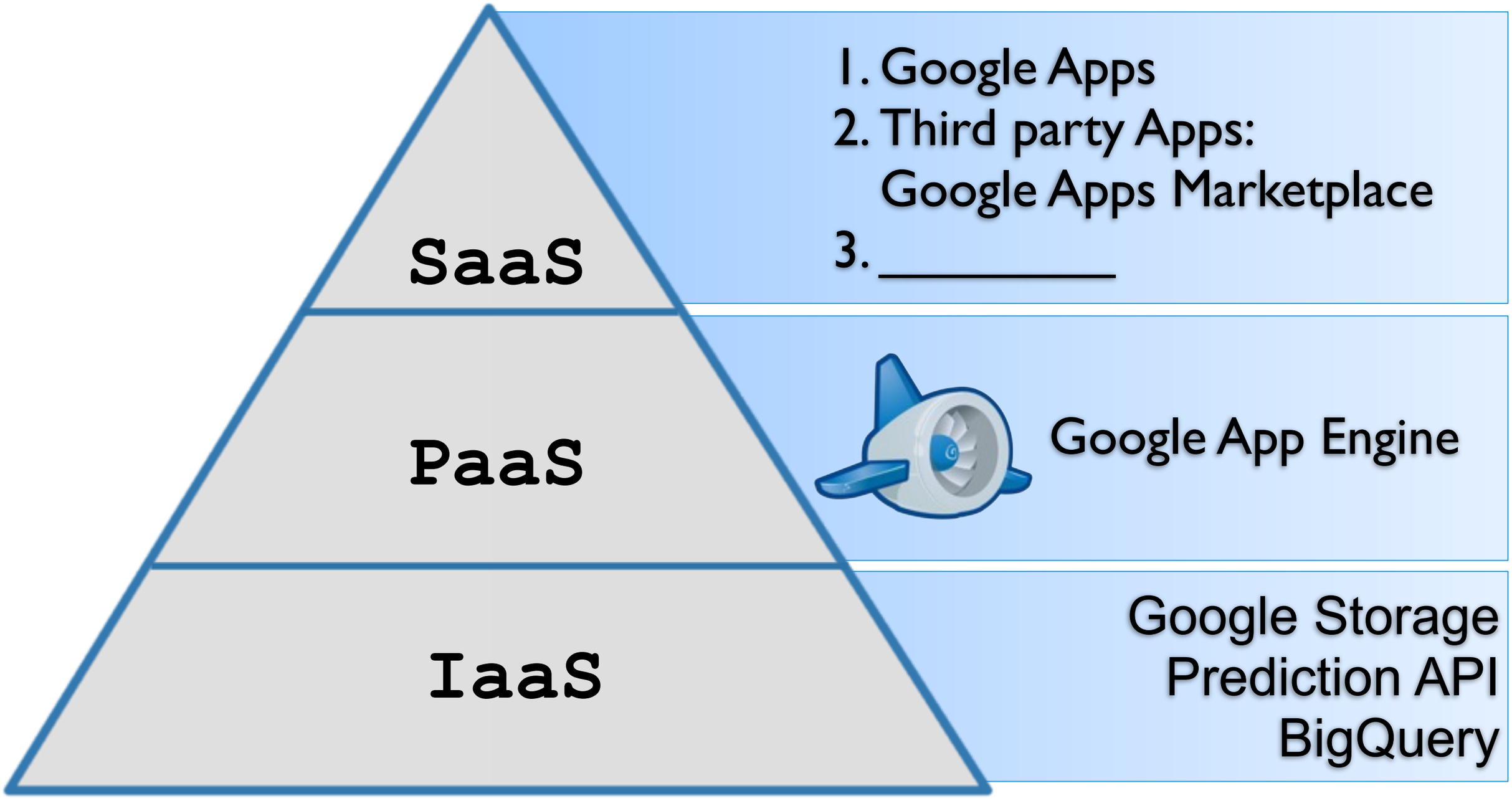
Source: Gartner AADI Summit Dec 2009

Cloud Computing Defined



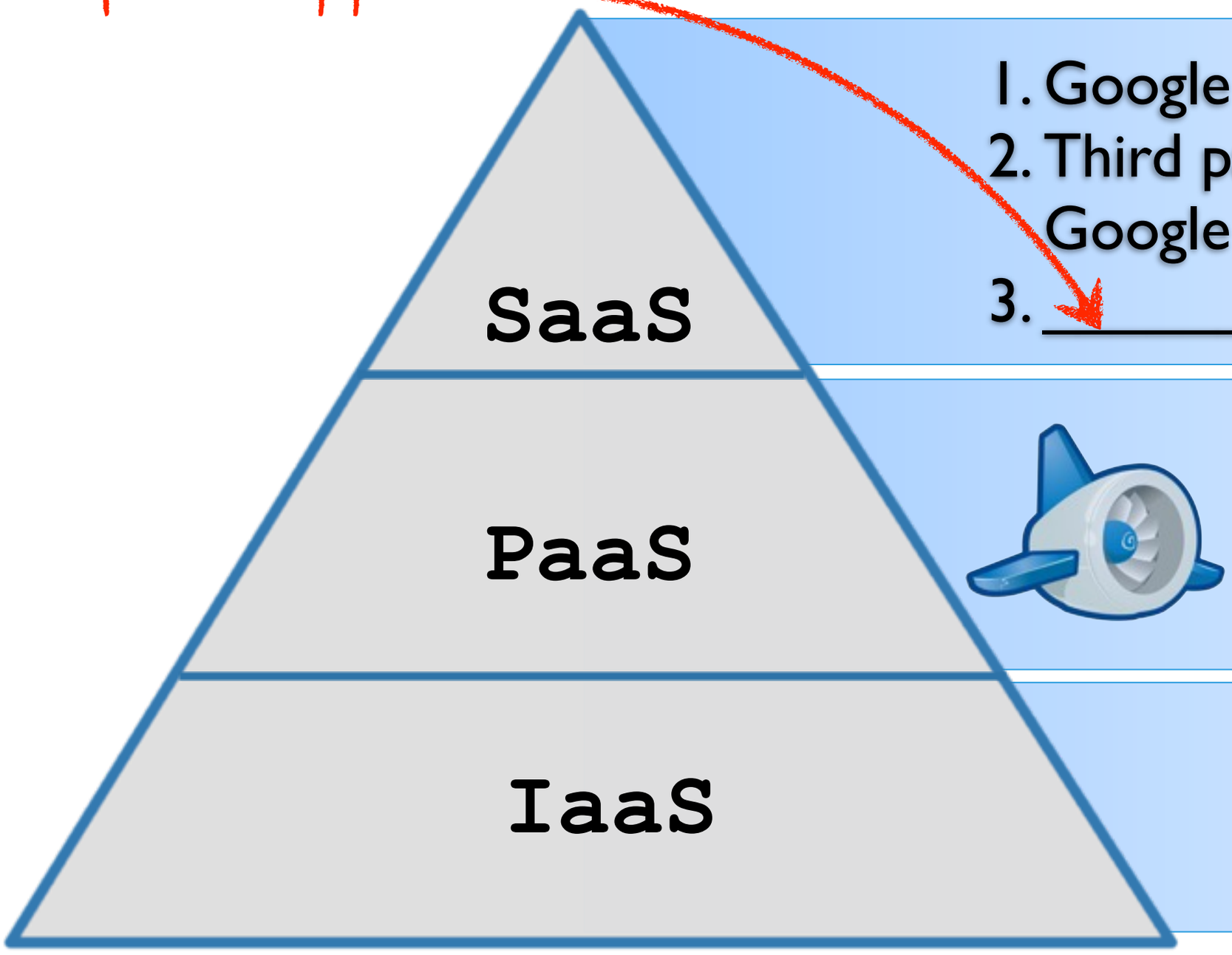
Source: Gartner AADI Summit Dec 2009

Google's Cloud Offerings



Google's Cloud Offerings

Your Apps

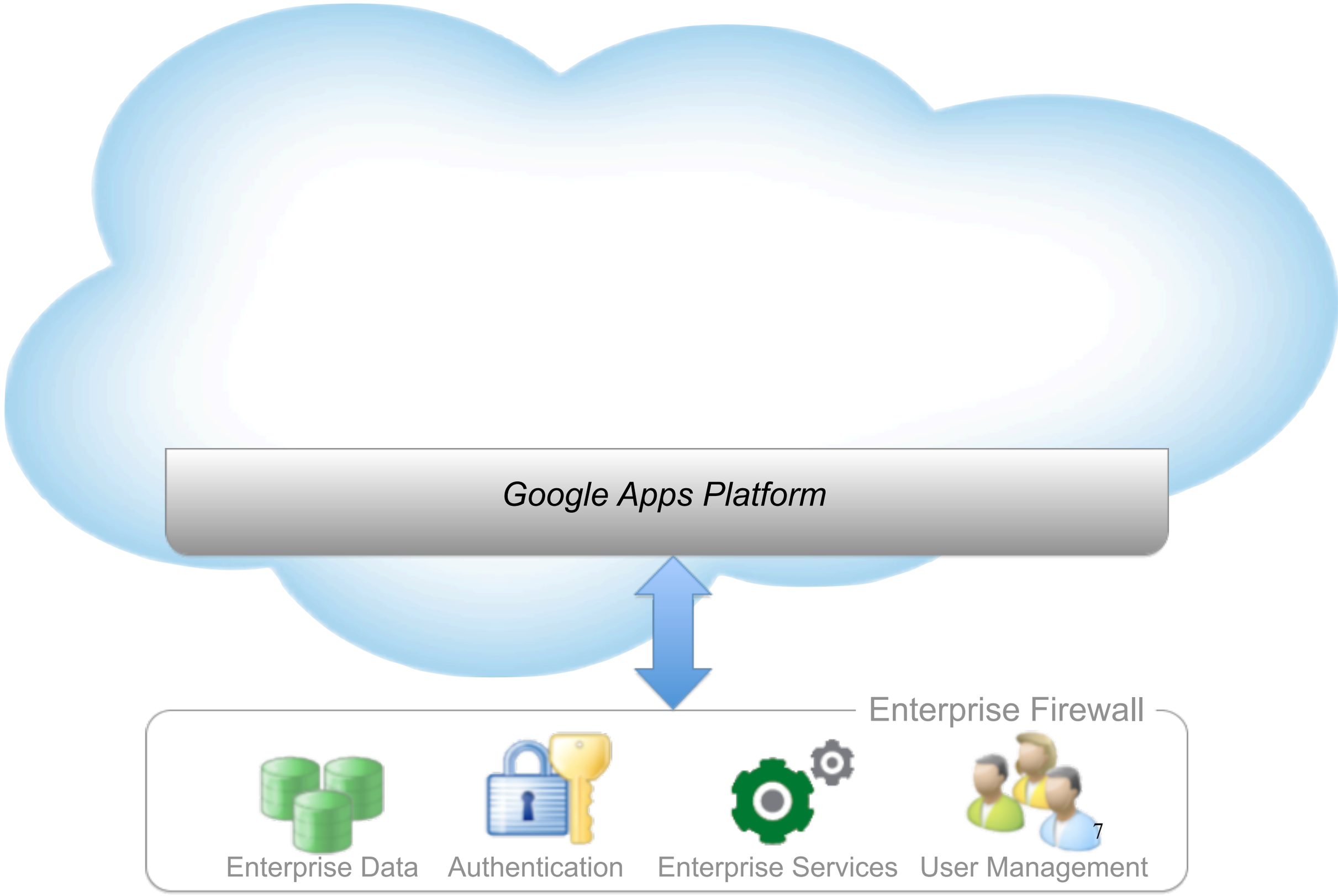


1. Google Apps
2. Third party Apps:
   Google Apps Marketplace
3. _____

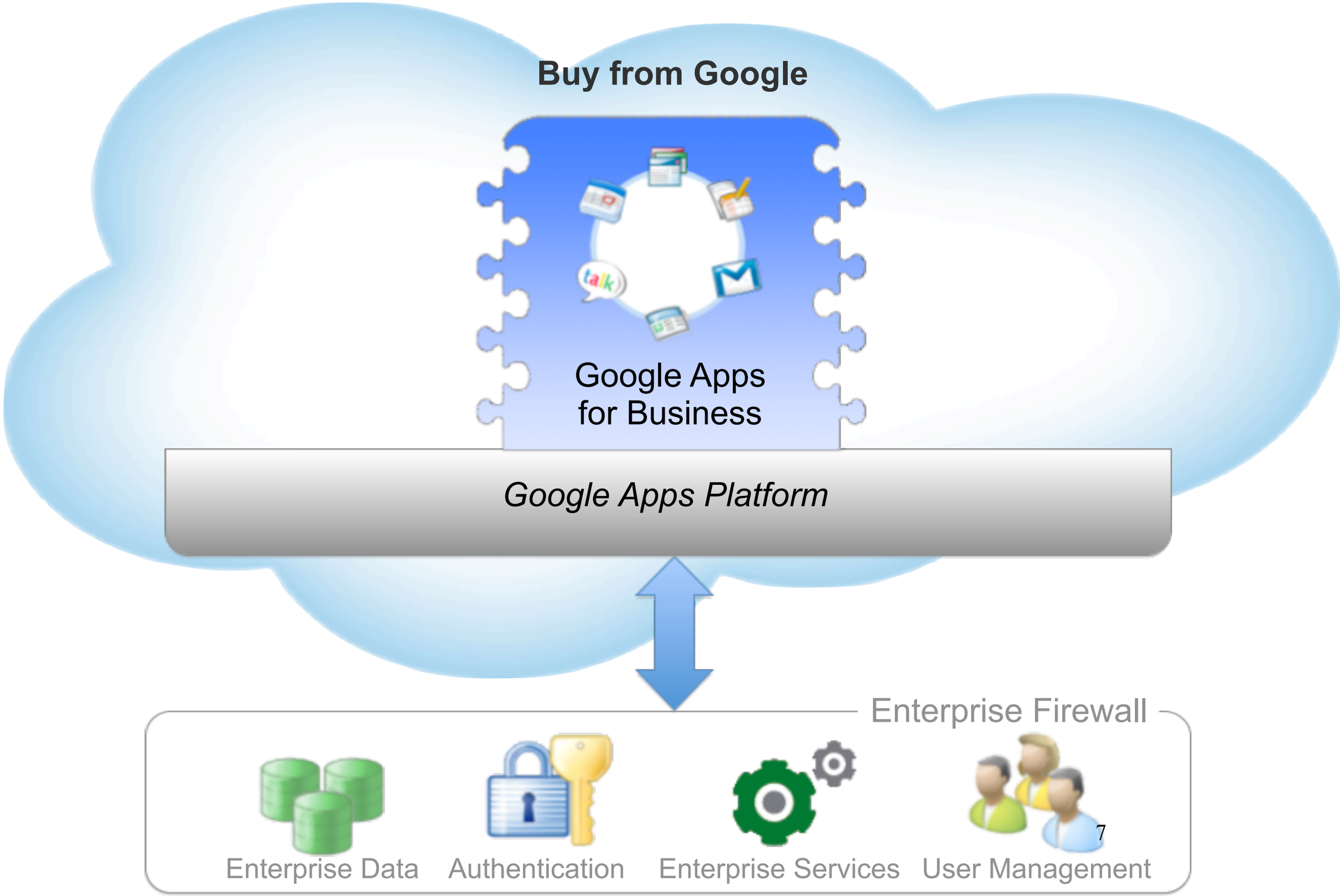
Google App Engine

Google Storage
Prediction API
BigQuery

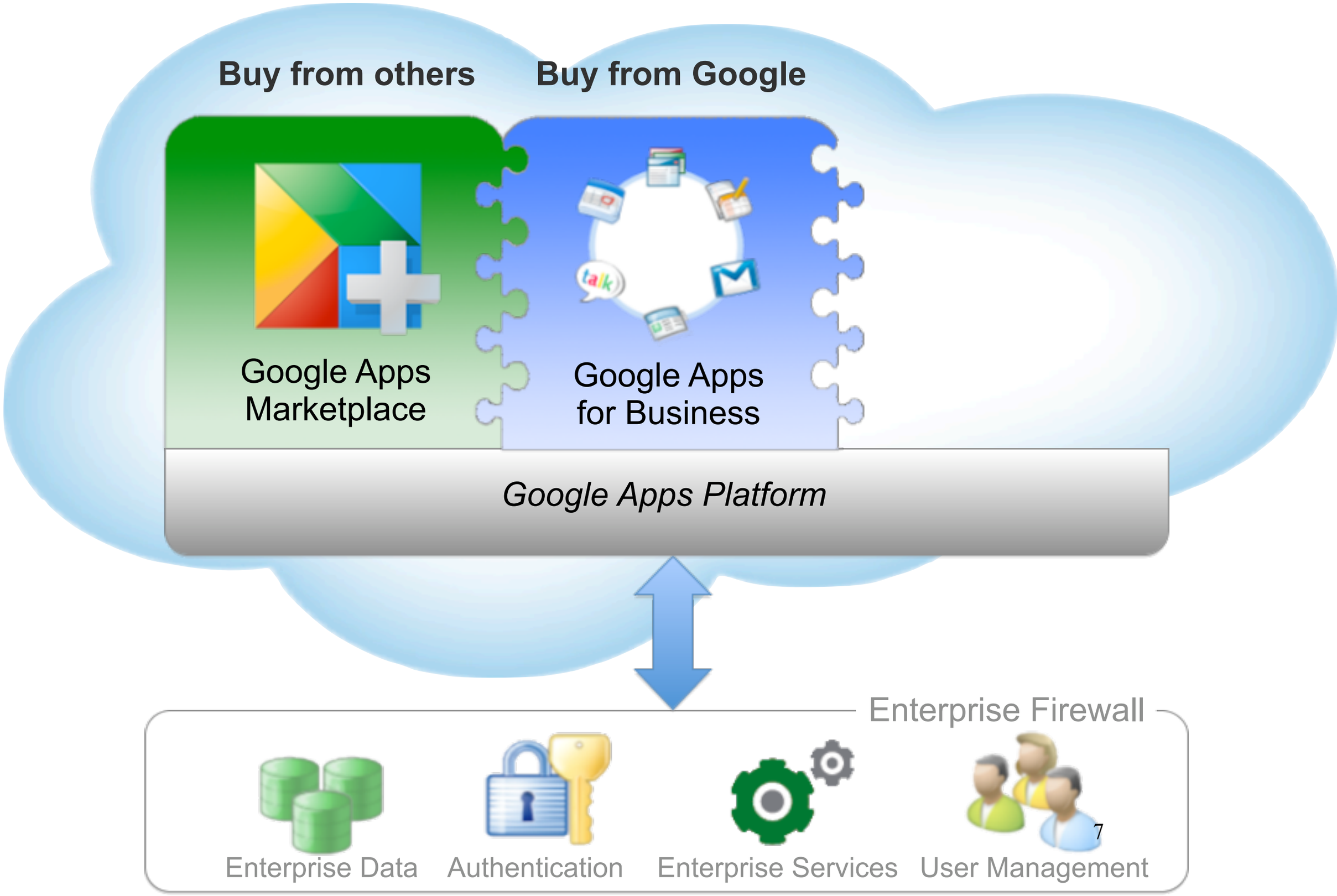
Build and Buy all your enterprise cloud apps...



