Data Transfer Crossroads:

Visualising LHC data and collaborations

Eirini Kolokytha
Interactive Experiences Engineer

14.09.2022

Idea^2,
Media Lab,
NCSR DEMOKRITOS Institute of Nuclear and Particle Physics

Media Lab

- Conceive, develop, deploy and support installations of interactive applications
- Testing new media
- Create pedagogical experiencies
- Unpacking physics

Computing in High Energy Physics and Data Needs

- February 1952, the Conseil Européen pour la Recherche Nucléaire is provisionally created.
- In the1980s and through the 1990s and beyond , the CERN's Program Library was in use for data analysis.
- In the 1990s, ROOT became available as well.
- As of 2009, LHC experiments went on-line, producing up to 100+ PetaBytes per year.

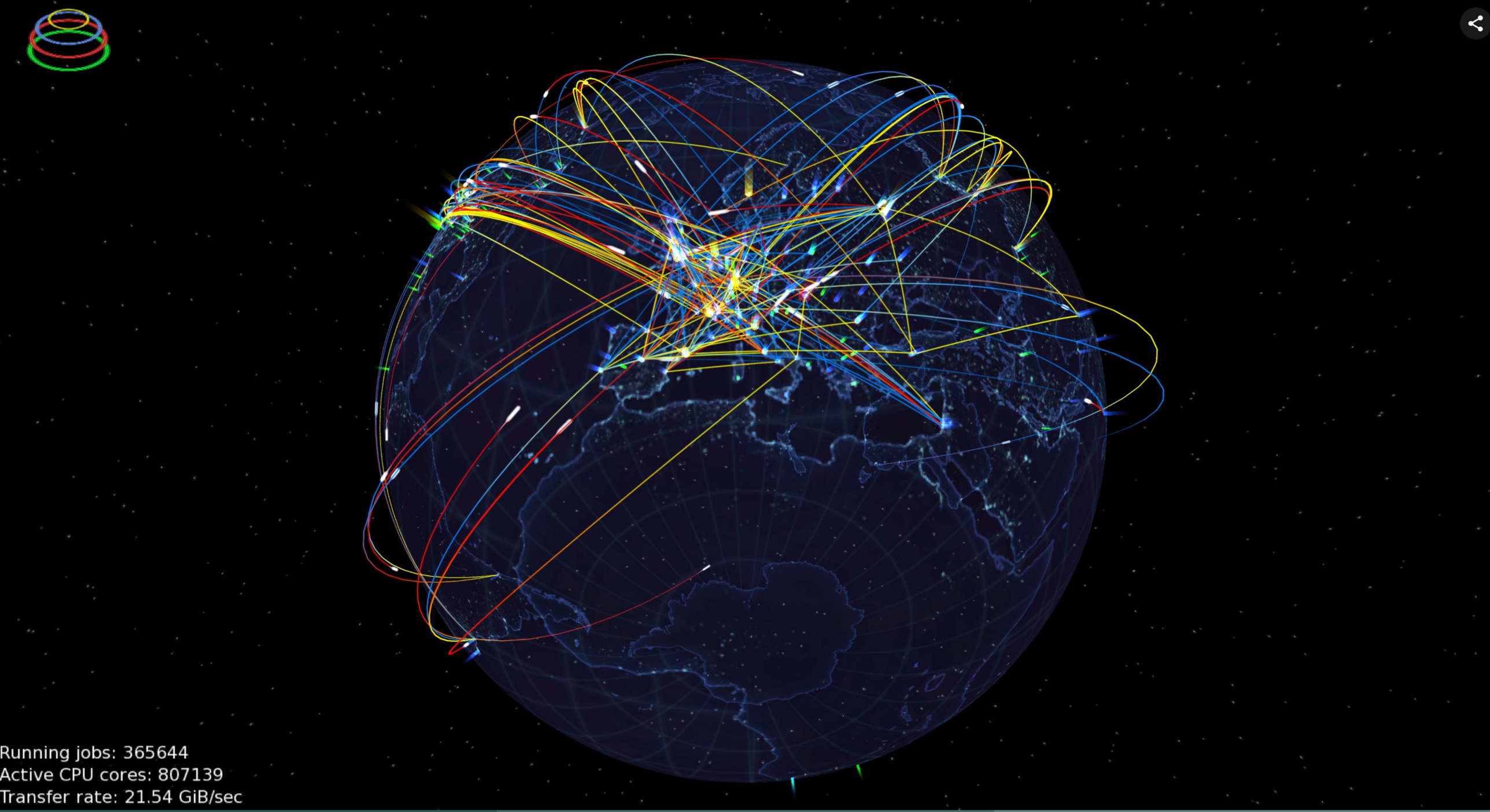
EGL - Explorer of Grid Load

- Monitoring the Worldwide LHC Computing Grid (WLCG)
- Visualizing data transfers of LHC experiments between collaborating institutes.
- Analyzing WLCG characteristics
- Detailed information about Sites

How it started?

