

# Google App Engine

---



# Outline

---

Cloud Computing (Brief)

What is Google App Engine (GAE)?

GAE Services

Storing Data and Data Processing

Real World Examples

---

# Cloud Computing

---

“Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.” -- NIST, The NIST Definition of Cloud Computing, September 2011.

---

# Cloud Computing

---

Key ideas:

- On-demand access
- Shared pool of resources
- Minimal service provider management/interaction

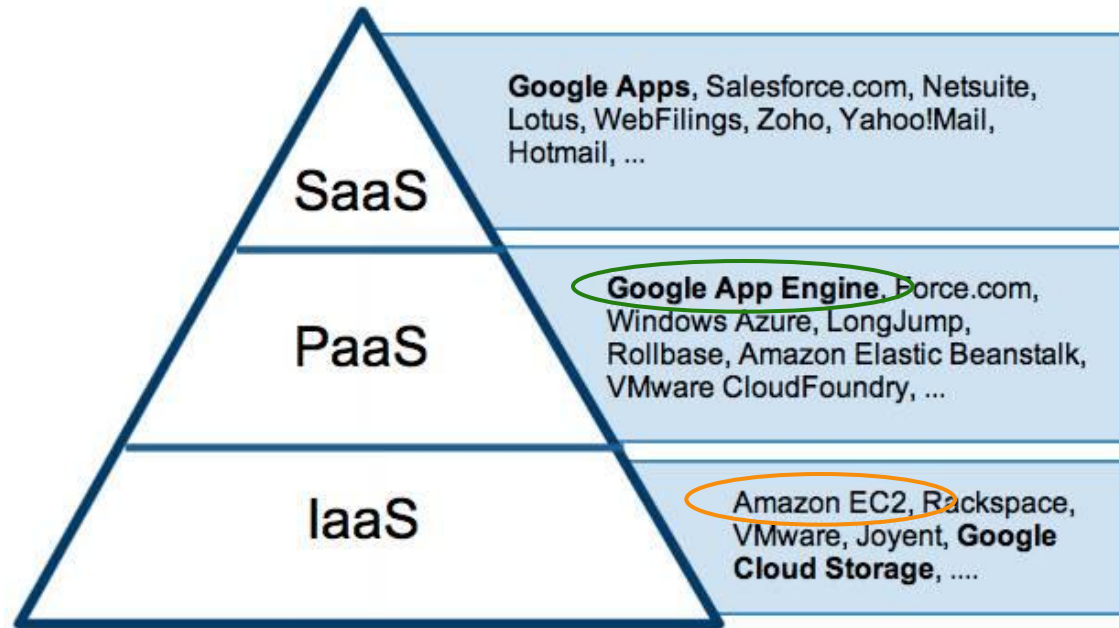
But, there's a spectrum of cloud products & services...

---

# Cloud Computing

---

## Cloud Computing as Gartner Sees It



Source: Gartner AADI Summit Dec 2009

# What Is Google App Engine (GAE)?

---

Is it a web server?

Is it an application development environment?

Is it a gateway to other Google products & services?

**YES**

---

# What Is Google App Engine (GAE)?

---

## **It's a web server**

- Runs on Google's infrastructure using WSGI
  - Can use a variety of Python development frameworks (Django, Jinja, etc.)
  - Scales automatically
-