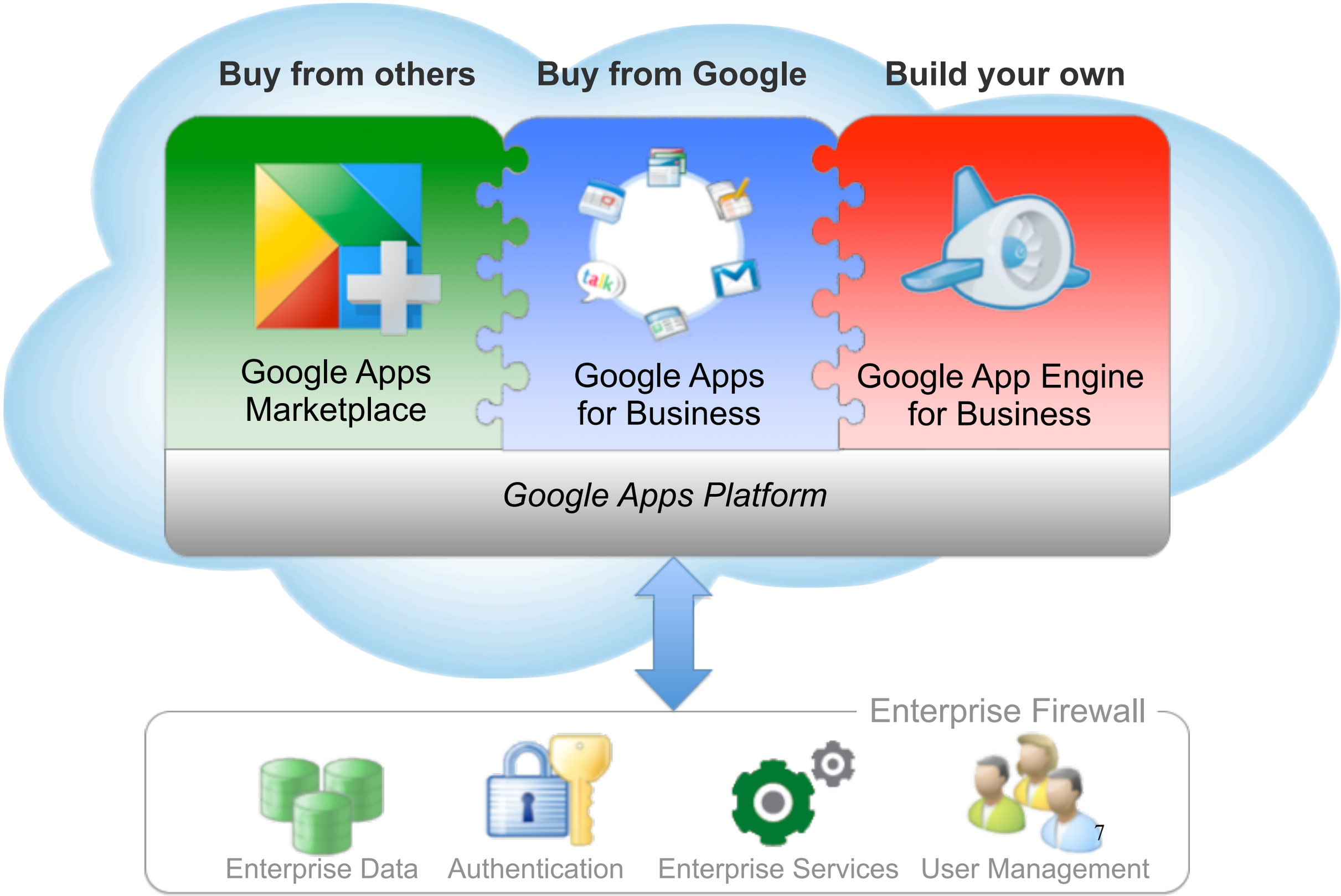
Build and Buy all your enterprise cloud apps...



Build and Buy all your enterprise cloud apps...

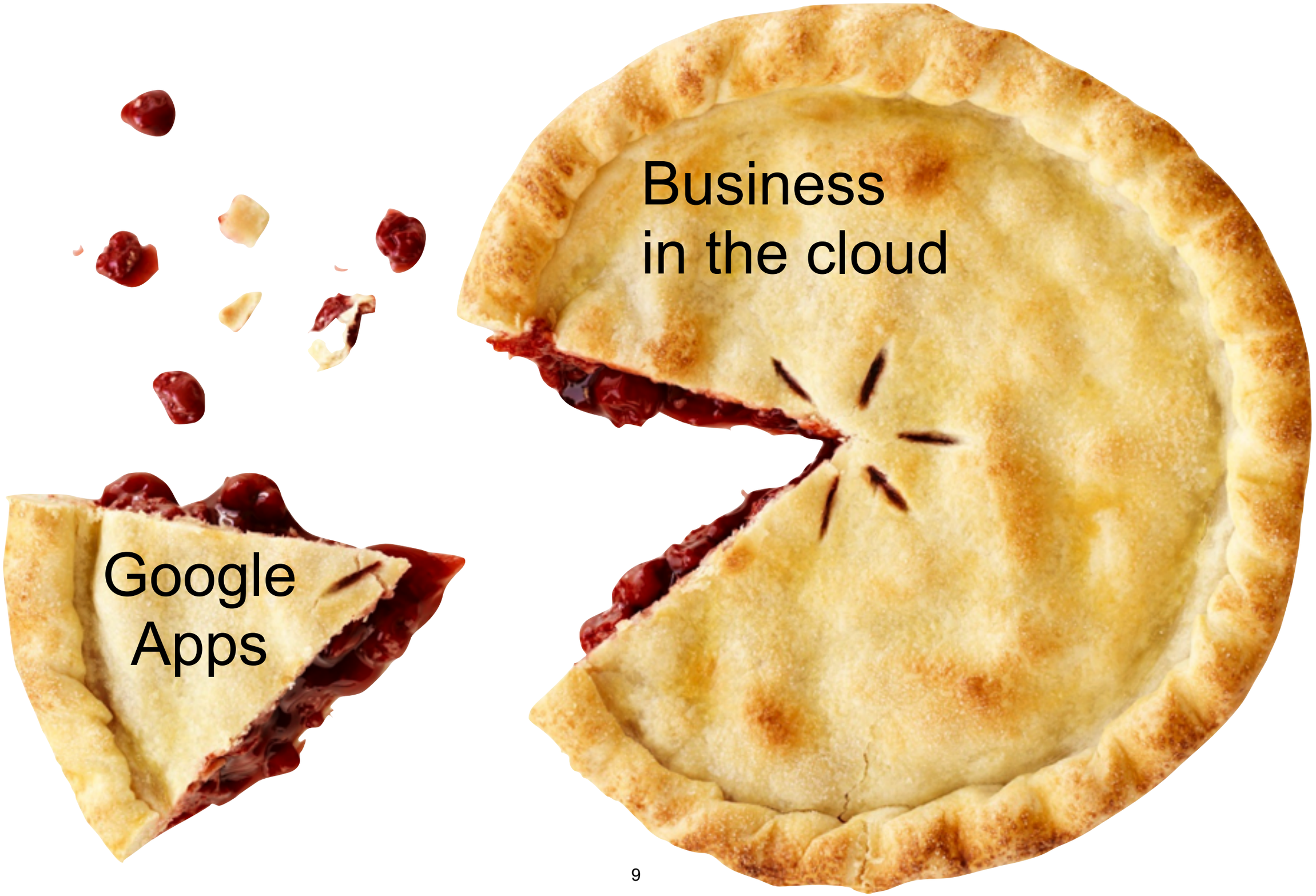


Build and Buy all your enterprise cloud apps...



Google Apps Marketplace

Customers want more Apps



Marketplace Overview

Google Apps is Focused on Messaging & Collaboration

Gives *every* employee powerful messaging and collaboration tools without the usual IT hassle and cost

These customers want more great cloud apps...

- [ ] Accounting & Finance
- [ ] Admin Tools
- [ ] Calendaring / Meetings
- [ ] Customer Management
- [ ] Document Management
- [ ] Productivity
- [ ] Project Management
- [ ] Sales & Marketing
- [ ] Security & Compliance
- [ ] Workflow

Google Apps Marketplace

Steps to sell your app to Google Apps customers

1. Build your app:

- with any tools and hosting provider you want

2. Integrate your app:

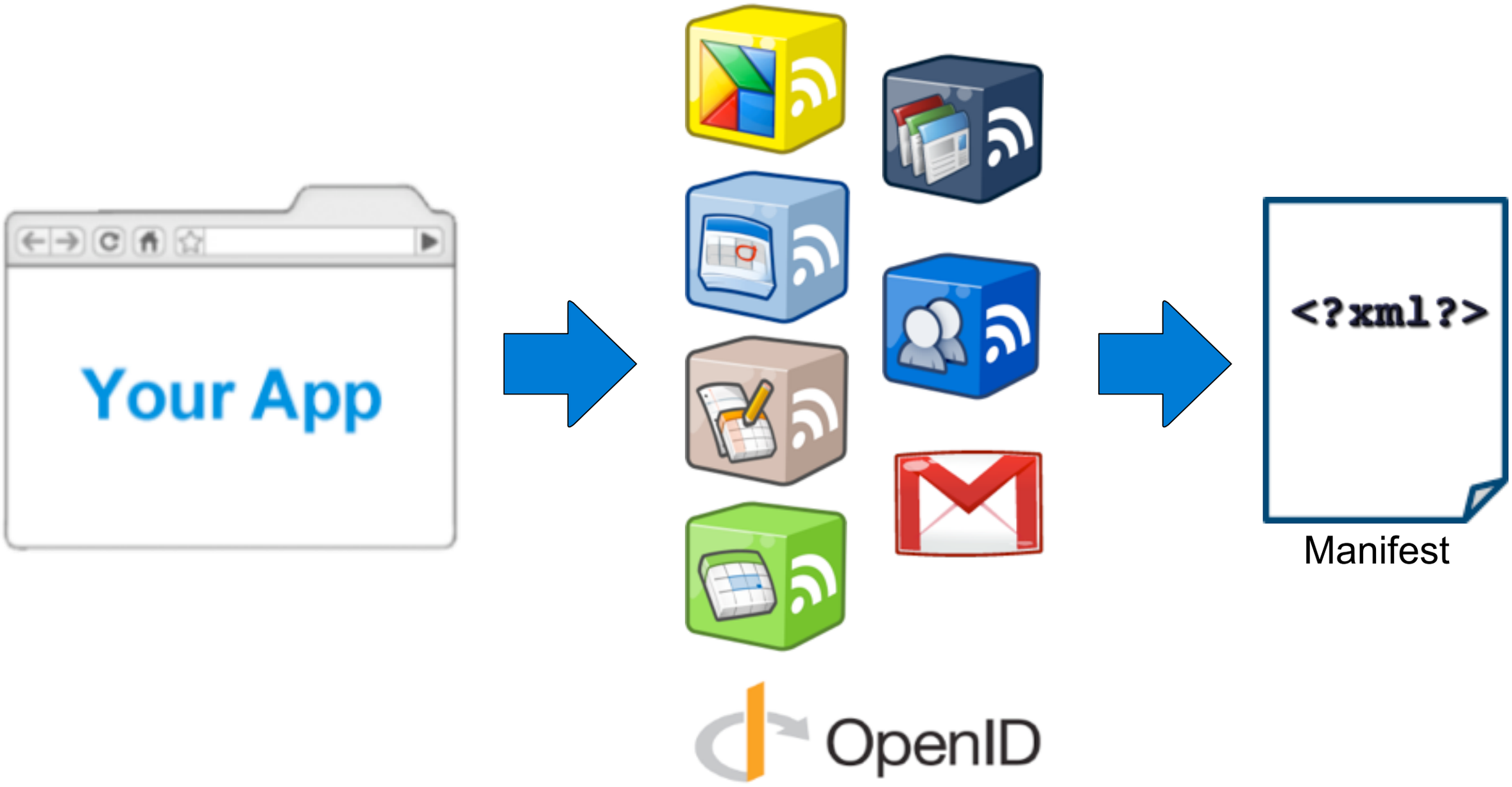
- add Single Sign On using OpenID (required)
- access over a dozen integration points from Calendar, Contacts, Docs, etc. using OAuth (optional)

3. Sell your app:

- to 3M+ businesses & 30M+ users
- through Google Apps resellers
- to wherever Google Apps goes
- **One-time fee of $100, 20% rev share starting 2H'10**

Everything in the Cloud

Integrating with Google Apps & the Marketplace



Demo Google Apps Marketplace

How it's done - Single Sign On

Single Sign On

OpenID with Google Apps

Sounds complicated, but not hard in practice!

Single Sign On

OpenID Libraries

Language	Libraries
Java	OpenID4Java, Step2
.NET	DotNetOpenAuth
PHP	php-openid, php-openid-apps-discovery
Ruby	ruby-openid, ruby-openid-apps-discovery
Any	RPX, Ping Identity

Single Sign On - Code

Step 1 of 3 - Initialize library

```php
function getOpenIDConsumer() {
  // Initialize client, storing associations
  // in memcache
  $store = new Auth_OpenID_MemcachedStore($MEMCACHE);
  $consumer = new Auth_OpenID_Consumer($store);

  // Enable Google Apps support
  GApps_OpenID_EnableDiscovery($consumer);
  return $consumer;
}
```

Single Sign On - Code

Step 2 of 3 - Make the request

```
// Create an auth request to the user's domain
$auth_request = getOpenIDConsumer()->begin($domain);

// Request email address & name during login
$ax = new Auth_OpenID_AX_FetchRequest;
$attr = Auth_OpenID_AX_AttrInfo::make(
    $AX_SCHEMA_EMAIL, 1, 1, 'email');
$attr = Auth_OpenID_AX_AttrInfo::make(
    $AX_SCHEMA_FIRSTNAME, 1, 1, 'first');
$attr = Auth_OpenID_AX_AttrInfo::make(
    $AX_SCHEMA_LASTNAME, 1, 1, 'last');
$auth_request->addExtension($ax);

// Render Javascript/form to post request
$form_html = $auth_request->htmlMarkup($REALM,
    $RETURN_TO, false,
    array('id' => 'openid_message'));
print $form_html;
```

Single Sign On - Code

Step 3 of 3 - Handle response

```php
// Parse the response from identity provider
$response = getOpenIDConsumer()->complete($RETURN_TO);
if ($response->status == Auth_OpenID_SUCCESS) {
  // Extract data from response
  $openid = $response->getDisplayIdentifier();
  $ax = new Auth_OpenID_AX_FetchResponse();
  $ax_resp = $ax->fromSuccessResponse($response);
  $email = $ax_resp->data[$AX_SCHEMA_EMAIL][0];
  $firstName = $ax_resp->data[$AX_SCHEMA_FIRSTNAME][0];
  $lastName = $ax_resp->data[$AX_SCHEMA_LASTTNAME][0];
  // Map to user in DB
  $user = fetch_user($openid, $email, $firstName,
     $lastName);
}
```

Single Sign On - Code

Manifest

<?xml?>

```
<ApplicationManifest>
  <Name>Sassy Voice</Name>
  <Description>Voice Mail &amp;amp; Messaging</Description>
  <Admin>
    <Link rel="setup">
http://voice.saasyapp.com/setup.php?domain=${DOMAIN_NAME}
    </Link>
  </Admin>
  <Extension id="navlink" type="link">
    <Name>SaasyVoice</Name>
    <Url>
http://voice.saasyapp.com/login.php?domain=${DOMAIN_NAME}
    </Url>
  </Extension>
  <Extension id="realm" type="openIdRealm">
    <Url>http://voice.saasyapp.com/</Url>
  </Extension>
</ApplicationManifest>
```

Single Sign On

What we learned about users

Claimed ID: http://example.com/openid?id=12345

Email: bob@example.com*
First Name: Bob
Last Name: Dobbs

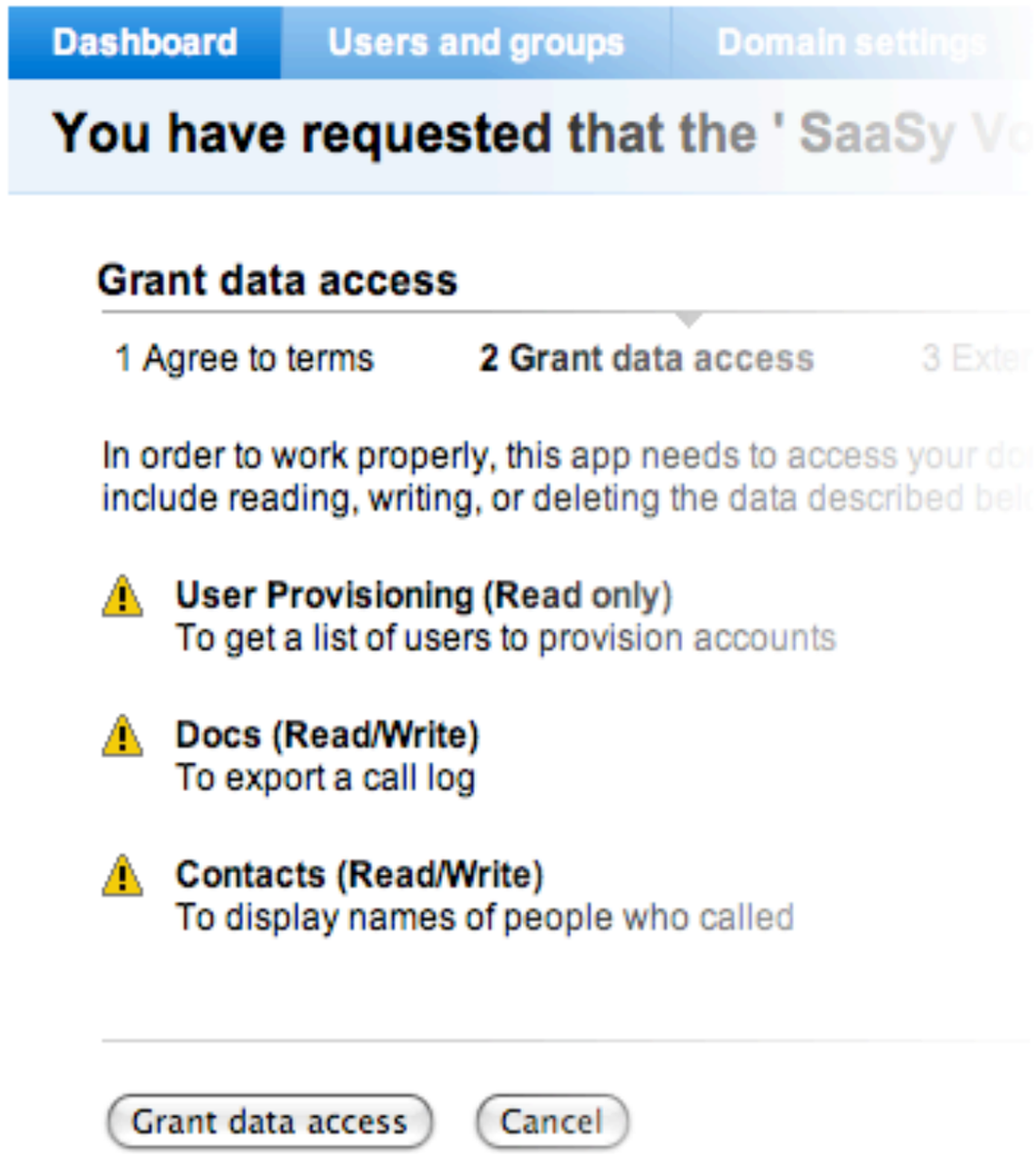
* Be sure to confirm or white-list trusted providers

How it's done - Data Access

2-Legged OAuth

Delegating data access

```
<ApplicationManifest>
  ...
  <Scope id="userFeed">
    <Url>https://apps-apis.google...
    <Reason>To get a list of user...
  </Scope>
  <Scope id="contactsFeed">
    <Url>https://www.google.com...
    <Reason>To display names of...
  </Scope>
  <Scope id="docsFeed">
    <Url>https://docs.google.com...
    <Reason>To export a call log...
  </Scope>
</ApplicationManifest>
```



Data Access - Code

Step 1 of 2 - Initialize the client

```
function getOauthClient() {
  $options = array(
    'consumerKey' => $CONSUMER_KEY,
    'consumerSecret' => $CONSUMER_SECRET,
    'signatureMethod' => 'HMAC-SHA1',
    'requestScheme' =>Zend_Oauth::REQUEST_SCHEME_HEADER,
    'version' => '1.0');
  $consumer = new Zend_Oauth_Consumer($options);
  $token = new Zend_Oauth_Token_Access();
  $httpClient = $token->getHttpClient($options);
  return $httpClient;
}
```

Data Access - Code

Step 2 of 2 - Fetching Users with Provisioning API

```
// Initialize client
$userClient = new Zend_Gdata_Gapps(getOauthClient());

// Query feed for current user's domain
$userQuery = new Zend_Gdata_Gapps_UserQuery(
  getCurrentUserDomain());
$usersFeed = $userClient->getUserFeed($userQuery);

// Extract data from user feed
$users = array();
foreach ($usersFeed as $userEntry) {
  $login = $userEntry->getLogin();
  $name = $userEntry->getName();
  $users[] = array(
   'username' => $login->getUsername(),
   'firstName' => $name->getGivenName(),
   'lastName' => $name->getFamilyName(),
   'admin' => $login->getAdmin());
}
```

