Ephemeral Hadoop Clusters in the Cloud

Greg Fodor, Etsy
gfodor@etsy.com
about me

gfodor@etsy.com
@gfodor
Data Wrangler
about etsy
the world’s handmade marketplace
total members: 9,000,000
total active shops: 800,000
items listed: 9.5M
page views per month: >1B
2010 sales: $314.3M
lots of data
about this talk
ephemeral?
“elastic” to the extreme
how did we get here?
wanted to dip our toes
stop hitting the database
stop grepping log files
2 data sources -> S3
database snapshots

input:
nightly diffs
(SELECT * FROM <table> WHERE update_date > 1 day ago)

output:
full tables as sequence files
visit logs

input:
akamai access logs
(event beacons)

output:
[visit_id, [event]]
processing the data
data flow
joins, group bys, etc.
cascading
Chris Wensel
http://www.cascading.org/
great implementation
Java syntax
cascading.jruby
Grégoire Marabout (Qualtera), Matt Walker (Etsy), Stefan Karpinski (Etsy), Steve Mardenfeld (Etsy)

github: http://bit.ly/o3DNTc
blog: http://etsy.me/cFytuL
cascade 'tag_counter', :input_prefix, :output_prefix do
  flow 'tag_counter' do
    tap_db_snapshot "listings"

    assembly "listings" do
      split_rows "tags", "tag",
      :pattern => /\./,
      :output => ["listing_id", "tag"]

      branch "tag_counts" do
        group_by "tag" do
          count "tag_count"
        end
      end
    end
  end
end

sink 'tag_counts', :scheme => text_line_scheme
end
“push” job binaries to S3

run on Elastic Map/Reduce
  starts cluster, runs, shuts down

access results on S3
next project:
shop recommendations
3 steps:

✔ data preparation - Cascading

✖ analysis/training

✖ prediction
sparse implementation of SVD
3 steps:

✔ data preparation - Cascading
✖ analysis/training - MATLAB
✖ prediction - MATLAB
“MATLAB, in my Hadoop cluster?”

It’s more likely than you think.

FREE PC CHECK!

CONTENTwatch™
hadoop streaming
arbitrary scripts for map & reduce
Swiss army knife

Full dataset analysis
Matlab, Ruby scripts

‘Artifact’ outputs
Tokyo Cabinet, Lucene, SQLite

Side-effects
MySQL, CloudFront
3 steps:
✔ data preparation - Cascading
✔ analysis/training - MATLAB
✔ prediction - MATLAB
Barnum
Sinatra web service on EC2
barnum starts job and passes callback URL

when job finishes, hadoop hits callback URL to barnum to proceed
Barnum Service
(Sinatra)

http://barnum.etsy.com
Barnum Service (Sinatra)

EMR Client

http://barnum.etsy.com
Barnum Service
(Sinatra)

EMR Client

http://barnum.etsy.com
Barnum constructs
Task 1

Job A
Cascading
45 c1.mediums
Task 1

Job A
Cascading
45 c1.mediums

Task 2

Job B

Job C
Chaining
3 steps:

✔ data preparation - Cascading
✔ analysis/training - MATLAB
✔ prediction - MATLAB
suggested_shops.yaml:

```
schedule: 15 07 * * *

transitions:
  START:
    SUCCEEDED: join_favorites
  join_favorites:
    SUCCEEDED: analysis
  analysis:
    SUCCEEDED: prediction
  prediction:
    SUCCEEDED: DONE

params:
  ALL:
    data_prefix: s3://data.etsy.com
    taste_prefix: s3://taste.etsy.com
  join_favorites:
    instance_type: c1.medium
    instance_count: 20
    min_listings: 3
  analysis:
    instance_type: m2.4xlarge
    instance_count: 1
  prediction:
    instance_type: m2.4xlarge
    instance_count: 10
    recommendations: 200
```

