



JanusGraph FOR

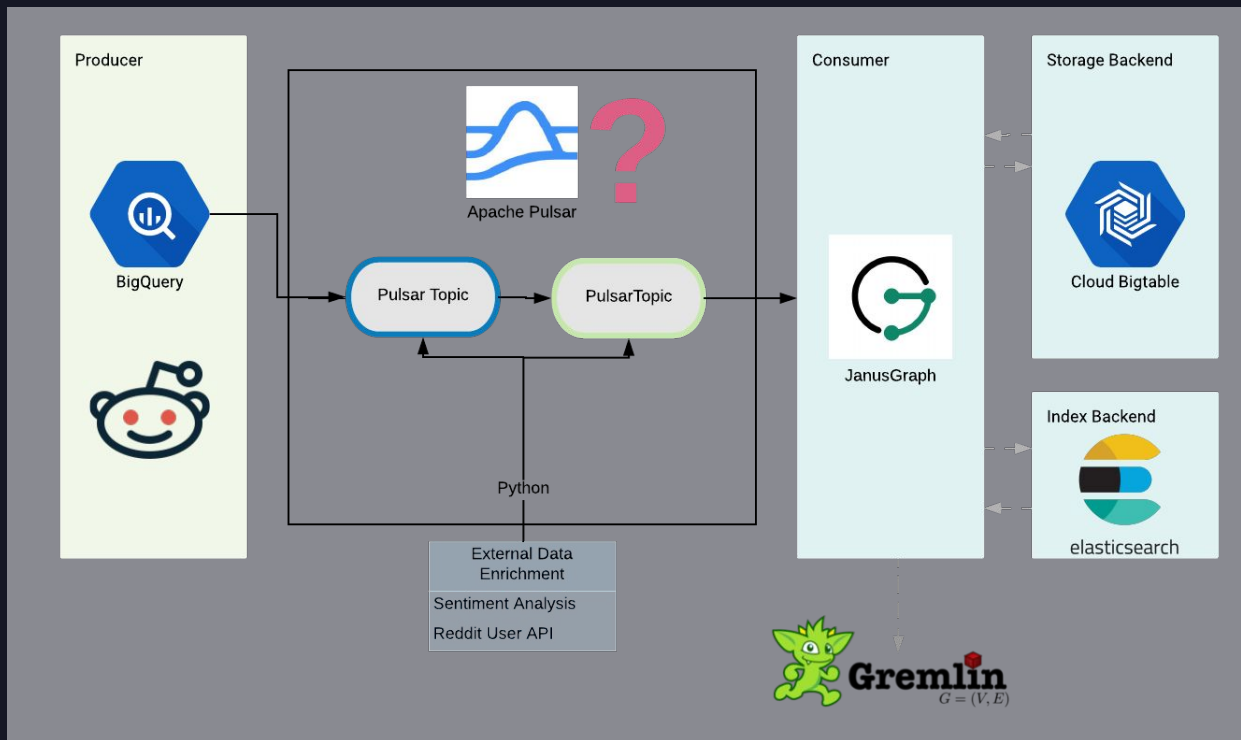


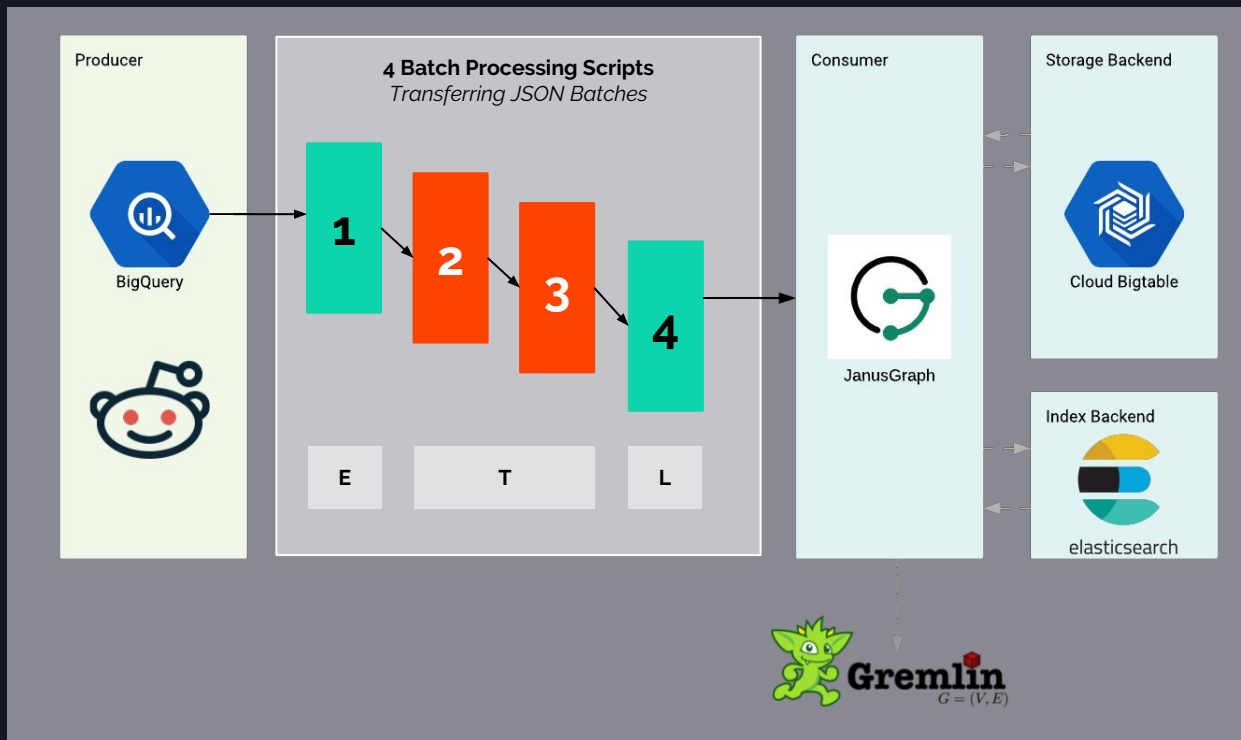
# reddit COMMENT SENTIMENT ANALYSIS



Presented by ***DATANAUTS INC.*** ✨







# STEP 1 (of 4): Extract Data From Google BigQuery



Extract

Transform

Load



## BigQuery Data Extraction Script



Google Cloud Platform  
Account / Credentials



google-cloud-bigquery  
Python Package

```
{
  "body": "Netflix are behind the excellent [RxJava](https://github.com/ReactiveX/RxJava) project and the other ReactiveX components. They definitely have some elite Java talent there.",
  "score_hidden": false,
  "archived": false,
  "name": "t1_crhah4d",
  "author": "kitd",
  "author_flair_text": null,
  "downs": 0,
  "created_utc": 1432288686,
  "subreddit_id": "t5_2fwo",
  "link_id": "t3_36uyww",
  "parent_id": "t3_36uyww",
  "score": 73,
  "retrieved_on": 1433204446,
  "controversiality": 0,
  "gilded": 0,
  "id": "crhah4d",
  "subreddit": "programming",
  "ups": 73,
  "distinguished": null,
  "author_flair_css_class": null
}
```

...

1 BigQuery

2 Google Cloud Platform

3 JSON File

## STEP 2 (of 4): Analyze Comment Sentiment



Extract

Transform

Load

```
{ "body": "Netflix are behind the excellent [RxJava](https://github.com/ReactiveX/RxJava) project and the other ReactiveX components. They definitely have some elite Java talent there.", "score_hidden": false, "archived": false, "name": "t1_crhah4d", "author": "kitd", "author_flair_text": null, "downs": 0, "created_utc": 1432288686, "subreddit_id": "t5_2fwo", "link_id": "t3_36uyvw", "parent_id": "t3_36uyvw", "score": 73, "retrieved_on": 1433204446, "controversiality": 0, "gilded": 0, "id": "crhah4d", "subreddit": "programming", "ups": 73, "distinguished": null, "author_flair_css_class": null }
...