Integration Recap

With not a lot of code we added:

- **Single Sign On** with OpenID
- **Quicker setup** with Attribute Exchange, Provisioning API
- **Integrated user data** with Contacts API
- **Uploading spreadsheets** with Docs API

Gadgets - your real estate in Google Apps

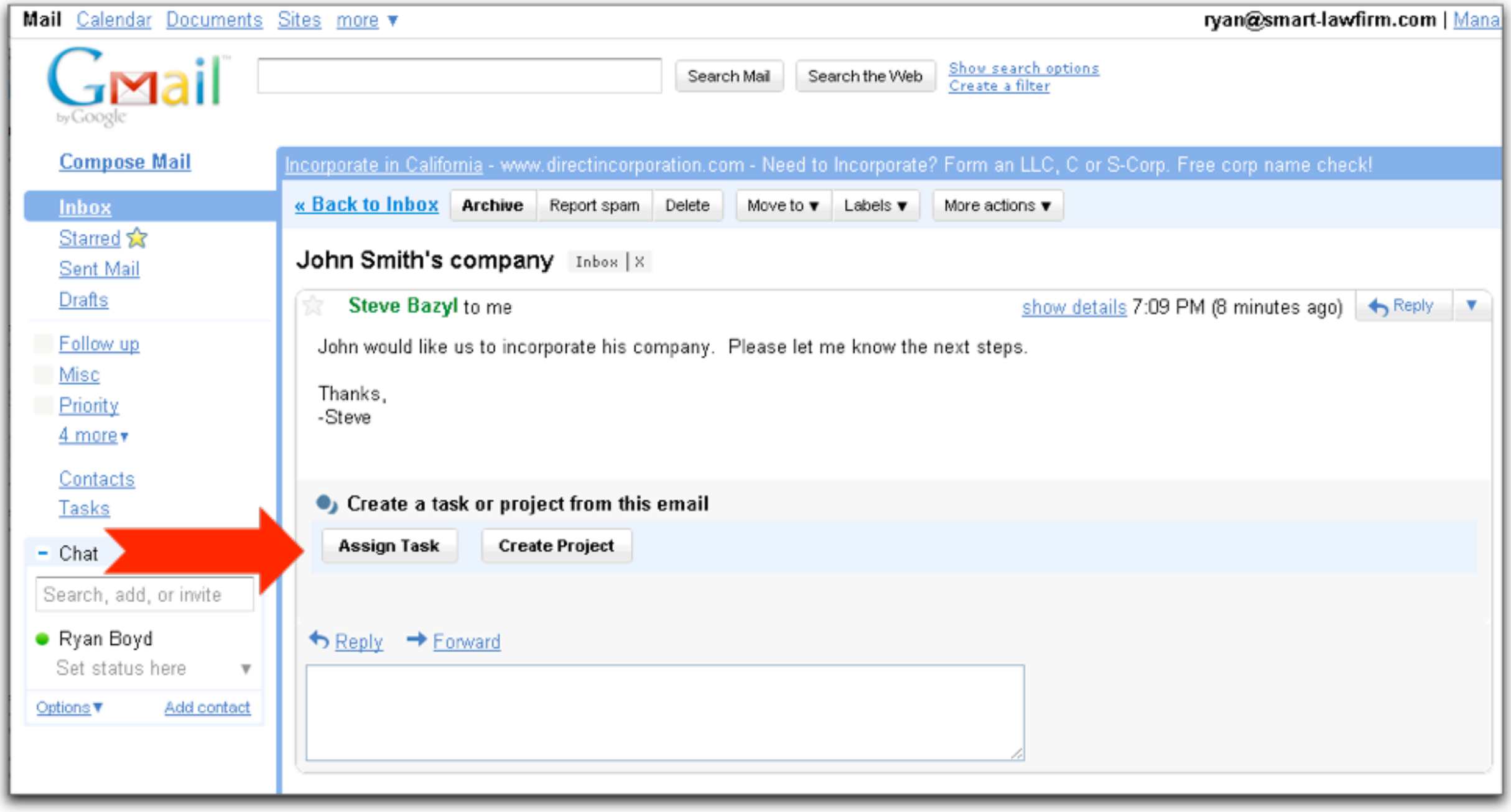
Gadgets

- Many types of gadgets
  - Gmail sidebar
  - Calendar sidebar
  - Sites
  - Spreadsheets
- Two new types of gadgets available this week!
  - Gmail contextual!
  - Wave!
- All use the OpenSocial gadgets specification

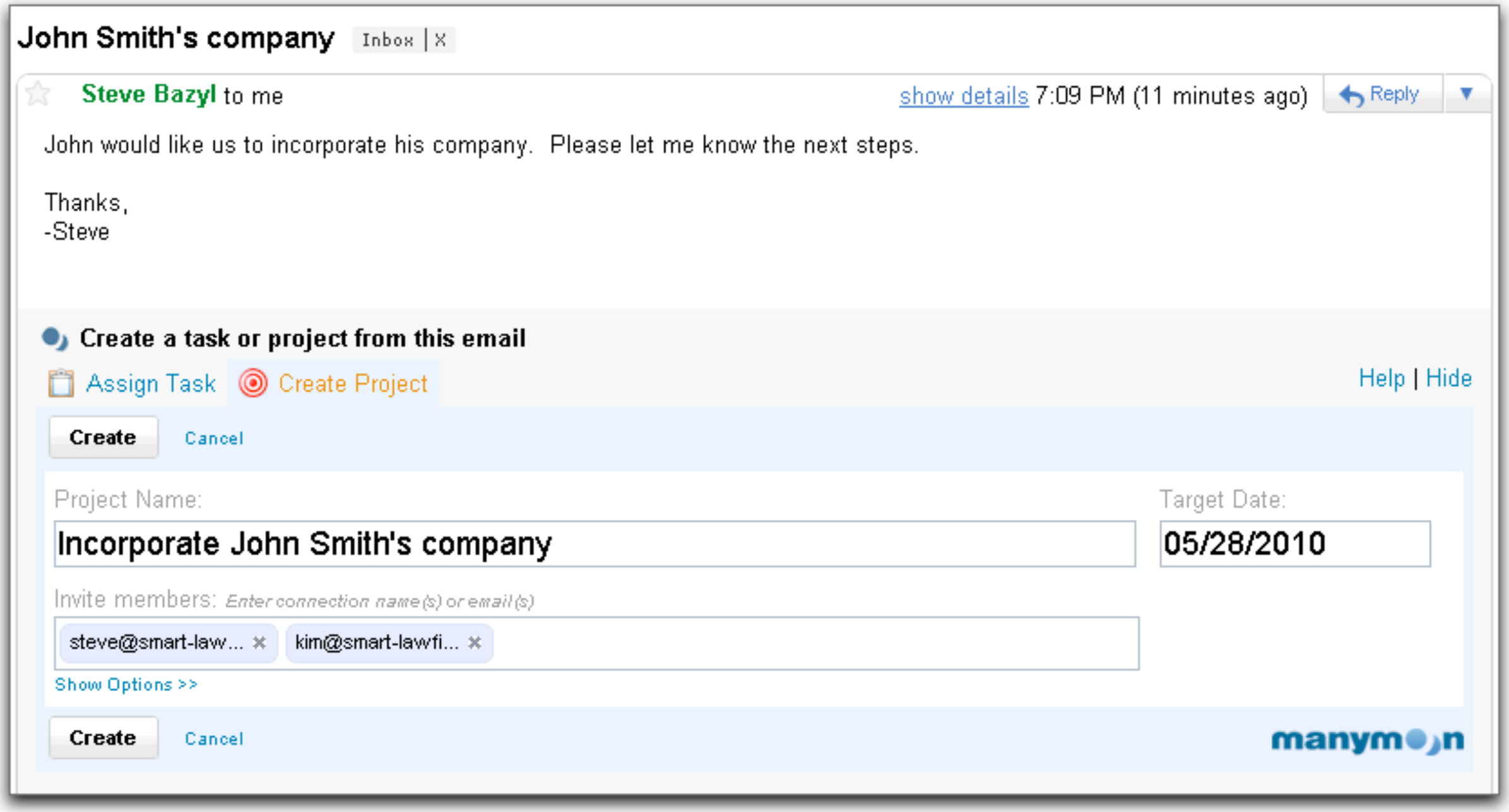
Gmail contextual gadgets

- Detect e-mail content via regular expressions
- Display and collect actionable business information directly in Gmail, below each message

Gmail contextual gadgets



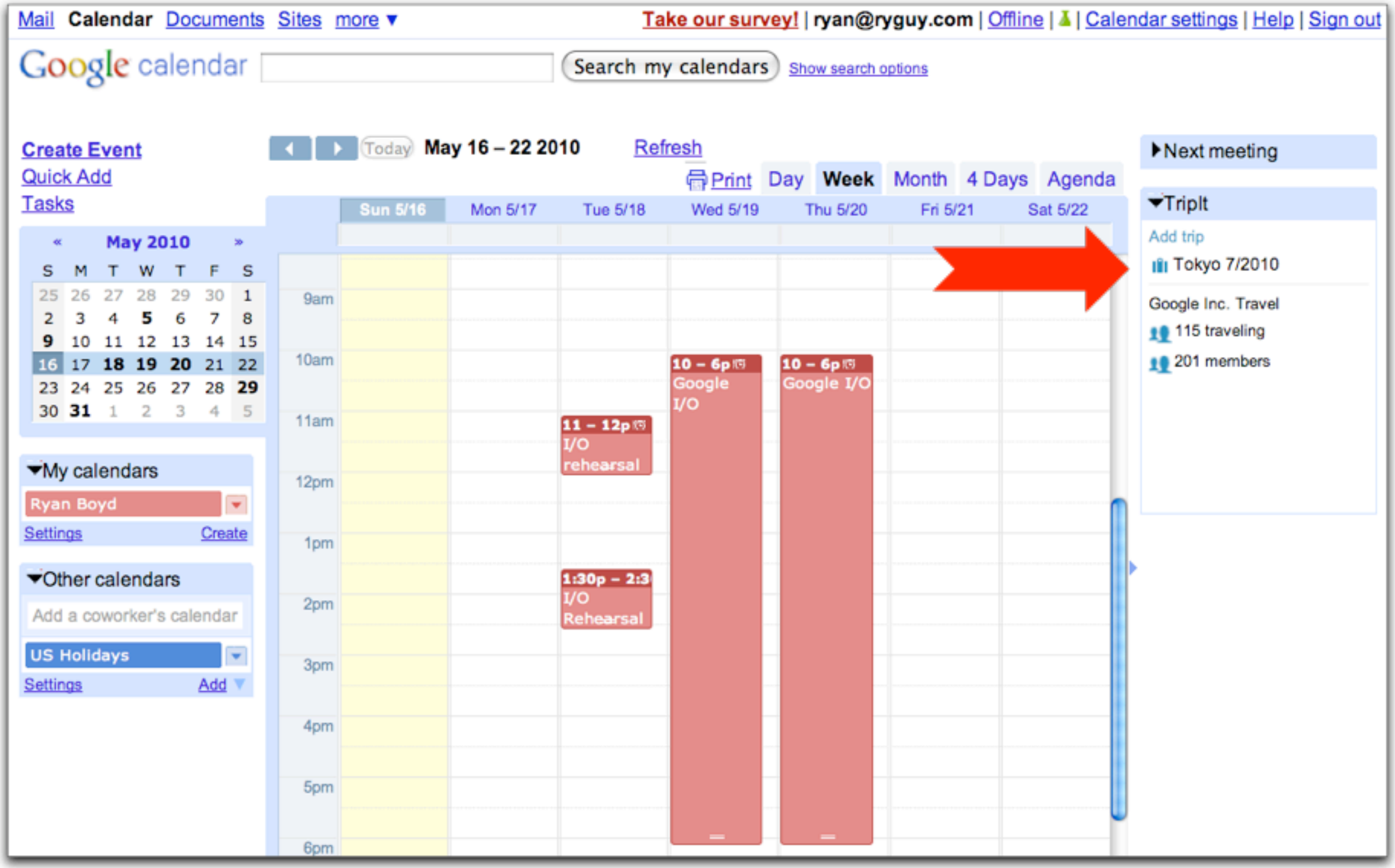
Gmail contextual gadgets



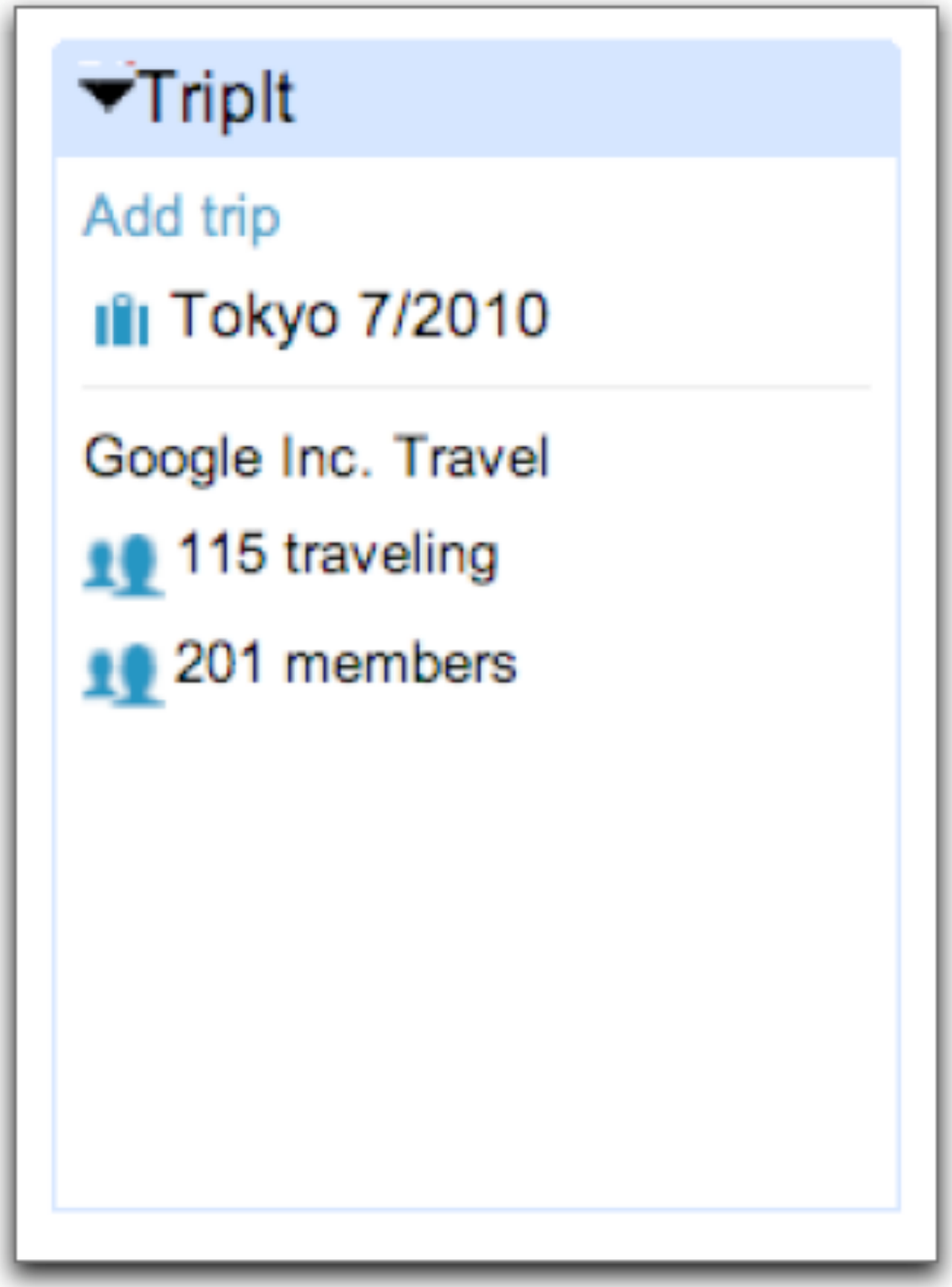
Calendar sidebar gadgets

- Can detect:
  - currently-displayed date range
  - currently-selected event
    - title
    - attendees
    - date/time

Calendar sidebar gadgets



Calendar sidebar gadgets



Sell

Complete Manifest

```
<ApplicationManifest xmlns="http://schemas.google.com/
ApplicationManifest/2009">
<Name>SaaSy Voice</Name>
  <Description>SaaSy Voice</Description>
  <Support>
    <Link rel="setup" href="https://.../setup.php?domain=${DOMAIN_NAME}"/
>
    <Link rel="manage" href="https://.../config.php?domain=$
{DOMAIN_NAME}"/>
    <Link rel="deletion-policy" href="http://.../deletion-policy.php"/>
    <Link rel="support" href="http://.../support.php"/>
  </Support>
```

Complete Manifest

(continued)

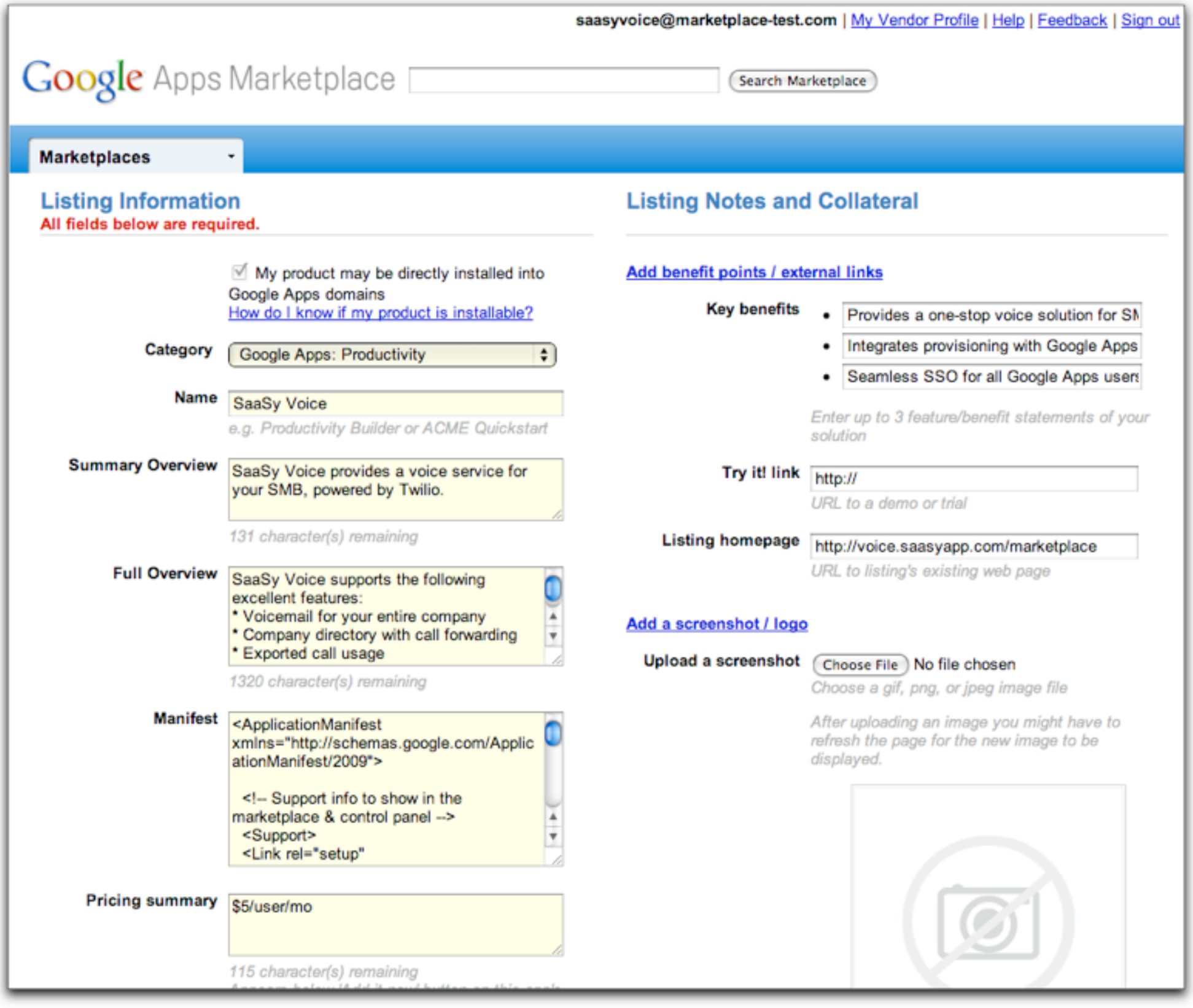
```
<!-- Link in universal navigation -->
<Extension id="oneBarLink" type="link">
  <Name>SaaSy Voice</Name>
  <Url>http://.../index.php?domain=${DOMAIN_NAME}</Url>
  <Scope ref="provisioningFeed"/>
  <Scope ref="contactsFeed"/>
  <Scope ref="docsFeed"/>
</Extension>
<!-- Declare our openid.realm -->
<Extension id="realm" type="openIdRealm">
  <Url>http://voice.saasyapp.com/marketplace/</Url>
</Extension>
```

