

DSA 8430

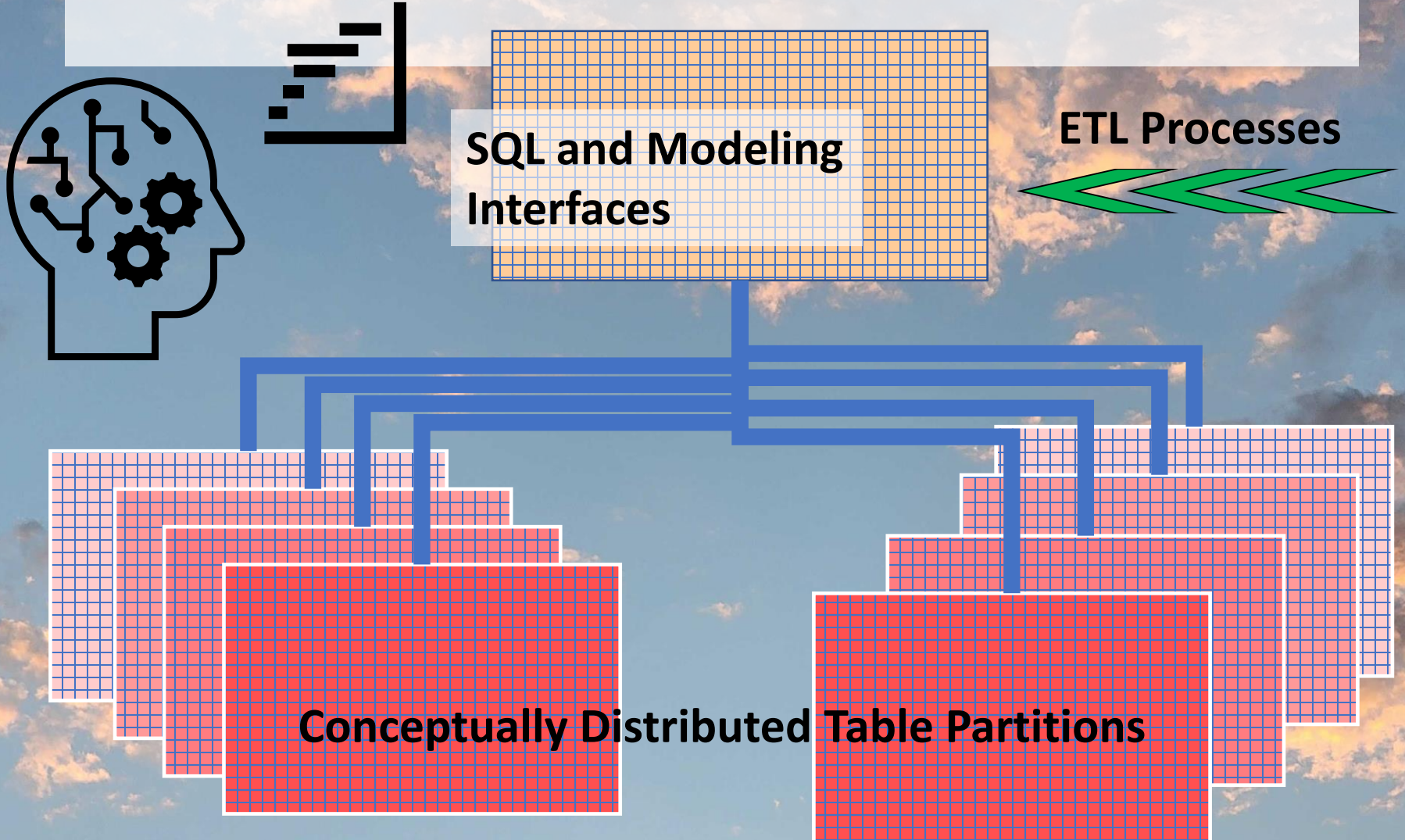
Parallel Computing for Data Analytics

Parallel Search
GCP Big Query
AWS Open Search

Module Topics

- Distributed Data Store Concepts
- GCP Big Query
- AWS Open Search

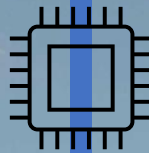
GCP Big Query



Elastic Search



Data Discovery, Mining, SQL Interfaces



Streaming Data

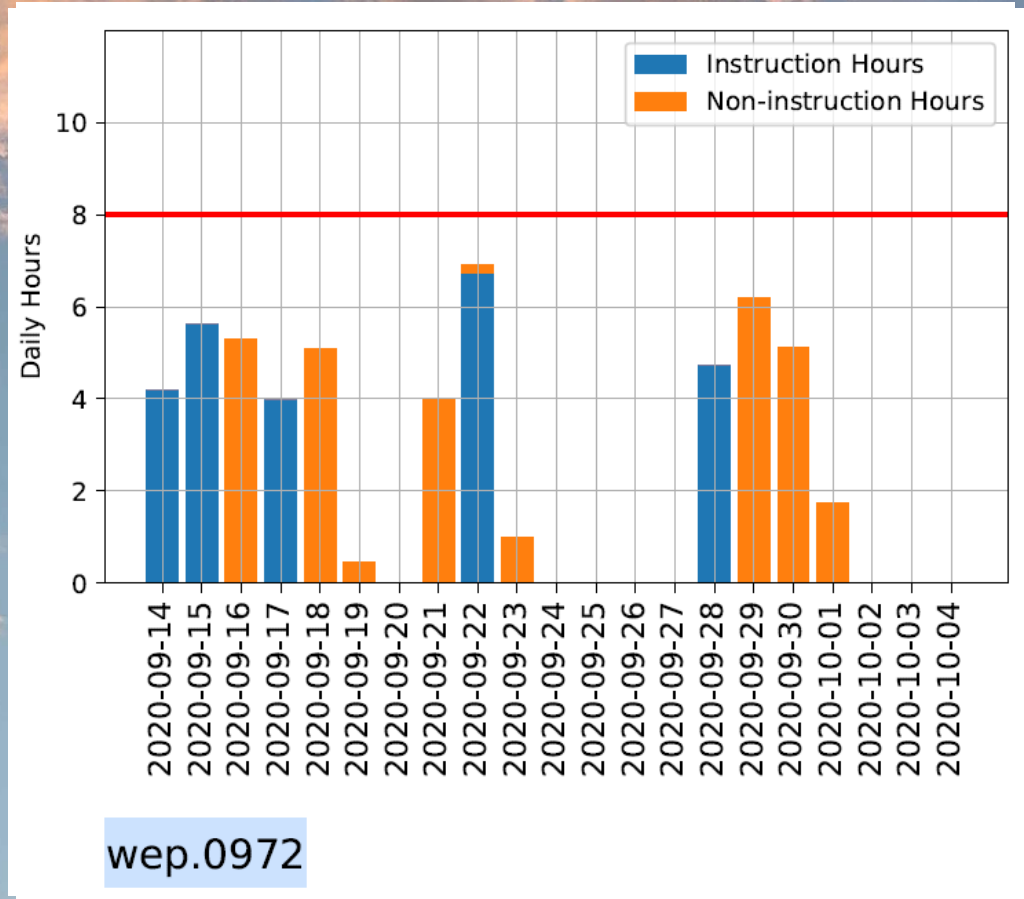


Distributed Indexes and Data Sinks

Elastic Search Use Case

DSA Europa & More

- Streaming Logs into local elastic search
 - Every notebook's save, load, close, and cell execution
- Can run analytics on per student, per notebook
- Course or Course Module Aggregates



AWS – Open Search (elastic search)

Not a recipe, a guide

- AWS Documentation is somewhat static, but the Console and API is constantly evolving!
- Carefully follow suggested “hints”
- Make sure you track on adaptations to CURL
- See notes on use the stream editor, sed

Tutorials Build to Exercises

search-scottgs-movies-ezsmqrqvcteirvzjoz6u46r74.us-east-1.es.amazonaws.com/_dashboards/app/discover/#/?_g=(filters:!(),refreshInterval:(pause:...

