

# Google App Engine Data Store

ae-10-datastore

[www.appenginelearn.com](http://www.appenginelearn.com)

open.michigan

Unless otherwise noted, the content of this course material is licensed under a Creative Commons Attribution 3.0 License.

<http://creativecommons.org/licenses/by/3.0/>.

Copyright 2009, Charles Severance, Jim Eng



# Data @ Google is BIG

- Google's main applications are their search and mail
- The entire Internet and everyone's mail is a lot of data
- Traditional data storage approaches such as a relational database just don't scale to the size at which Google applications operate
- Sharded, sorted, array with hierarchical keys

<http://labs.google.com/papers/bigtable.html>

# Advanced Stuff

[All Sessions](#)

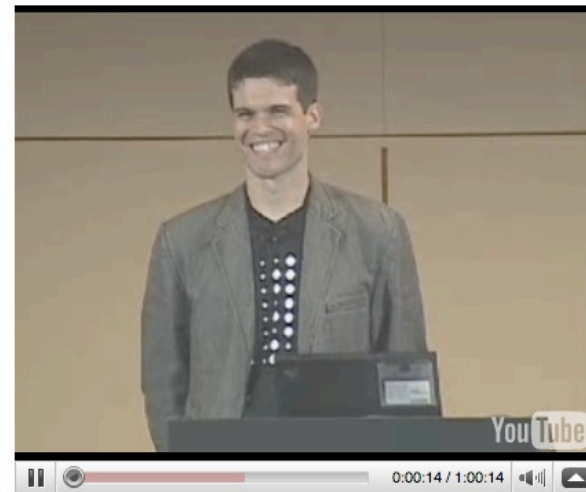
[Google I/O Website](#)

## Under the Covers of the Google App Engine Datastore

Ryan Barrett (Google)

Ever wonder why you can't do joins in the Google App Engine datastore? Why your app is seeing deadlines so often? Why it's so hard to tell whether a query will need an index? Why we offer both parent/child relationships and reference properties? Or why list properties don't seem to make any sense at all? This talk will explain how the datastore itself works, why these seeming peculiarities (and many others!) exist, and what you can do about them.

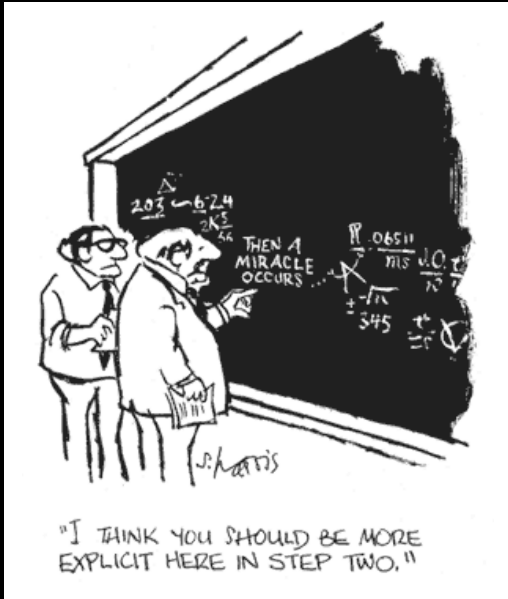
[Presentation Slides](#)



<http://sites.google.com/site/io/under-the-covers-of-the-google-app-engine-datastore>

# Model-View-Controller

Design Pattern



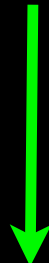
"I THINK YOU SHOULD BE MORE EXPLICIT HERE IN STEP TWO."