Complete Manifest

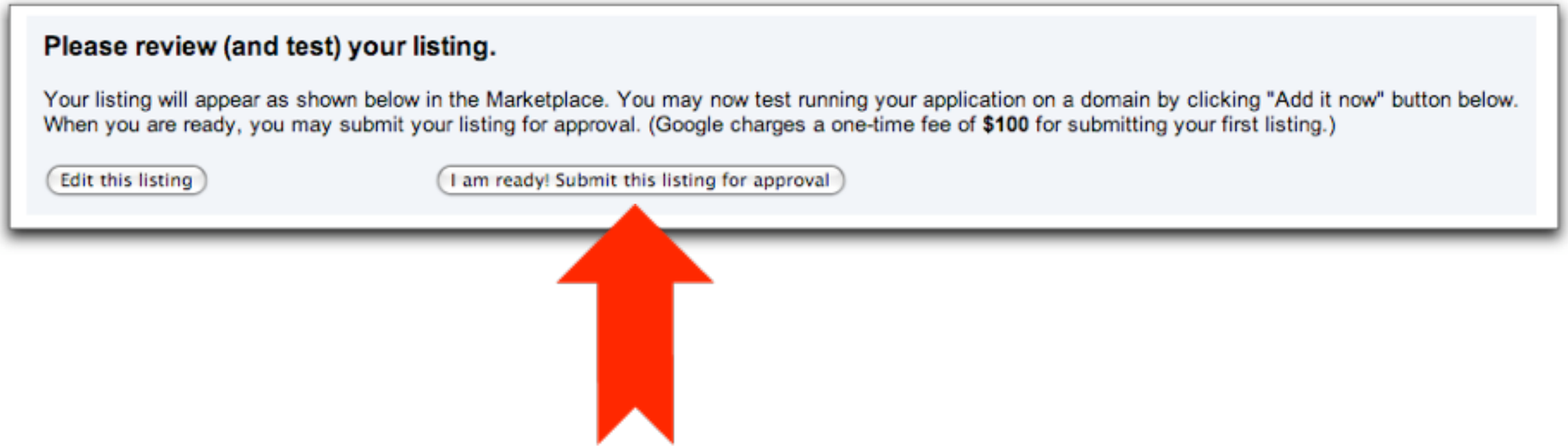
(continued)

```
<Scope id="provisioningFeed">
  <Url>https://apps-apis.google.com/a/feeds/user/#readonly</Url>
  <Reason>To get a list of users to provision accounts</Reason>
</Scope>
<Scope id="contactsFeed">
  <Url>https://www.google.com/m8/feeds/</Url>
  <Reason>To display names of people who called</Reason>
</Scope>
<Scope id="docsFeed">
  <Url>https://docs.google.com/feeds/</Url>
  <Reason>To export a call log</Reason>
 </Scope>
</ApplicationManifest>
```

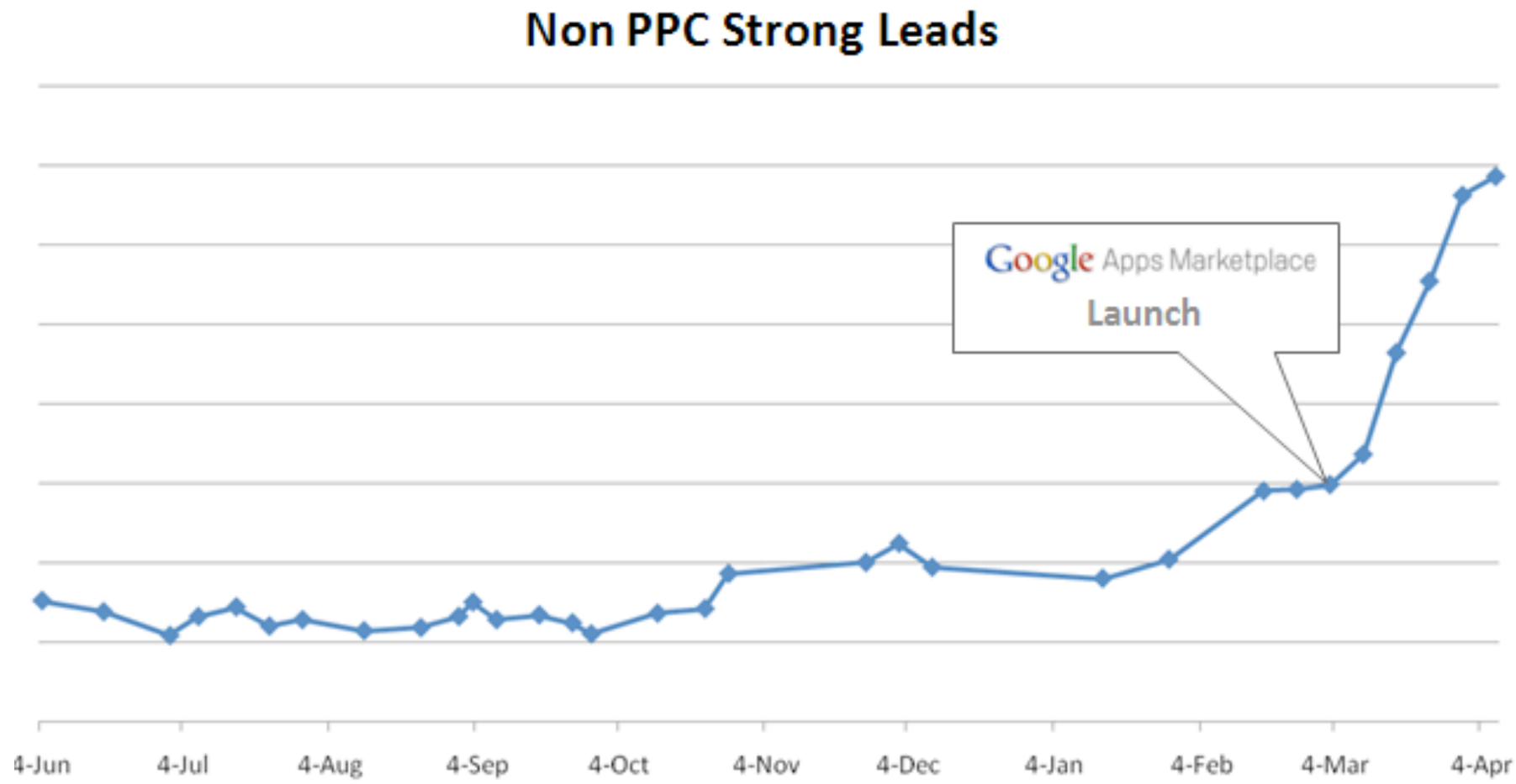
Listing in the Google Apps Marketplace



Listing in the Google Apps Marketplace



Reach 30 million Google Apps Users



"We've seen a very meaningful increase in high quality, non-paid lead flow directly attributable to Google Apps customers...Our customers cite Smartsheet's tight integration with Google's Data APIs as a key factor in their decision to purchase."

Brent Frei, Smartsheet

Learn More

Resources

- **Business and Marketing**
  - http://developer.googleapps.com/marketplace
- **Technical and Code**
  - http://code.google.com/googleapps/marketplace
- **Shopping!**
  - http://www.google.com/appsmarketplace
- **Don't have Google Apps?**
  - http://www.google.com/a

Google App Engine

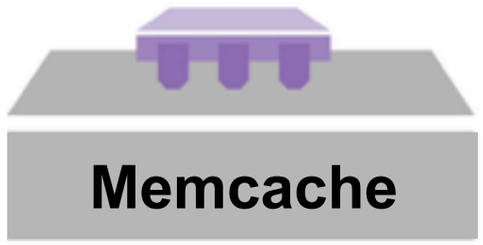
Google App Engine

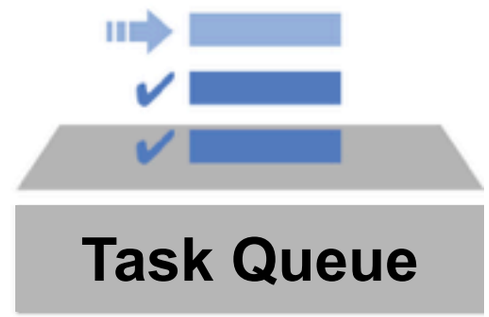
- Easy to **build**
- Easy to **maintain**
- Easy to **scale**

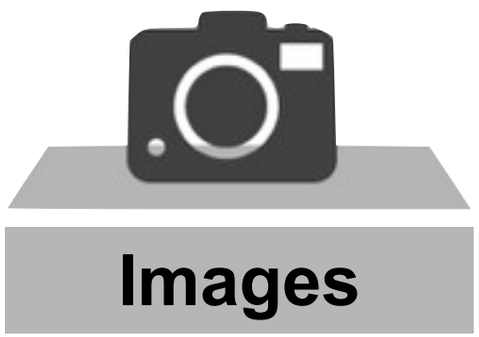
Cloud development in a box

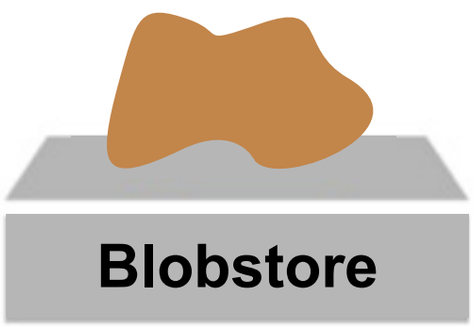
- SDK & “The Cloud”
- Hardware
- Networking
- Operating system
- Application runtime
  - Java, Python
- Static file serving
- Services
- Fault tolerance
- Load balancing

App Engine Services









Always free to get started

~5M pageviews/month

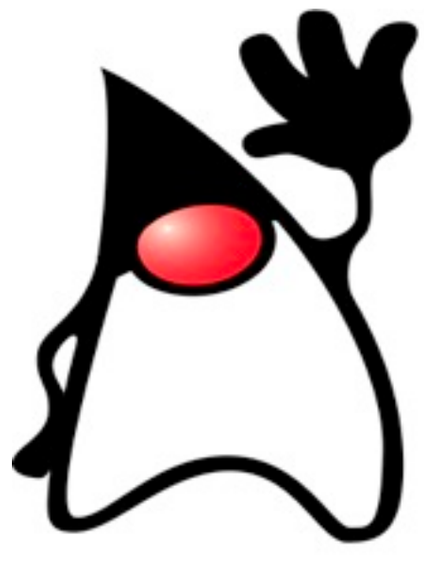
- 6.5 CPU hrs/day
- 1 GB storage
- 650K URL Fetch calls/day
- 2,000 recipients emailed
- 1 GB/day bandwidth
- 100,000 tasks enqueued
- 650K XMPP messages/day

Purchase additional resources *

*** free monthly quota of ~5 million page views still in full effect**

Extended Language support through JVM

- Java
- Scala
- JRuby (Ruby)
- Groovy (Gaelyk Framework)
- Quercus (PHP)
- Rhino (JavaScript)
- Jython (Python)

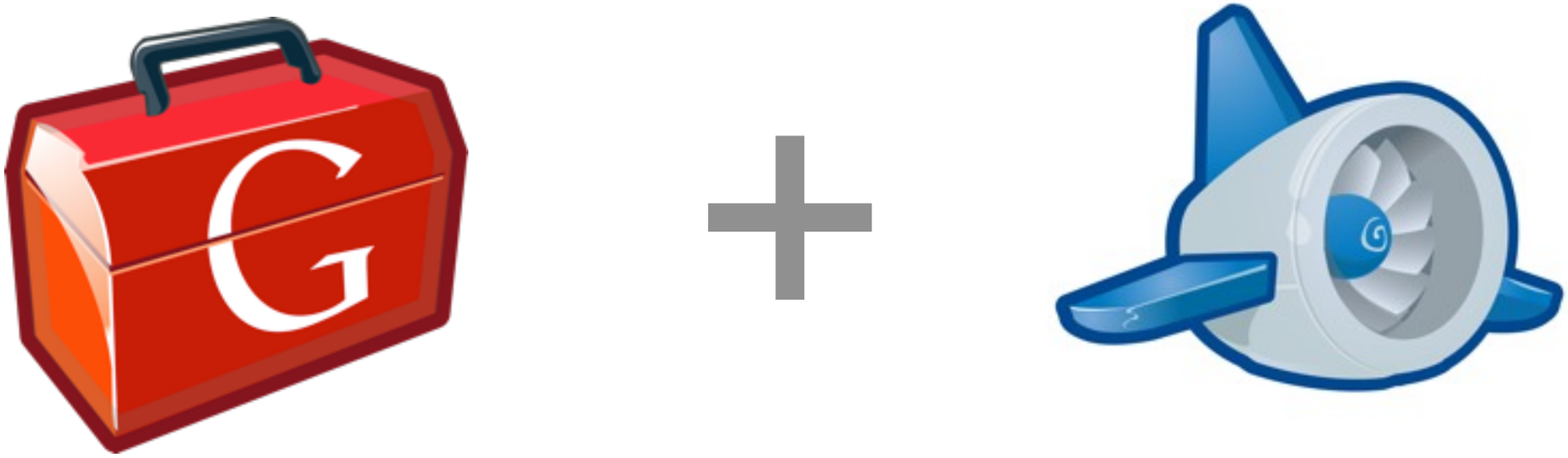


Duke, the Java mascot

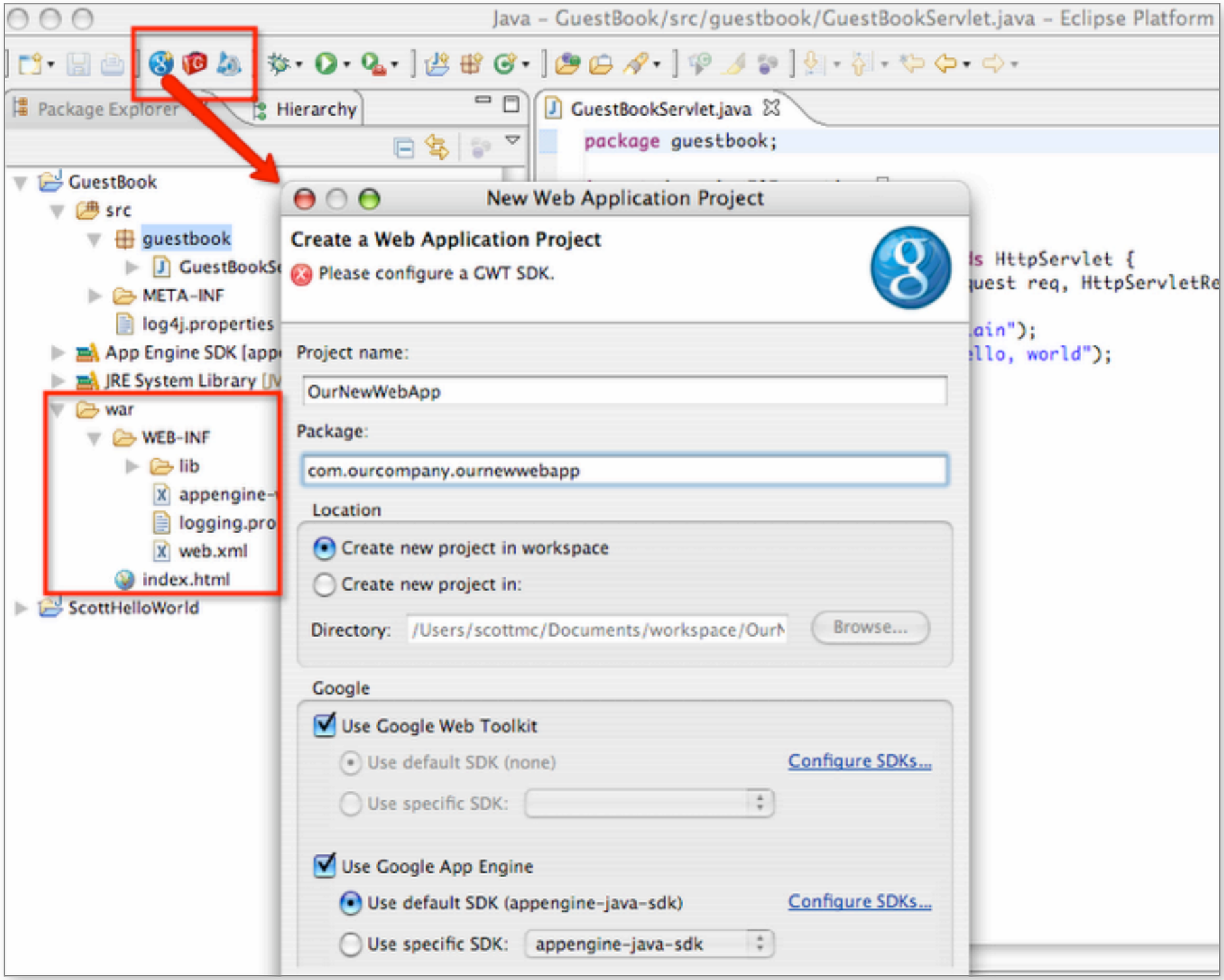
Google

Development Tools for Java App Engine

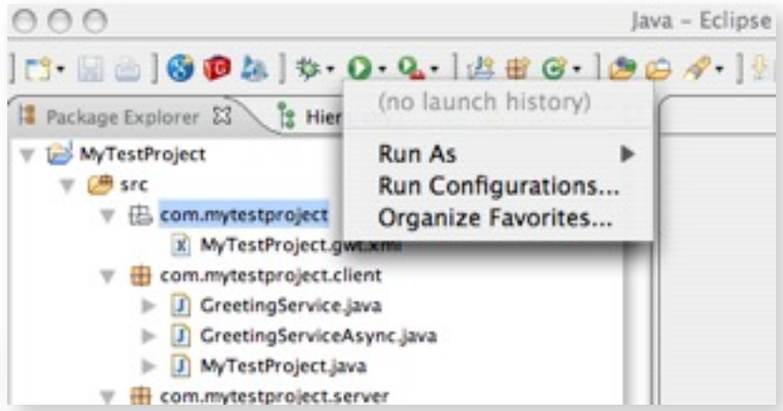
Google's Complete Java Development Stack