OpenSearch Dashboards

Discover New Save Open

Search

+ Add filter

scottgs-movies* 5,004 hits

Search field names

Filter by type 0

Selected fields

_source

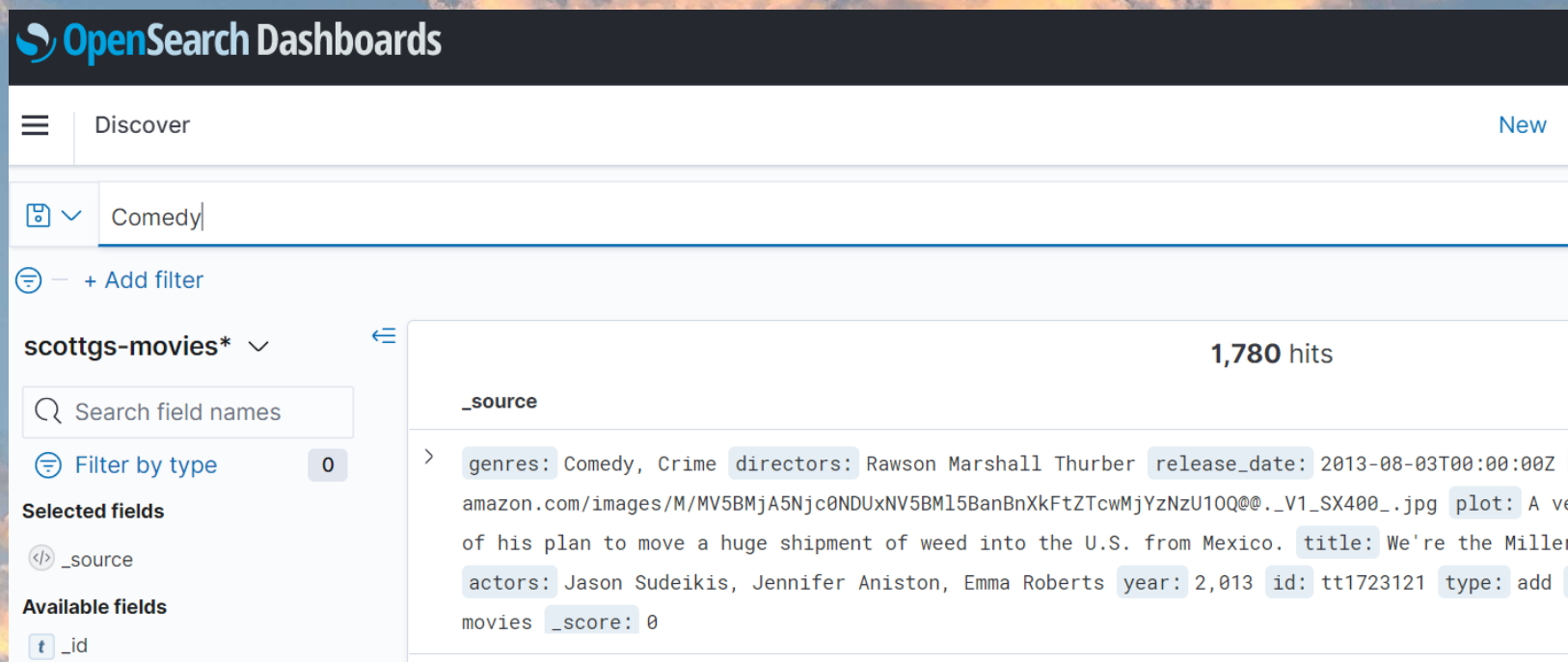
Available fields

- _id
- _index
- _score
- _type
- actor
- actors
- director
- directors
- genre
- genres

