Welcome!
We will get started shortly!

elastic

Agenda

1. Welcome
2. Introduction
3. Elastic and JPMC
4. Empowering Data-Driven Decisions: Analyzing transactions with Elastic Kibana and Observability Tools
5. Elasticsearch Essentials: Data Loading with Python

elastic

Jessica Garson

@JessicaGarson

@JessicaGarson@macaw.social

Senior Developer Advocate

elastic

What’s new with Elastic?

8.15 is out!

elastic

semantic_text

Making RAG development much easier with simplified mapping, type and automatic chunking.

elastic

8.14

```
POST semantic-starwars/_doc
{
  "quote": "These are <em>not</em> the droids you are looking for.",
  "quote_e5": [ 0.5, 10, 6, ...]
}
GET semantic-starwars/_search
{
  "query": {
    "knn": {
      "field": "quote_e5"
      "k" : 10,
      "num_candidates": 100,
      "query_vector_builder": {
        "text_embedding": {
          "model_id": "e5-small-multilingual",
          "model_text": "search for an android"
}}}}}
```

elastic

8.15

```
POST semantic-starwars/_doc
{
  "quote": "These are <em>not</em> the droids you are looking for."
}
GET semantic-starwars/_search
{
  "query": {
    "semantic": {
      "field": "quote_e5"
      "query" : "search for an android"
}}}
```

elastic

What else is new?

- We've improved vector search with new features like hamming distance, bit vectors, and memory-efficient quantization, alongside SIMD optimization.
- LogsDB index mode 3 boosts log efficiency with better sorting and storage.
- Enhanced OpenTelemetry integration in Elastic Cloud simplifies log collection and host health monitoring.
- New API connectors for AI tasks include Google AI Studio and Vertex AI.
- ES|QL now supports advanced geospatial searches and more robust cross-cluster functionality.
- The latest Logstash monitoring provides detailed performance dashboards.

elastic

Learn more about 8.15

https://www.elastic.co/blog/whats-new-elastic-8-15-0

https://discuss.elastic.co/t/what-s-new-in-elastic-8-15/364628

elastic


Elasticsearch Essentials:
Data Loading with Python
for Interplanetary Insights
elastic