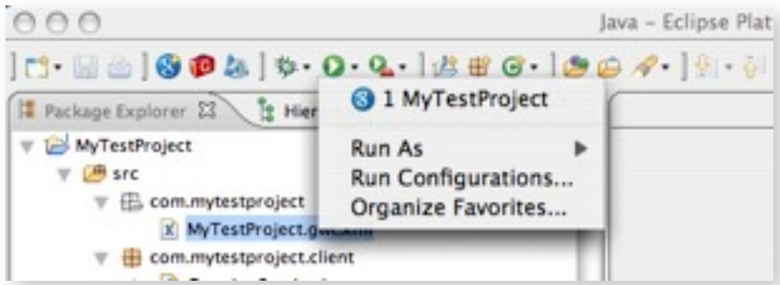


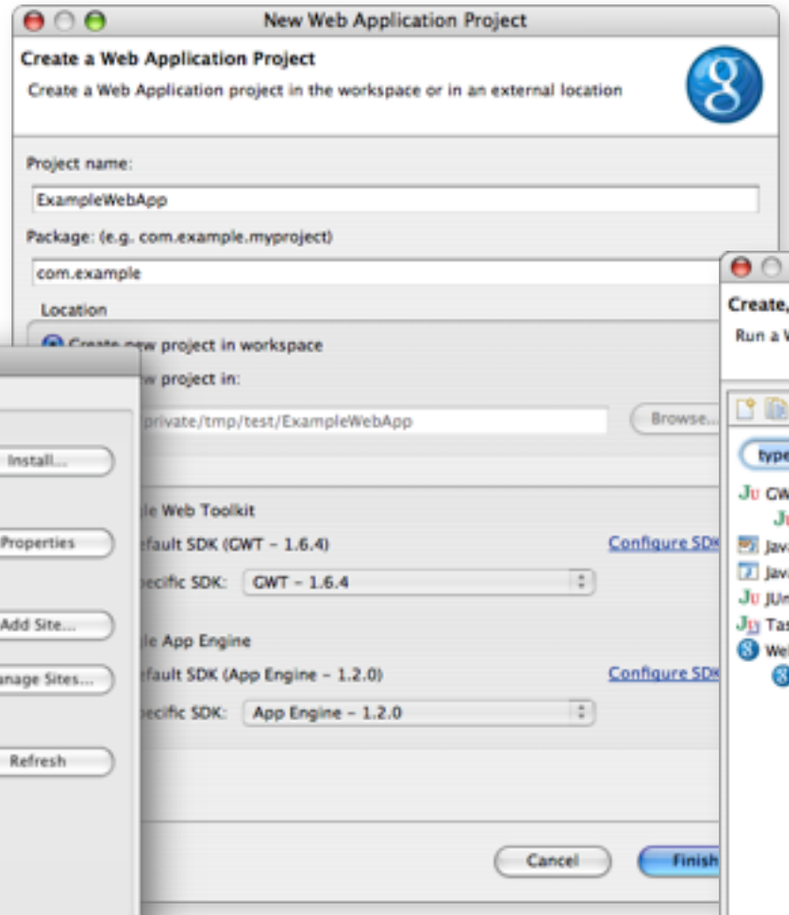
Google Plugin for Eclipse

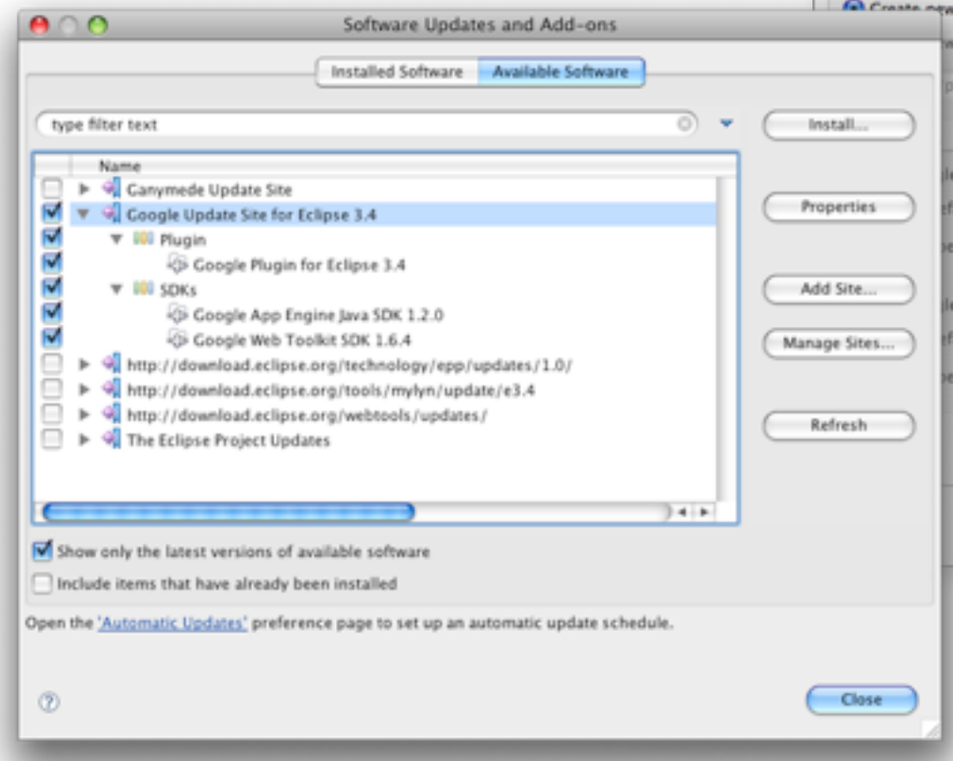


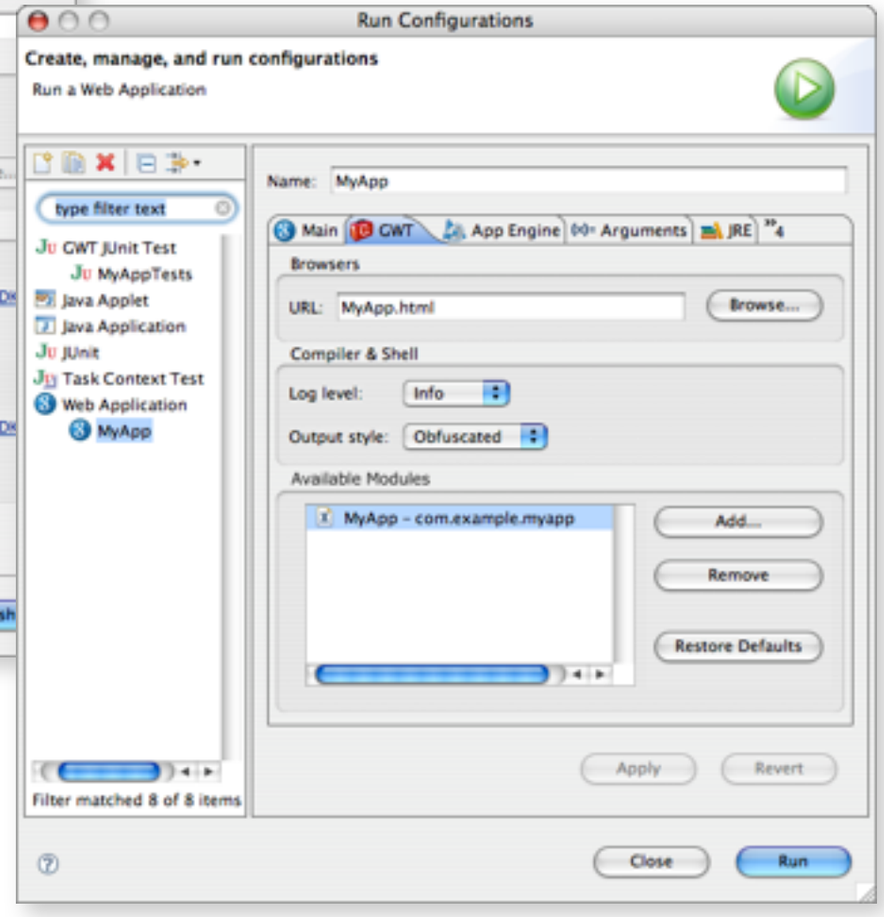
Google Plugin for Eclipse







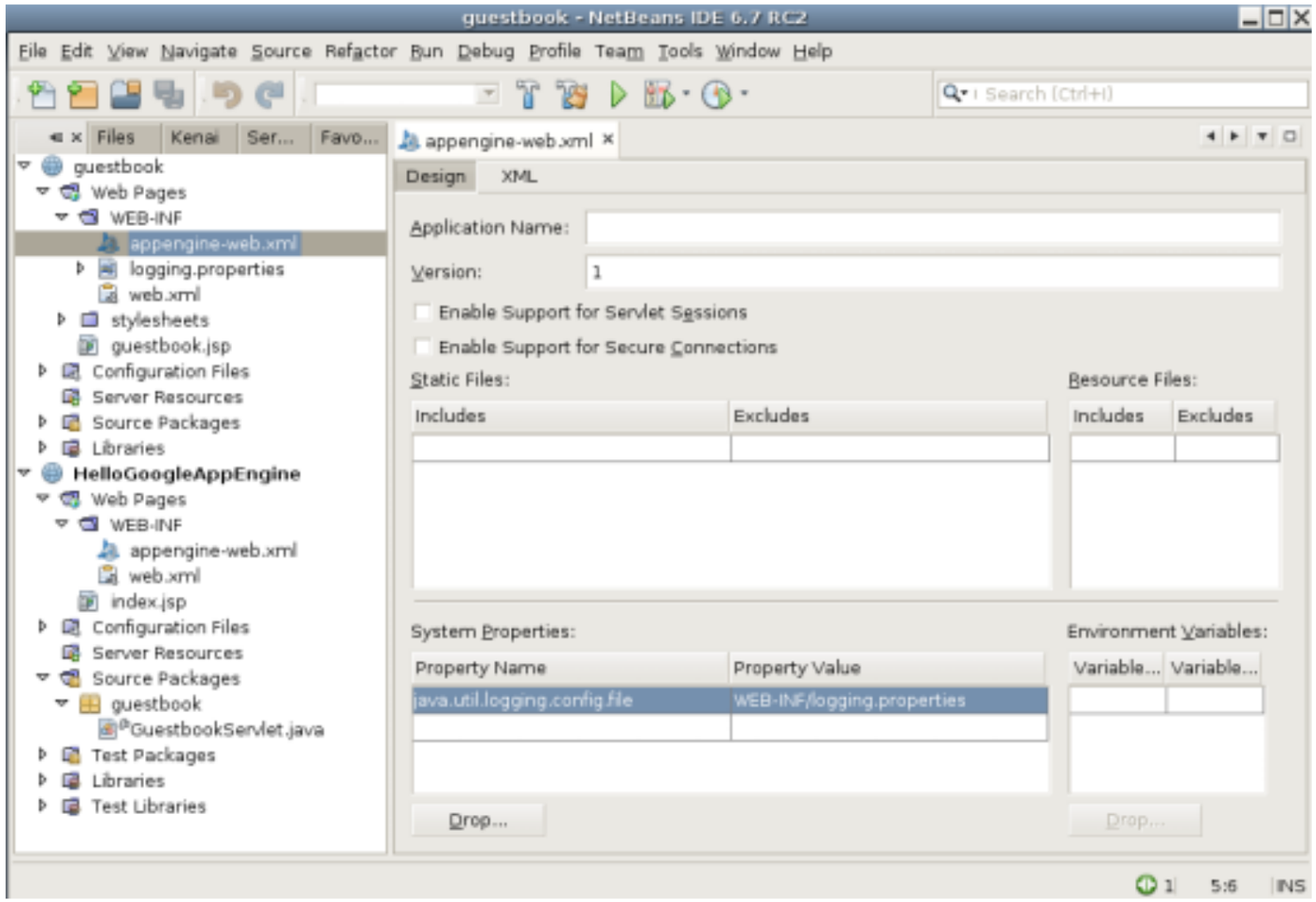




NeatBeans Plugin for Java App Engine

Google

http://kenai.com/projects/nbappengine/




F

Gaelyk

Gaelyk

http://gaelyk.appspot.com

Gaelyk

http://gaelyk.appspot.com

Google App Engine For Business

Google App Engine for Business

Same scalable cloud hosting platform. Designed for the enterprise.

- **Enterprise application management**
  - Centralized domain console
- **Enterprise reliability and support**
  - 99.9% Service Level Agreement
  - Premium Developer Support
- **Hosted SQL**
  - Managed relational SQL database in the cloud
- **SSL on your domain**
  - Including "naked" domain support
- **Secure by default**
  - Integrated Single Sign On (SSO)
- **Pricing that makes sense**
  - Pay only for what you use

* Hosted SQL and SSL on your domain available later this year

Domain Console

Google app engine

patrick@chanezon.com | My Account | Help | Sign out

Domain
- Applications
- **Settings**
- Billing

Domain Settings for chanezon.com

Mapping:
ext.chanezon.com

User Settings:
- Domain users must opt-in before using domain features
- Domain users can view the list of all domain applications
- Domain users can create domain applications

Domain:
chanezon.com

Application Identifier Prefix:
chanezon.com:

The application indentifier, or app_id, for every domain application will be prefixed with **chanezon.com:**

Update

Domain Administrators

Add Domain Administrator:

Complete email address, for example user@chanezon.com. **Note:** Google Apps administrators for the domain always have administrator access.

Add Administrator

App Engine for Business Pricing

Intranet apps:
Each app costs $8 / active user / month
Capped at $1,000 / month (i.e. users above 125 are free)
Apps are auth-restricted to domain users
Development is free
Overage charges on Background Analysis/Storage

Non intranet apps (external/public/ISV apps):
Pricing TBD

Postpaid (i.e. billed at the end of month)

App Engine for Business Support

Paid Support
- Email based
- 1000$/month
- 1h response time on operational issues
- 8h on development issues

SLA
- 99.9% uptime
- Service credits from 10% to 100% refund of monthly bill

Google Developer Qualification

Chrome Extensions

Gadgets

Search

App Engine

JS Maps API

KML

3D

App Engine for Data Crunchers

- High Performance Image Serving
- OpenId/Oauth integration
- Increased quotas
  - > 1k entities per query
  - 10” task queues
- Async UrlFetch
- Mapper API (Reduce coming soon)
- Channel API
- Matcher API

Mapper API

- First component of App Engine's MapReduce toolkit
- Large scale data manipulation
- Examples include:
  - Report generation
  - Computing statistics and metrics …
- Python Example:
  - http://blog.notdot.net/2010/05/Exploring-the-new-mapper-API
- Java Example:
  - http://ikaisays.com/2010/07/09/using-the-java-mapper-framework-for-app-engine/

Channel API

- Allows for Server Push (Comet) to browser
  - Blog post announcement:
    - http://googleappengine.blogspot.com/2010/05/app-engine-at-google-io-2010.html
  - External coverage:
    - Sneak Peak from an early trusted tester
    - http://bitshaq.com/2010/09/01/sneak-peak-gae-channel-api/
- Demo code for Dance Dance Robot available here:
  - http://code.google.com/p/dance-dance-robot/
  - Also see: https://groups.google.com/group/google-appengine-java/browse_thread/thread/6fa09953ffae2cd3/c1db7de5fdb82b65?pli=1#

Matcher API

- Allows an app to register a set of queries to match against a stream of documents
  - Trustes Testers, Python only
  - Group post announcement:
    - http://groups.google.com/group/google-appengine/msg/40021537e2e58962
  - Docs:
    - http://code.google.com/p/google-app-engine-samples/wiki/AppEngineMatcherService
- Demo code:
  - http://code.google.com/p/google-app-engine-samples/source/browse/#svn/trunk/matcher-sample

Google Storage for Developers

Store your data in Google's cloud

Wednesday, February 16, 2011

What Is Google Storage?

- Store your data in Google's cloud
  - any format, any amount, any time
- You control access to your data
  - private, shared, or public
- Access via Google APIs or 3rd party tools/libraries

Google Storage Technical Details

RESTful API

- Verbs: `GET`, `PUT`, `POST`, `HEAD`, `DELETE`
- Resources: identified by URI, like:
  `http://commondatastorage.googleapis.com/bucket/object`
- Compatible with S3

Buckets

- Flat containers (no bucket hierarchy)

Performance and Scalability

Object types and size

- Objects of any type and 100GB+ / Object
- Unlimited numbers of objects, 1000s of buckets
- Range-get support for data retrieval

Replication

- All data replicated to multiple US data centers
- Leveraging Google's worldwide network for data delivery

Consistency

- “Read-your-writes” data consistency

Security and Privacy Features

Authenticated downloads from a web browser

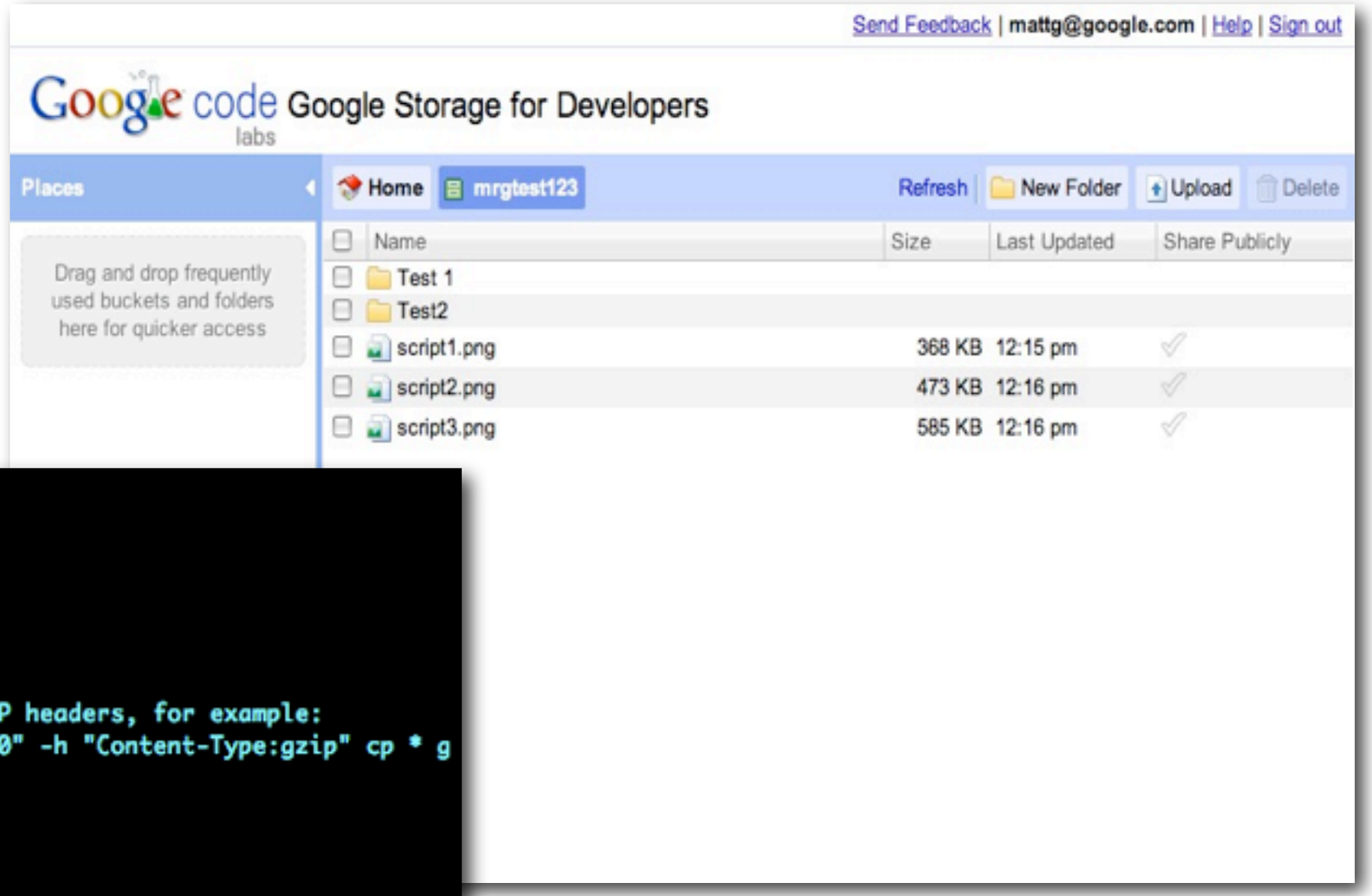
- Sharing with individuals
- Group sharing via Google Groups
- Sharing with Google Apps domains

Permissions set on Buckets or Objects

- READ (an object, or list a bucket's contents)
- WRITE (applicable to buckets, allows upload/delete/etc)
- FULL_CONTROL (read/write ACLs on objects or buckets)

Tools

Google Storage Manager



gsutil

```
dhcp-172-19-3-109:~ wferrell$ gsutil

SYNOPSIS
  gsutil [-d] [-h header]... command args

  -d option shows HTTP protocol detail.

  -h option allows you to specify additional HTTP headers, for example:
     gsutil -h "Cache-Control:public,max-age=3600" -h "Content-Type:gzip" cp * g
s://bucket

  Commands:
    Concatenate object content to stdout:
      cat [-h] uri...
        -h  Prints short header for each object.
    Copy objects:
      cp [-a canned_acl] [-t] [-z ext1,ext2,...] src_uri dst_uri
        - or -
      cp [-a canned_acl] [-t] [-z extensions] uri... dst_uri
        -a Sets named canned_acl when uploaded objects created (list below).
        -t Sets MIME type based on file extension.
        -z 'txt,html' Compresses file uploads with the given extensions.
    Get ACL XML for a bucket or object (save and edit for "setacl" command):
```

Google Storage Benefits

High Performance and Scalability

Backed by Google infrastructure

Strong Security and Privacy

Control access to your data

Easy to Use

Get started fast with Google & 3rd party tools

Some Early Google Storage Adopters

vmware

syncplicity

TECH

APPIRIO

SnapABug

UNITED STATES NAVY

MEDIABEACON

ViVU

Cloud Sherpas

memeo

widgetbox

theguardian

socialwok

XYLABS

Google Storage usage within Google

Google BigQuery

Google Prediction API

Panoramio from Google

picnik

google.org
Haiti Relief Imagery

Google patents
USPTO data

dL Google

double click
Partner Reporting

YouTube™
Partner Reporting

Google Storage - Availability

Limited preview in US* currently

- 100GB free storage and network per account
- Sign up for wait list at
    - http://code.google.com/apis/storage/

* Non-US preview available on case-by-case basis

Google Prediction API

Google's prediction engine in the cloud

Introducing the Google Prediction API

- Google's sophisticated machine learning technology
- Available as an on-demand RESTful HTTP web service

A virtually endless number of applications...