Web Server

HTTP Request



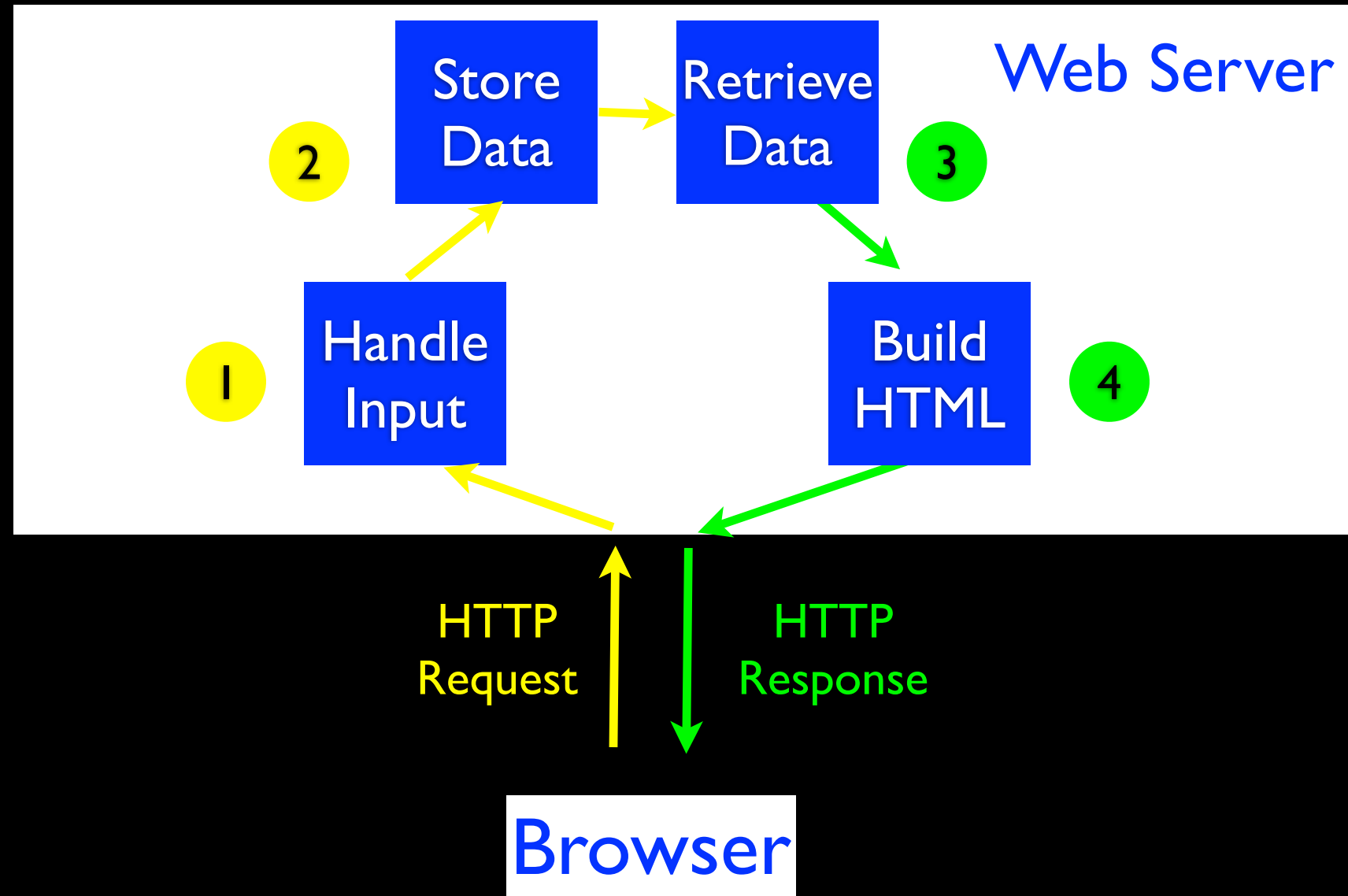
HTTP Response



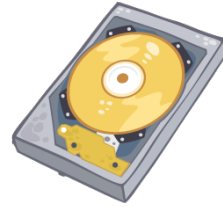
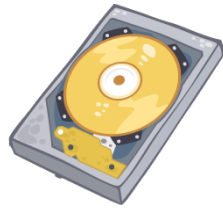
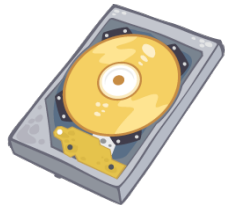
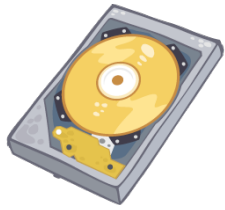
Browser

# Tasks Inside the Server

- Process any form input - possibly storing it in a database or making some other change to the database such as a delete
- Decide which screen to send back to the user
- Retrieve any needed data
- Produce the HTML response and send it back to the browser







Data Server

2

Store Data



Retrieve Data

3

1

Handle Input



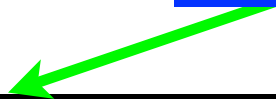
Build HTML

4

Web Server

Request

Response



# Terminology

- We call the Data bit - the “**Model**” or Data Model
- We call the “making the next HTML” bit the “**View**” or “Presentation Layer”
- We call the handling of input and the general orchestration of it all the “**Controller**”

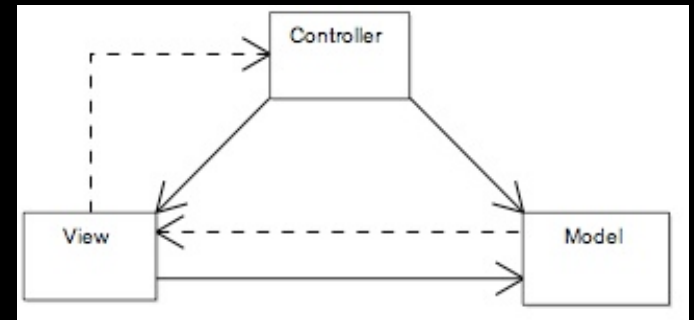
# Model View Controller

- We name the three basic functions of an application as follows
  - **Controller** - The Python code that does the thinking and decision making
  - **View** - The HTML, CSS, etc. which makes up the look and feel of the application
  - **Model** - The persistent data that we keep in the data store