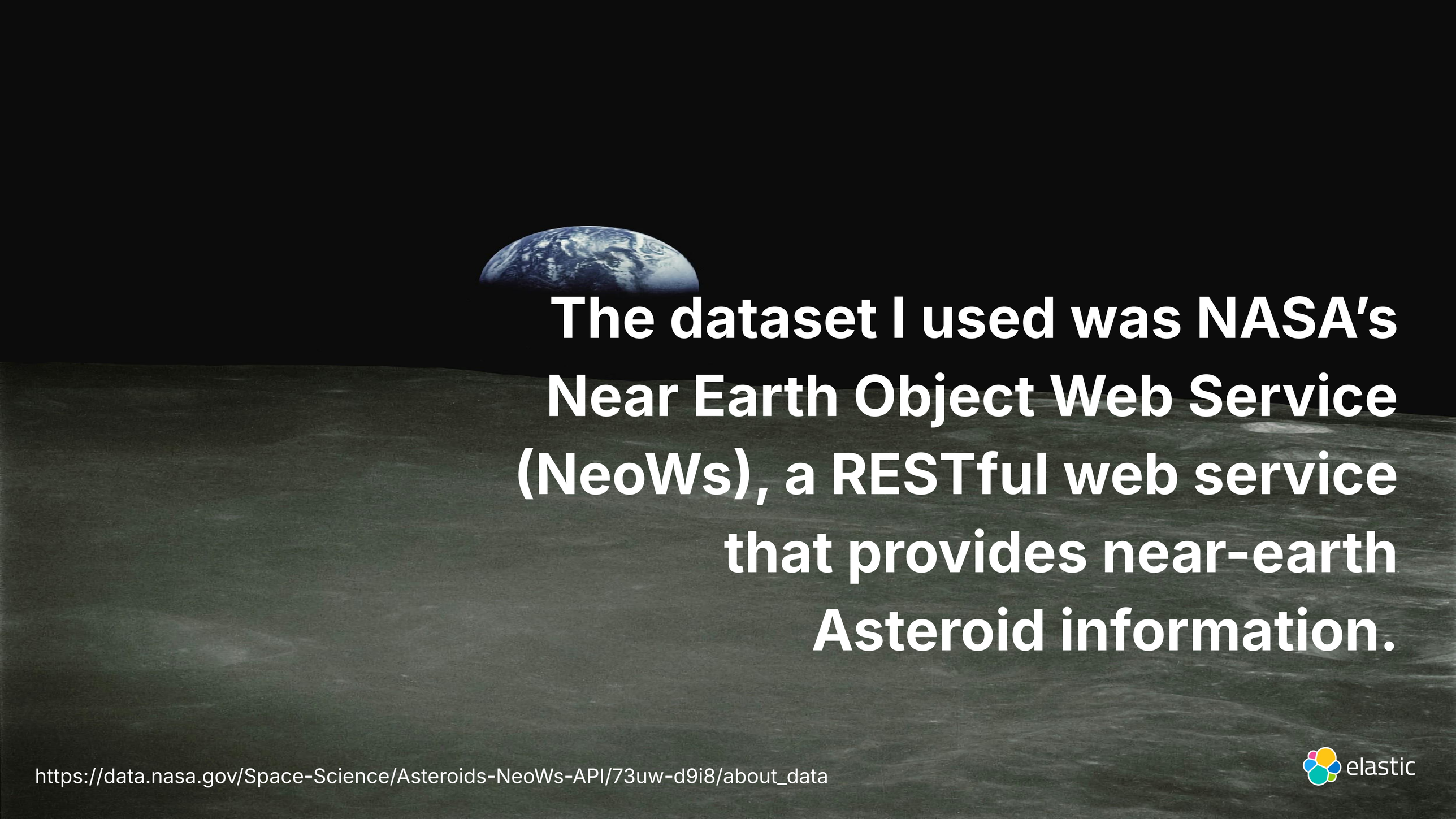
Code and slides

elastic

This talk will walk you through how to use the Python client as a toolkit 🧰

https://www.elastic.co/guide/en/elasticsearch/client/python-api/current/index.html

elastic



The dataset I used was NASA's Near Earth Object Web Service (NeoWs), a RESTful web service that provides near-earth Asteroid information.

https://data.nasa.gov/Space-Science/Asteroids-NeoWs-API/73uw-d9i8/about_data

elastic

What is an index?

An index inside Elasticsearch is a data structure where you can store your data in documents.

https://www.elastic.co/blog/what-is-an-elasticsearch-index

elastic

Schema definition

Each index has a mapping that defines the structure of the documents, specifying field types and how they should be indexed and stored.

https://www.elastic.co/blog/what-is-an-elasticsearch-index

elastic

Scalability and Performance

Elasticsearch indexes are designed to handle large volumes of data, offering fast search and retrieval capabilities through distributed architecture and advanced indexing techniques

https://www.elastic.co/blog/what-is-an-elasticsearch-index

elastic


Let's take a look at the dataset in Postman
https://www.postman.com/
elastic

What you need to get started

- Python > 3.8
- A NASA API key
- Elasticsearch > 8.x

https://wiki.python.org/moin/BeginnersGuide/Download
https://api.nasa.gov/
https://www.elastic.co/guide/en/elasticsearch/reference/current/getting-started.html

elastic

```
pip install requests pandas elasticsearch notebook
```

elastic

Demo

- Uploading data into a DataFrame
- Creating an index
- Loading our DataFrame into an index

elastic

Parallel Bulk

parallel_bulk() is a tool that enhances the bulk() API by using multiple threads to process tasks simultaneously.

I had a problem

When I first started working with Elasticsearch my data in my index would quickly go out of date.


Oh wait, I know how
to solve this!

Why use a time based solution for this problem?

Maintaining up-to-date data is crucial, especially when dealing with frequently changing dynamic datasets.