Search Results:

- Hit 1:**
_source: {
 director: Baird, Stuart
 genre: Action, Crime, Thriller
 year: 1,998
 actor: Downey Jr., Robert, Jones, Tommy Lee Jones, Robert Patrick, Michael, Pantoliano, Joe, Jacob, Irène, Nelligan, Kate, Roebuck, Daniel, Malahide, Patrick, Richardson, LaTanya, Wood, Nick, Minkoff, Robert, Brown, Spitfire, Foster, Reese, Spielberg, Bruce, Mukherji, Kevin, Cray, Ed, Fordham, John, ...
 title: U.S. Marshals
 _id: 3
 _type: _doc
 _index: scottgs-movies
 _score: 0
}
- Hit 2:**
_source: {
 directors: James Wan
 release_date: 2013-09-13T00:00:00Z
 rating: 7.1
 genres: Horror, Thriller
 image_url: https://m.media-amazon.com/images/M/MV5BMTg0OTA5ODIxNF5BM15BanBnXkFtZTcwNTUzNDg4OQ@@._V1_SX400_.jpg
 plot: The haunted Lambert family has a mysterious childhood secret that has left them dangerously connected to the spirit world.
 title: Insidious: Chapter 3
 running_time_secs: 6360
 actors: Patrick Wilson, Rose Byrne, Barbara Hershey
 year: 2,013
 id: tt2226417
 type: add
 _type: _doc
 _index: scottgs-movies
 _score: 0
}
- Hit 3:**
_source: {
 directors: Michael Bay
 release_date: 2014-06-25T00:00:00Z
 genres: Action, Adventure, Sci-Fi
 image_url: https://m.media-amazon.com/images/M/MV5BMTQyMDA5Nzg0N15BM15BanBnXkFtZTgwNzA4NDcxMDE@._V1_SX400_.jpg
 plot: A mechanic and his friends bring down Autobots and Decepticons - and a paranoid government official - on them.
 title: Transformers: Age of Extinction
 year: 2,014
 actors: Mark Wahlberg, Nicola Peltz, Jack Reynor
 id: tt2109248
 type: add
 _id: tt2109248
 _type: _doc
 _score: 0
}

Tip of the iceberg for Open Search



The screenshot displays the OpenSearch Dashboards interface. At the top, the 'Discover' tab is active. A search filter 'Comedy' is applied. The left sidebar shows the 'scottgs-movies*' index pattern with 0 filters and 0 selected fields. The main panel shows 1,780 hits for the '_source' field. The first hit is expanded, showing a JSON document with fields like genres, directors, release_date, plot, title, actors, year, id, type, and _score.

OpenSearch Dashboards

Discover [New](#)

Comedy

+ Add filter

scottgs-movies* [v](#)