```

### Sentiment Analysis Script

<http://sentiment.nrc.ca/lexicons-for-research/>



```
{ "score_hidden": false, "archived": false, "name": "t1_crhah4d", "author": "kitd", "author_flair_text": null, "downs": 0, "created_utc": 1432288686, "subreddit_id": "t5_2fwo", "link_id": "t3_36uyvw", "parent_id": "t3_36uyvw", "score": 73, "retrieved_on": 1433204446, "controversiality": 0, "gilded": 0, "id": "crhah4d", "subreddit": "programming", "ups": 73, "distinguished": null, "author_flair_css_class": null, "sentiment_score": 0.13043478260869565 }
...

```

### “Sentiment Score”

= average sentiment of comment

$$= \frac{\text{sum of scores of words}}{\text{num of words}}$$

Range: (-1, 1)

(- 1.0) ← Neutral → (+ 1.0)  
Negative      Neutral      Positive

## STEP 3 (of 4): Supplement User Data



Extract

Transform

Load

```
{ "score_hidden": false, "archived": false, "name": "t1_cr49kci", "author": "wampum",  
  "author_flair_text": null, "downs": 0, "created_utc": 1431234873, "subreddit_id": "t5_vf2",  
  "link_id": "t3_35gpf9", "parent_id": "t3_35gpf9", "score": 40, "retrieved_on": 1432873482,  
  "controversiality": 0, "gilded": 0, "id": "cr49kci", "subreddit": "nsfw", "ups": 40,  
  "distinguished": null, "author_flair_css_class": null, "sentiment_score": 0.0 }  
...
```

### Supplementary Reddit Data App

**PRAW** =  
Python Reddit API Wrapper



```
{ "score_hidden": false, "archived": false, "name": "t1_cr49kci", "author": "wampum",  
  "author_flair_text": null, "downs": 0, "created_utc": 1431234873, "subreddit_id": "t5_vf2",  
  "link_id": "t3_35gpf9", "parent_id": "t3_35gpf9", "score": 40, "retrieved_on": 1432873482,  
  "controversiality": 0, "gilded": 0, "id": "cr49kci", "subreddit": "nsfw", "ups": 40,  
  "distinguished": null, "author_flair_css_class": null, "sentiment_score": 0.0,  
  "user_created_utc": 1299040426.0, "added_user_created_utc": true }  
...
```

usnm →

reddit's  
API

User found

Success

User found, SUSPENDED

Partial Data

User not found, DELETED

Dropped

## STEP 4 (of 4): Populate JanusGraph



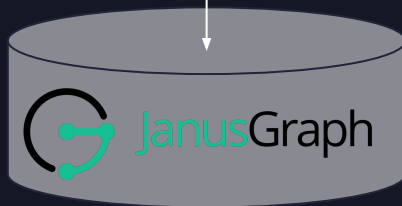
Extract

Transform

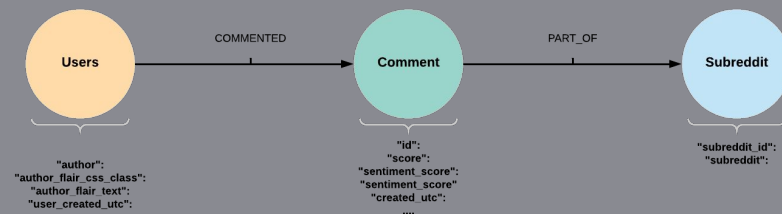
Load

```
{ "score_hidden": false, "archived": false, "name": "t1_cr49kci", "author": "wampum",  
  "author_flair_text": null, "downs": 0, "created_utc": 1431234873, "subreddit_id": "t5_vf2",  
  "link_id": "t3_35gpf9", "parent_id": "t3_35gpf9", "score": 40, "retrieved_on": 1432873482,  
  "controversiality": 0, "gilded": 0, "id": "cr49kci", "subreddit": "nsfw", "ups": 40,  
  "distinguished": null, "author_flair_css_class": null, "sentiment_score": 0.0,  
  "user_created_utc": 1299040426.0, "added_user_created_utc": true }  
...
```

### Data Writing Script



### JanusGraph Schema:









**TIME  
SINCE  
JOINED**

**COMMENT  
SENTIMENT**

**REDDIT  
USER**

**SUBREDDIT**