User table snapshot (S3)
Favorites table snapshot (S3)
suggested_shops.yaml:

```
schedule: 15 07 * * *

transitions:
  START:
    SUCCEEDED: join_favorites
  join_favorites:
    SUCCEEDED: analysis
  analysis:
    SUCCEEDED: prediction
  prediction:
    SUCCEEDED: DONE

params:
  ALL:
    data_prefix: s3://data.etsy.com
    taste_prefix: s3://taste.etsy.com
  join_favorites:
    instance_type: c1.medium
    instance_count: 20
    min_listings: 3
  analysis:
    instance_type: m2.4xlarge
    instance_count: 1
  prediction:
    instance_type: m2.4xlarge
    instance_count: 10
    recommendations: 200
```

- User table snapshot (S3)
- Favorites table snapshot (S3)

```
join_favorites(min_listings:3)
  Cascading
  20 c1.mediums
```
suggested_shops.yaml:

```
schedule: 15 07 ***

transitions:
  START:
    SUCCEEDED: join_favorites
  join_favorites:
    SUCCEEDED: analysis
  analysis:
    SUCCEEDED: prediction
  prediction:
    SUCCEEDED: DONE

params:
  ALL:
    data_prefix: s3://data.etsy.com
    taste_prefix: s3://taste.etsy.com
  join_favorites:
    instance_type: c1.medium
    instance_count: 20
    min_listings: 3
  analysis:
    instance_type: m2.4xlarge
    instance_count: 1
  prediction:
    instance_type: m2.4xlarge
    instance_count: 10
    recommendations: 200
```

User table snapshot (S3)
Favorites table snapshot (S3)

```
join_favorites(min_listings:3)
   Cascading
   20 c1.mediums
```
suggested_shops.yaml:

```
schedule: 15 07 * * *
transitions:
  START:
    SUCCEEDED: join_favorites
join_favorites:
  SUCCEEDED: analysis
analysis:
  SUCCEEDED: prediction
prediction:
  SUCCEEDED: DONE
params:
  ALL:
    data_prefix: s3://data.etsy.com
taste_prefix: s3://taste.etsy.com
join_favorites:
  instance_type: c1.medium
  instance_count: 20
  min_listings: 3
analysis:
  instance_type: m2.4xlarge
  instance_count: 1
prediction:
  instance_type: m2.4xlarge
  instance_count: 10
  recommendations: 200
```

- User table snapshot (S3)
- Favorites table snapshot (S3)
- join_favorites(min_listings:3) Cascading 20 c1.mediums
- Joined Favorite Data (S3)
suggested_shops.yaml:

```yaml
schedule: 15 07 * * *

transitions:
  START:
    SUCCEEDED: join_favorites
  join_favorites:
    SUCCEEDED: analysis
  analysis:
    SUCCEEDED: prediction
  prediction:
    SUCCEEDED: DONE

params:
  ALL:
    data_prefix: s3://data.etsy.com
    taste_prefix: s3://taste.etsy.com
  join_favorites:
    instance_type: c1.medium
    instance_count: 20
    min_listings: 3
  analysis:
    instance_type: m2.4xlarge
    instance_count: 1
  prediction:
    instance_type: m2.4xlarge
    instance_count: 10
    recommendations: 200
```

- User table snapshot (S3)
- Favorites table snapshot (S3)
- join_favorites(min_listings:3)
  - Cascading
  - 20 c1.mediums
- Joined Favorite Data (S3)
suggested_shops.yaml:

```
schedule: 15 07 ***

transitions:
    START:
    SUCCEEDED: join_favorites
    join_favorites
    SUCCEEDED: analysis
    analysis:
        SUCCEEDED: prediction
    prediction:
        SUCCEEDED: DONE

params:
    ALL:
        data_prefix: s3://data.etsy.com
        taste_prefix: s3://taste.etsy.com
    join_favorites:
        instance_type: c1.medium
        instance_count: 20
        min_listings: 3
    analysis:
        instance_type: m2.4xlarge
        instance_count: 1
    prediction:
        instance_type: m2.4xlarge
        instance_count: 10
        recommendations: 200
```