# What Is Google App Engine (GAE)?

---

**It's an application development environment**

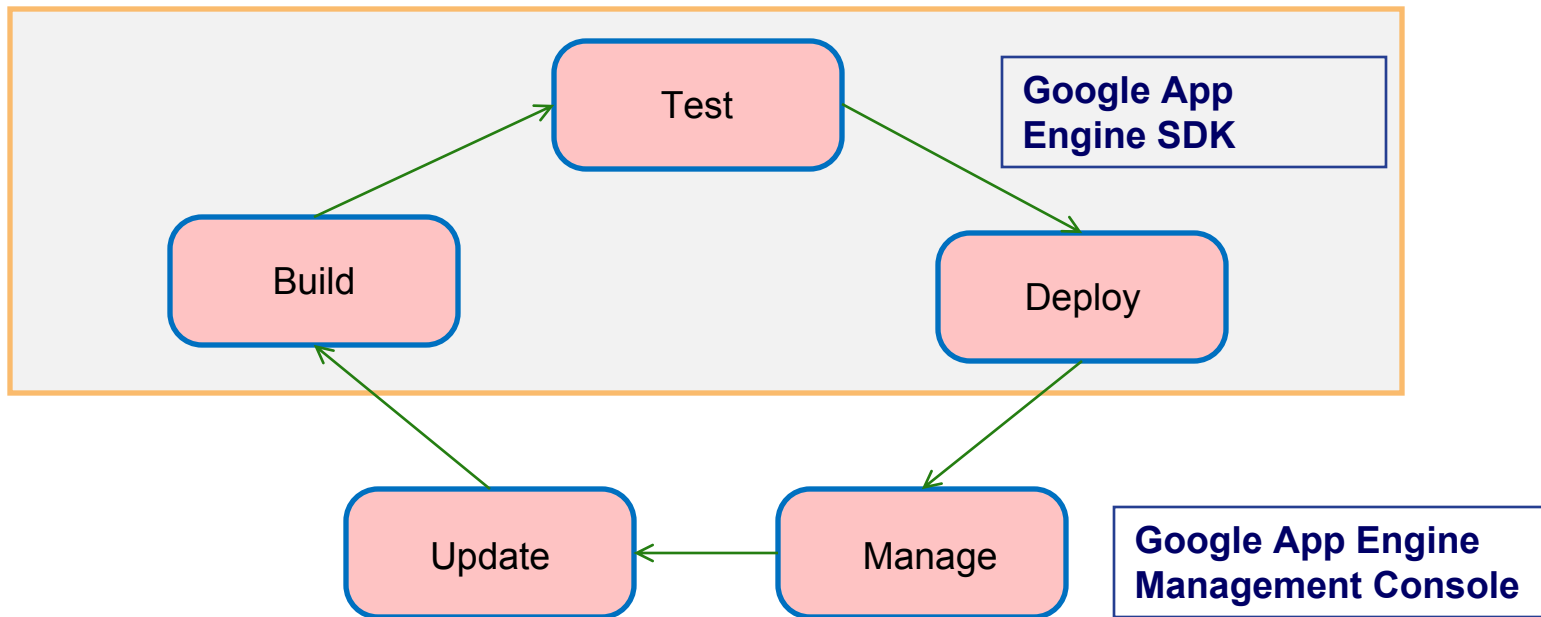
- Has its own SDK
  - Build & test locally, deploy globally
  - SDKs for Java, Python, Go & PHP
  - Can develop on PCs, Mac OS or Linux
  - Right price, too (FREE!)
-