<http://en.wikipedia.org/wiki/Model-view-controller>

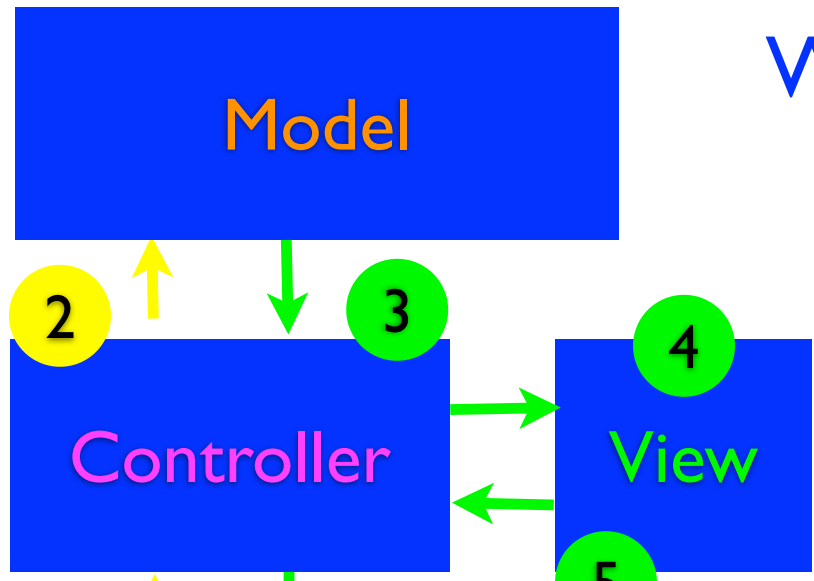
# Model-View-Controller

*“In MVC, the **model** represents the information (the data) of the application and the business rules used to manipulate the data; the **view** corresponds to elements of the user interface such as text, checkbox items, and so forth; and the **controller** manages details involving the communication to the model of user actions.”*



<http://en.wikipedia.org/wiki/Model-View-Controller>

Web Server



HTTP Request

HTTP Response

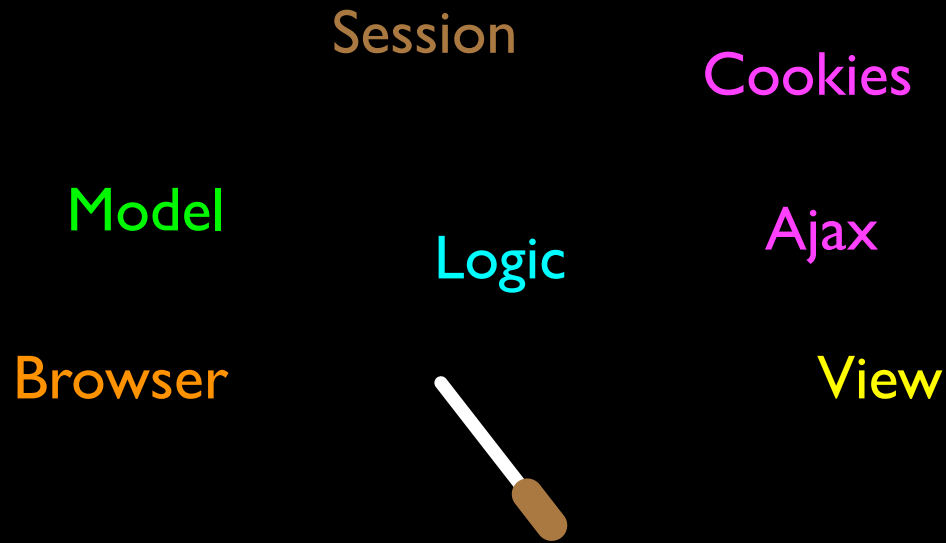
Browser

# Our Architecture: MVC

- **Model** - Holds the permanent data which stays long after the user has closed their web browsers
- **View** - Produces the HTML Response
- **Controller** - Receives each request and handles input and orchestrates the other elements

# Controller “Orchestrates”

<http://www.kapralova.org/MORAL.htm>



The controller is the conductor of all of the other aspects of MVC.

# Adding Models to our Application

ae-10-datastore

<http://code.google.com/appengine/docs/datastore/>



# Django Models



- Thankfully we use a very simple interface to define **objects** (a.k.a. **Models**) and store them in BigTable
- Google's BigTable is where the **models** are stored
- We don't need to know the details
- The pattern of these **models** is taken from the Django project

<http://docs.djangoproject.com/en/dev/ref/models/instances/?from=olddocs>

# A Simple Model

```
from google.appengine.ext import db
```

```
# A Model for a User  
class User(db.Model):  
    acct = db.StringProperty()  
    pw = db.StringProperty()  
    name = db.StringProperty()
```

Each model is a Python **class** which extends the **db.Model** class.

```
newuser = User(name="Chuck", acct="csev", pw="pw")  
newuser.put()
```

# Property Types

- **StringProperty** - Any string
- **IntegerProperty** - An Integer Number
- **DateTimeProperty** - A date + time
- **BlobProperty** - File data
- **ReferenceProperty** - A reference to another model instance

<http://code.google.com/appengine/docs/datastore/>

Property class	Value type	Sort order
<a href="#">StringProperty</a>	<a href="#">str</a> <a href="#">unicode</a>	Unicode ( <a href="#">str</a> is treated as ASCII)
<a href="#">BooleanProperty</a>	<a href="#">bool</a>	<a href="#">False</a> < <a href="#">True</a>
<a href="#">IntegerProperty</a>	<a href="#">int</a> <a href="#">long</a>	Numeric
<a href="#">FloatProperty</a>	<a href="#">float</a>	Numeric
<a href="#">DateTimeProperty</a> <a href="#">DateProperty</a> <a href="#">TimeProperty</a>	<a href="#">datetime.datetime</a>	Chronological
<a href="#">ListProperty</a> <a href="#">StringListProperty</a>	<a href="#">list</a> of a supported type	If ascending, by least element; if descending, by greatest element
<a href="#">ReferenceProperty</a> <a href="#">SelfReferenceProperty</a>	<a href="#">db.Key</a>	By path elements (kind, ID or name, kind, ID or name...)
<a href="#">UserProperty</a>	<a href="#">users.User</a>	By email address (Unicode)
<a href="#">BlobProperty</a>	<a href="#">db.Blob</a>	(not orderable)
<a href="#">TextProperty</a>	<a href="#">db.Text</a>	(not orderable)
<a href="#">CategoryProperty</a>	<a href="#">db.Category</a>	Unicode