Customer Sentiment

Transaction Risk

Species Identification

Message Routing

Diagnostics

Churn Prediction

Legal Docket Classification

Suspicious Activity

Work Roster Assignment

Inappropriate Content

Recommend Products

Political Bias

Uplift Marketing

Email Filtering

Career Counseling

... and many more ...

How does it work?

1. TRAIN
The Prediction API finds relevant ***features*** in the sample data during training.

"english"	**The** quick brown fox jumped over **the** lazy dog.
"english"	**To** err **is** human, but **to** really foul things up you need a computer.
"spanish"	**No** hay mal **que por** bien **no** venga.
"spanish"	**La** tercera **es la** vencida.

2. PREDICT
The Prediction API later searches for those ***features*** during prediction.

?	**To** be or not **to** be, that **is the** question.
?	**La** fe mueve montañas.

Google Apps Marketplace

Search Marketplace

Marketplaces

Google Apps

Products

- Accounting & Finance
- Admin Tools
- Calendar & Scheduling
- Customer Management
- Document Management
- EDU
- Productivity
- Project Management
- Sales & Marketing
- Security & Compliance
- Workflow

Professional Services

- Archiving & Discovery Implementation
- Custom Application Development
- EDU Specialists
- Google Analytics
- Medium-Large Business Implementation
- Small Business Implementation
- Support & Managed Services
- Training & Change Management

Enterprise Search

Products

- Content Connectors
- OneBox Modules
- Search Extensions

Professional Services

- GSA Deployment
- Google Mini Deployment
- Custom Development
- Training
- GeoSpatial Solutions

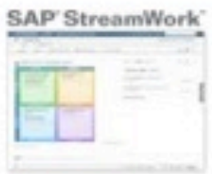
The Google Apps Marketplace offers products and services designed for Google users, including installable apps that integrate directly with Google Apps. Installable apps are easy to use because they include single sign-on, Google's universal navigation, and some even include features that integrate with your domain's data.

Featured Apps

Concur Breeze – Free Mobile and Web Expense Reporting

Concur Breeze is designed specifically to help small and mid-sized businesses take the hassle out of expense reporting, allowing your employees to spend more time making your business successful.

Try popular & notable apps



SAP StreamWork

SAP StreamWork is a collaborative decision-making solution that brings together the people, information, and proven business approaches to drive fast, meaningful results.

ERPLY

ERPLY offers web-based software for managing your points of sale, inventory, relationships and billing.

Gantter Project

Gantter.com is a powerful, web-based Project Management Tool that requires no software to be installed and it completely integrates with Google Docs.

"Tops" In Google Apps

Top Installed this week

1. Insightly: Free simple CRM and Project Management
   ★★★★★ 321 reviews
2. Manymoon: Free Social Productivity, Project Management & Task Management
   ★★★★★ 194 reviews
3. Zoho CRM (3 users free)
   ★★★★★ 41 reviews
4. MailChimp
   ★★★★★ 23 reviews
5. myERP.com (Free 2 users) Accounting CRM Sales Inventory
   ★★★★★ 72 reviews

Top installed

1. Manymoon: Free Social Productivity, Project Management & Task Management
   ★★★★★ 194 reviews
2. Insightly: Free simple CRM and Project Management
   ★★★★★ 321 reviews
3. Aviary Design Suite (Free)
   ★★★★★ 14 reviews
4. OffiSync - [FREE] Integrate Microsoft Office with Google Apps
   ★★★★★ 572 reviews
5. Zoho CRM (3 users free)
   ★★★★★ 41 reviews

Recently added

- Zoho Books - Accounting and bookkeeping software

A Prediction API Example

Automatically determine application recommendations

- **Goal**: Increase relevancy on the Apps Marketplace via recommendations
- **Customers**: Businesses of various sizes and industries using Google Apps around the world
- **Data**: Sampling of previous installs of applications
- **Outcome**: Predict applications which would be appropriate for a new customer visiting the site

Using the Prediction API

A simple three step process...

1. Upload

Upload your training data to **Google Storage**

2. Train

Build a model from your data

3. Predict

Make new predictions

Step 1: Upload

Upload your training data to Google Storage

- Training data: **outputs** and **input features**
- Data format: comma separated value format (CSV), result in first column

```
"SlideRocket","EDUCATION","us","en","10","5"
"MailChimp","BUSINESS","us","en","7","0"
"MailChimp","STANDARD","se","sv","1","0"
"Smartsheet","BUSINESS","us","en","13","4"
```

Upload to Google Storage

```
gsutil cp installs gs://appdata/
```

Step 2: Train

Create a new model by training on data

To train a model:

```
POST prediction/v1.1/training?data=appdata%2Finstalls
```

Training runs asynchronously. To see if it has finished:

```
GET prediction/v1.1/training/appdata%2Finstalls

{"data":{
   "data":"appdata/installs",
   "modelinfo":"estimated accuracy: 0.xx"}}}
```

Step 3: Predict

Apply the trained model to make predictions on new data

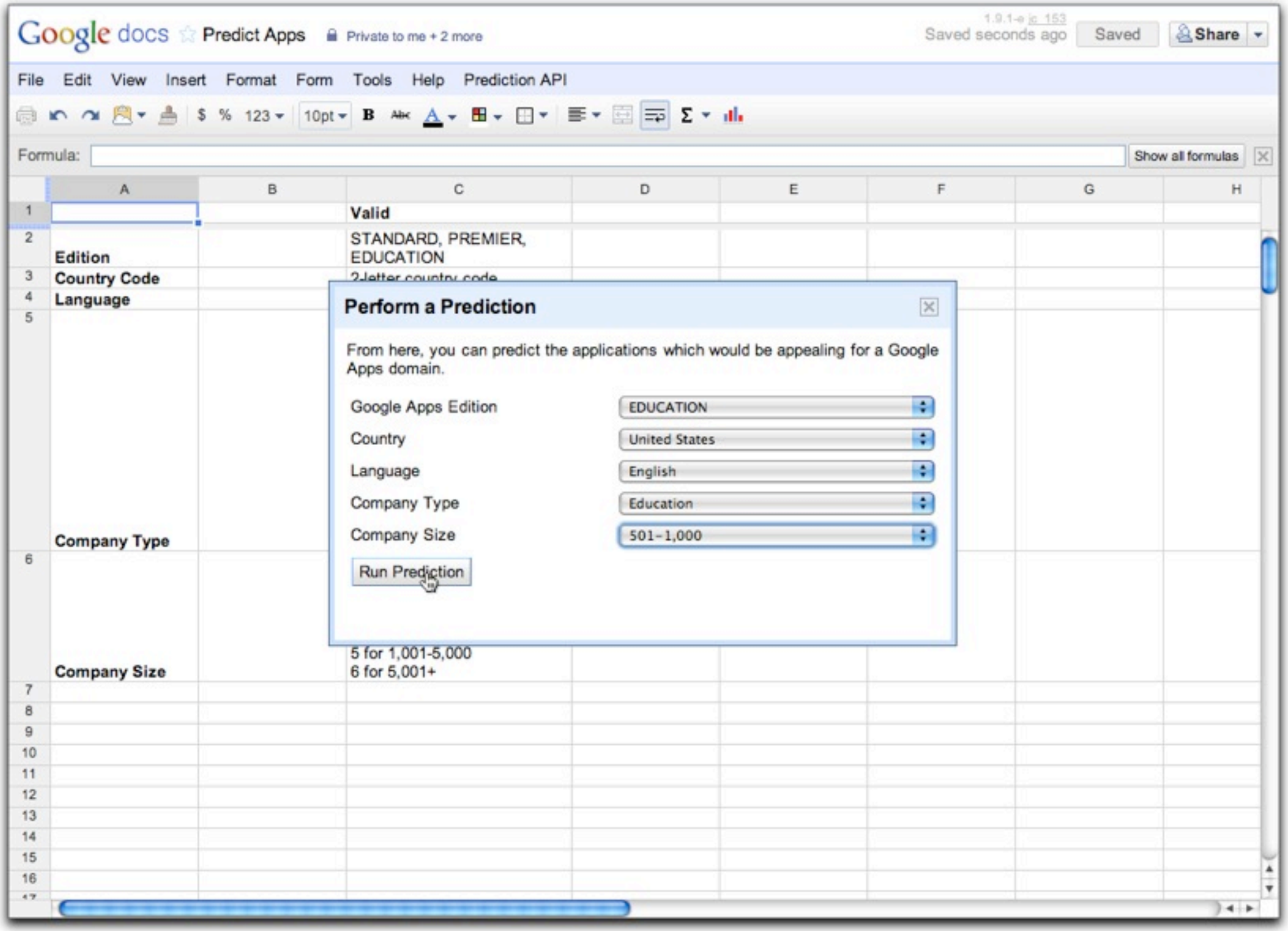
```
POST prediction/v1.1/query/appdata%2Finstalls/predict

{ "data":{
    "input": { "mixture" : [
      "EDUCATION","us","en","10","0" ]}}}


{ data : {
  "kind" : "prediction#output",
  "outputLabel":"Manymoon",
  "outputMulti" :[
     {"label":"OffiSync", "score": x.xx}
     {"label":"Zoho CRM", "score": x.xx}
     {"label":"MailChimp", "score": x.xx}]}}
```

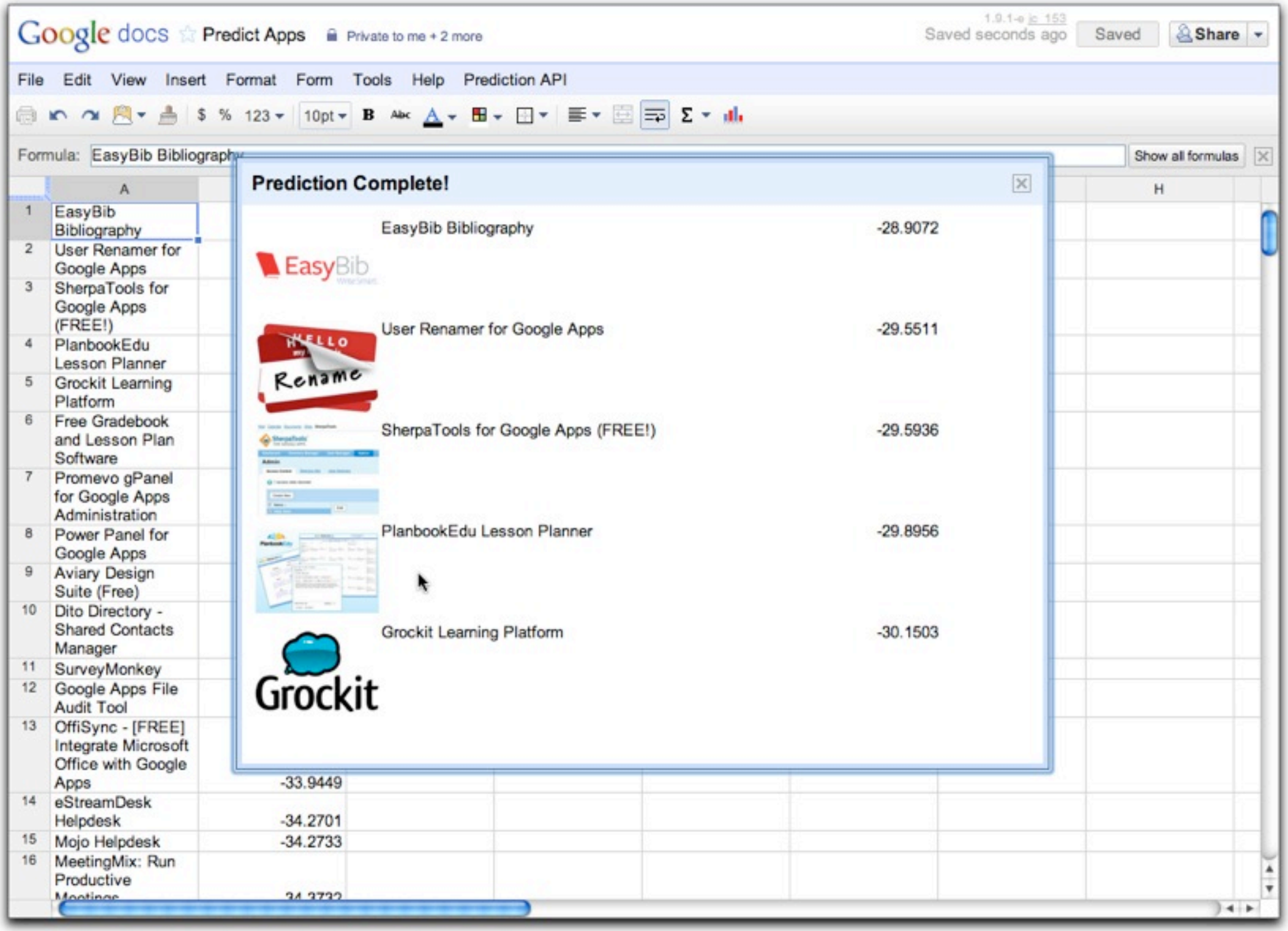
Demo!

Demo Screenshots



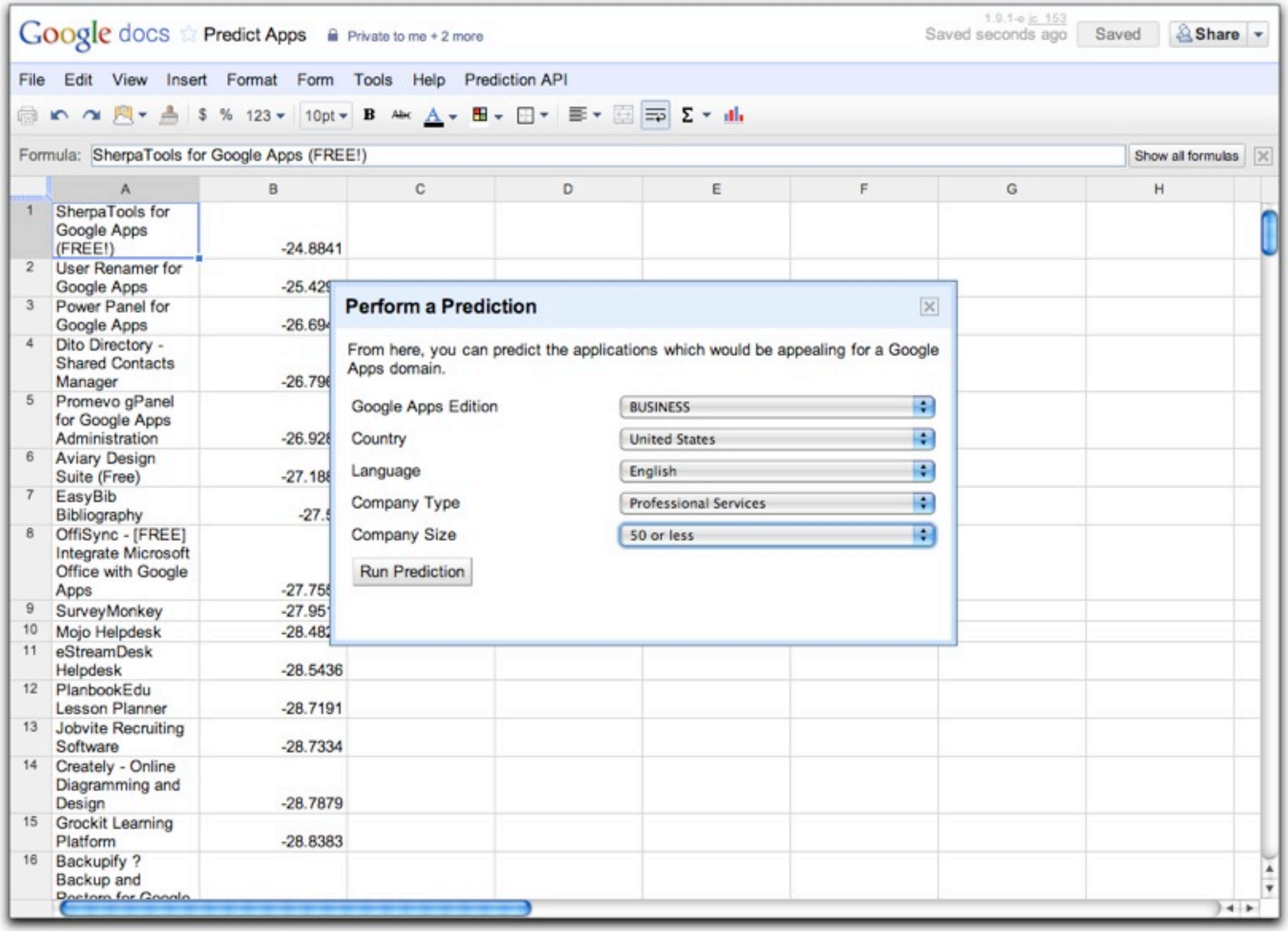
Predicting apps for a 501-1,000 seat educational institution

Demo Screenshots



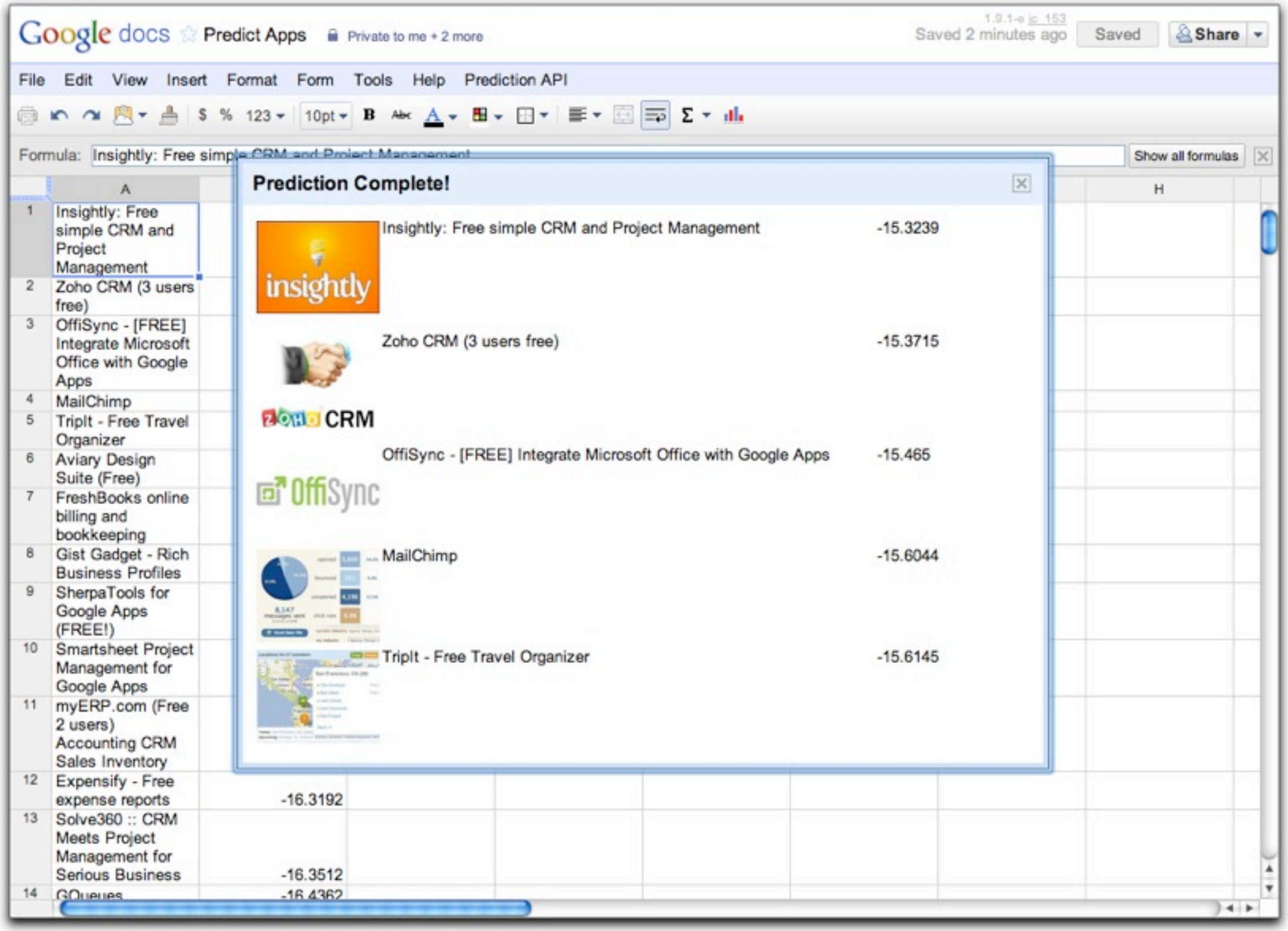
Predicting apps for a 501-1,000 seat educational institution

