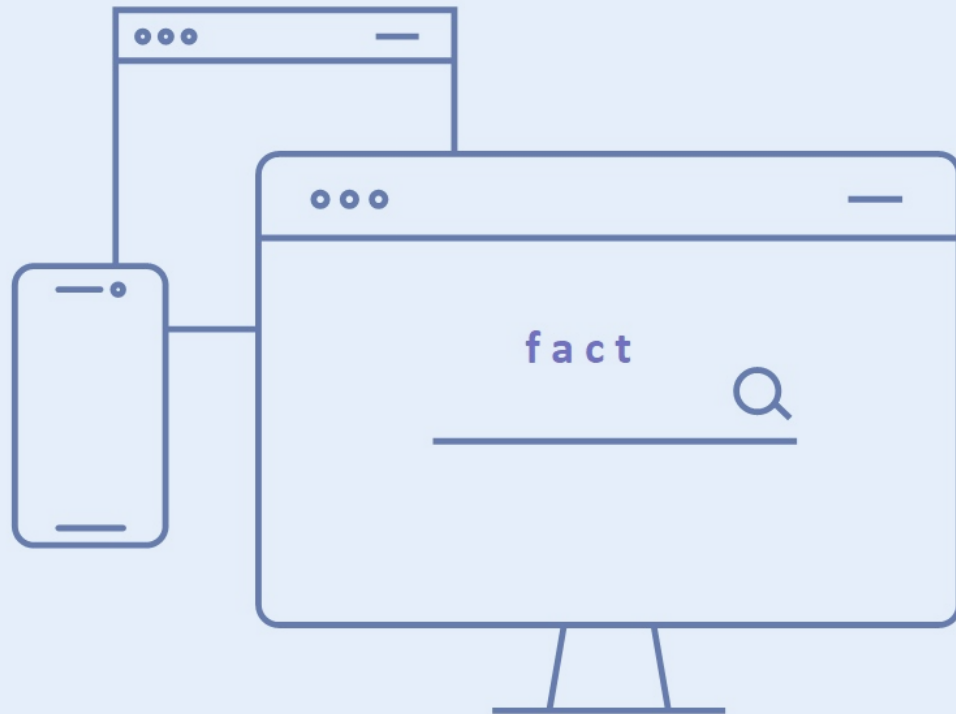




What is ***FACT***?

Fast **A**ccess to **C**linical **T**rial insights

AD08

Julia Gnatek, Senior Statistical Programmer, Biostatistics, Novo Nordisk

Steffen Nørgaard Vansgaard, Biostatistics Manager, Biostatistics, Novo Nordisk

DISCLAIMER

Views and opinions expressed are those of the speakers and not necessarily Novo Nordisk

What is ***FACT***?

Technical overview

DEMO

Conclusions

Questions

Agenda

FACT development team



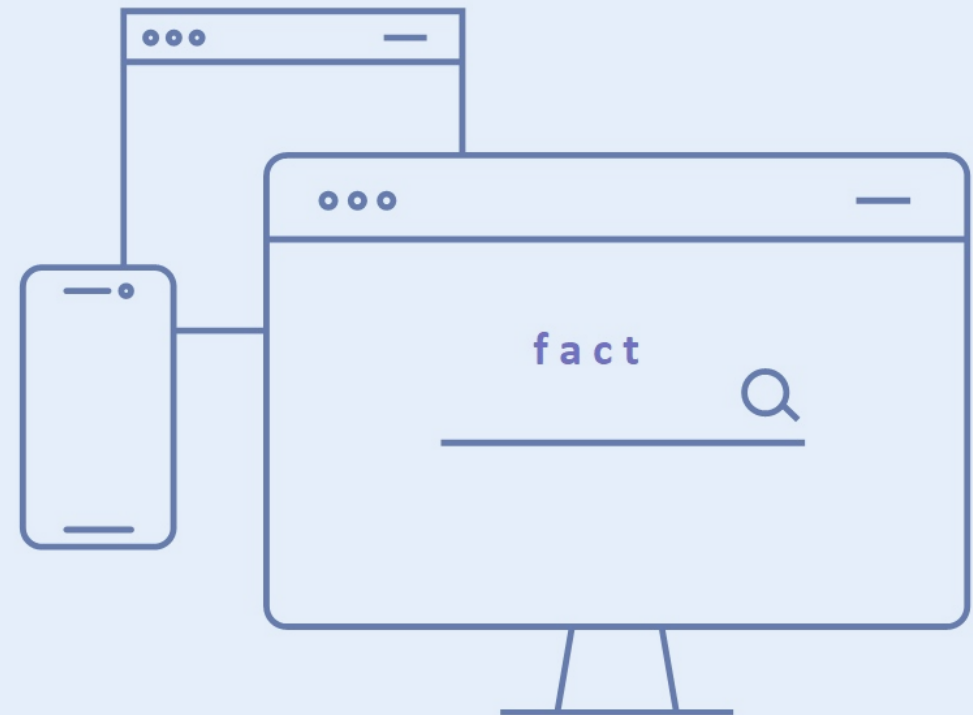
damvad • ANALYTICS

- Collaboration between Novo Nordisk and Damvad analytics

What is ***FACT***?

Fast **A**ccess to **C**linical **T**rial insights:

- Is a cloud-based search engine
- Allows easy access and navigation of analysis outputs and study documents





F_{ast} A_{ccess} To C_{linical} T_{rial} Insights

Enter search...

Use AND between words if you only want results that contain all words

 [ADVANCED](#)

 [PREVIOUS SEARCHES](#)

[SEARCH](#)

FACT: Technologies used

Agility

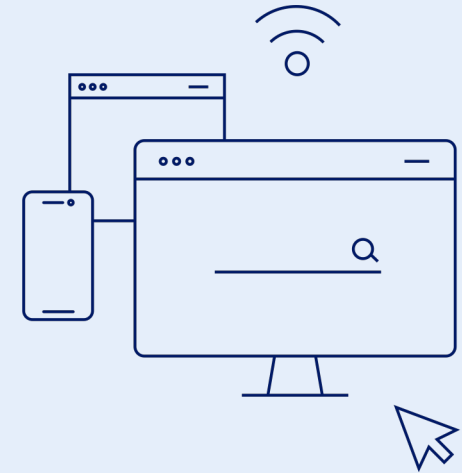
Access to broad range of technologies



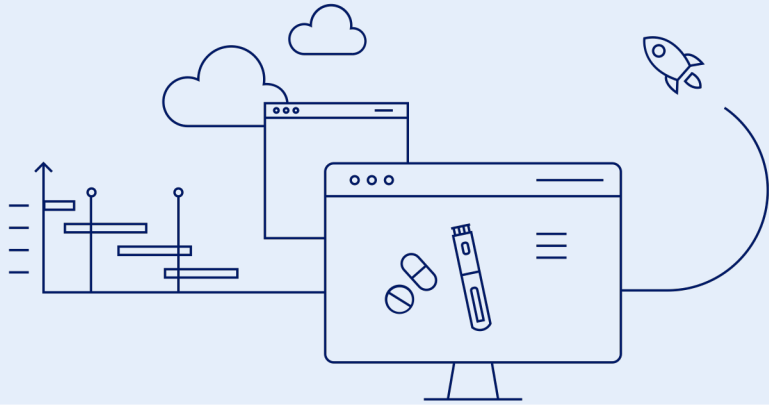
Cloud Computing

Elasticity

Grow and shrink capacity as needed

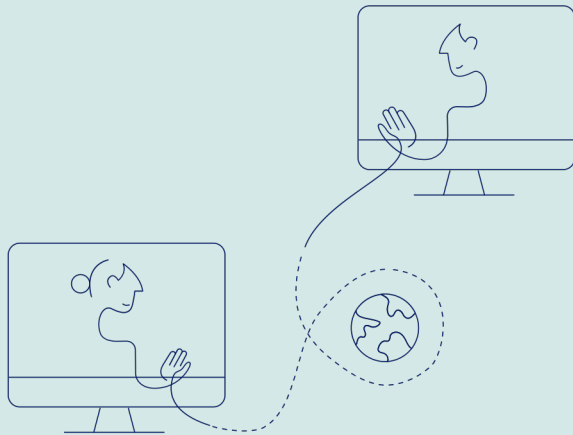


FACT: Technologies used



Amazon Web Services:

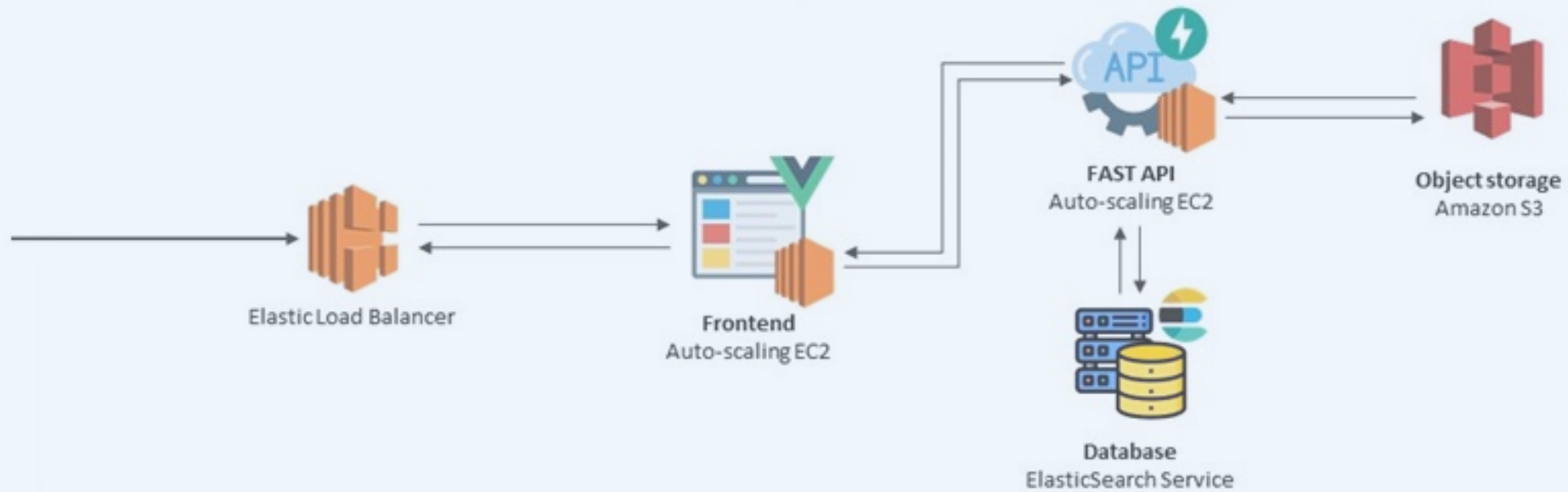
- All FACT components are running on AWS servers
- AWS service enabling the use of Natural Language Processing
- FACT insights are stored in OpenSearch database



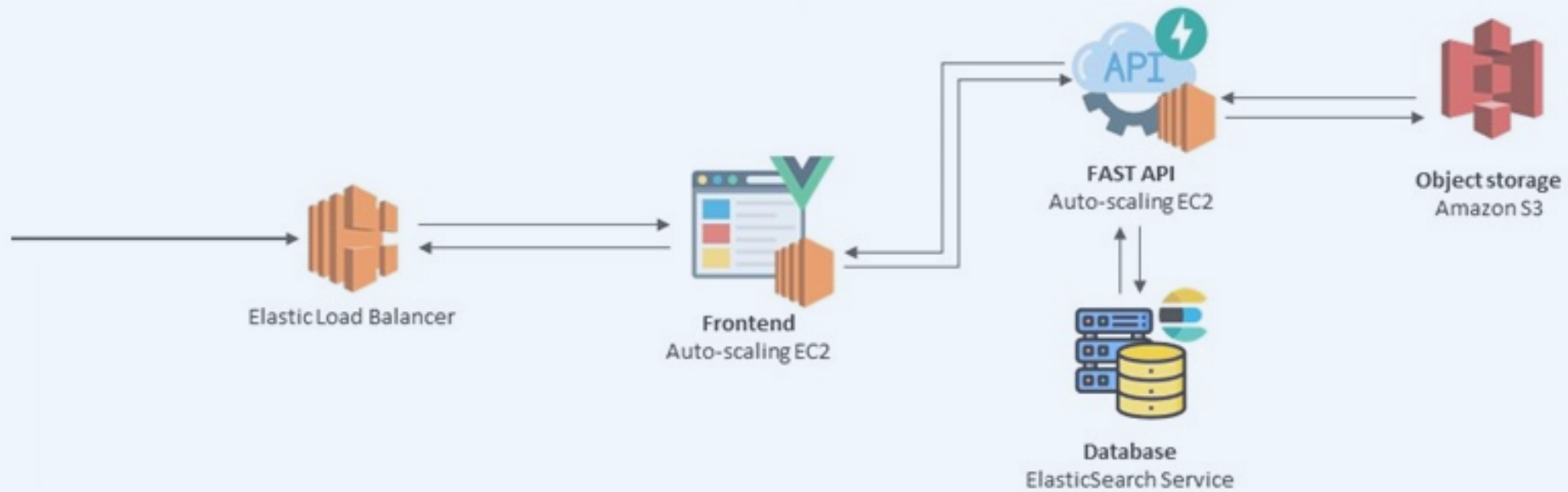
Python programming language

- High-level programming language
- Big selection of packages and support libraries
- Used to process files and extract metadata
- Connects front-end search page to AWS server

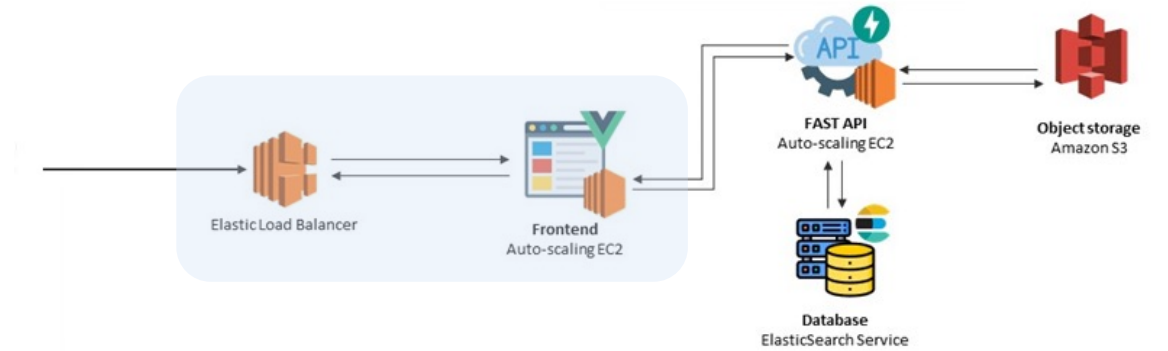
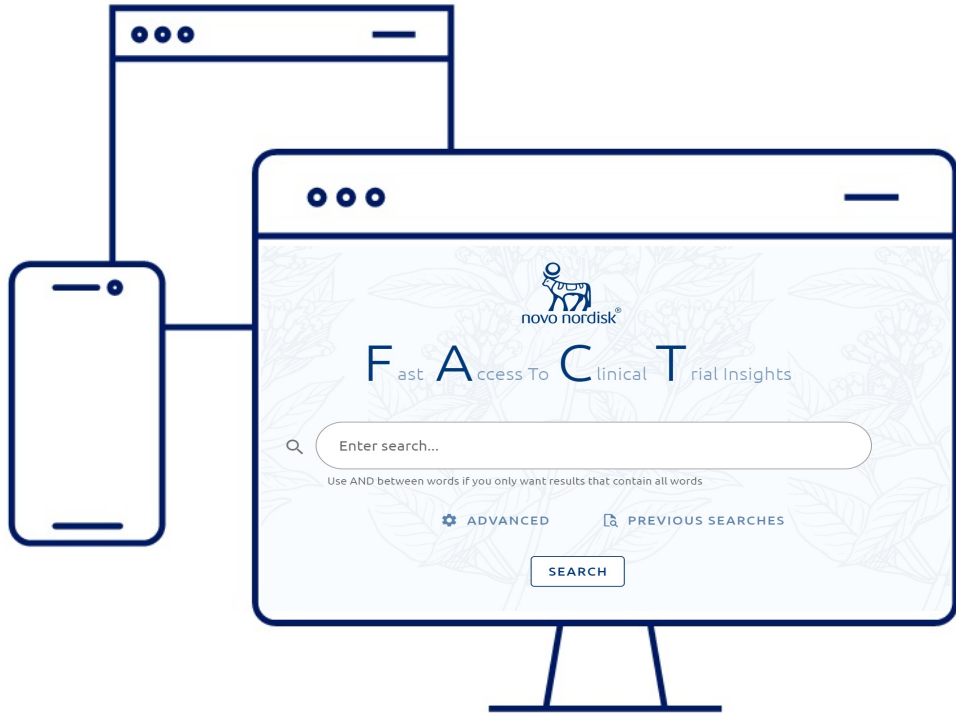
FACT technical overview



