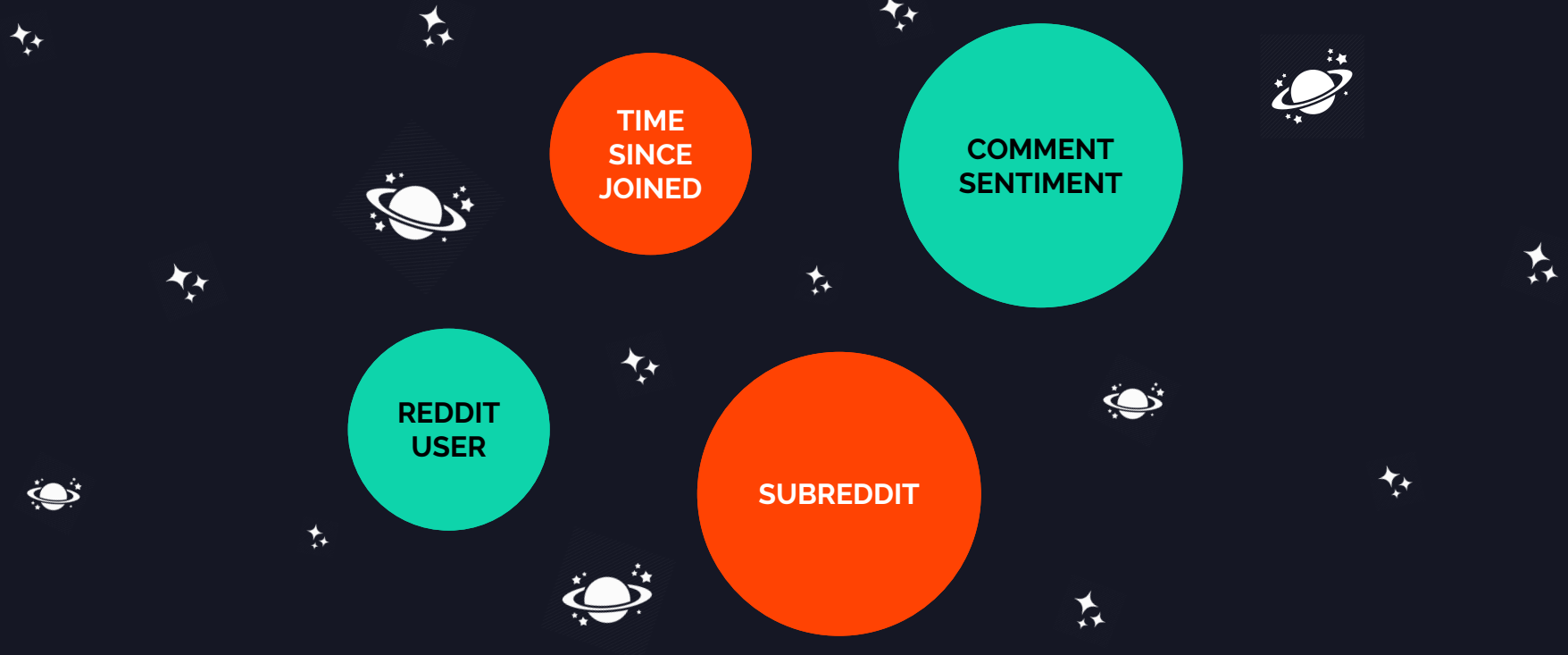
JanusGraph FOR

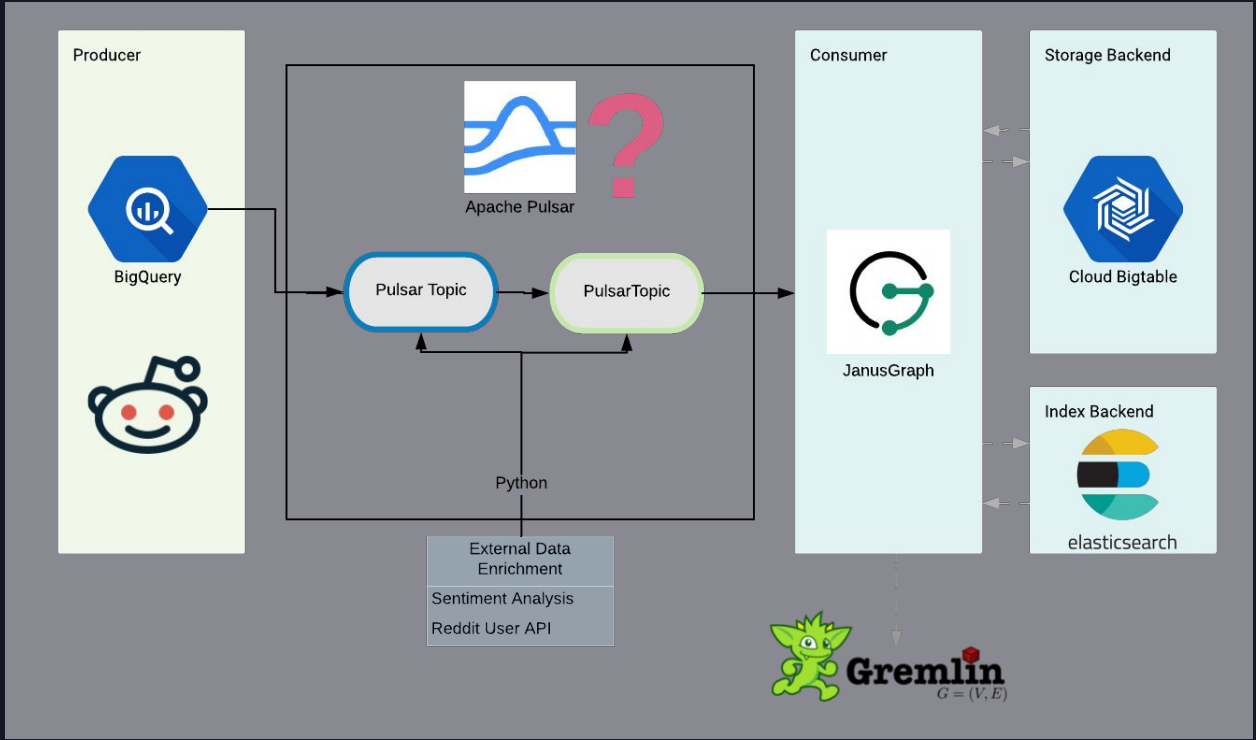
reddit COMMENT SENTIMENT ANALYSIS