# Keep it simple for a while

```
from google.appengine.ext import db
```

```
# A Model for a User  
class User(db.Model):  
    acct = db.StringProperty()  
    pw = db.StringProperty()  
    name = db.StringProperty()
```

Each model is a Python **class** which extends the **db.Model** class.

```
newuser = User(name="Chuck", acct="csev", pw="pw");  
newuser.put();
```

# Inserting a User and listing Users

App Engine - HTML

http://localhost:8080/login

App Engine [Sites](#) [Topics](#) [Login](#)

## Please Log In

Please enter your id and password to log in to this site.

Account:

Password:

App Engine - HTML

http://localhost:8080/apply

App Engine [Sites](#) [Topics](#) [Login](#)

## New Account Request

Please enter your information below:

Name:

Account:

Password:

App Engine - HTML

http://localhost:8080/memb...

App Engine [Sites](#) [Topics](#) [Members](#) [Logout \(csev\)](#)

## Members

Name	Account	Password
Chuck	csev	pw

```
class ApplyHandler(webapp.RequestHandler):
```

```
def post(self):
```

```
    self.session = Session()
    xname = self.request.get('name')
    xacct = self.request.get('account')
    xpw = self.request.get('password')
```

```
    # Check for a user already existing
    que = db.Query(User).filter("acct =",xacct)
    results = que.fetch(limit=1)
```

```
    if len(results) > 0 :
        doRender(self,"apply.htm",{ 'error' : 'Account Already Exists' } )
        return
```

```
    newuser = User(name=xname, acct=xacct, pw=xpw);
    newuser.put();
    self.session['username'] = xacct
    doRender(self,"index.htm",{ })
```

Get Session

Form  
Data

App Engine - HTML

http://localhost:8080/apply

App Engine [Sites](#) [Topics](#) [Login](#)

## New Account Request

Please enter your information below:

Name:

Account:

Password:

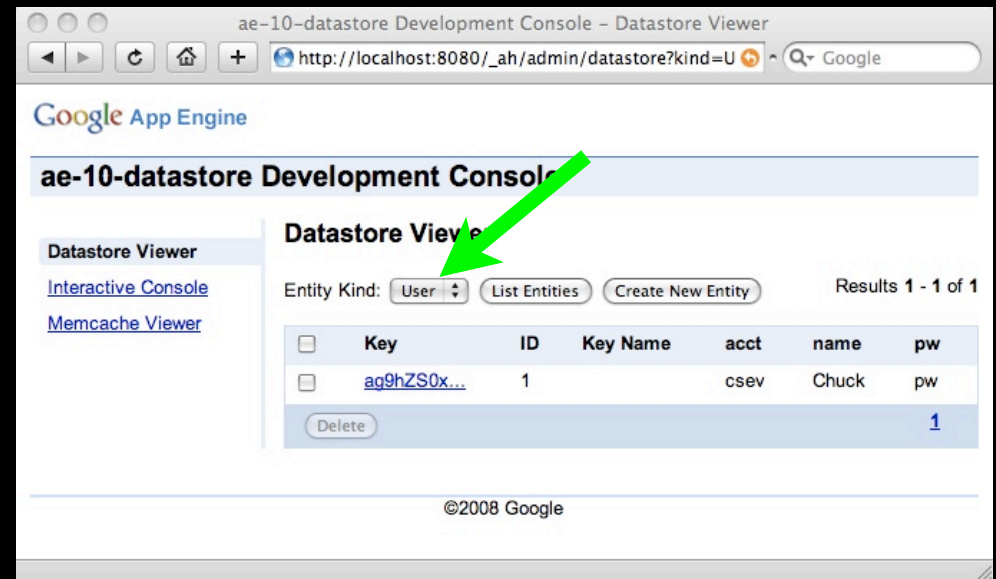
Check for  
existing user.

Insert User

Update Session

http://localhost:8080/\_ah/admin/

Using the Developer console we can see the results of the `put()` operation as a new `User` object is now in the data store.