FACT technical overview



FACT Frontend page



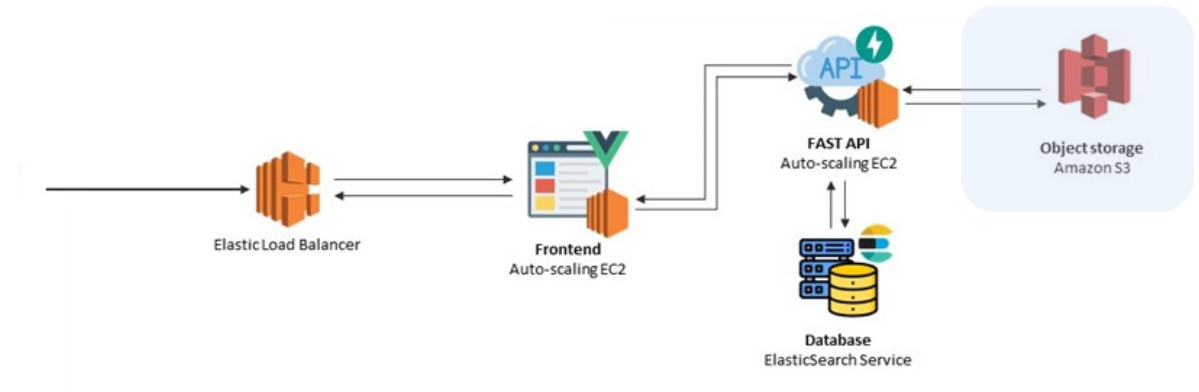
Frontend

What is it?
Where is it?
What is it made of?

FACT S3 Storage



S3 is a cloud file storage



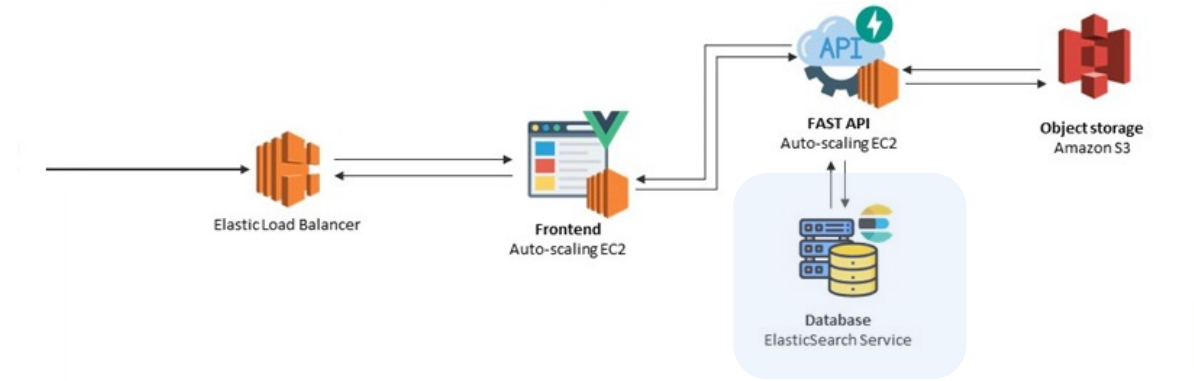
Amazon Simple Storage Service (Amazon S3)

What is it?

How are we using it?

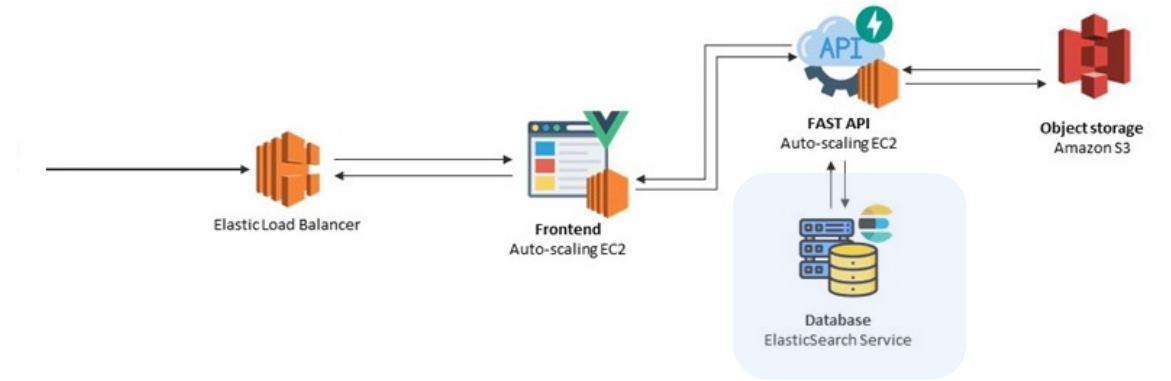
Why do we need it?

FACT OpenSearch



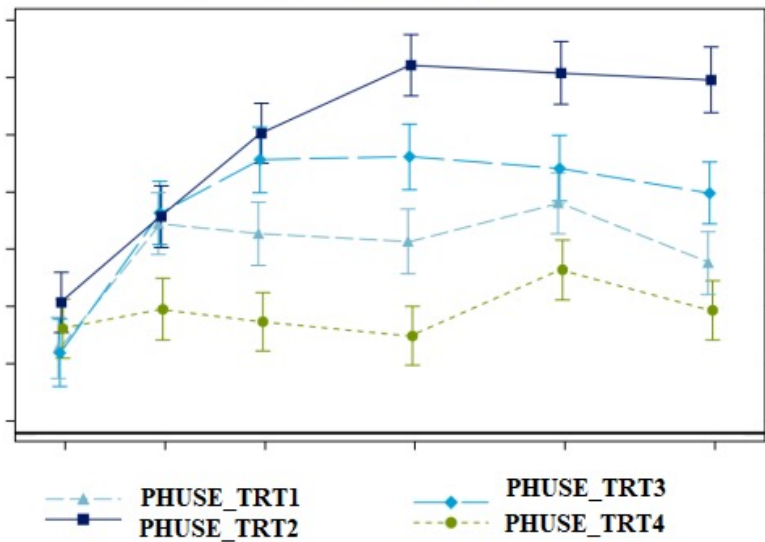
- A distributed search and analytics engine
- Inverted-index data structure
- Scores documents based on ranking function
- Key-value pairs can be boosted to optimize the search

FACT OpenSearch



PhuseStudy2 example plot

MEAN PLOT - PARAMETER_1



```

"_source" : {
  "modified_date" : "2022-10-12T08:55:08Z",
  "outputtype" : "figure",
  "id" : "some_id",
  "project" : "phuseproject1",
  "trial" : "phusestudy2",
  "instance" : "some_instance",
  "level" : "2",
  "titlelines" : "1",
  "fontsize" : "9",
  "paper" : "port",
  "gwidth" : "defaultcm",
  "gheight" : "defaultcm",
  "filename" : "p1s2_mplt1.png",
  "figuredataset" : "some_dataset.sas7bdat",
  "programname" : "p1s2_mplt1.sas",
  "title" : "mean plot parameter_1>",
  "meta_file" : "p_drive/PhuseProject1/PhuseStudy2/some_instance/output/p1s2_mplt1.xml",
  "output" : "PhuseStudy2 example plot MEAN PLOT - PARAMETER_1 4 PHUSE_TRT1 +-- PHUSE_TRT3 --== pHOUSE TRI?
  --6-- PHUSE_TRT4",
  "output_file" : "p_drive/PhuseProject1/PhuseStudy2/some_instance/output/p1s2_mplt1.png"
}
  
```

* In production, metadata is enriched with info from different systems, to ensure that study and project information can be used for search