Presented by ***DATANAUTS INC.***

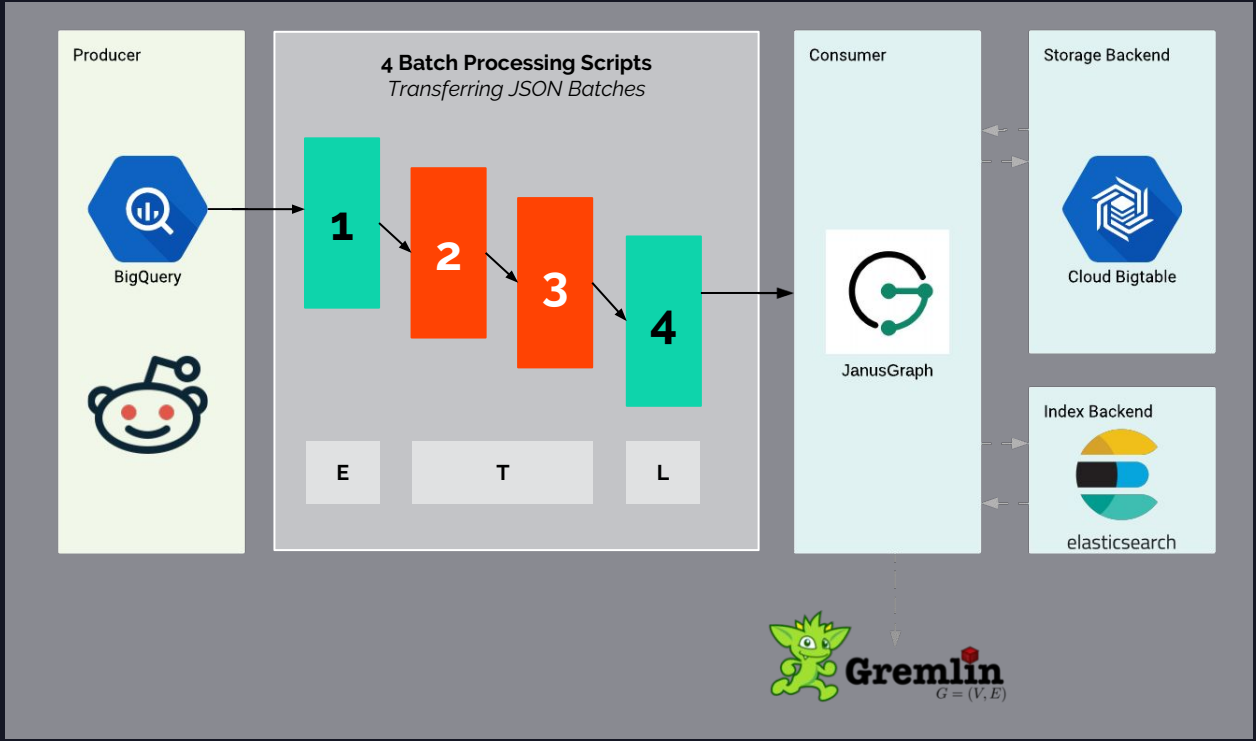
DATANAUTS INC. Explored Reddit User, Sentiment, and Subreddit Relationships