The screenshot shows the Google App Engine Datastore Viewer interface. The browser address bar displays `http://localhost:8080/_ah/admin/datastore?kind=User`. The page title is "ae-10-datastore Development Console - Datastore Viewer". The main content area shows "Datastore Viewer" with a sidebar containing "Interactive Console" and "Memcache Viewer". The "Datastore Viewer" section displays "Entity Kind: User" and "Results 1 - 1 of 1". A table lists the entity details:

Key	ID	Key Name	acct	name	pw
<a href="#">aq9hZS0x...</a>	1		csev	Chuck	pw

A green arrow points to the "Entity Kind: User" dropdown menu. Below the table is a "Delete" button and a page number "1". The footer shows "©2008 Google".

```
newuser = User(name=xname, acct=xacct, pw=xpw);  
newuser.put();
```



```
class MembersHandler(webapp.RequestHandler):
```

```
    def get(self):
```

```
        que = db.Query(User)
```

```
        user_list = que.fetch(limit=100)
```

```
        doRender(self, 'memberscreen.htm',  
                 {'user_list': user_list})
```

We simply construct a **query for the User objects**, and **fetch the first 100 User Objects**. Then we pass this list into the **memberscreen.htm** template as a **context variable** named **'user\_list'**.

```
{% extends "_base.htm" %}
{% block bodycontent %}
  <h1>Members</h1>
  <p>
  <table>
  <tr><th>Name</th><th>Account</th><th>Password</th></tr>
  {% for user in user_list %}
    <tr>
      <td>{{ user.name }}</td>
      <td>{{ user.acct }}</td>
      <td>{{ user.pw }}</td>
    </tr>
  {% endfor %}
  </table>
{% endblock %}
```

templates/members.htm

In the template, we use the `for` directive to loop through each user in the `user_list` variable in the context. For each user we construct a table row with their name, account, and pw.

# Google App Engine References

ae-ll-chat

App Engine - HTML

http://localhost:8080/chat

App Engine [Sites](#) [Topics](#) Chat [Members](#) [Logout \(sally\)](#)

# Appengine Chat

Yes, it was surprisingly easy - make sure to look at the key() method (sally) Sat 22 Nov 2008

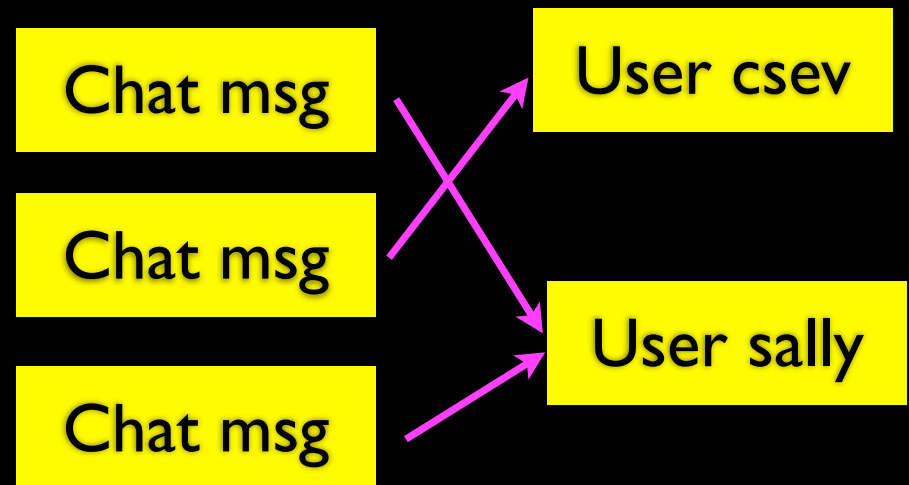
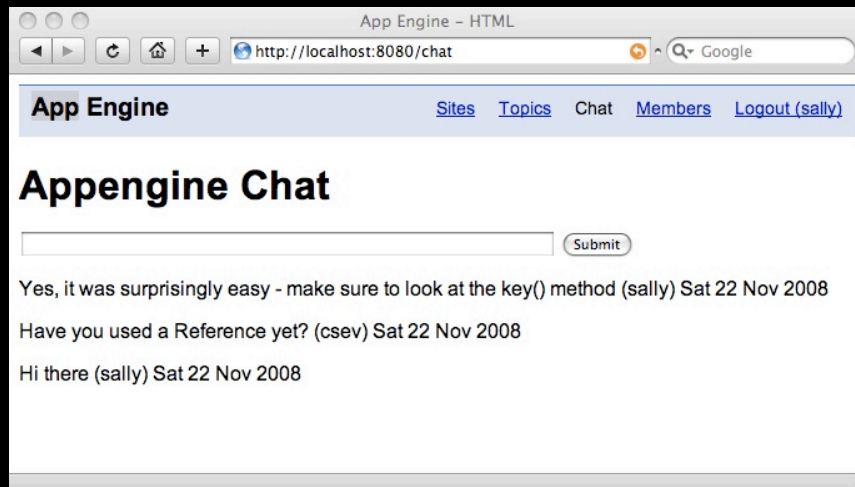
Have you used a Reference yet? (csev) Sat 22 Nov 2008

Hi there (sally) Sat 22 Nov 2008

Which user posted each message?