	PHUSE_TRT1	PHUSE_TRT2	PHUSE_TRT3	Total
Number of patients	XX	XX	XX	XX
Age (years)				
N	X	X	X	X
Mean (SD)	XX.X (X.X)	XX.X (X.X)	XX.X (X.X)	XX.X (X.X)
Median	XX.X	XX.X	XX.X	XX.X
Min ; Max	XX ; XX	XX ; XX	XX ; XX	XX ; XX
Country, N (%)				
N	X (XXX.X)	X (XXX.X)	X (XXX.X)	X (XXX.X)
Australia	X (XX.X)	X (XX.X)	X (XX.X)	X (XX.X)
Brazil	X (XX.X)	X (XX.X)	X (XX.X)	X (XX.X)
...	X (XX.X)	X (XX.X)	X (XX.X)	X (XX.X)
Ethnicity, N (%)				
N	X (XXX.X)	X (XXX.X)	X (XXX.X)	X (XXX.X)
Hispanic or Latino	X (XX.X)	X (XX.X)	X (XX.X)	X (XX.X)
Not Hispanic or Latino	X (XX.X)	X (XX.X)	X (XX.X)	X (XX.X)
NA	X (XX.X)	X (XX.X)	X (XX.X)	X (XX.X)
Race, N (%)				
N	X (XXX.X)	X (XXX.X)	X (XXX.X)	X (XXX.X)
American Indian or Alaska Native	X (XX.X)	X (XX.X)	X (XX.X)	X (XX.X)
Asian	X (XX.X)	X (XX.X)	X (XX.X)	X (XX.X)
Black or African American	X (XX.X)	X (XX.X)	X (XX.X)	X (XX.X)
Native Hawaiian or Other Pacific Islander	X (XX.X)	X (XX.X)	X (XX.X)	X (XX.X)
White	X (XX.X)	X (XX.X)	X (XX.X)	X (XX.X)
Other	X (XX.X)	X (XX.X)	X (XX.X)	X (XX.X)
NA	X (XX.X)	X (XX.X)	X (XX.X)	X (XX.X)
Blood type, N (%)				
N	X (XXX.X)	X (XXX.X)	X (XXX.X)	X (XXX.X)
A	X (XX.X)	X (XX.X)	X (XX.X)	X (XX.X)
B	X (XX.X)	X (XX.X)	X (XX.X)	X (XX.X)
AB	X (XX.X)	X (XX.X)	X (XX.X)	X (XX.X)
O	X (XX.X)	X (XX.X)	X (XX.X)	X (XX.X)
NA	X (XX.X)	X (XX.X)	X (XX.X)	X (XX.X)

N: number of participants.
NA: Not applicable or not assessed.
SD: Standard deviation.

PhuseProject2/PhuseStudy1
some_date - p2s1_demo.sas/p2s1_demo.txt

```
"_source" : {
  "modified_date" : "2022-10-12T08:57:24Z",
  "outputtype" : "table",
  "id" : "some_id",
  "project" : "phuseproject2",
  "trial" : "phusestudy3",
  "instance" : "some_instance",
  "level" : "2",
  "titlelines" : "1",
  "fontsize" : "9",
  "paper" : "port",
  "gwidth" : "defaultcm",
  "gheight" : "defaultcm",
  "filename" : "p2s3_demo.txt",
  "figuredataset" : "some_dataset.sas7bdat",
  "programname" : "p2s3_demo.sas",
  "title" : "phuse baseline demographics>",
  "meta_file" : "p_drive/PhuseProject2/PhuseStudy3/some_instance/output/p2s3_demo.xml",
  "output_file" : "p_drive/PhuseProject2/PhuseStudy3/some_instance/output/p2s3_demo.txt",
  "output" : "PHUSE Baseline demographics
```

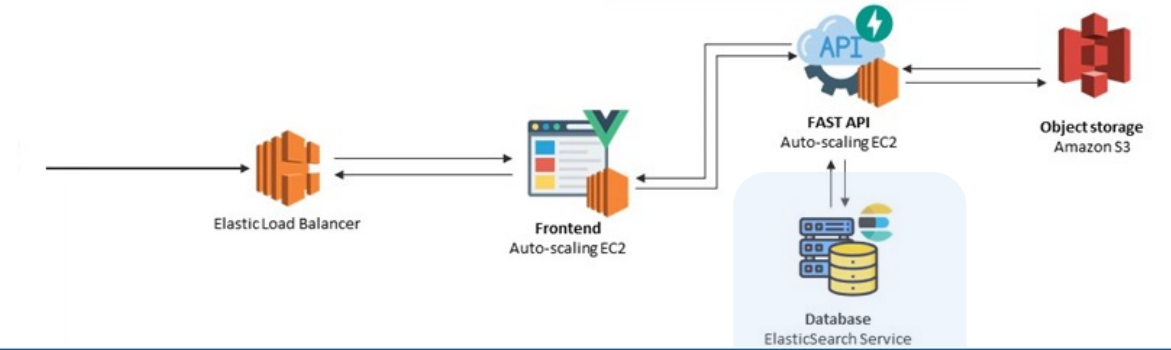
PHUSE_TRT1 PHUSE_TRT2 PHUSE_TRT3 Total

Number of patients XX XX XX XX Age (years) N X X X X Mean (SD) XX.X (X.X) XX.X (X.X) XX.X (X.X) XX.X (X.X) Median XX.X XX.X XX.X XX.X Min ; Max XX ; XX XX ; XX XX ; XX XX ; XX Country, N (%) N X (XXX.X) X (XXX.X) X (XXX.X) X (XXX.X) Australia X (XX.X) X (XX.X) X (XX.X) X (XX.X) Brazil X (XX.X) X (XX.X) X (XX.X) X (XX.X) ... X (XX.X) X (XX.X) X (XX.X) X (XX.X) Ethnicity, N (%) N X (XXX.X) X (XXX.X) X (XXX.X) Hispanic or Latino X (XX.X) X (XX.X) X (XX.X) X (XX.X) Not Hispanic or Latino X (XX.X) X (XX.X) X (XX.X) X (XX.X) NA X (XX.X) X (XX.X) X (XX.X) X (XX.X) Race, N (%) N X (XXX.X) X (XXX.X) X (XXX.X) X (XXX.X) American Indian or Alaska Native X (XX.X) X (XX.X) X (XX.X) X (XX.X) Asian X (XX.X) X (XX.X) X (XX.X) Black or African American X (XX.X) X (XX.X) X (XX.X) Native Hawaiian or Other Pacific X (XX.X) X (XX.X) X (XX.X) Islander X (XX.X) X (XX.X) X (XX.X) White X (XX.X) X (XX.X) X (XX.X) Other X (XX.X) X (XX.X) X (XX.X) X (XX.X) NA X (XX.X) X (XX.X) X (XX.X) X (XX.X) Blood type, N (%) N X (XXX.X) X (XXX.X) X (XXX.X) X (XXX.X) A X (XX.X) X (XX.X) X (XX.X) X (XX.X) B X (XX.X) X (XX.X) X (XX.X) X (XX.X) AB X (XX.X) X (XX.X) X (XX.X) X (XX.X) O X (XX.X) X (XX.X) X (XX.X) X (XX.X) NA X (XX.X) X (XX.X) X (XX.X) X (XX.X)

N:
number of participants. NA: Not applicable or not assessed. SD: Standard deviation. PhuseProject2 /PhuseStudy1 some_date - p2s1_demo.sas/p2s1_demo.txt"