User table snapshot (S3)
Favorites table snapshot (S3)
join_favorites(min_listings:3)
    Cascading
    20 c1.mediums
Joined Favorite Data (S3)
suggested_shops.yaml:

```
schedule: 15 07 * * *

transitions:
  START:
    SUCCEEDED: join_favorites
  join_favorites:
    SUCCEEDED: analysis
  analysis:
    SUCCEEDED: prediction
  prediction:
    SUCCEEDED: DONE

params:
  ALL:
    data_prefix: s3://data.etsy.com
    taste_prefix: s3://taste.etsy.com
  join_favorites:
    instance_type: c1.medium
    instance_count: 20
    min_listings: 3
  analysis:
    instance_type: m2.4xlarge
    instance_count: 1
  prediction:
    instance_type: m2.4xlarge
    instance_count: 10
    recommendations: 200
```

Diagram:
- User table snapshot (S3)
- Favorites table snapshot (S3)
- join_favorites(min_listings:3)
  - Cascading
  - 20 c1.mediums
- Joined Favorite Data (S3)
- analysis()
  - Hadoop Streaming (Matlab)
  - 1 m2.4xlarge
suggested_shops.yaml:

```yaml
schedule: 15 07 ***

transitions:
  START:
    SUCCEEDED: join_favorites
  join_favorites:
    SUCCEEDED: analysis
  analysis:
    SUCCEEDED: prediction
  prediction:
    SUCCEEDED: DONE

params:
  ALL:
    data_prefix: s3://data.etsy.com
    taste_prefix: s3://taste.etsy.com
  join_favorites:
    instance_type: c1.medium
    instance_count: 20
    min_listings: 3
  analysis:
    instance_type: m2.4xlarge
    instance_count: 1
  prediction:
    instance_type: m2.4xlarge
    instance_count: 10
    recommendations: 200
```

Diagram:
- User table snapshot (S3)
- Favorites table snapshot (S3)
- join_favorites(min_listings:3)
  - Cascading
  - 20 c1.mediums
- analysis()
  - Hadoop Streaming (Matlab)
  - 1 m2.4xlarge
- Joined Favorite Data (S3)
suggested_shops.yaml:

```yaml
schedule: 15 07 ***

transitions:
  START:
    SUCCEEDED: join_favorites
  join_favorites:
    SUCCEEDED: analysis
  analysis:
    SUCCEEDED: prediction
  prediction:
    SUCCEEDED: DONE

params:
  ALL:
    data_prefix: s3://data.etsy.com
    taste_prefix: s3://taste.etsy.com
  join_favorites:
    instance_type: c1.medium
    instance_count: 20
    min_listings: 3
  analysis:
    instance_type: m2.4xlarge
    instance_count: 1
  prediction:
    instance_type: m2.4xlarge
    instance_count: 10
    recommendations: 200
```

Diagram:
- User table snapshot (S3)
- Favorites table snapshot (S3)
- `join_favorites(min_listings:3)`
  - Cascading
  - 20 c1.mediums
- `analysis()`
  - Hadoop Streaming (Matlab)
  - 1 m2.4xlarge
- Joined Favorite Data (S3)
- Matlab BIN matrix files (S3)
suggested_shops.yaml:

```
schedule: 15 07 * * *

transitions:
  START:
    SUCCEEDED: join_favorites
  join_favorites:
    SUCCEEDED: analysis
  analysis:
    SUCCEEDED: prediction
  prediction:
    SUCCEEDED: DONE

params:
  ALL:
    data_prefix: s3://data.etsy.com
    taste_prefix: s3://taste.etsy.com
  join_favorites:
    instance_type: c1.medium
    instance_count: 20
    min_listings: 3
  analysis:
    instance_type: m2.4xlarge
    instance_count: 1
  prediction:
    instance_type: m2.4xlarge
    instance_count: 10
    recommendations: 200
```

Diagram:

- User table snapshot (S3)
- Favorites table snapshot (S3)
- join_favorites(min_listings:3)
  - Cascading
  - 20 c1.mediums
  - analysis()
    - Hadoop Streaming (Matlab)
    - 1 m2.4xlarge
  - Joined Favorite Data (S3)
  - Matlab BIN matrix files (S3)
suggested_shops.yaml:

```
schedule: 15 07 * * *

transitions:
  START:
    SUCCEEDED: join_favorites
    join_favorites:
      SUCCEEDED: analysis
  analysis:
    SUCCEEDED: prediction
  prediction:
    SUCCEEDED: DONE

params:
  ALL:
    data_prefix: s3://data.etsy.com
    taste_prefix: s3://taste.etsy.com
  join_favorites:
    instance_type: c1.medium
    instance_count: 20
    min_listings: 3
  analysis:
    instance_type: m2.4xlarge
    instance_count: 1
  prediction:
    instance_type: m2.4xlarge
    instance_count: 10
    recommendations: 200
```