Demo Screenshots



Predicting apps for a small business

Demo Screenshots



Predicting apps for a small business

Prediction API Capabilities

Data

- Input Features: numeric or unstructured text
- Output: up to hundreds of discrete categories, or continuous values

Training

- Many machine learning techniques
- Automatically selected
- Performed asynchronously

Access from many platforms:

- Web app from Google App Engine
- Apps Script (e.g. from Google Spreadsheet)
- Desktop app

Prediction API - Pricing

Free Quota in trial/development

- 100 predictions/day, 5MB trained/day
- Available for 6 months

Paid Usage

- $10/month per project includes 10,000 predictions
- Additional predictions are $0.50 per 1,000
- Absolute limit of 60,000 predictions per day
- $0.002 per MB trained (max size per dataset is 100MB)

Google Storage - Availability

Limited preview in US* currently

- Sign up for wait list at
    - http://code.google.com/apis/predict/

* Non-US preview available on case-by-case basis

Google BigQuery

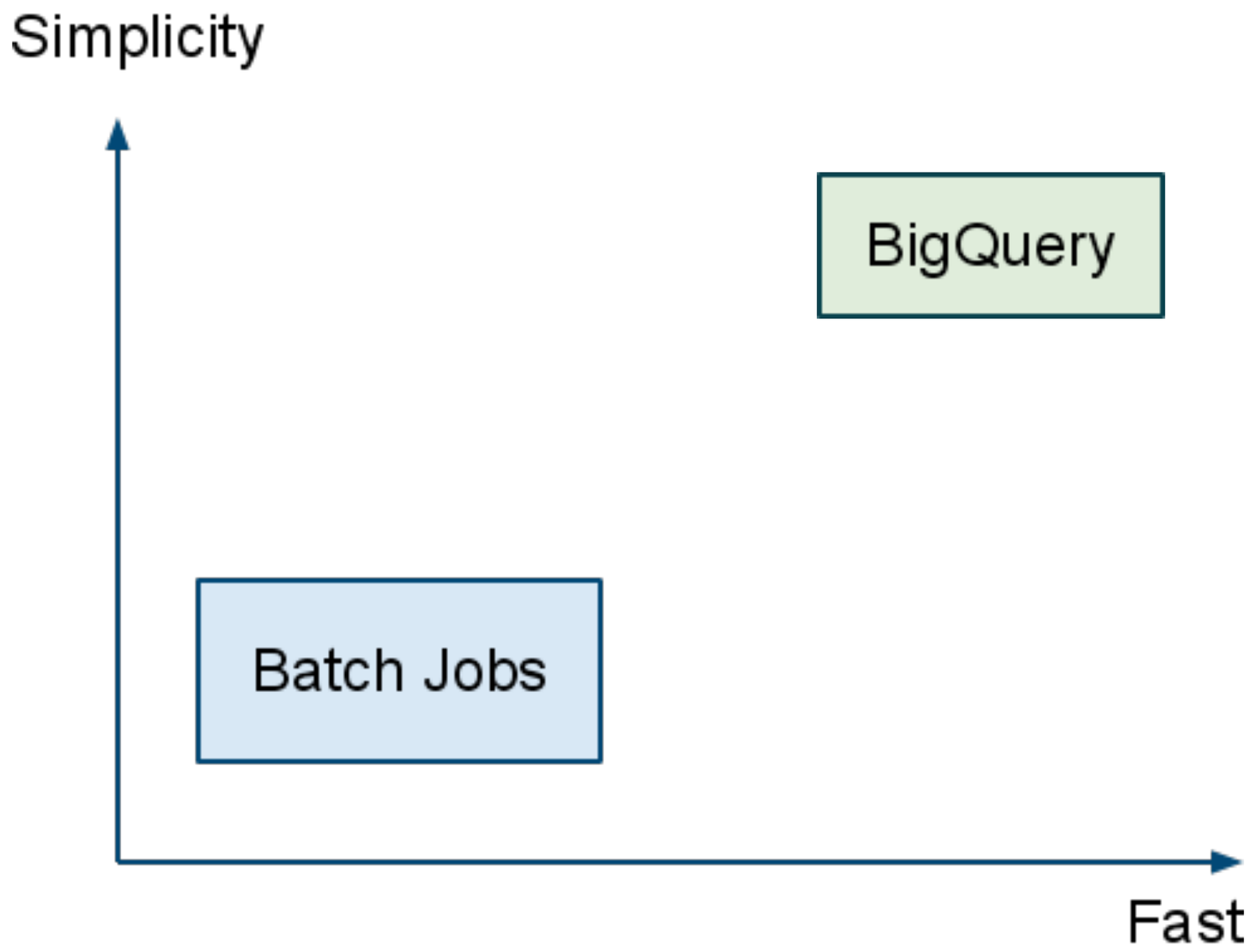
Interactive analysis of large datasets in Google's cloud

Introducing Google BigQuery

- Google's large data adhoc analysis technology
  - Analyze massive amounts of data in seconds
- Simple SQL-like query language
- Flexible access
  - REST APIs, JSON-RPC, Google Apps Script

Why BigQuery?

Working with large data is a challenge



Many Use Cases …

Interactive Tools

Spam

Trends Detection

Web Dashboards

Network Optimization

Key Capabilities of BigQuery

- **Scalable**: Billions of rows
- **Fast**: Response in seconds
- **Simple**: Queries in SQL

- Web Service
  - REST
  - JSON-RPC
  - Google App Scripts

Using BigQuery

Another simple three step process...

1. Upload

Upload your raw data to **Google Storage**

2. Import

Import raw data into **BigQuery table**

3. Query

Perform **SQL queries** on table

Writing Queries

Compact subset of SQL

- **`SELECT ... FROM ... WHERE ... GROUP BY ... ORDER BY ... LIMIT ...;`**

Common functions

- **`Math, String, Time, ...`**

Additional statistical approximations

- **`TOP`**
- **`COUNT DISTINCT`**

BigQuery via REST

GET /bigquery/v1/tables/{table name}

GET /bigquery/v1/query?q={query}

Sample JSON Reply:

```
{
  "results": {
    "fields": { [
      {"id":"COUNT(*)","type":"uint64"}, ... ]
    },
    "rows": [
      {"f":[{"v":"2949"}, ...]},
      {"f":[{"v":"5387"}, ...]}, ... ]
  }
}
```

Also supports JSON-RPC

Security and Privacy

Standard Google Authentication

- Client Login
- OAuth
- AuthSub

HTTPS support

- protects your credentials
- protects your data

Relies on Google Storage to manage access

Large Data Analysis Example

Wikimedia Revision History

BigQuery

Help Report a problem Send feedback Sign out

Enter SQL query

```
SELECT TOP(title, 5), COUNT(*)
FROM [wikipedia]
WHERE wp_namespace = 0;
```

Run

Wikimedia Revision history data from:
http://download.wikimedia.org/enwiki/latest/enwiki-latest-pages-meta-history.xml.7z

Large Data Analysis Example

Wikimedia Revision History

BigQuery

Help Report a problem Send feedback Sign out

Enter SQL query

```
SELECT TOP(title, 5), COUNT(*)
FROM [wikipedia]
WHERE wp_namespace = 0;
```

Run

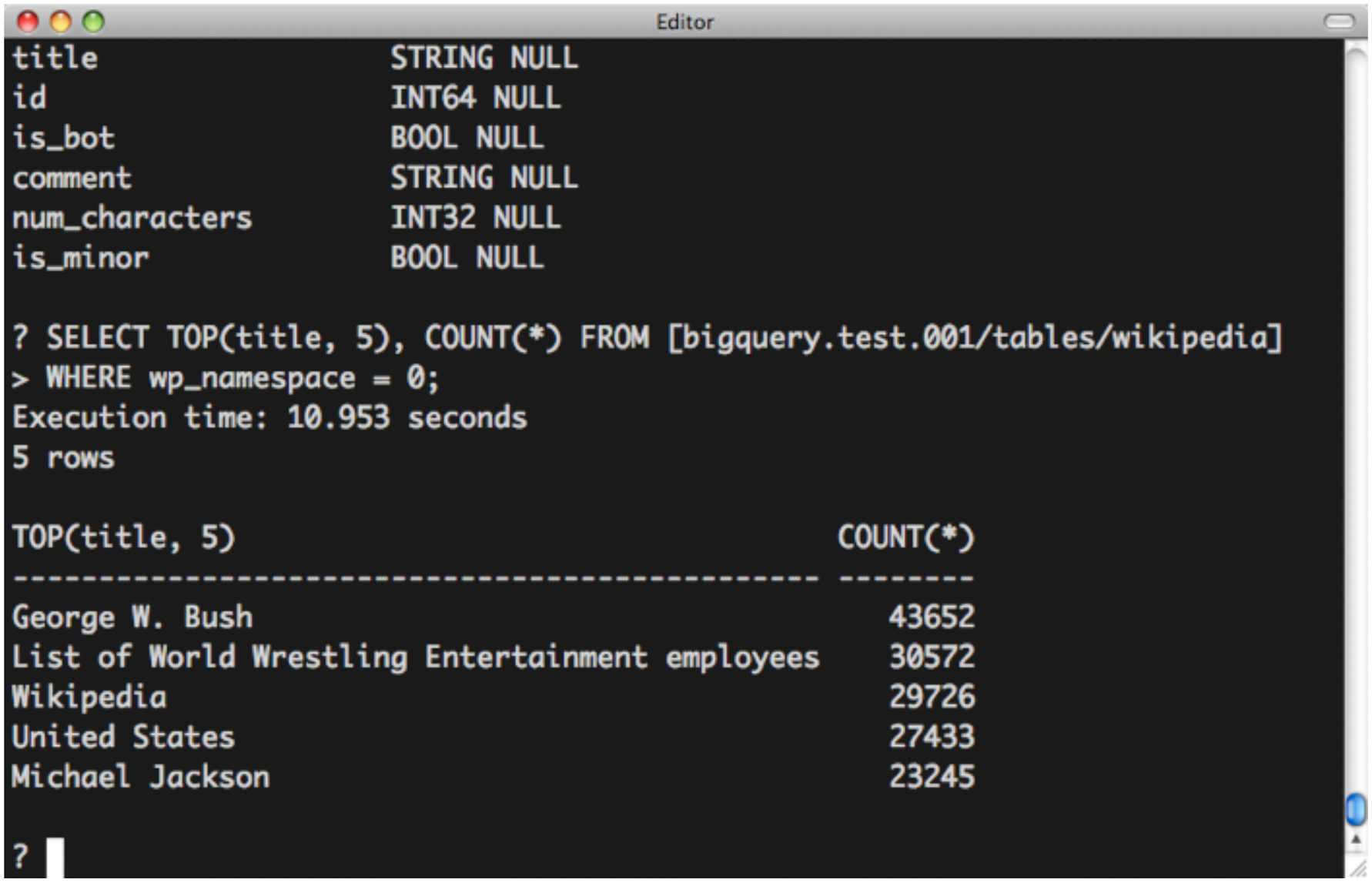
TOP(title, 5)	COUNT(*)
George W. Bush	43652
List of World Wrestling Entertainment emplo	30571
Wikipedia	29722
United States	27432
TOP(title, 5)	**COUNT(*)**

Wikimedia Revision history data from:
http://download.wikimedia.org/enwiki/latest/enwiki-latest-pages-meta-history.xml.7z

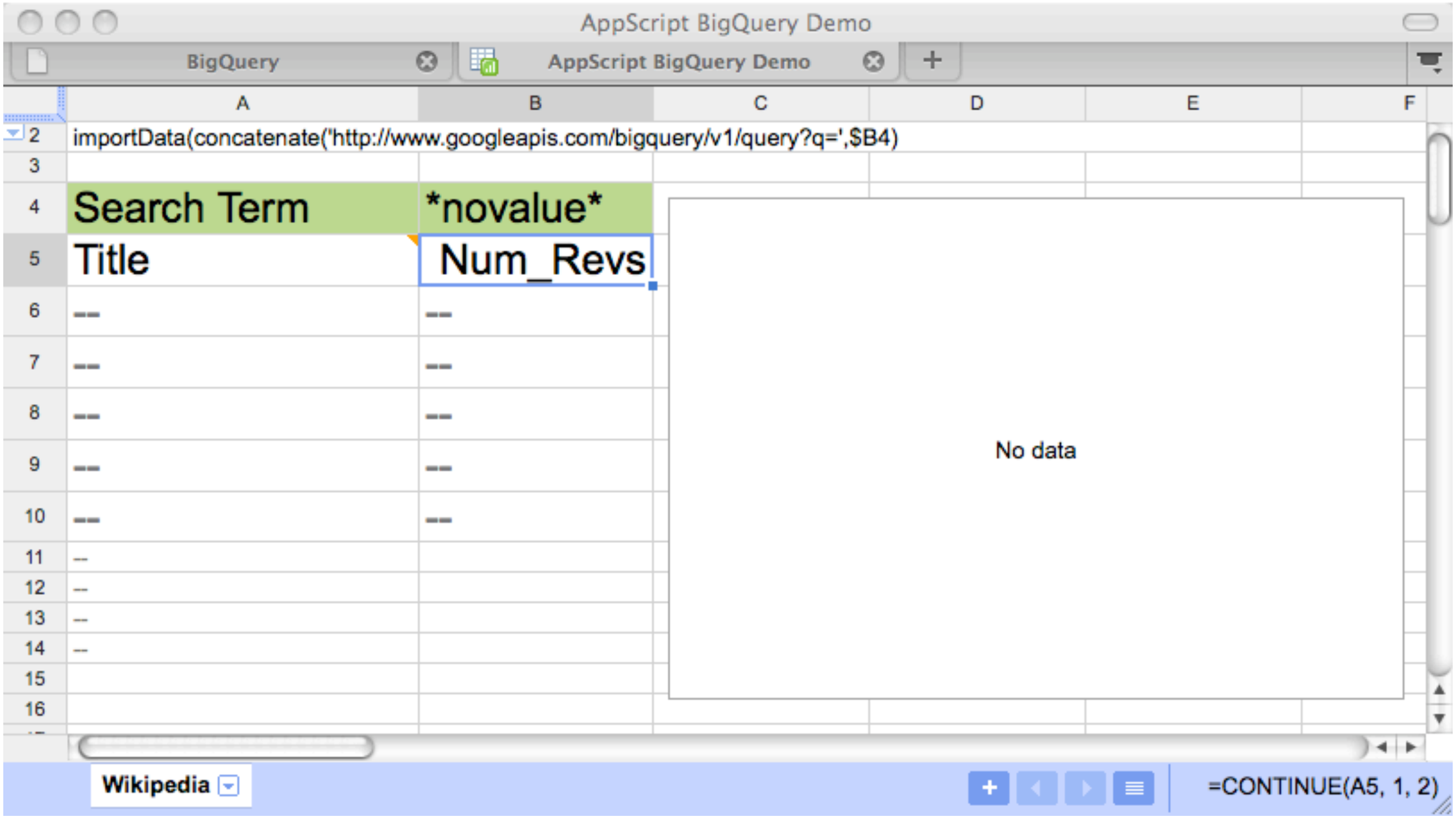
Using BigQuery Shell

Python DB API 2.0 + B. Clapper's **`sqlcmd`**

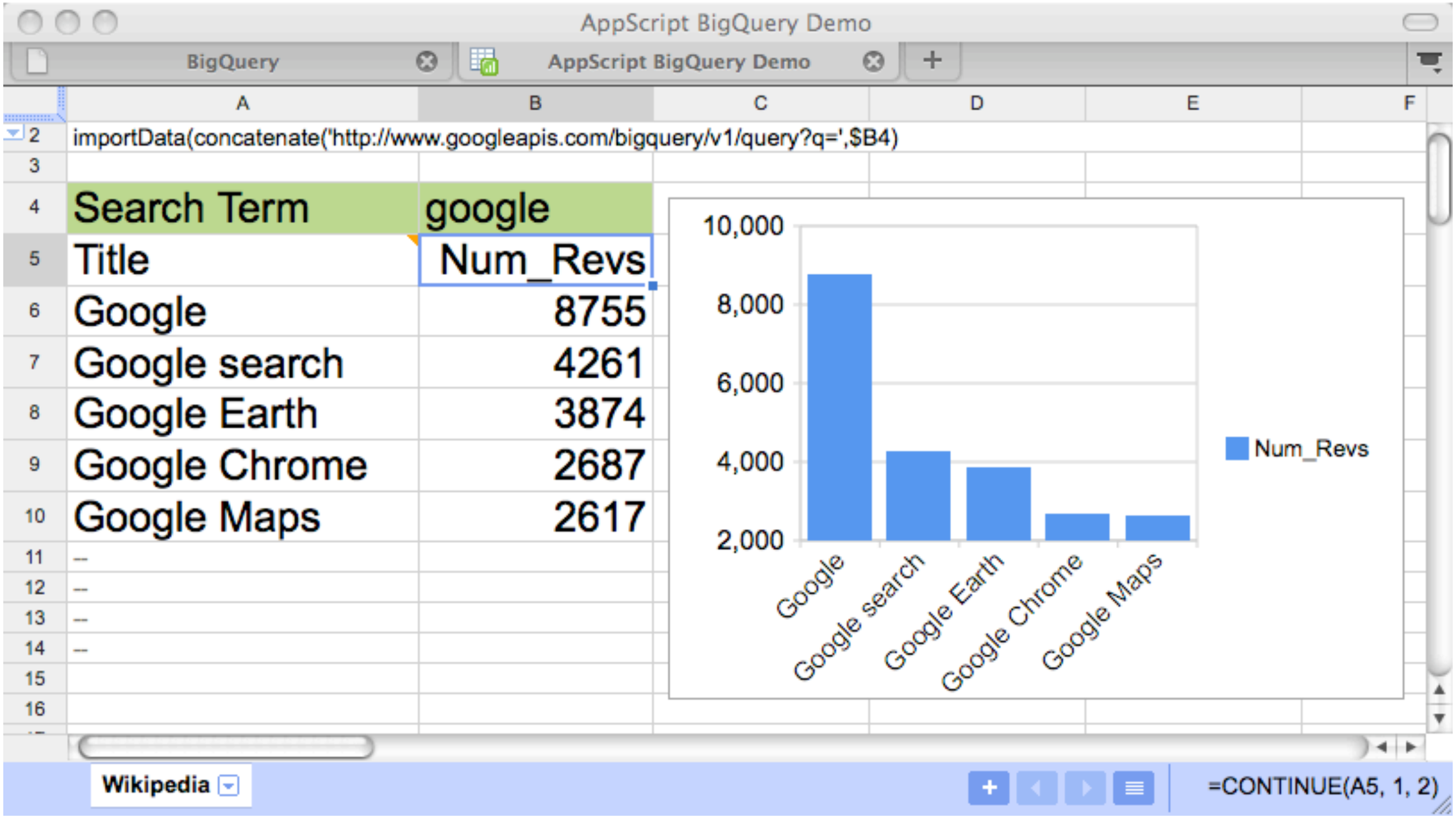
http://www.clapper.org/software/python/sqlcmd/