}

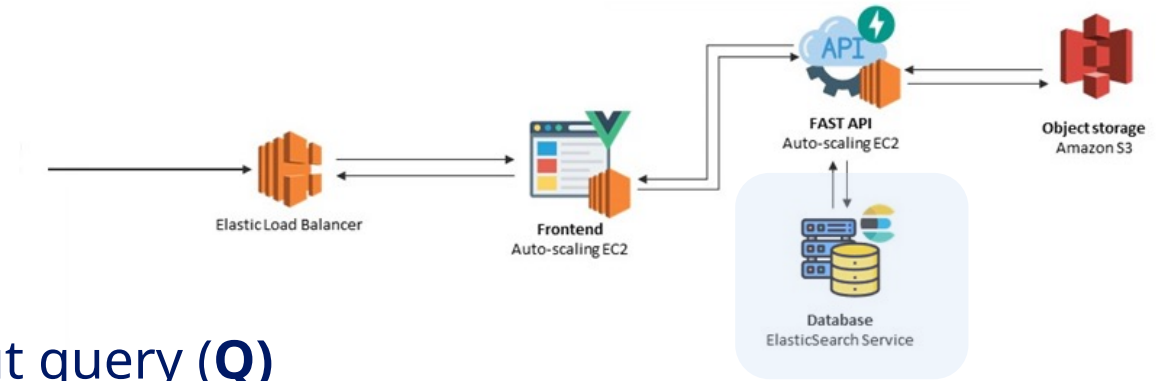
FACT OpenSearch



Term	Document Count	Frequency	Document ID
adverse	2	3	p_drive/phuseproject1/phusestudy1/some_instance/output/p1s1_dispo.txt#p_drive/phuseproject1/phusestudy1/some_instance/output/p1s1_dispo.xml p_drive/phuseproject2/phusestudy3/some_instance/output/p2s3_dispo.txt#p_drive/phuseproject2/phusestudy3/some_instance/output/p2s3_dispo.xml
ae	1	1	p_drive/phuseproject1/phusestudy1/some_instance/output/p1s1_dispo.txt#p_drive/phuseproject1/phusestudy1/some_instance/output/p1s1_dispo.xml
african	1	1	p_drive/phuseproject2/phusestudy3/some_instance/output/p2s3_demo.txt#p_drive/phuseproject2/phusestudy3/some_instance/output/p2s3_demo.xml
age	1	1	p_drive/phuseproject2/phusestudy3/some_instance/output/p2s3_demo.txt#p_drive/phuseproject2/phusestudy3/some_instance/output/p2s3_demo.xml
alaska	1	1	p_drive/phuseproject2/phusestudy3/some_instance/output/p2s3_demo.txt#p_drive/phuseproject2/phusestudy3/some_instance/output/p2s3_demo.xml
american	1	2	p_drive/phuseproject2/phusestudy3/some_instance/output/p2s3_demo.txt#p_drive/phuseproject2/phusestudy3/some_instance/output/p2s3_demo.xml
			p_drive/phuseproject1/phusestudy1/some_instance/document/ADRG.pdf p_drive/phuseproject1/phusestudy1/some_instance/document/protocol.pdf p_drive/phuseproject1/phusestudy1/some_instance/output/p1s1_dispo.txt#p_drive/phuseproject1/phusestudy1/some_instance/output/p1s1_dispo.xml p_drive/phuseproject1/phusestudy2/some_instance/document/ADRG.pdf p_drive/phuseproject1/phusestudy2/some_instance/document/protocol.pdf p_drive/phuseproject2/phusestudy3/some_instance/document/ADRG.pdf p_drive/phuseproject2/phusestudy3/some_instance/document/protocol.pdf
analysis	8	16	p_drive/phuseproject2/phusestudy3/some_instance/output/p2s3_dispo.txt#p_drive/phuseproject2/phusestudy3/some_instance/output/p2s3_dispo.xml p_drive/phuseproject2/phusestudy3/some_instance/output/p2s3_demo.txt#p_drive/phuseproject2/phusestudy3/some_instance/output/p2s3_demo.xml
applicable	2	2	p_drive/phuseproject2/phusestudy3/some_instance/output/p2s3_dispo.txt#p_drive/phuseproject2/phusestudy3/some_instance/output/p2s3_dispo.xml
asian	1	1	p_drive/phuseproject2/phusestudy3/some_instance/output/p2s3_demo.txt#p_drive/phuseproject2/phusestudy3/some_instance/output/p2s3_demo.xml
assessed	2	2	p_drive/phuseproject2/phusestudy3/some_instance/output/p2s3_demo.txt#p_drive/phuseproject2/phusestudy3/some_instance/output/p2s3_demo.xml p_drive/phuseproject2/phusestudy3/some_instance/output/p2s3_dispo.txt#p_drive/phuseproject2/phusestudy3/some_instance/output/p2s3_dispo.xml

FACT OpenSearch

- Default similarity function, BM25 is used
- It is scoring document (**D**) based on the input query (**Q**)



$$score(D, Q) = \sum_{i=1}^n IDF(q_i) \cdot \frac{f(q_i, D) \cdot (k_1 + 1)}{f(q_i, D) + k_1 \cdot (1 - b + b \cdot \frac{fieldLen}{avgFieldLen})}$$