Original System Design



Current System Design



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_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}
...
```


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_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, "sentiment_score": 0.13043478260869565}
...
```


"Sentiment Score"

= average sentiment of comment

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

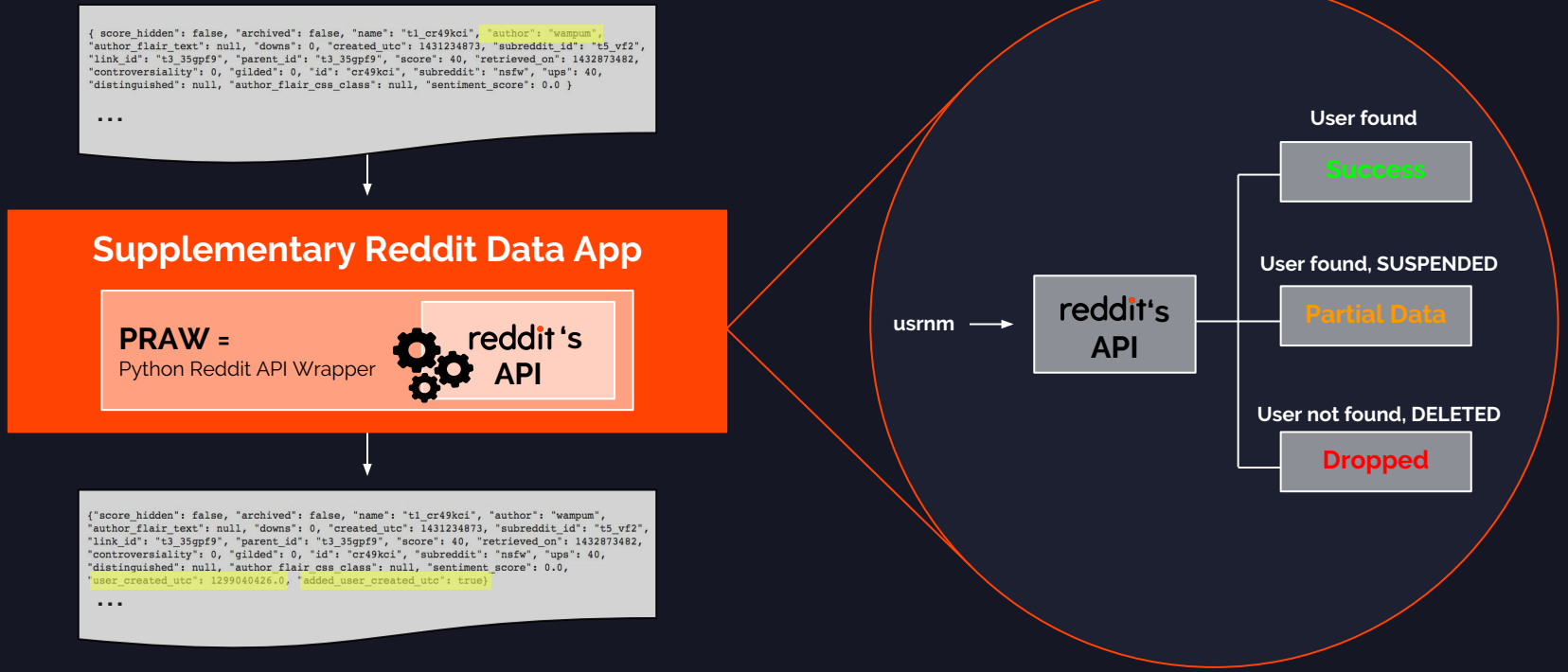
Range: (-1, 1)

(- 1.0) ⟷ (+ 1.0)