Let’s update our data once

elastic

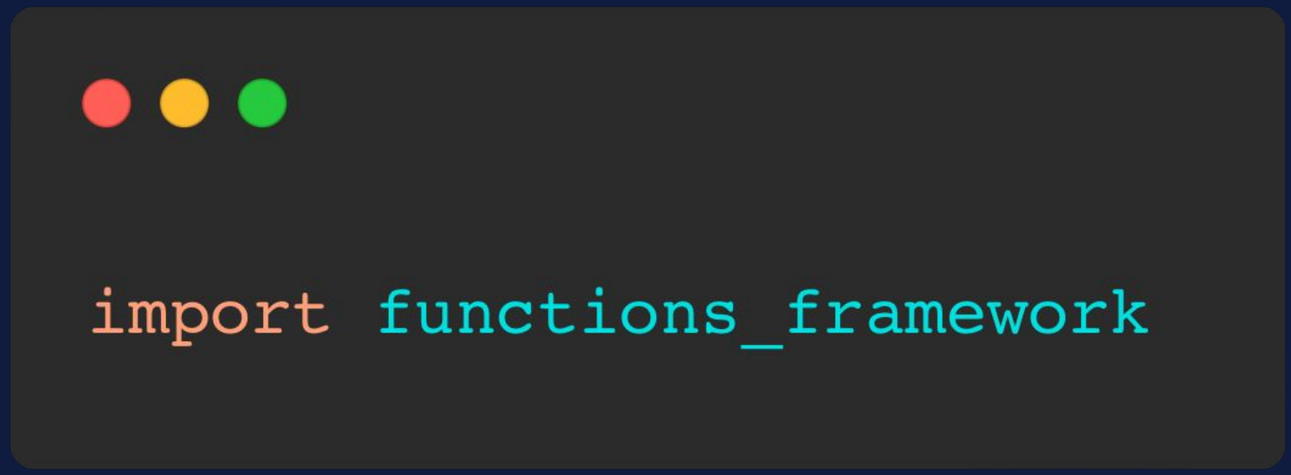

Using Google
Cloud Platform
elastic

Solution overview

https://cloud.google.com/scheduler
https://cloud.google.com/functions
https://www.elastic.co/search-labs/blog/keeping-your-elasticsearch-index-current-with-python-and-google-cloud-platform-functions

elastic

Import statement



elastic

Like a main function

```python
@functions_framework.cloud_event
def hello_pubsub(cloud_event):
    index_name = "asteroid_data_set"
    es = connect_to_elastic()
    last_update_date = updated_last(es, index_name)
    print(last_update_date)
    response = connect_to_nasa(last_update_date)
    print(response)
    df = create_df(response)
    try:
        if df is None:
            raise ValueError("DataFrame is None. There may be a problem.")
        update_new_data(df, es, last_update_date, index_name)
        print(updated_last(es, index_name))
    except Exception as e:
        print(f"An error occurred: {e}")
```

elastic

Demo

- Show configuration of the function
- Show the code
- Trigger a run
- Check in Elastic

You can also use an Azure Function App or AWS Lambda

https://github.com/JessicaGarson/Elasticsearch-Essentials-Data-Loading-with-Python-for-Interplanetary-Insights/tree/main/Azure%20and%20AWS

elastic

	GCP	Azure	AWS
Advantages	Similar to how I write code locally	Scheduling happens in the in the code itself.	Widely used in the community
Drawbacks	Not as widely used.	The process of creating a function took me the longest	You need to create a deployment zip.

Next steps

elastic

Using an ingest pipeline for Elasticsearch can be a natural next step for optimizing data uploads into an index, mainly if you deal with large volumes or complex data transformations.

https://www.elastic.co/guide/en/elasticsearch/reference/current/ingest.html

elastic

Depending on the size and frequency of data updates, consider batching data to reduce the number of API calls and to enhance performance.

Let me know if this talk inspires you to build anything. I'm @JessicaGarson on most platforms.

My email is jess.garson@elastic.co

elastic

Elastic Contributors

elastic.co/community/contributor

This community program is designed to recognize and reward the hard work of our awesome contributors. Join this friendly competition and earn points for:

Code contributions | Presentations | Video tutorials | Event organization
Translations | Technical Q&A | Written content | Content validation

The top contributors will win cool prizes such as **Elastic trainings**, **certificate exams, cloud credits** and more. Check out the rules on our website and start submitting today!

elastic

Get involved

The community would love to hear from you!

If you have a cool use case to share at a meetup, please let us know! We would love to make that happen.

And if your company would like to host a meetup - we have user groups in other cities as well, please let us know.

Send an email to meetups@elastic.co or ully@elastic.co

elastic

Let us know if you have any questions on our Discuss forums and the community Slack channel.

https://discuss.elastic.co/
https://communityinviter.com/apps/elasticstack/elastic-community

elastic


Thank you!
elastic