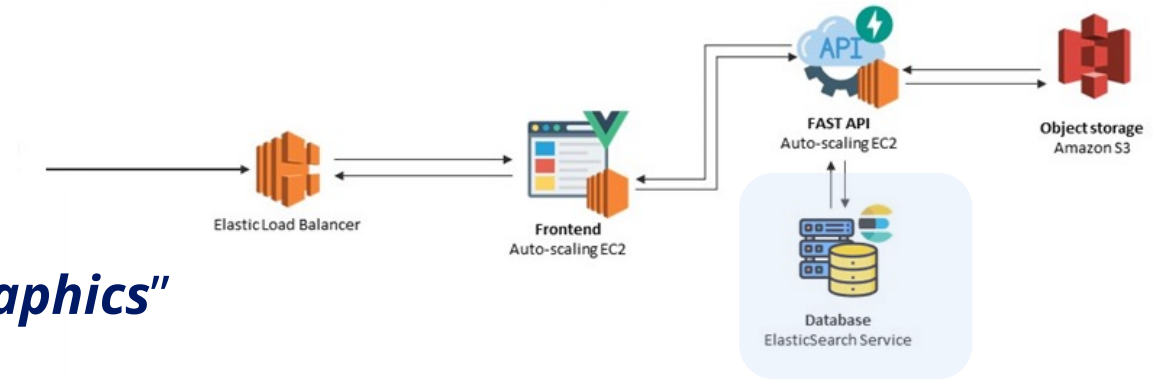
*inverse document
frequency*

*keywords
 $q_1, \dots, q_n$
given query Q*

*how many times
does the q_i occur in
document D*

*how long document
is relative to average
document length*

FACT OpenSearch

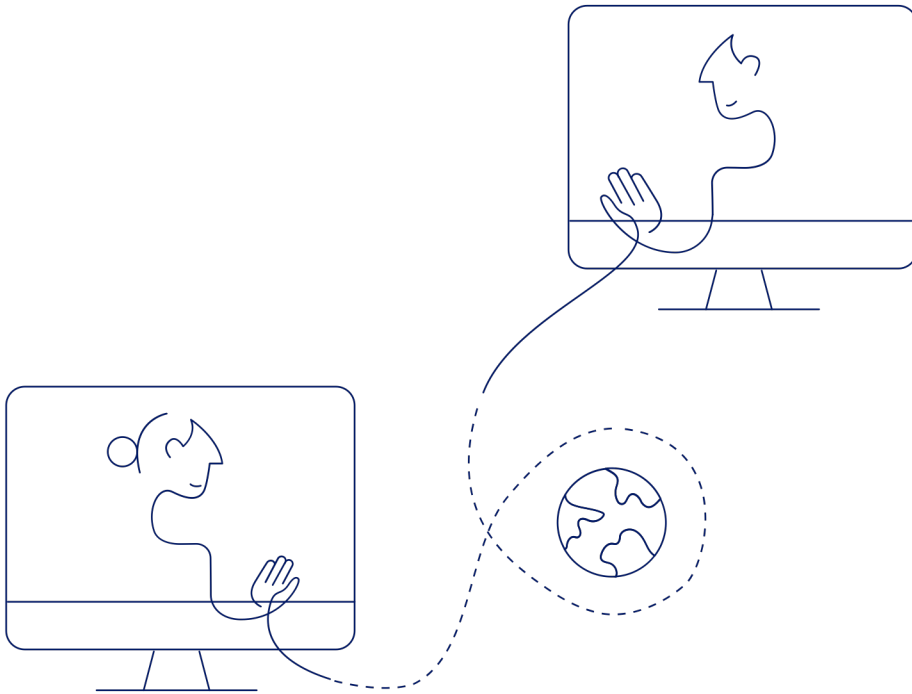


Document scores, when searching for "**demographics**"

```

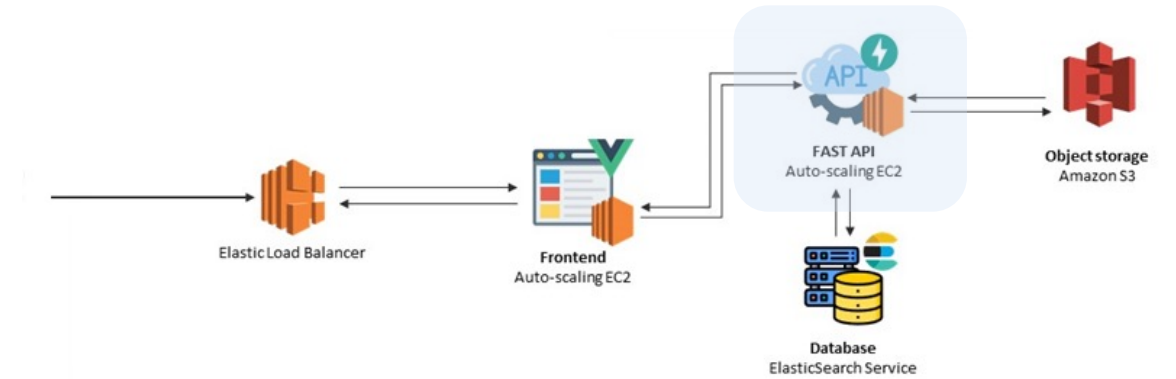
"hits" : [
  {
    "_index" : "jygn-test-index",
    "_type" : "_doc",
    "_id" : "p_drive/phuseproject2/phusestudy3/some_instance/statprog/p2s3_demo.sas",
    "_score" : 2.6347983,
    "fields" : {
      "output" : [
        "/*-----*/ /* PHUSE example program */ /*-----*/ /*-----*/ /* PhuseProject2/PhuseStudy3 */ /* Demographics output program */
        /*-----*/ data _null_; put "Hello PHUSE! Let's assume this is a demographics output program."; run;""
      ]
    }
  },
  {
    "_index" : "jygn-test-index",
    "_type" : "_doc",
    "_id" : "p_drive/phuseproject2/phusestudy3/some_instance/output/p2s3_demo.txt#p_drive/phuseproject2/phusestudy3/some_instance/output/p2s3_demo.xml",
    "_score" : 0.65576994,
    "fields" : {
      "output" : [
        "PHUSE Baseline demographics _____ PHUSE_TRT1 PHUSE_TRT2 PHUSE_TRT3 Total
        _____ Number of patients XX XX XX XX Age (years) N X X X X Mean (SD) XX.X ( X.X) XX.X ( X.X) XX.X ( X.X) XX.X ( X.X)
        Median XX.X XX.X XX.X XX.X Min ; Max XX ; XX XX ; XX XX ; XX XX ; XX Country, N (%) N X (XXX.X) X (XXX.X) X (XXX.X) X (XXX.X) Australia X ( XX.X) X ( XX.X) X ( XX.X) X ( XX.X) Brazil X ( XX.X) X ( XX.X) X ( XX.X)
        ) X ( XX.X) ... X ( XX.X) X ( XX.X) X ( XX.X) X ( XX.X) Ethnicity, N (%) N X (XXX.X) X (XXX.X) X (XXX.X) X (XXX.X) Hispanic or Latino X ( XX.X) X ( XX.X) X ( XX.X) X ( XX.X) Not Hispanic or Latino X ( XX.X) X ( XX.X)
        .X) X ( XX.X) X ( XX.X) NA X ( XX.X) X ( XX.X) X ( XX.X) X ( XX.X) Race, N (%) N X (XXX.X) X (XXX.X) X (XXX.X) X (XXX.X) American Indian or Alaska Native X ( XX.X) X ( XX.X) X ( XX.X) X ( XX.X) Asian X ( XX.X) X
        ( XX.X) X ( XX.X) X ( XX.X) Black or African American X ( XX.X) X ( XX.X) X ( XX.X) X ( XX.X) Native Hawaiian or Other Pacific X ( XX.X) X ( XX.X) X ( XX.X) X ( XX.X) Islander White X ( XX.X) X ( XX.X) X ( XX.X)
        X ( XX.X) Other X ( XX.X) X ( XX.X) X ( XX.X) NA X ( XX.X) X ( XX.X) X ( XX.X) X ( XX.X) Blood type, N (%) N X (XXX.X) X (XXX.X) X (XXX.X) X (XXX.X) A X ( XX.X) X ( XX.X) X ( XX.X) X ( XX.X) B X ( XX.X)
        ) X ( XX.X) X ( XX.X) X ( XX.X) AB X ( XX.X) X ( XX.X) X ( XX.X) X ( XX.X) O X ( XX.X) X ( XX.X) X ( XX.X) X ( XX.X) NA X ( XX.X) X ( XX.X) X ( XX.X) X ( XX.X)
        _____ N: number of participants. NA: Not applicable or not assesed. SD: Standard deviation. PhuseProject2/PhuseStudy1
        some_date - p2s1_demo.sas/p2s1_demo.txt"
      ]
    }
  }
]
  
```

FACT API

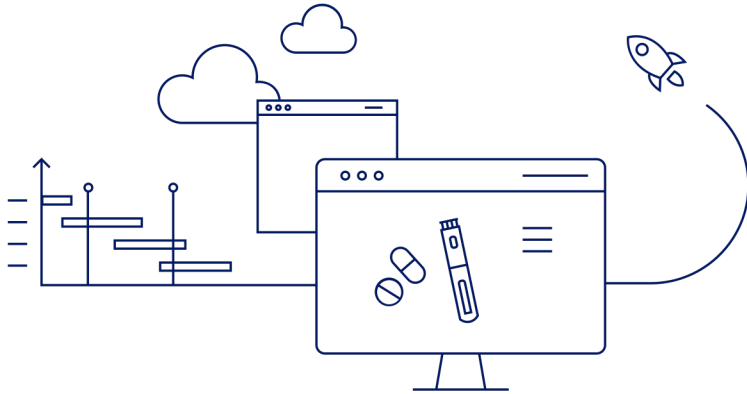


APPLICATION PROGRAMMING INTERFACE (API)

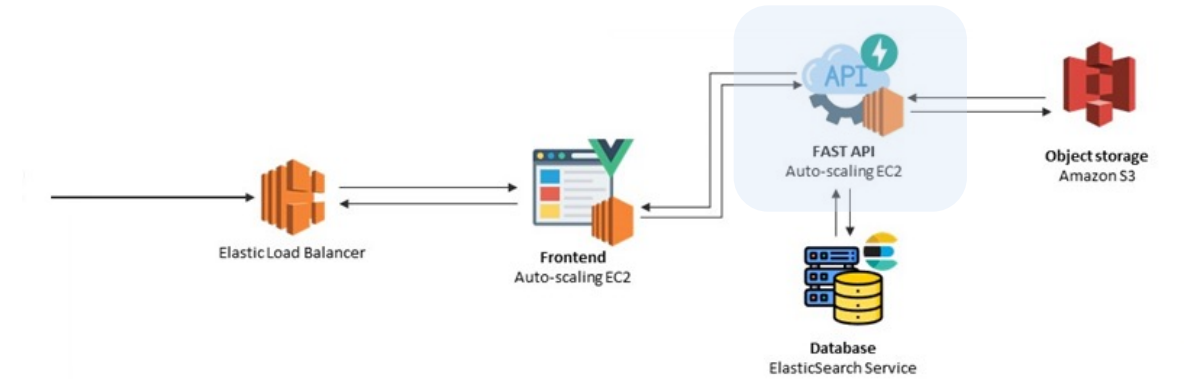
What is an API?
Why do we need one?
What is an endpoint?



FACT API



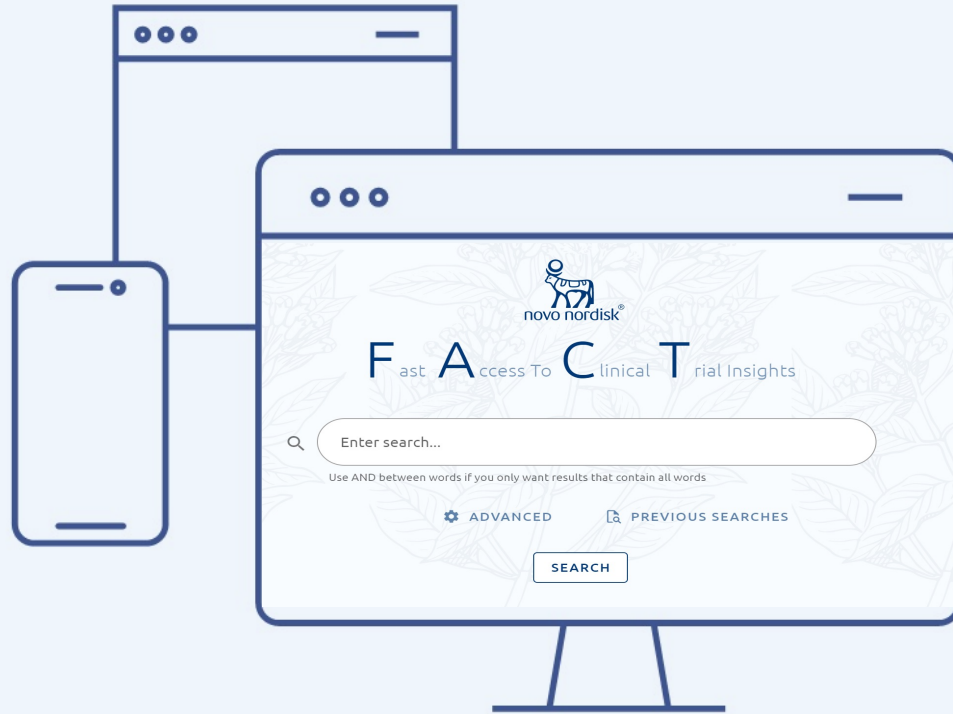
API retrieves data from the databases and passes it to the frontend page