# Relationships

- We need to create a new model for Chat messages and then relate Chat messages by marking them as belonging to a particular user



# Three Kinds of Keys

- **Logical Key** - What we use to look something up from the outside world - usually unique for a model
- **Primary Key** - Some “random” number which tells the database where it put the data - also unique - and opaque
- **Reference** - When we have a field that points to the primary key of another model (a.k.a. **Foreign Key**)

```
class User(db.Model):  
    acct = db.StringProperty()  
    pw = db.StringProperty()  
    name = db.StringProperty()
```

User
acct
pw
name

```
class User(db.Model):  
    acct = db.StringProperty()  
    pw = db.StringProperty()  
    name = db.StringProperty()
```

User
key()
acct
pw
name

```
newuser = User(name=name, acct=acct, pw=pw)  
newuser.put()  
self.session['username'] = acct  
self.session['userkey'] = newuser.key()
```



```
class User(db.Model):  
    acct = db.StringProperty()  
    pw = db.StringProperty()  
    name = db.StringProperty()
```

User
key()
acct
pw
name

```
newuser = User(name=name, acct=acct, pw=pw)  
key = newuser.put();  
self.session['username'] = acct  
self.session['userkey'] = key
```

# Fast Lookup By Primary Key

- Lookup by **primary key** is faster than by **logical key** - because the **primary key** is about “where” the object is placed in the data store and there is *\*only one\**
- So we put it in **session** for later use...

```
newuser = User(name=name, acct=acct, pw=pw);  
key = newuser.put();  
self.session['username'] = acct  
self.session['userkey'] = key
```

# When we log in...

```
que = db.Query(User).filter("acct =",acct).filter("pw = ",pw)
results = que.fetch(limit=1)
if len(results) > 0 :
    user = results[0]
    self.session['username'] = acct
    self.session['userkey'] = user.key()
    doRender(self,"index.htm",{ } )
else:
    doRender(self,"loginscreen.htm",
             {'error' : 'Incorrect login data'} )
```

# When we log Out...

```
class LogoutHandler(webapp.RequestHandler):  
  
    def get(self):  
        self.session = Session()  
        self.session.delete_item('username')  
        self.session.delete_item('userkey')  
        doRender(self, 'index.htm')
```

When we log out - we make sure to remove the key from the session as well as the account name.

# Making References

# References

- When we make a new object that needs to be associated with or related to another object - we call this a “Reference”
- Relational Databases call these “Foreign Keys”

App Engine - HTML

http://localhost:8080/chat

App Engine [Sites](#) [Topics](#) Chat [Members](#) [Logout \(sally\)](#)

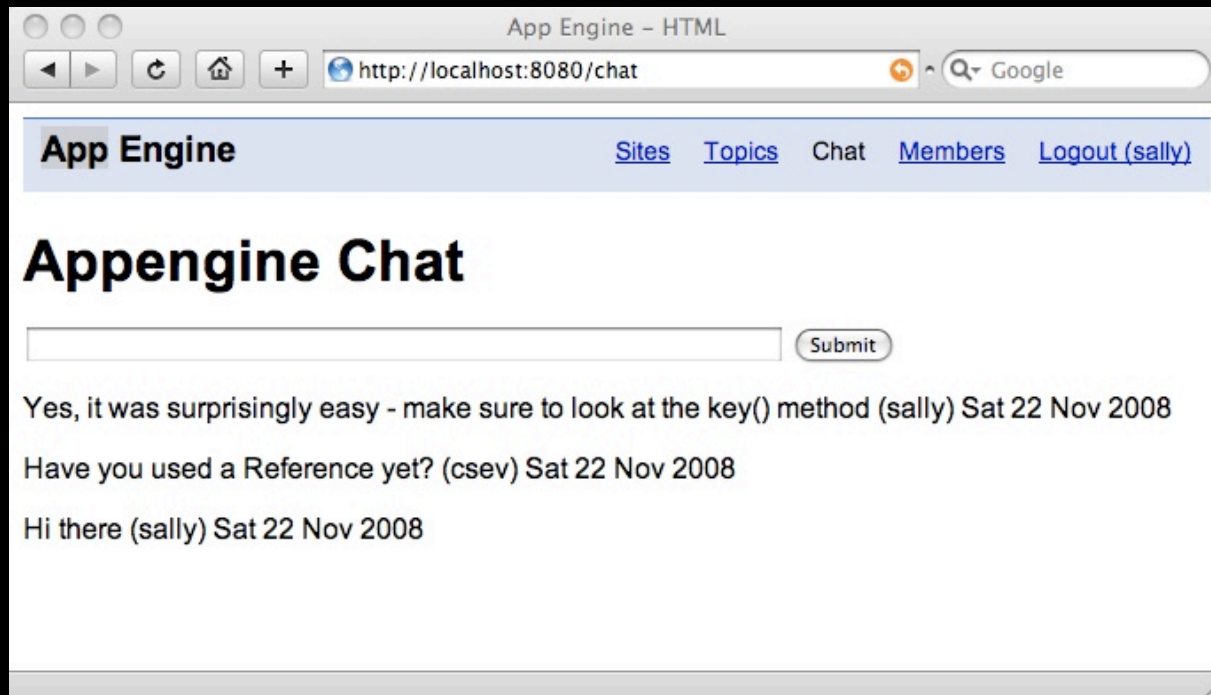
# Appengine Chat

Yes, it was surprisingly easy - make sure to look at the key() method (sally) Sat 22 Nov 2008

Have you used a Reference yet? (csev) Sat 22 Nov 2008

Hi there (sally) Sat 22 Nov 2008

Who posted each message?



## Database Normalization

We could just store the **account** strings in each chat message. This is bad practice generally - particularly if we might want to know more detail about the User later. We don't like to make multiple copies of anything except **primary keys**.