Negative **Neutral** **Positive**

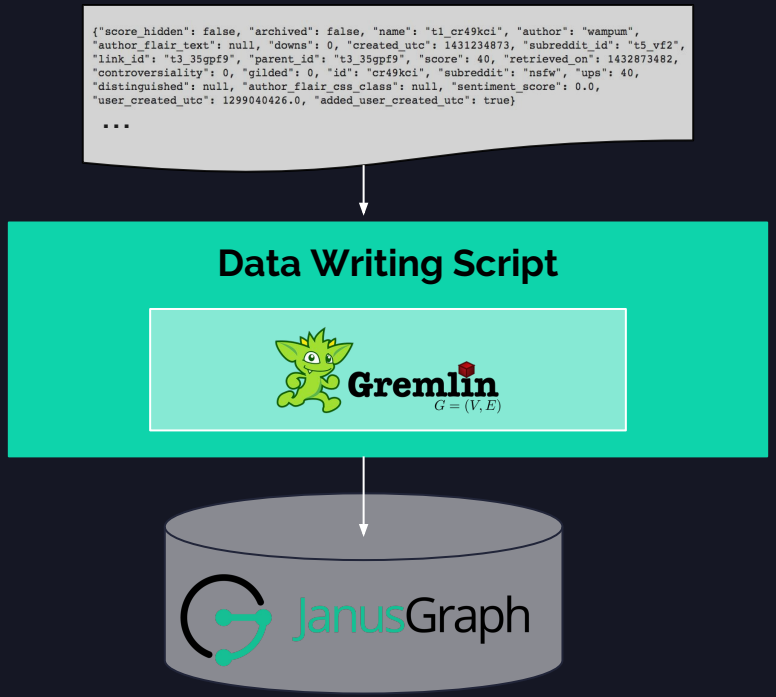
STEP 3 (of 4): Supplement User Data

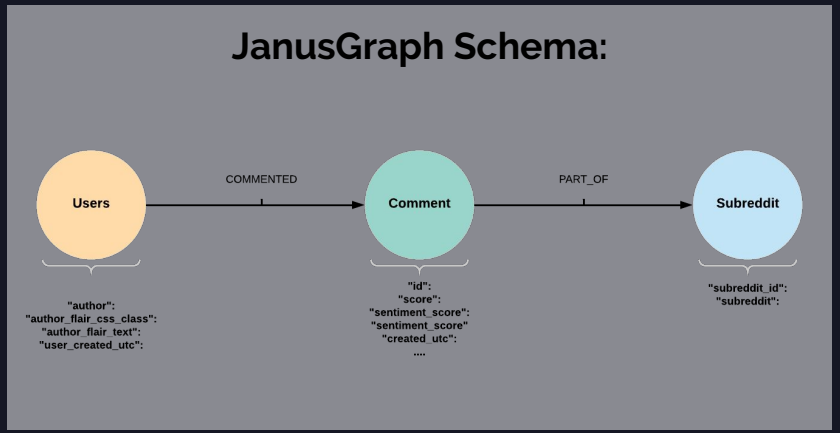
Extract | Transform | Load