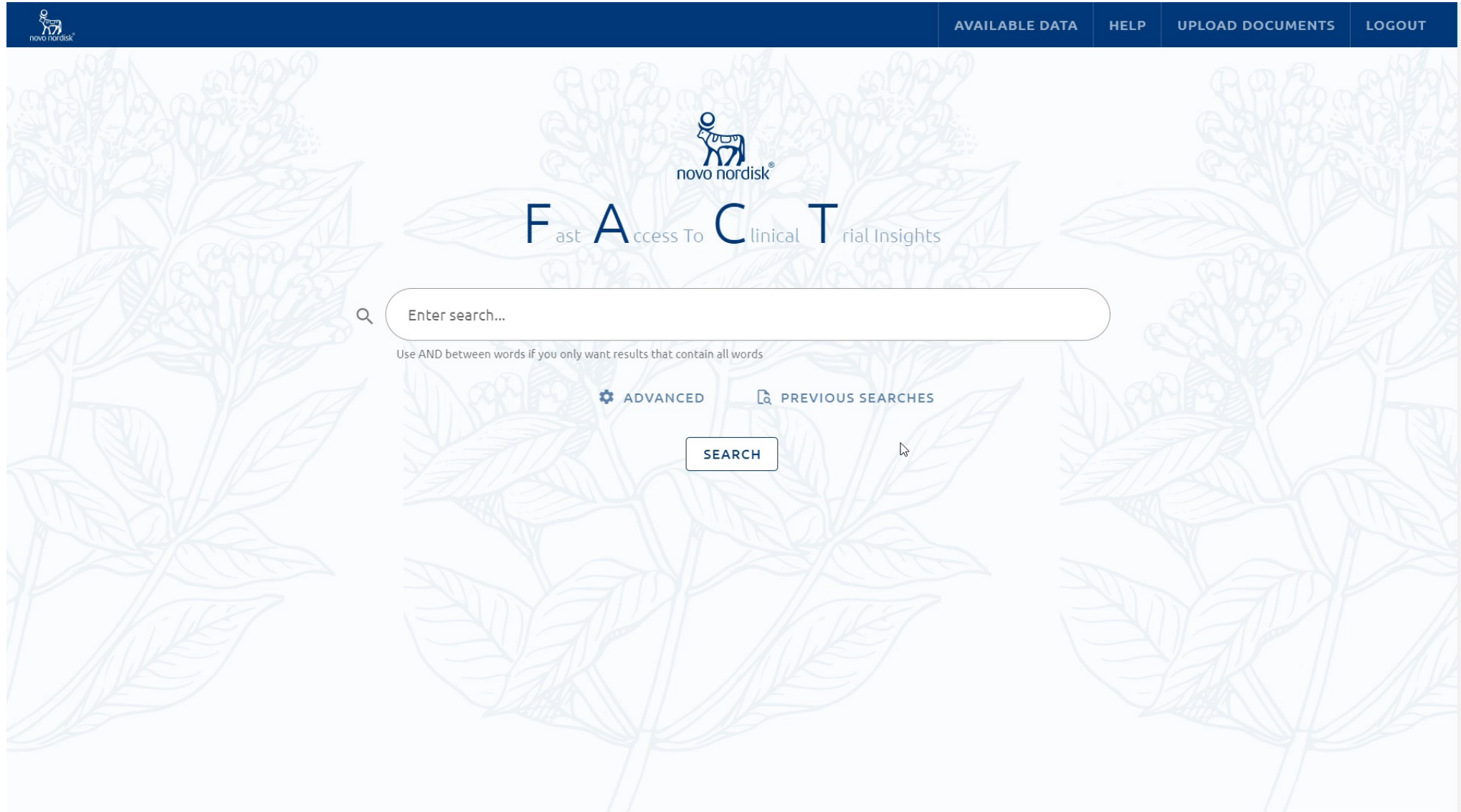
FACT API

- Implemented in python with FastAPI package
- Key endpoints
- Async API
- Docker use
- Amazon Beanstalk

FACT DEMO



FACT: How to use it



novo nordisk®

AVAILABLE DATA | HELP | UPLOAD DOCUMENTS | LOGOUT

novo nordisk®

F_{ast} A_{ccess To} C_{linical} T_{rial Insights}

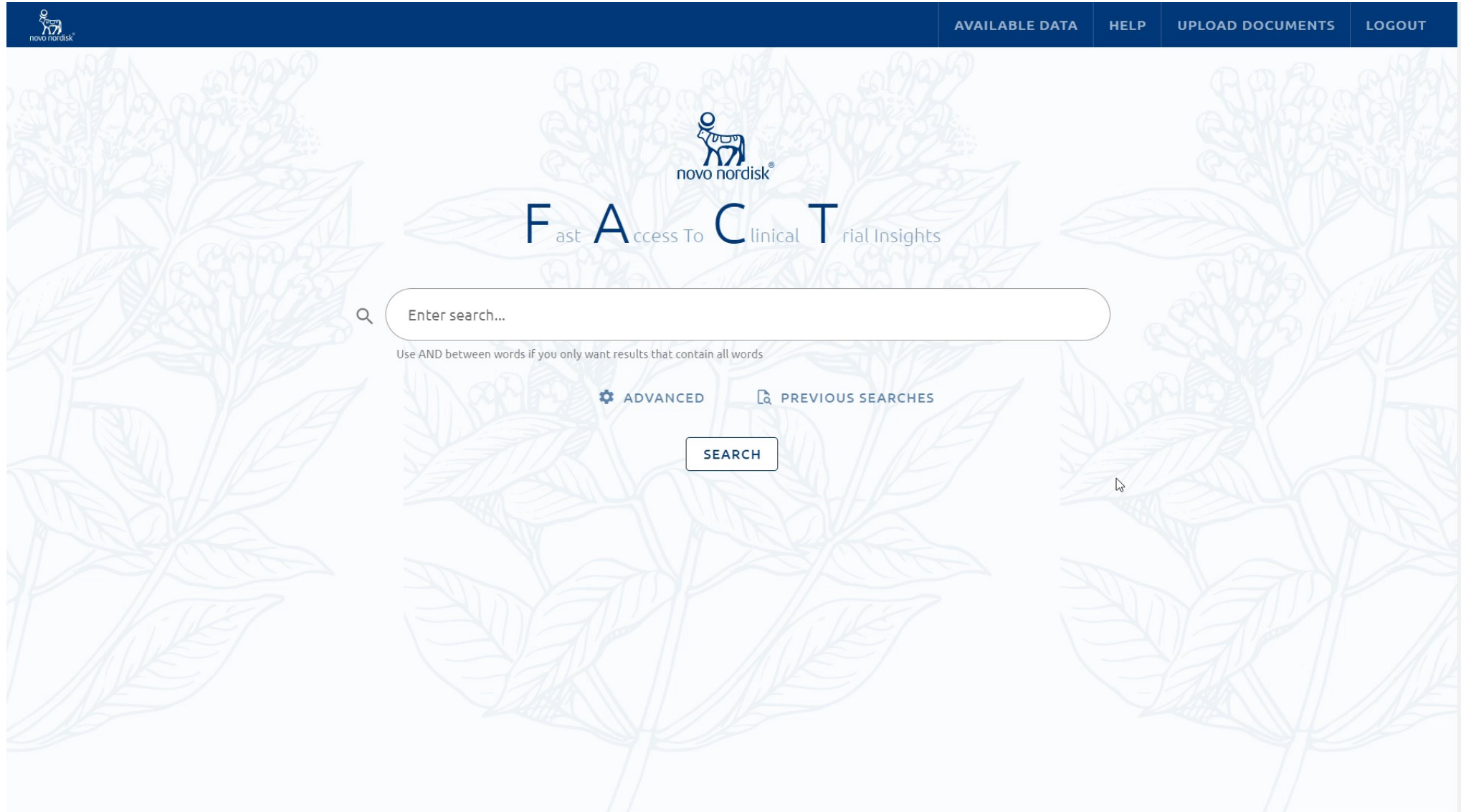
Enter search...