[http://en.wikipedia.org/wiki/Database\\_normalization](http://en.wikipedia.org/wiki/Database_normalization)

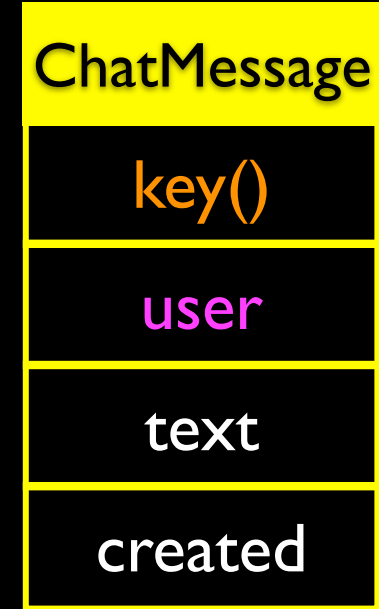
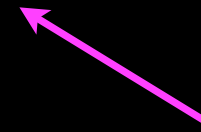
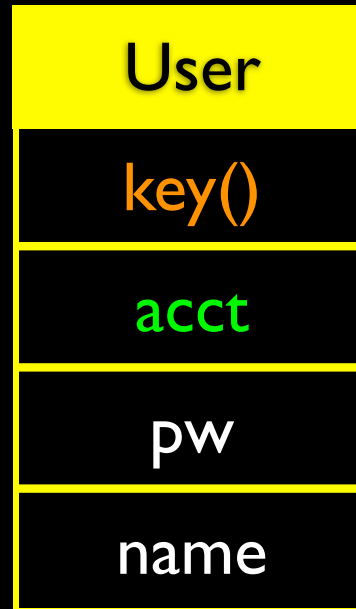


```
class ChatMessage(db.Model):  
    user = db.ReferenceProperty()  
    text = db.StringProperty()  
    created = db.DateTimeProperty(auto_now=True)
```

So we make a reference property in our Chat message model. The property does *\*not\** need to be named “user” - but it is a convenient pattern. Also note the created field that we let the data store auto-populate.

## Relating Models

```
class User(db.Model):  
    acct = db.StringProperty()  
    pw = db.StringProperty()  
    name = db.StringProperty()
```



```
class ChatMessage(db.Model):  
    user = db.ReferenceProperty()  
    text = db.StringProperty()  
    created = db.DateTimeProperty(auto_now=True)
```

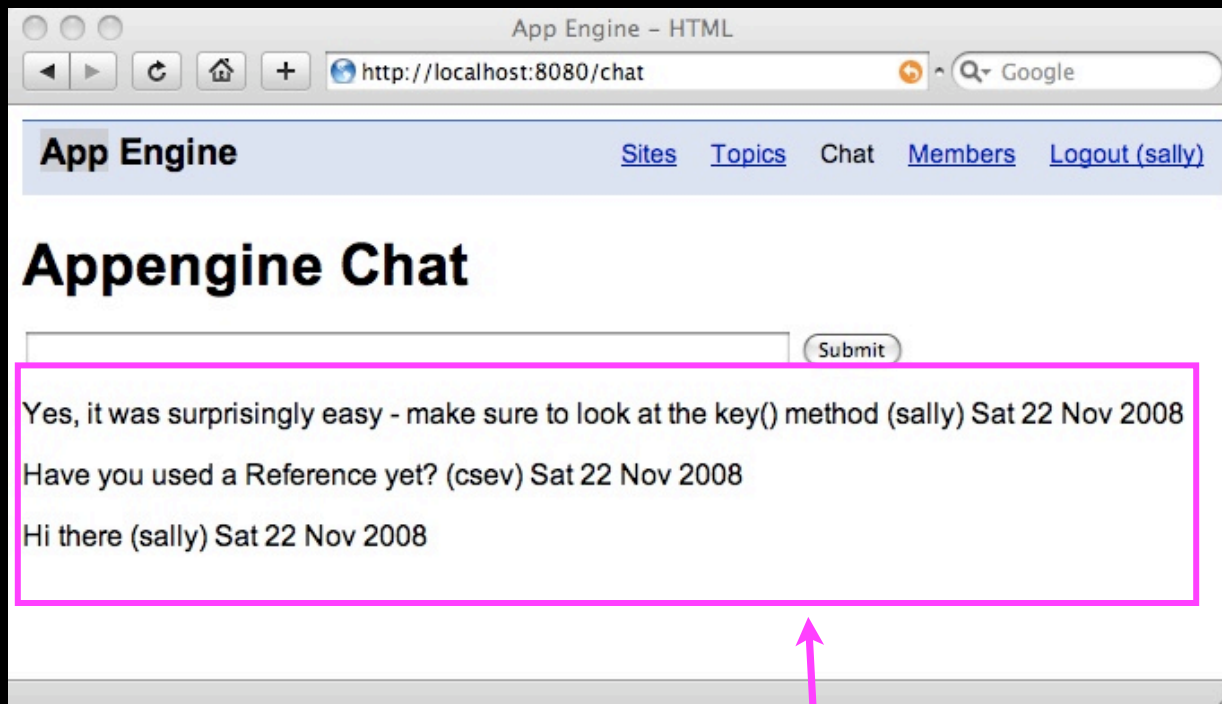
## Populating References

```
class ChatMessage(db.Model):
    user = db.ReferenceProperty()
    text = db.StringProperty()
    created = db.DateTimeProperty(auto_now=True)

def post(self):
    self.session = Session()

    msg = self.request.get('message')
    newchat = ChatMessage(user = self.session['userkey'], text=msg)
    newchat.put();
```

When we create a ChatMessage, we get the **message text** from the chatscreen.htm form, and then **user reference is the key of the current logged in user** taken from the Session. Note: Some error checking removed from this example.