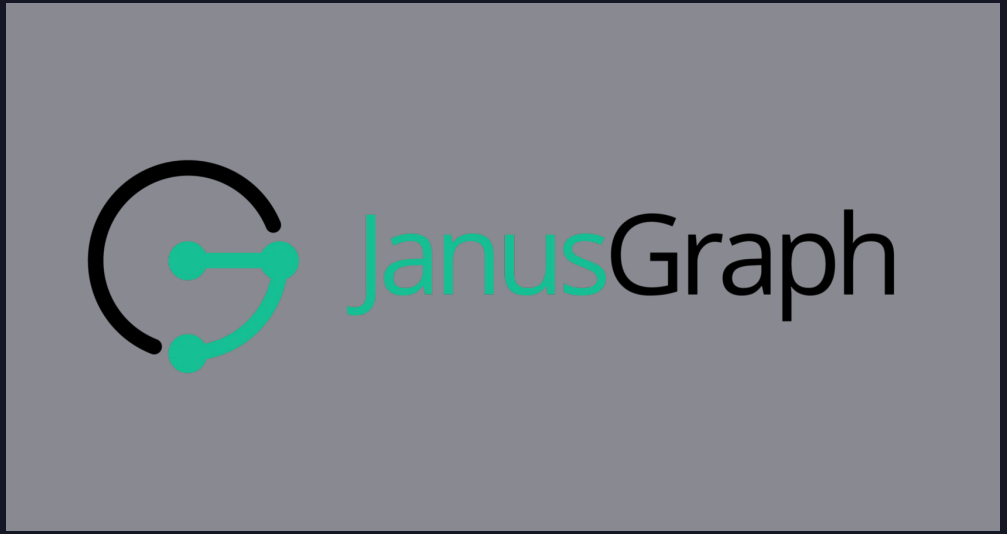
STEP 4 (of 4): Populate JanusGraph

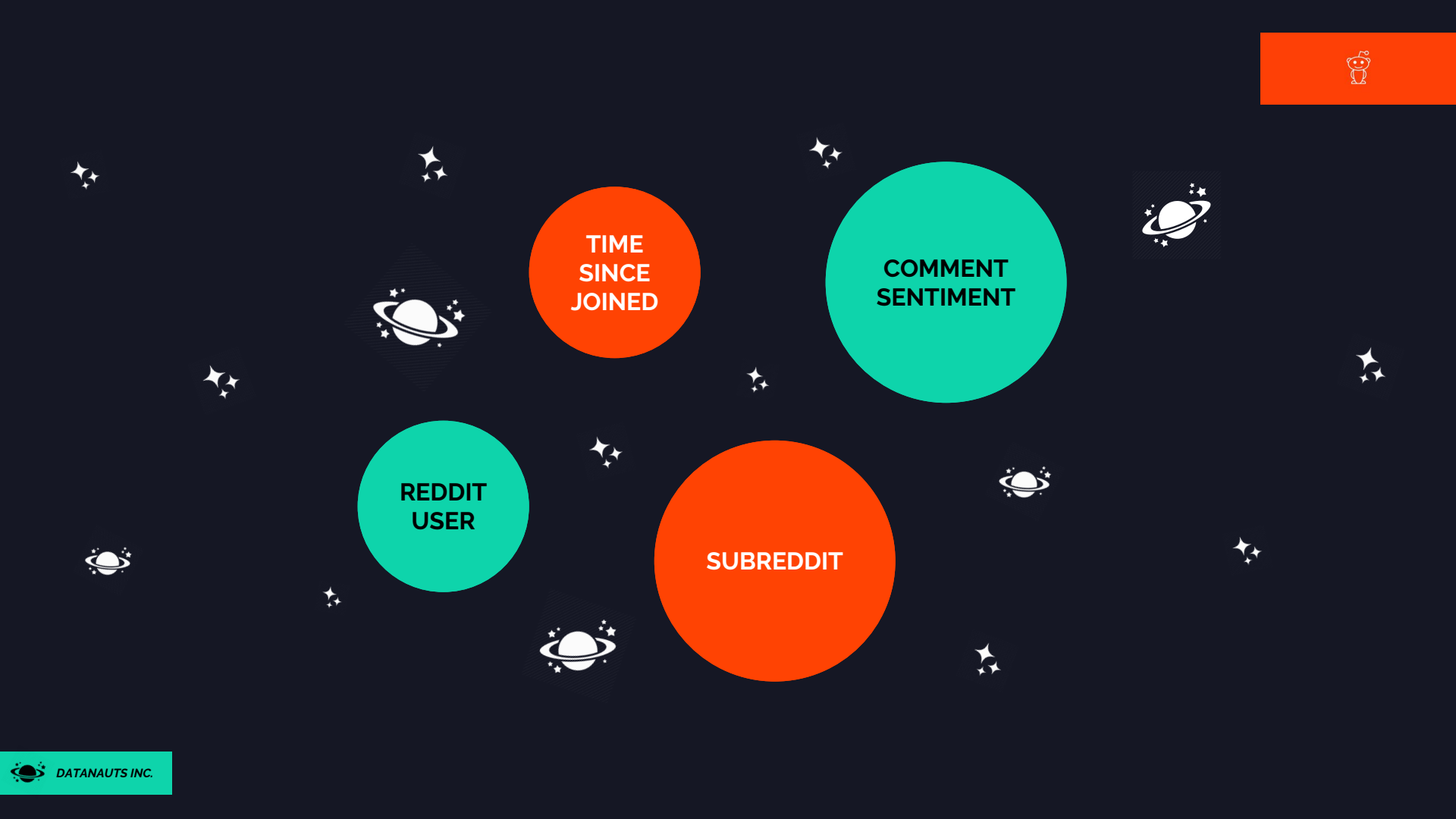
Extract | Transform | Load





DEMO




TIME SINCE JOINED
COMMENT SENTIMENT
REDDIT USER
SUBREDDIT