Use AND between words if you only want results that contain all words

⚙️ ADVANCED | 🔍 PREVIOUS SEARCHES

SEARCH

FACT: What's in there?



novo nordisk®

AVAILABLE DATA | HELP | UPLOAD DOCUMENTS | LOGOUT

novo nordisk®

F_{ast} A_{ccess To} C_{linical} T_{rial Insights}

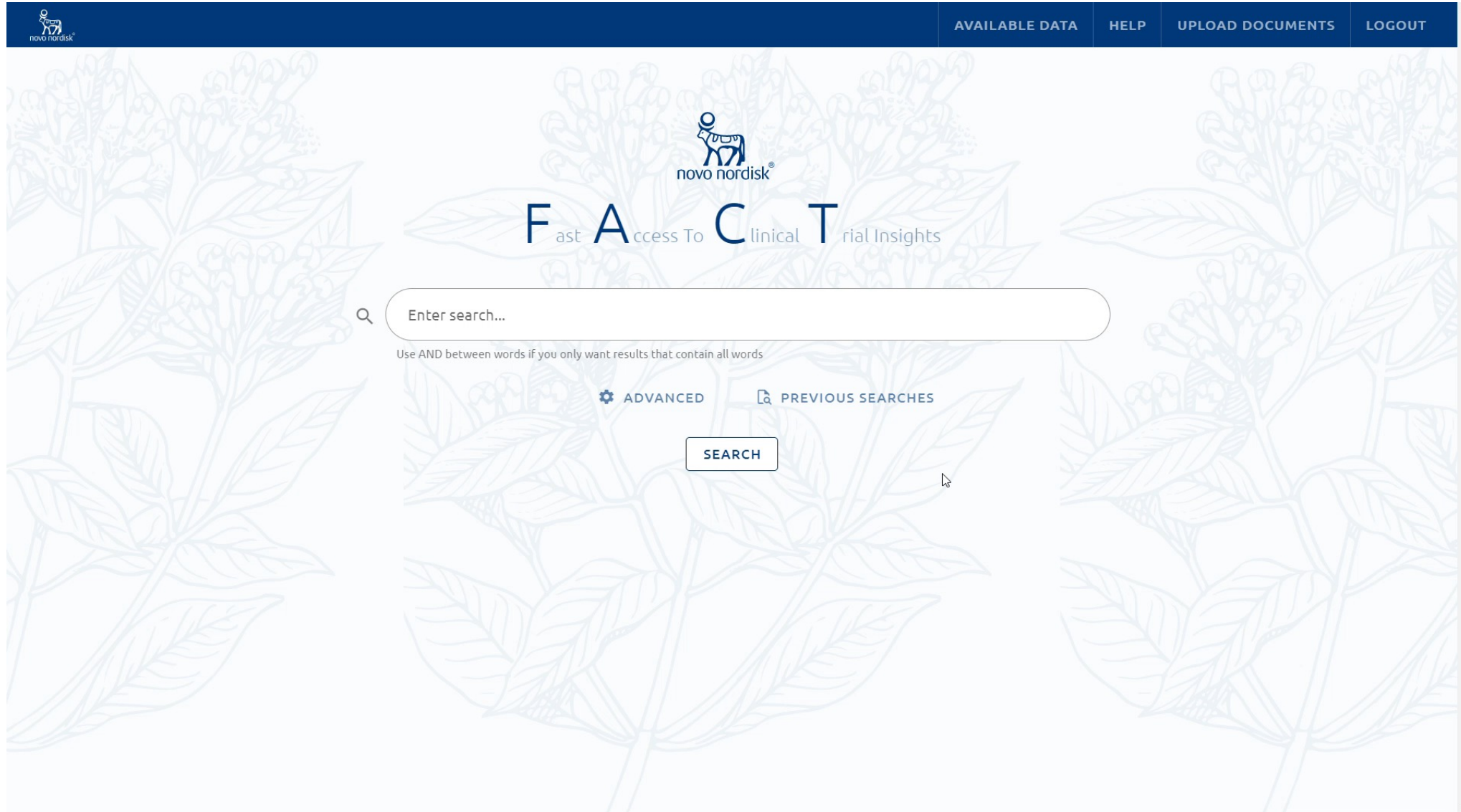
Enter search...

Use AND between words if you only want results that contain all words

⚙️ ADVANCED | 🔍 PREVIOUS SEARCHES

SEARCH

FACT: Use cases



novo nordisk®

AVAILABLE DATA | HELP | UPLOAD DOCUMENTS | LOGOUT

novo nordisk®

F_{ast} A_{ccess To} C_{linical} T_{rial Insights}

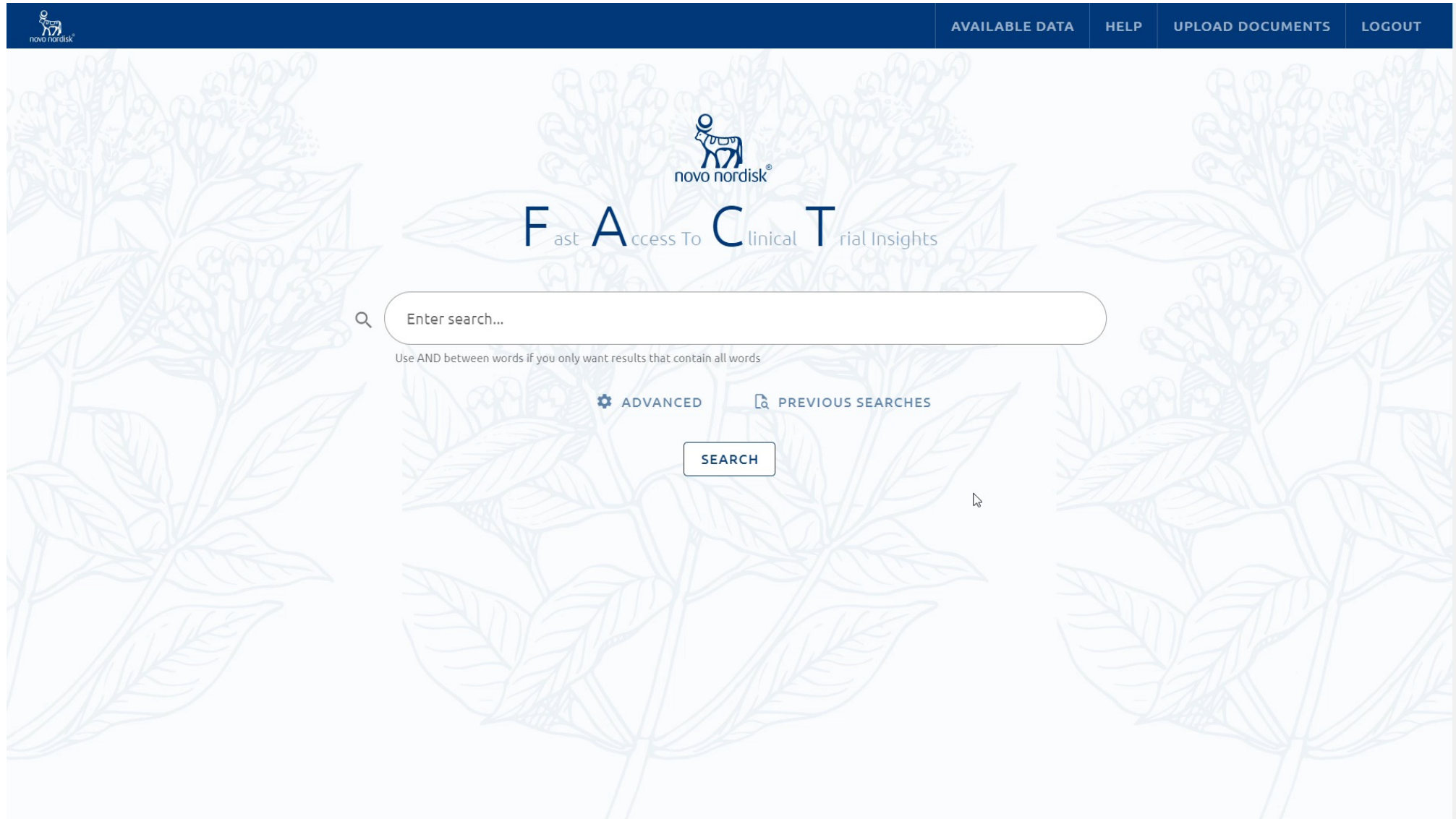
Enter search...

Use AND between words if you only want results that contain all words

⚙️ ADVANCED | 🔍 PREVIOUS SEARCHES

SEARCH

FACT: Use cases



novo nordisk®

AVAILABLE DATA | HELP | UPLOAD DOCUMENTS | LOGOUT

novo nordisk®

Fast Access To Clinical Trial Insights

Enter search...

Use AND between words if you only want results that contain all words

ADVANCED | PREVIOUS SEARCHES

SEARCH

FACT: GDPR considerations



- Listings containing IDs of study participants are not included in this tool
- End-of-text (EOT) documents containing listings are also excluded
- Inclusion of the IDs could theoretically be used to identify study participants

FACT: Conclusions

- Cloud based solution
- Search engine
- Easily retrieving relevant files
- Quickly finding answers to scientific questions



FACT: Q&A