# What Is Google App Engine (GAE)?

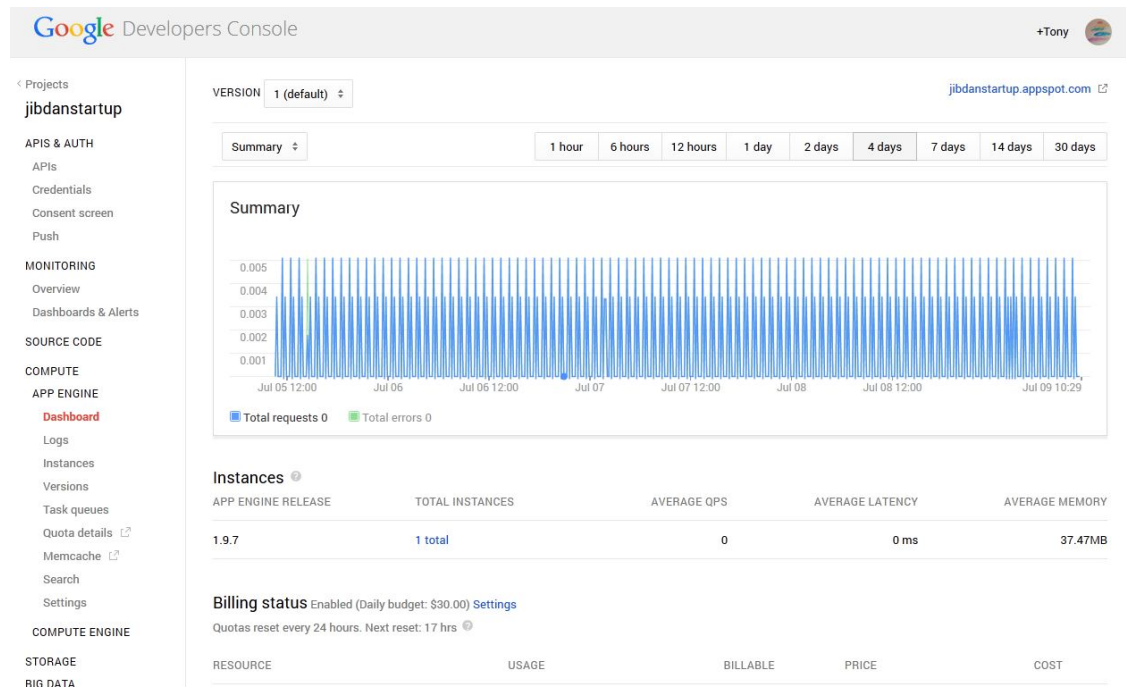
---

**It's an application development environment**



# What Is Google App Engine (GAE)?

It's an application development environment



# What Is Google App Engine (GAE)?

---

**It's a gateway to other Google products**

All Cloud Services



## App Engine

Develop your app easily using built-in services that make you more productive.



## Compute Engine

Deploy and run virtual machines on Google Cloud Platform.



## BigQuery

Analyze terabytes of data in seconds, and load data with ease.



## Hadoop on Google Cloud Platform

Run Apache Hadoop along with your favorite community tools on Google Cloud Platform.



## Cloud SQL

Store and manage data using a relational MySQL database.



## Cloud Storage

Use a durable and highly available object storage service with global edge-caching and versioning.



## Cloud Datastore

Use a managed, NoSQL, schemaless database for storing non-relational data.



## Cloud DNS

A highly available and scalable DNS service to route end users to Internet apps and services.

# GAE Services

---

APIs galore:

- Memcache
- URLFetch
- Datastore
- SMS & Voice
- Mail
- OAuth
- Search

And many others including other Google products

---

# GAE Services

---

## URLFetch

Similar to urllib but optimized to run in Google's environment

### urllib2

```
import urllib2

url = "http://www.google.com/"
try:
    result = urllib2.urlopen(url)
    doSomethingWithResult(result)
except urllib2.URLError, e:
    handleError(e)
```

### urlfetch

```
from google.appengine.api import urlfetch

url = "http://www.google.com/"
result = urlfetch.fetch(url)
if result.status_code == 200:
    doSomethingWithResult(result.content)
```

# GAE Services

---

## Datastore

No-SQL database built-in to the framework:

```
from google.appengine.ext import ndb

class Account(ndb.Model):
    username = ndb.StringProperty()
    userid = ndb.IntegerProperty()
    email = ndb.StringProperty()
```

