Creating an Effective Data Platform

Kurt Brown
Context
400 Billion Events / Day
Philosophy
Who chooses the technologies?
The North Star
Netflix Expense Policy

“Act in Netflix’s Best Interest”
Impact vs. Perfection
Question Everything
Does it really matter?
Whose Data Platform is it?

- mine
- yours
Whose Data Platform is it?

- [ ] mine
- [x] yours
Visibility
I want....
I want....
The “FLESS theorem”

Flexibility

Speed

Scale
Flexibility and Scale

E.g.
Flexibility and Speed

In-memory DBs

E.g. MySQL
Speed and Scale

E.g.

* cassandra
* druid
Or more balanced trade-offs

E.g.

presto

TERADATA
Get your head in the cloud
Cost
Cost
Elasticity

- Grow naturally
- Spikes
- Reservations
- Spot
Services

• EMR, SQS, S3,…
• Ready for action
• Focus!
Redundancy / Global Presence

- Multi-DC
- International
Why not?
Cloud Mindset

- Self-service
- Distributed systems
- Auto-scaling
- ....
Cost?
Black Boxes and Building Blocks

Input

Black Box

Output
SELECT * FROM dse.geo_country_d
select * from dse.geo_country_d

Query Text:
select * from dse.geo_country_d

Parameters:

Status: RUNNING

Job Type: HiveJob
Genie Job Id: 9a3b8auc-fea9-11e4-8201-12f997ac03e5
Start Time: 5/19/2015 9:35:12 PM
Update Time: 5/19/2015 9:35:12 PM

Links:

select * from dse.geo_country_d

Query Text:
select * from dse.geo_country_d

Parameters:
To reattach to this job using Kragle:

```python
import kragle as kg
job = kg.genie.reattach_job('9a3b8aac-fea9-11e4-8201-12f997ac03a5')
```

To fetch the first 100 rows of data from genie stdout:

```python
kg.transport.Transporter().source(kg.genie.reattach_job('9a3b8aac-fea9-11e4-8201-12f997ac03a5')).execute().row_map(limit=100)
```

To resubmit this query as a new job:

```python
job = kg.genie.HiveJob().script('"select * from dse.geo_country_d"
running_job = job.execute()
```
```python
In[1]: import krags as kg

In[2]: job = kg.genie.PrestoJob()
    .script('select * from dse.geo_country_d')
    .headers()
    .execute()

In[3]: job.pandas()
```

Out[3]:
```
<table>
<thead>
<tr>
<th>country_iso_code</th>
<th>country_sk</th>
<th>country_desc</th>
<th>country_full_desc</th>
<th>company_code</th>
<th>subregion_sk</th>
<th>subregion_pd_sk</th>
<th>content_subregion_sk</th>
<th>content</th>
</tr>
</thead>
<tbody>
<tr>
<td>FR</td>
<td>8728</td>
<td>France</td>
<td>France</td>
<td>65</td>
<td>14</td>
<td>14</td>
<td>15</td>
<td>France</td>
</tr>
<tr>
<td>CM</td>
<td>8700</td>
<td>Cameroon</td>
<td>Cameroon</td>
<td>-1</td>
<td>9</td>
<td>3</td>
<td>9</td>
<td>Other</td>
</tr>
<tr>
<td>AR</td>
<td>8582</td>
<td>Argentina</td>
<td>Argentina</td>
<td>33</td>
<td>8</td>
<td>4</td>
<td>8</td>
<td>Latin A</td>
</tr>
<tr>
<td>MD</td>
<td>8788</td>
<td>Moldova</td>
<td>Moldova, Republic</td>
<td>-1</td>
<td>9</td>
<td>3</td>
<td>9</td>
<td>Other</td>
</tr>
</tbody>
</table>
```
Who do we hire?

- “That guy” / “That woman”
- Not married to a tech
- Been knocked around some
- A hires A, B hires C
- Pay for the top
Netflix Culture Deck

Google Search: “Netflix Culture Deck”