Yes, it was surprisingly easy - make sure to look at the key() method (sally) Sat 22 Nov 2008  
Have you used a Reference yet? (csev) Sat 22 Nov 2008  
Hi there (sally) Sat 22 Nov 2008

We need to display the list of the most recent  
ChatMessage objects on the page.

```
def post(self):
    self.session = Session()
    msg = self.request.get('message')
    newchat = ChatMessage(user = self.session['userkey'], text=msg)
    newchat.put();

    que = db.Query(ChatMessage).order("-created");
    chat_list = que.fetch(limit=10)
    doRender(self, "chatscreen.htm",
             { 'chat_list': chat_list })
```

We retrieve the list of chat messages, and pass them into the template as **context variable** named “chat\_list” and then render “chatscreen.htm”.

ChatMessage
key()
user
text
created

## chatscreen.htm

```
{% extends "_base.htm" %}
{% block bodycontent %}
    <h1>Appengine Chat</h1>
    <p>
        <form method="post" action="/chat">
            <input type="text" name="message" size="60"/>
            <input type="submit" name="Chat"/>
        </form>
    </p>
    {% ifnotequal error None %}
        <p>
            {{ error }}
        </p>
    {% endifnotequal %}
    {% for chat in chat_list %}
        <p>{{ chat.text }} ({{ chat.user.acct}})
            {{ chat.created|date:"D d M Y"}}</p>
    {% endfor %}
{% endblock %}
```

In the chatscreen.htm template, we **loop through the context variable and process each chat message.**

## chatscreen.htm

```
{% extends "_base.htm" %}
{% block bodycontent %}
    <h1>Appengine Chat</h1>
    <p>
        <form method="post" action="/chat">
            <input type="text" name="message" size="60"/>
            <input type="submit" name="Chat"/>
        </form>
    </p>
    {% ifnotequal error None %}
        <p>
            {{ error }}
        </p>
    {% endifnotequal %}
    {% for chat in chat_list %}
        <p>{{ chat.text }} ({{ chat.user.acct }})
            {{ chat.created|date:"D d M Y"}}</p>
    {% endfor %}
{% endblock %}
```

In the chatscreen.htm template, we **loop through the context variable and process each chat message.**

For a reference value we access the **.user attribute and then the .acct attribute within the .user related to this chat message.**

# Walking a reference

- The `chat_list` contains a list of chat objects
- The iteration variable `chat` is each chat object in the list
- `chat.user` is the associated user object (follow the reference)
- `chat.user.acct` is the user's account

```
{% for chat in chat_list %}  
  <p>{{ chat.text }} ({{chat.user.acct}})  
    {{chat.created|date:"D d M Y"}}</p>  
{% endfor %}
```



ae-11-chat Development Console - Datastore Viewer - Edit Entity

Google App Engine

ae-11-chat Development Console

Datastore Viewer

Interactive Console  
Memcache Viewer  
Task Queues  
Cron Jobs  
XMPP  
Inbound Mail

**Edit Entity** chat

Entity Kind: ChatMessage

Entity Key: agphZS0xMS1jaGF0chELEgtDaGF0TWVzc2FnZRglDA

ID: 37

text (string): Have you used a Reference yet?

user (Key): agphZS0xMS1jaGF0cgoLEgRVc2VyGAgM

created (datetime): 2008-11-22 13:49:55

Save Changes Delete

©2009 Google

ae-11-chat Development Console - Datastore Viewer - Edit Entity

Google App Engine

ae-11-chat Development Console

Datastore Viewer

Interactive Console  
Memcache Viewer  
Task Queues  
Cron Jobs  
XMPP  
Inbound Mail

**Edit Entity** chat.user

Entity Kind: User

Entity Key: agphZS0xMS1jaGF0cgoLEgRVc2VyGAgM

ID: 8

account (string): csev

password (string): secret

name (string): Chuck

Save Changes Delete

©2009 Google

```

{% for chat in chat_list %}
  <p>{{ chat.text }} ({{chat.user.acct}})
  {{chat.created|date:"D d M Y"}}</p>
{% endfor %}

```

## chatscreen.htm

```
{% extends "_base.htm" %}
{% block bodycontent %}
    <h1>Appengine Chat</h1>
    <p>
        <form method="post" action="/chat">
            <input type="text" name="message" size="60"/>
            <input type="submit" name="Chat"/>
        </form>
    </p>
    {% ifnotequal error None %}
        <p>
            {{ error }}
        </p>
    {% endifnotequal %}
    {% for chat in chat_list %}
        <p>{{ chat.text }} ({{ chat.user.acct}})
            {{ chat.created|date:"D d M Y"}}</p>
    {% endfor %}
{% endblock %}
```

To make the date format a little nicer we use a **date:formatter** which shows the day of week, day of month, month, and year.

# Summary

- All objects stored in the data store are given a **primary key** which we get from either the **put()** call or the **key()** call
- We place these keys in **ReferenceProperty** values to connect one model to another
- When an attribute is a reference property, we use syntax like **chat.user.acct** - to look up fields in the referenced object