Search field names

Filter by type 0

Selected fields

_source

Available fields

_id

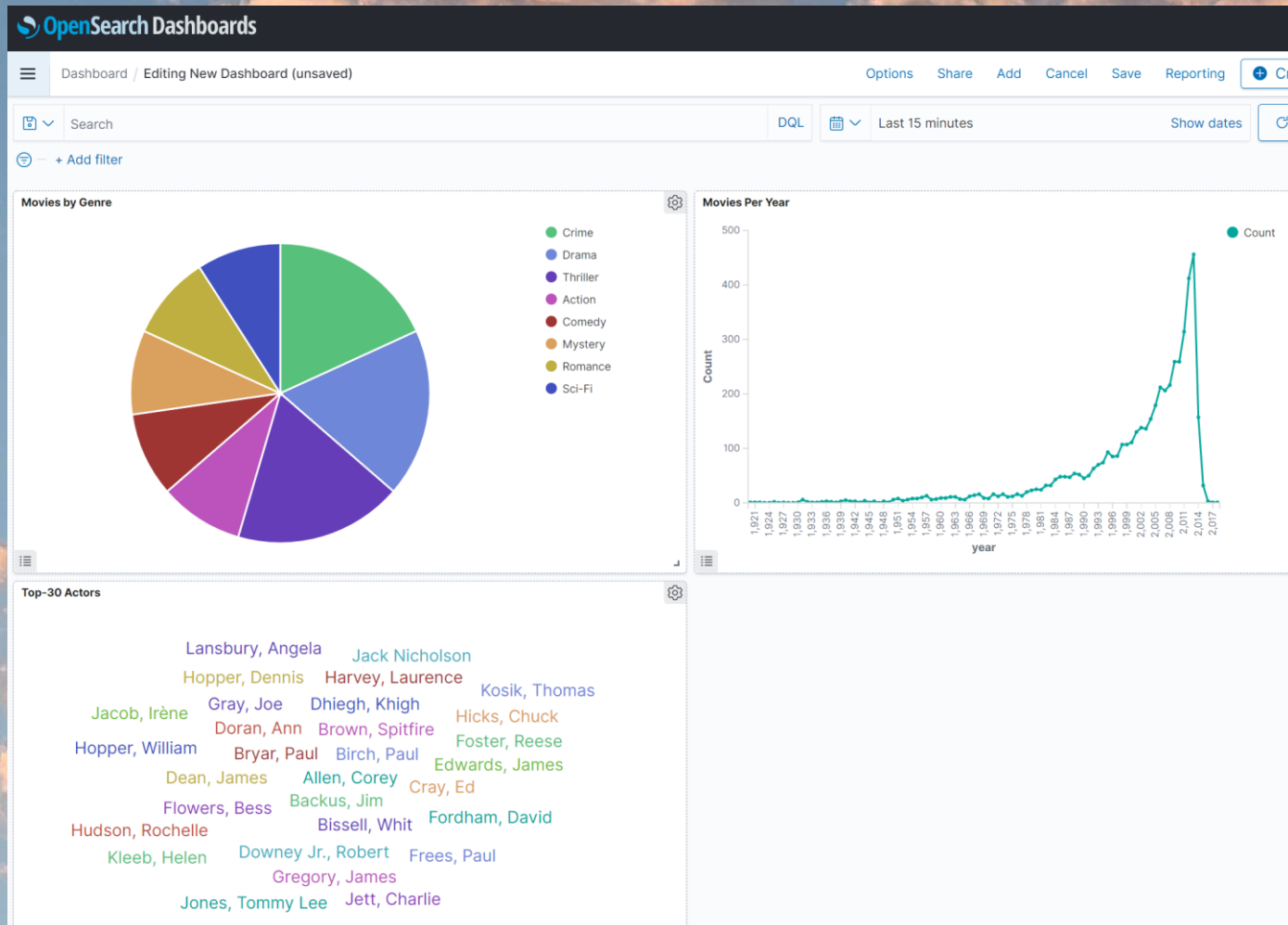
1,780 hits

_source

```
> {
  "genres": "Comedy, Crime",
  "directors": "Rawson Marshall Thurber",
  "release_date": "2013-08-03T00:00:00Z",
  "amazon.com/images/M/MV5BMjA5Njc0NDUxNV5BMl5BanBnXkFtZTcwMjYzNzU10Q@@._V1_SX400_.jpg": "plot: A ve",
  "plot": "of his plan to move a huge shipment of weed into the U.S. from Mexico.",
  "title": "We're the Miller",
  "actors": "Jason Sudeikis, Jennifer Aniston, Emma Roberts",
  "year": "2,013",
  "id": "tt1723121",
  "type": "add",
  "movies": "_score: 0"
}
```


AWS Exercise

You must explore & experiment



Pay Attention to Detail!

Start early, work methodically!

Take notes as you work through things!