Query:

```
qry = Account.query(Account.userid == 42)
```

---

# GAE Services

## Datastore

Google Developers Console

+Tony

COMPUTE

APP ENGINE

Dashboard

Logs

Instances

Versions

Task queues

Quota details

Memcache

Search

Settings

COMPUTE ENGINE

VM instances

Disks

Snapshots

Images

Networks

Network load balancing

Metadata

Zones

Operations

Quotas

STORAGE

CLOUD STORAGE

CLOUD DATASTORE

Dashboard

Query

Indexes

Cloud SQL

Create Entity

KIND streamdata Filters

streamdata Entities

	NAME/ID	created_at	querydate	source	text	ticker	tweet_id	user
<input type="checkbox"/>	<a href="#">id=17755</a>	Thu May 30 13:41:08 +0000 2013	2013-05-31 23:37:47	Twitter		\$WFC	340100565860249600	AmericanEtrader
<input type="checkbox"/>	<a href="#">id=24757</a>	Thu May 30 13:26:50 +0000 2013	2013-06-01 00:55:36	Twitter		\$C	340096967940136960	xtranger51
<input type="checkbox"/>	<a href="#">id=30259</a>	Fri May 31 16:33:34 +0000 2013	2013-06-03 14:15:50	Twitter		\$INTC	340506348787732480	Wisdom_Folly
<input type="checkbox"/>	<a href="#">id=39126</a>	Wed May 29 13:51:09 +0000 2013	2013-06-04 00:25:13	Twitter		\$GM	339740699023011840	LargeVoidBot
<input type="checkbox"/>	<a href="#">id=46105</a>	2013-05-16T20:21:22Z	2013-05-31 23:10:29	StockTwits		RWT	13625258	cctranscripts
<input type="checkbox"/>	<a href="#">id=80693</a>	Wed May 29 19:52:44 +0000 2013	2013-06-01 01:12:03	Twitter		\$CHK	339831694519574528	CramersShirt
<input type="checkbox"/>	<a href="#">id=85145</a>	Thu May 30 19:10:11 +0000 2013	2013-06-01 00:47:54	Twitter		\$AMZN	340183372729028608	mqoptions
<input type="checkbox"/>	<a href="#">id=85146</a>	Thu May 30 15:39:58 +0000 2013	2013-06-01 00:47:56	Twitter		\$AMZN	340130470186602496	SeekingAlpha
<input type="checkbox"/>	<a href="#">id=85147</a>	Thu May 30 13:57:06 +0000 2013	2013-06-01 00:47:57	Twitter		\$AMZN	340104585865748481	SPEQL8R
<input type="checkbox"/>	<a href="#">id=85148</a>	Thu May 30 11:22:36 +0000 2013	2013-06-01 00:47:58	Twitter		\$AMZN	340065703900114944	AmazonByMaaya
<input type="checkbox"/>	<a href="#">id=85149</a>	Thu May 30 04:46:32 +0000 2013	2013-06-01 00:47:58	Twitter		\$AMZN	339966031328067584	AAinslie
<input type="checkbox"/>	<a href="#">id=85150</a>	Thu May 30 02:49:54 +0000 2013	2013-06-01 00:47:59	Twitter		\$AMZN	339936680184274944	AmazonByMaaya
<input type="checkbox"/>	<a href="#">id=85151</a>	Thu May 30 00:46:20 +0000 2013	2013-06-01 00:48:02	Twitter		\$AMZN	339905582280622081	SaintsSense

Return to original console Send feedback Privacy & Terms

# GAE Services

## Datastore

Google Developers Console

+Tony

COMPUTE

APP ENGINE

Dashboard

Logs

Instances

Versions

Task queues

Quota details

Memcache

Search

Settings

COMPUTE ENGINE

VM instances

Disks

Snapshots

Images

Networks

Network load balancing

Metadata

Zones

Operations

Quotas

STORAGE

CLOUD STORAGE

CLOUD DATASTORE

Dashboard

Query

Indexes

KIND

All Kinds

STORAGE SPACE BY ENTITY KIND

All Entity statistics

Last updated: 1 day 6 hr ago

Total size:	8.81GB	32.98GB	4.44GB
Entry count:	25,222,925	333,638,242	25,216,218

Breakdown by Property Type

TYPE	SIZE	INDEX SIZE
Blob	1.53KB	—
Boolean	41.61MB	536.68MB
Date/Time	625.3MB	4.91GB
Integer	70.88KB	11.43KB
Key	70.95KB	392.07KB
NULL	2.47MB	35.05MB
String	3.29GB	25.51GB
Text	2.76GB	—
Metadata	4.09GB	—