Diagram:
- User table snapshot (S3)
- Favorites table snapshot (S3)
- join_favorites(min_listings:3)
  - Cascading
  - 20 c1.mediums
- analysis()
  - Hadoop Streaming (Matlab)
  - 1 m2.4xlarge
- Joined Favorite Data (S3)
- Matlab BIN matrix files (S3)
suggested_shops.yaml:

```yaml
schedule: 15 07 ***

transitions:
  START:
    SUCCEEDED: join_favorites
  join_favorites:
    SUCCEEDED: analysis
  analysis:
    SUCCEEDED: prediction
  prediction:
    SUCCEEDED: DONE

params:
  ALL:
    data_prefix: s3://data.etsy.com
    taste_prefix: s3://taste.etsy.com
  join_favorites:
    instance_type: c1.medium
    instance_count: 20
    min_listings: 3
  analysis:
    instance_type: m2.4xlarge
    instance_count: 1
    min_listings: 3
  prediction:
    instance_type: m2.4xlarge
    instance_count: 1
    recommendations: 200
```

**Diagram:**

- **User table snapshot (S3)**
- **Favorites table snapshot (S3)**
- **join_favorites(min_listings:3)**
  - **Cascading**
  - **20 c1.mediums**
- **analysis()**
  - **Hadoop Streaming (Matlab)**
  - **1 m2.4xlarge**
- **prediction(recommendations:200)**
  - **Hadoop Streaming (Matlab)**
  - **10 m2.4xlarge**
  - **Joined Favorite Data (S3)**
  - **Matlab BIN matrix files (S3)**
suggested_shops.yaml:

```yaml
schedule: 15 07 ***

transitions:
  START:
    SUCCEEDED: join_favorites
  join_favorites:
    SUCCEEDED: analysis
  analysis:
    SUCCEEDED: prediction
  prediction:
    SUCCEEDED: DONE

params:
  ALL:
    data_prefix: s3://data.etsy.com
    taste_prefix: s3://taste.etsy.com
  join_favorites:
    instance_type: c1.medium
    instance_count: 20
    min_listings: 3
  analysis:
    instance_type: m2.4xlarge
    instance_count: 1
    min_listings: 3
  prediction:
    instance_type: m2.4xlarge
    instance_count: 10
    recommendations: 200
```
suggested_shops.yaml:

```
schedule: 15 07 * * *

transitions:
    START:
        SUCCEEDED: join_favorites
    join_favorites:
        SUCCEEDED: analysis
    analysis:
        SUCCEEDED: prediction
    prediction:
        SUCCEEDED: DONE

params:
    ALL:
        data_prefix: s3://data.etsy.com
        taste_prefix: s3://taste.etsy.com
    join_favorites:
        instance_type: c1.medium
        instance_count: 20
        min_lists: 3
    analysis:
        instance_type: m2.4xlarge
        instance_count: 1
    prediction:
        instance_type: m2.4xlarge
        instance_count: 10
        recommendations: 200
```

Diagram:
- User table snapshot (S3)
- Favorites table snapshot (S3)
- join_favorites(min_lists:3)
  - Cascading
  - 20 c1.mediums
- analysis()
  - Hadoop Streaming (Matlab)
  - 1 m2.4xlarge
- prediction(recommendations:200)
  - Hadoop Streaming (Matlab)
  - 10 m2.4xlarge
- Predictions (S3)
suggested_shops.yaml:

schedule: 15 07 * * *

transitions:

START:
  SUCCEEDED: join_favorites
join_favorites:
  SUCCEEDED: analysis
analysis:
  SUCCEEDED: prediction
prediction:
  SUCCEEDED: DONE

params:

ALL:
  data_prefix: s3://data.etsy.com
  taste_prefix: s3://taste.etsy.com
join_favorites:
  instance_type: c1.medium
  instance_count: 20
  min_listings: 3
analysis:
  instance_type: m2.4xlarge
  instance_count: 1
prediction:
  instance_type: m2.4xlarge
  instance_count: 10
  recommendations: 200
getting data back to web stack?
ad-hoc shell scripts

TSV into unsharded MySQL

not re-usable
Hadoop Cluster
Hadoop Cluster

S3

AWS

Etsy

Etsy API

Gearman

MySQL

Memcache

CREATE TABLE + Bulk INSERT
datasets are versioned based upon job execution time
## Enqueue load

- **Target:**
- **Bucket:**
- **Path:**
- **Time Path:**
- **Pre-warm:**
  - Yes
  - Go

## Loads

<table>
<thead>
<tr>
<th>ID</th>
<th>Target</th>
<th>Time Path</th>
<th>Status</th>
<th>Total Jobs</th>
<th>Completed Jobs</th>
</tr>
</thead>
<tbody>
<tr>
<td>265643610</td>
<td>Metrics.KeywordDistribution</td>
<td>2011_06_27/19_25_41</td>
<td>available</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>265622642</td>
<td>Metrics.StemmedTags</td>
<td>2011_06_27/19_27_28</td>
<td>available</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>265352260</td>
<td>Metrics.KeywordDistribution</td>
<td>2011_06_27/15_01_30</td>
<td>available</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>265332236</td>
<td>Metrics.StemmedTags</td>
<td>2011_06_27/15_03_24</td>
<td>available</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>266271519</td>
<td>Metrics.Shop</td>
<td>2011_06_27/13_00_02</td>
<td>available</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>265175322</td>
<td>UserSimilarity</td>
<td>2011_06_27/07_18_03</td>
<td>available</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>265109940</td>
<td>TasteTest.ItemTermIdf</td>
<td>2011_06_27/07_16_03</td>
<td>available</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>266169519</td>
<td>TasteTest.ShopTastemakerMale</td>
<td>2011_06_27/07_16_03</td>
<td>available</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>265103860</td>
<td>TasteTest.PathEdge</td>
<td>2011_06_27/07_16_03</td>
<td>available</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>
MySQL Tables:

Memcache Cluster:
Output dataset <-> ORM Model
class EtsyModel_TasteTest_RecommendableFavoriteListing
    extends EtsyModel_DatasetRowBase {
PHP:

class EtsyModel_TasteTest_RecommendeableFavoriteListing
    extends EtsyModel_DatasetRowBase {

Cascading:

dataset "TasteTest_RecommendeableFavoriteListing" do
    project "user_id", "shop_id", "listing_id"
    unique "user_id", "shop_id", "listing_id"
    group_output_by "user_id"
end
PHP:

class EtsyModel_TasteTest_RecommnendableFavoriteListing extends EtsyModel_DatasetRowBase {

Cascading:

dataset "TasteTest_RecommnendableFavoriteListing" do
  project "user_id", "shop_id", "listing_id"
  unique "user_id", "shop_id", "listing_id"
  group_output_by "user_id"
end

PHP:

EtsyModel_Dataset::tap(
  "TasteTest_RecommnendableFavoriteListing"
);
Old tables regularly dropped
how we’re using this stack

analytics  (internal)

products  (external)
analytics
products
search quality recommendations
May 2011:
4,926 successful job runs
scale up from zero
isolation
isolation across runs
fresh machine each time
isolation between developers

no toe-stepping
heterogeneous clusters
big RAM when you need it
(but not when you don’t)
need one machine?
use one machine.
writing jobs
PHENOMENAL COSMIC POWERS
prototyping
run slow, unoptimized version on 500 machine for < $100
parameter tuning
Try N=1, 2, 5, 10 and see which results in best output
questions?
photo credits

[2] by Dan4th http://www.flickr.com/photos/43264265@N00/5371557240/
[3] by mandolux http://www.flickr.com/photos/73935252@N00/34418046/
[4] by The Suss-Man http://www.flickr.com/photos/8692813@N06/4580254188/
[5] by Stephen Rees http://www.flickr.com/photos/60142746@N00/214461223/
[8] by ViaMoi http://www.flickr.com/photos/12187843@N07/3343619603/
[10] clipart (really)