BigQuery from a Spreadsheet



BigQuery from a Spreadsheet

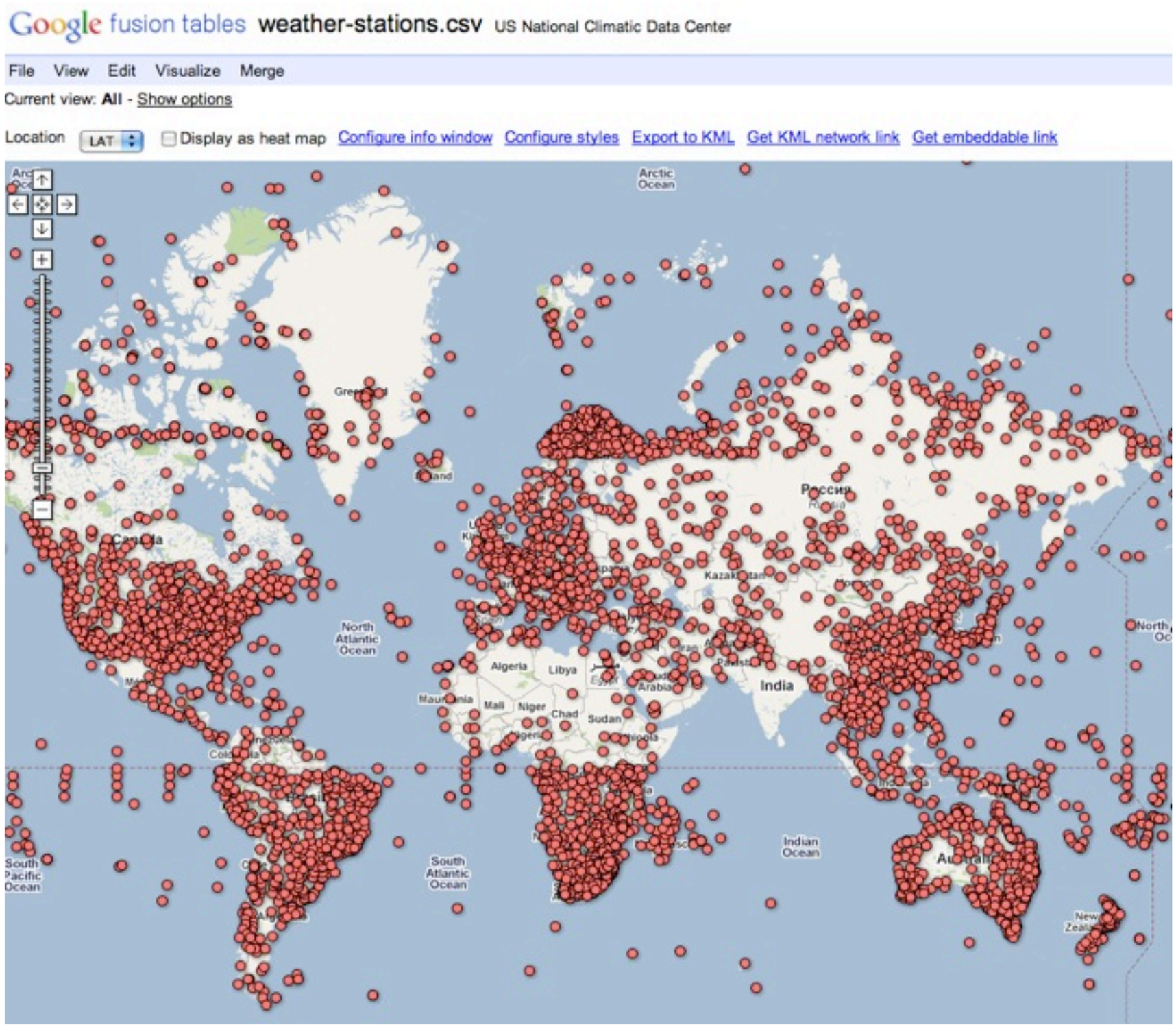


Google Fusion Tables

Google Fusion Tables

- Manage large collections of tabular data in the cloud
  - 100 Mb tables
  - Filters, Aggregation, Merge
  - ACL, Collaboration, Discuss Data
  - Visualizations
- REST API
  - Geo queries
- Maps Integration
  - FusionTablesLayer

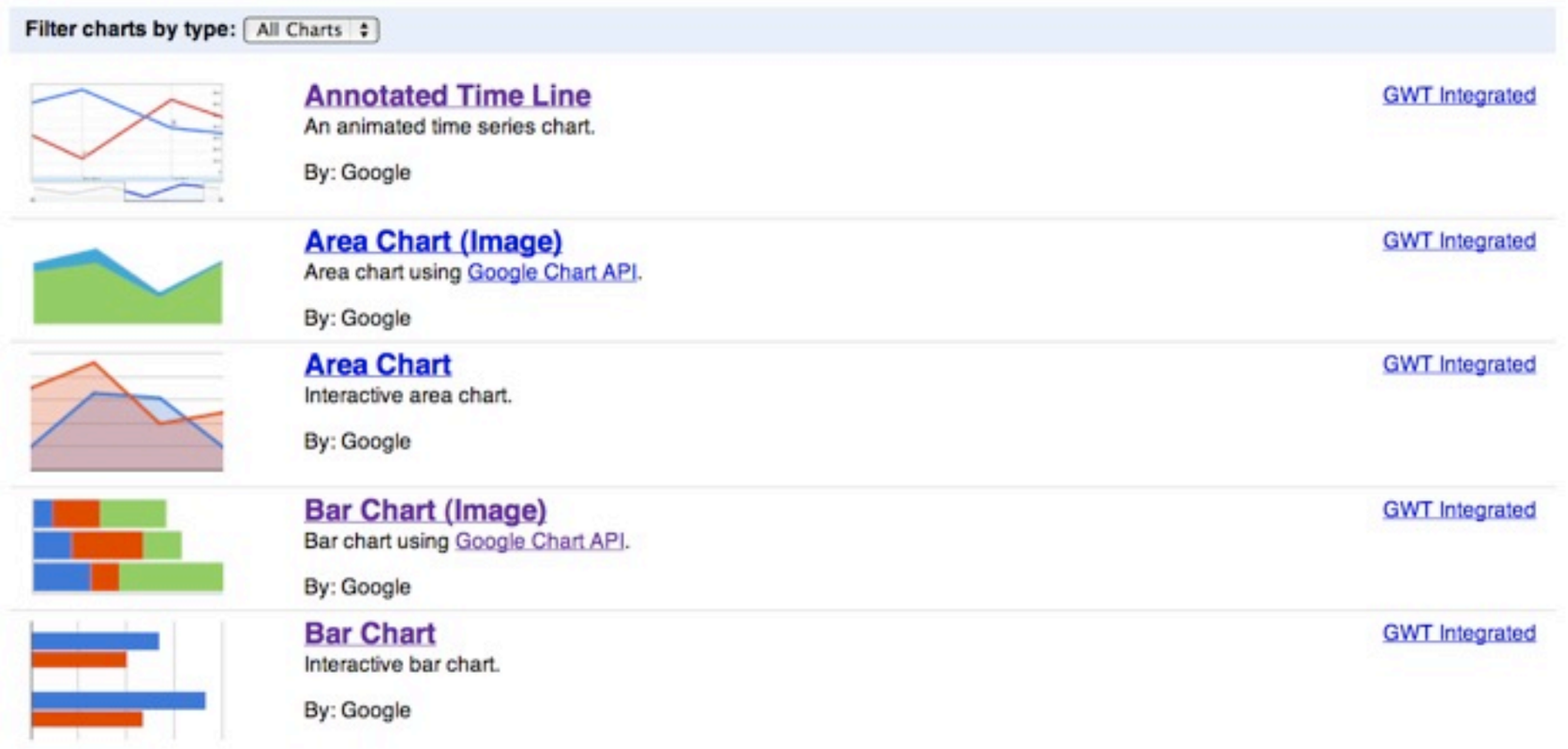
Google Fusion Tables



Google Visualization API

Google Visualization API

- Collection of JavaScript Visualization components
  - Some from Google (Chart Tools)
  - Some from other developers
  - Share the same wire protocol for Data Sources

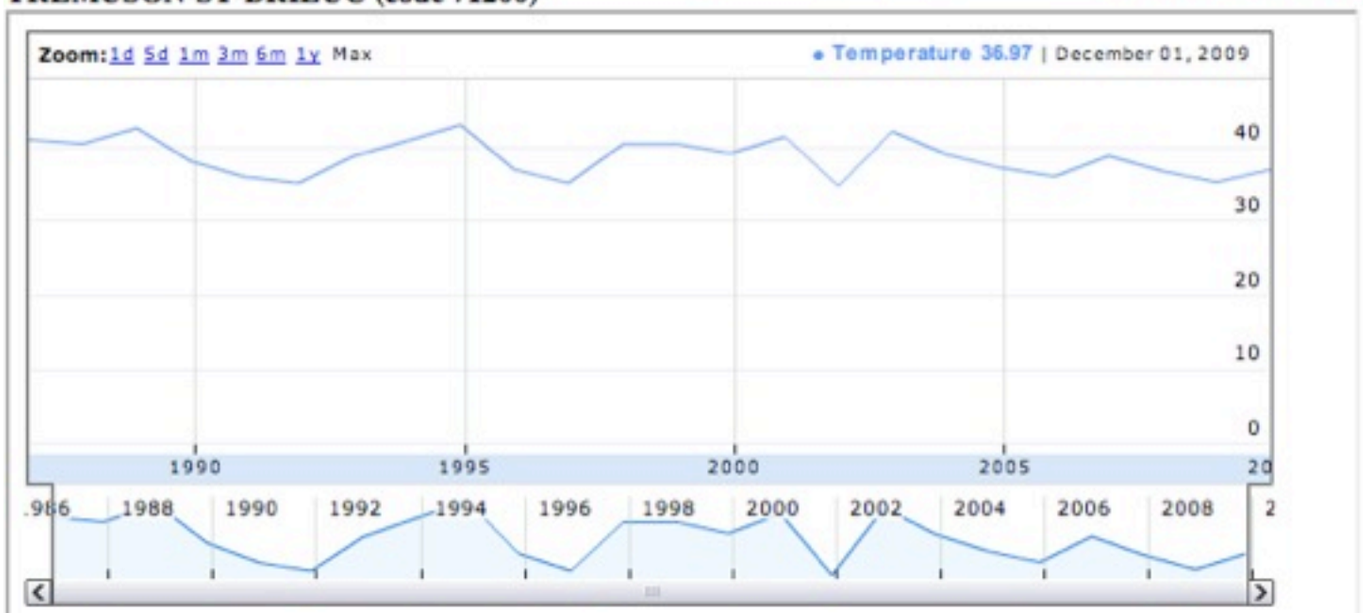


Example: Weather data

- US National Climatic Data Center
  - weather data at stations around the globe since 1929
  - Stored in Google Storage
  - Created a Table for Bigquery
  - Upload Weather Station coordinates in Fusion Tables
  - App Engine App
    - Maps API to display weather station Maps
    - Bigquery to query average temperature in January
    - A bit of Python to create a JSON Data Source
    - Visualization API
- Just an example: rince, repeat, enhance!

Example: Weather data

TREMUSON-ST-BRIEUC (code 71200)

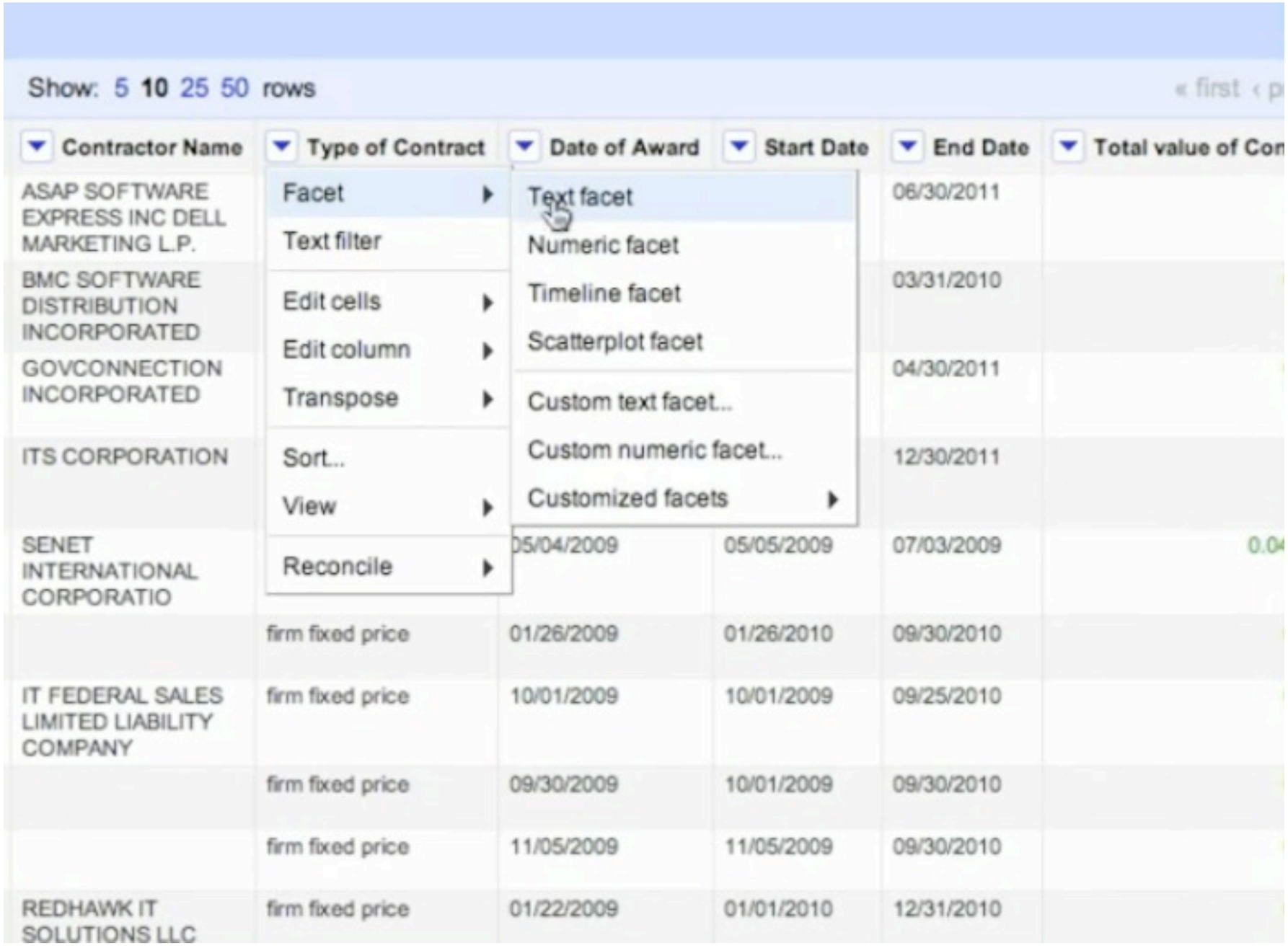


Google Refine

Google Refine

- Power tool for working with messy data
  - Cleanup
  - Transform
  - Augment
  - (Link with FreeBase)
- Desktop software for now
- http://code.google.com/p/google-refine/

Google Refine

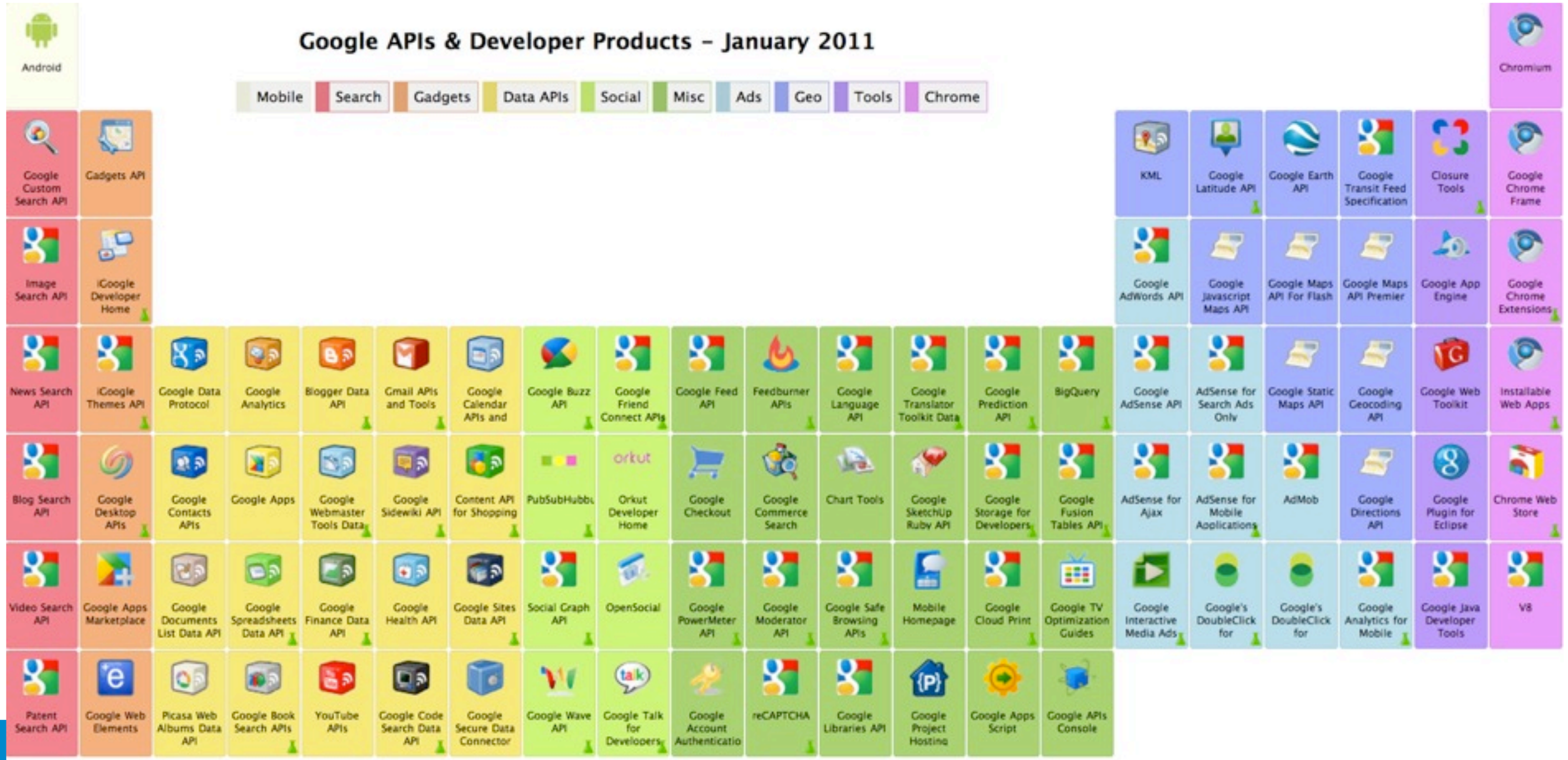


Recap

- **Google App Engine**
  - Easy to build, deploy and manage web apps
- **Google Storage**
  - High speed data storage on Google Cloud
- **Prediction API**
  - Google's machine learning technology
- **BigQuery**
  - Interactive analysis of very large data sets
- **Google Fusion Tables**
  - Manage collections of tabular data in the cloud
- **Google Refine**
  - Power tool for working with messy data
- **Google Visualization**
  - Collection of JavaScript Visualization

More information

http://code.google.com/apis/
http://code.google.com/more/table/



Jfokus 2011

Google Cloud for Java Developers: Platform and Monetization

Patrick Chanezon, Developer Advocate, Cloud
@chanezon, chanezon@google.com