String

Date/Time

Composite Index

Metadata

Text

Other



# GAE Services

---

## Cron jobs

- Great for repetitive tasks
  - “Every 10 minutes, run this Python module.”
- Managed via the cron.yaml file:

```
cron:  
- description: daily summary job  
  url: /tasks/summary  
  schedule: every 24 hours  
- description: monday morning mailout  
  url: /mail/weekly  
  schedule: every monday 09:00  
  timezone: Australia/NSW  
- description: new daily summary job  
  url: /tasks/summary  
  schedule: every 24 hours  
  target: beta
```

# Storing Data and Data Processing

---

- Beyond the Datastore
  - Several Options:
    - Google Cloud Storage
    - Google Cloud SQL
    - Google BigQuery
    - Hadoop on Google Cloud (Requires Google Compute Engine)
-

# Storing Data and Data Processing

---

## Google Cloud Storage

- Store massive amounts of data offline, cheaply
  - Ideal for log files, old app data that doesn't need to be retrieved by app users.
- Example: 106GB ~\$3/month on Google Cloud Storage. 45GB ~\$8/month on GAE

# Storing Data and Data Processing

---

## Google Cloud SQL

- Run your own MySQL instances
  - No server configuration, encryption, replication, patch management or backups to set up. Google does it all.
-

# Storing Data and Data Processing

## Google BigQuery

- Tool to analyze large data sets

The screenshot displays the Google BigQuery web interface. On the left, a sidebar shows the project 'jibdanstartup' with a list of datasets including 'publicdata:samples', 'github\_nested', 'github\_timeline', 'gsod', 'natality', 'shakespeare', 'trigrams', and 'wikipedia'. The 'natality' dataset is selected. The main area is titled 'New Query' and contains a SQL query: 

```
1 SELECT year, state, is_male, child_race, avg(weight_pounds)
2 FROM [publicdata:samples.natality]
3 GROUP BY state, is_male, child_race
```

 Below the query, a green status bar indicates 'Valid: This query will process 3.47 GB when run.' and a 'RUN QUERY' button is visible. The 'Table Details: natality' section is expanded, showing a 'Description' field and a 'Table Info' table. The 'Table Info' table lists the following details: Table ID (publicdata:samples.natality), Table Size (21.9 GB), Number of Rows (137,826,763), Creation Time (7:47pm, 1 May 2012), and Last Modified (7:47pm, 1 May 2012). A 'Preview' section at the bottom shows the first row of data with columns: Row, source\_year, year, month, day, wday, state, is\_male, child\_race, weight\_pounds, plurality, apgar\_1min, apgar\_5min, and mother\_resid.

Google bigquery

COMPOSE QUERY

Query History  
Job History

jibdanstartup

- gniptest
- jibdantweetstream
- publicdata:samples
  - github\_nested
  - github\_timeline
  - gsod
  - natality**
  - shakespeare
  - trigrams
  - wikipedia

New Query

```
1 SELECT year, state, is_male, child_race, avg(weight_pounds)
2 FROM [publicdata:samples.natality]
3 GROUP BY state, is_male, child_race
```

Valid: This query will process 3.47 GB when run.

RUN QUERY Save Query Save View Enable Options Query complete (1.4s elapsed, cached)

Table Details: natality

Schema Details Query Table

Description

Describe this table...

Table Info

Table ID	publicdata:samples.natality
Table Size	21.9 GB
Number of Rows	137,826,763
Creation Time	7:47pm, 1 May 2012
Last Modified	7:47pm, 1 May 2012

Preview

Row	source_year	year	month	day	wday	state	is_male	child_race	weight_pounds	plurality	apgar_1min	apgar_5min	mother_resid
-----	-------------	------	-------	-----	------	-------	---------	------------	---------------	-----------	------------	------------	--------------

# Storing Data and Data Processing

---

## Hadoop on Google Cloud

- Processing data vs. analyzing data
  - Usually involves deploying a cluster of VMs, hence used on Google Compute Engine
  - Simpler App Engine solutions: MapReduce Python library, Managed VMs
-

# Real World Examples

---