- In August 2016, EGL started as the project of Mayank Sharma for CERN’ s openlab Summer Student Programme, supervised by Joao Pequenao.
- Mellisa Gaillard and Edward Karavakis contributed with their input.
- The application was using Google Earth KML tool.


Running jobs: 365118
Active CPU cores: 795836
Transfer rate: 18.35 GiB/sec


Running jobs: 365644
Active CPU cores: 807139
Transfer rate: 21.54 GiB/sec

Methodology

- Conceiving the application
- Graphics were made in Unity and have been updated throughout the years.
- In the initial project, genuine data were missing.
- Collaboration with IT department

Sources of information

- Transfer information for all VOs is being fetched by the central MONIT infrastructure.
- Job information is being fetched from different sources:
- For ATLAS and CMS the respective dashboards (in the central monitoring infrastructure) are being used.
- For ALICE http://alimonitor.cern.ch is the source of information
- For LHCb the information currently comes from a static file. Dynamic information is expected to become available in the course of next year.
- The topology and details of Sites and Federations are being fetched from CRIC https://wlcg-cric.cern.ch

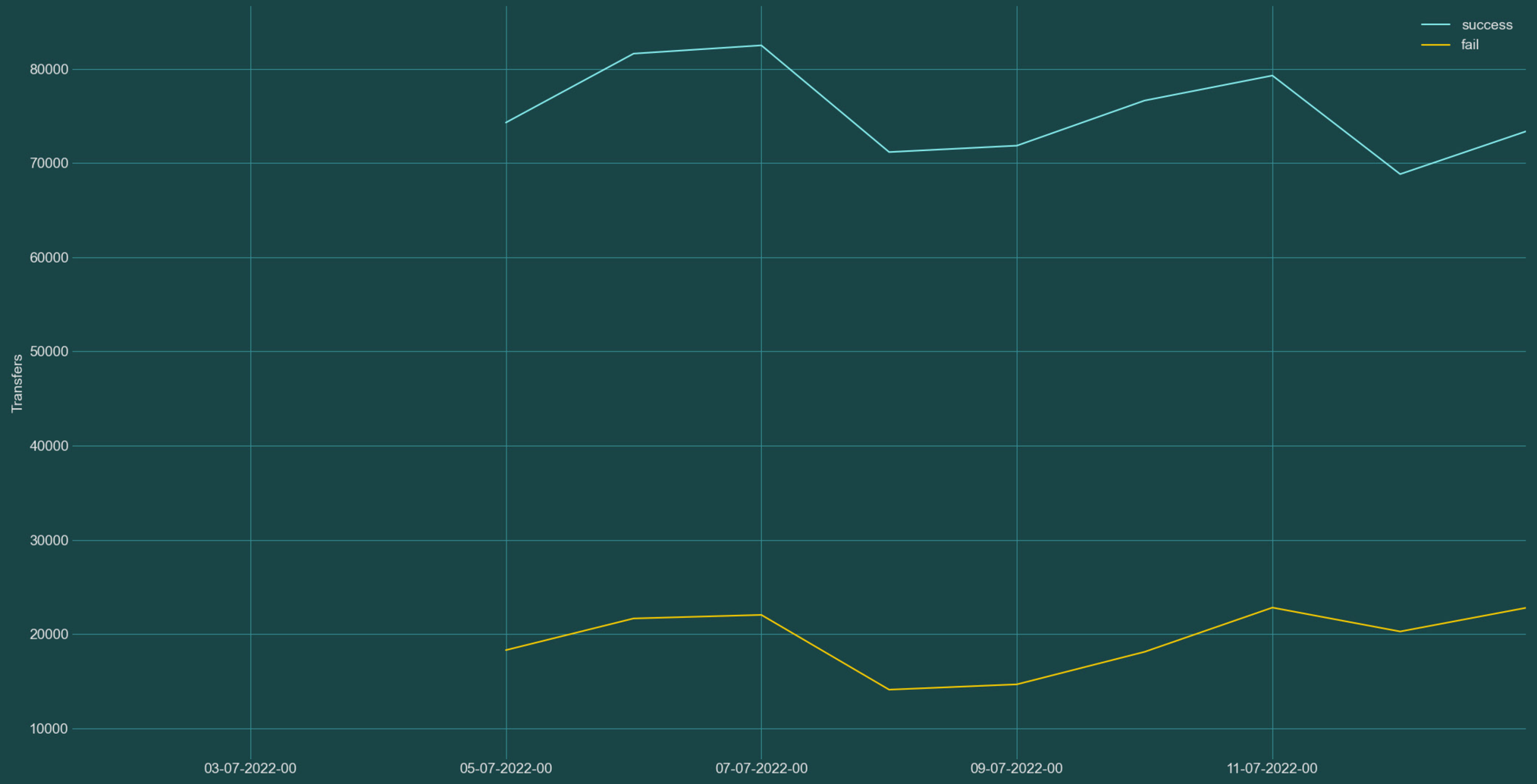
Technical information

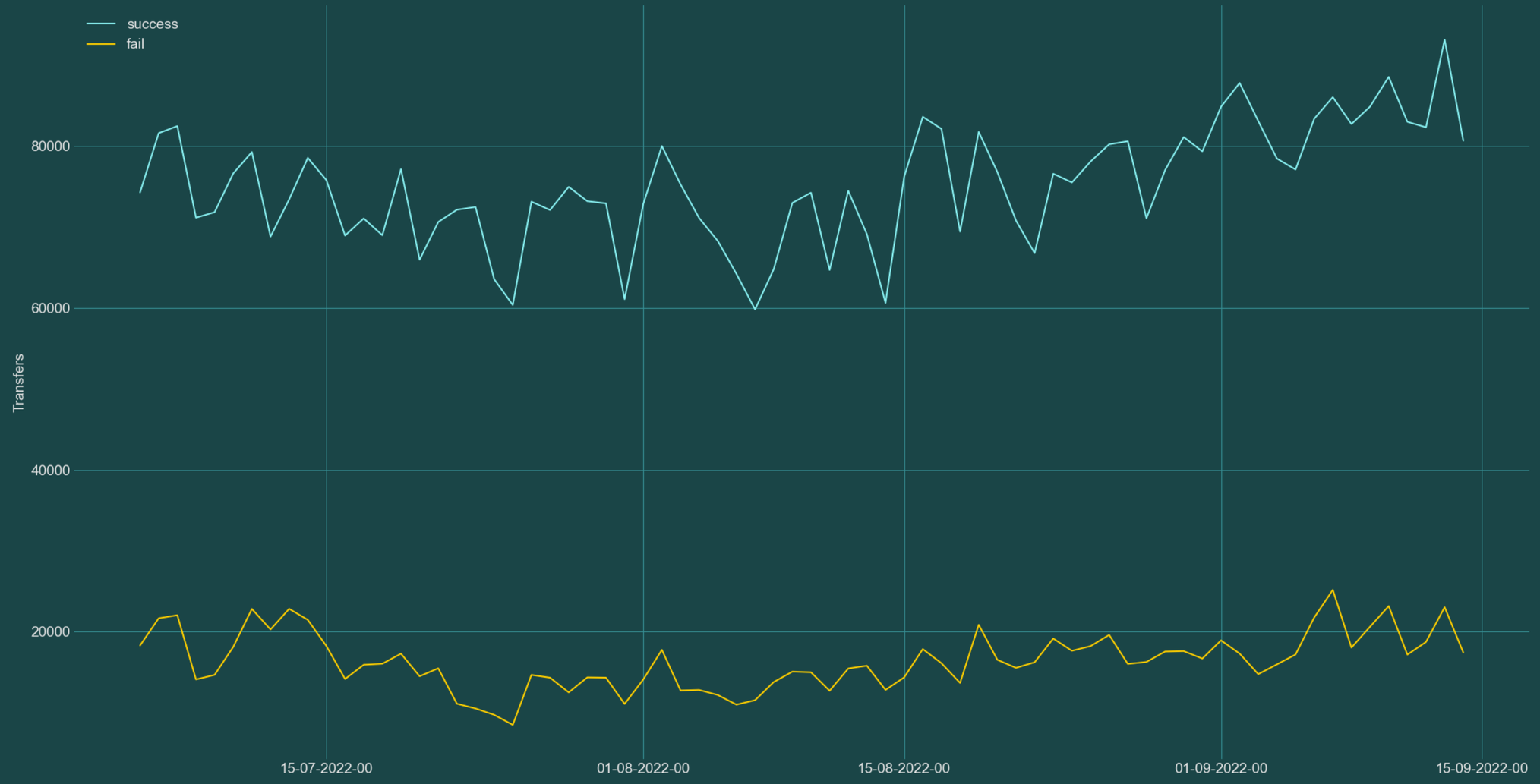
- A Django based application has been developed to fetch the information, aggregate it and provide it for visualisation, through REST APIs.
- The application is containerised and deployed on Openshift (exposed only inside the CERN network).
- The data are regularly fetched in the application and there is also a cron job that is backing up the results of our JSON REST APIs in S3.
- S3 storage is used both as a backup and also as a way to browse through the history of data.

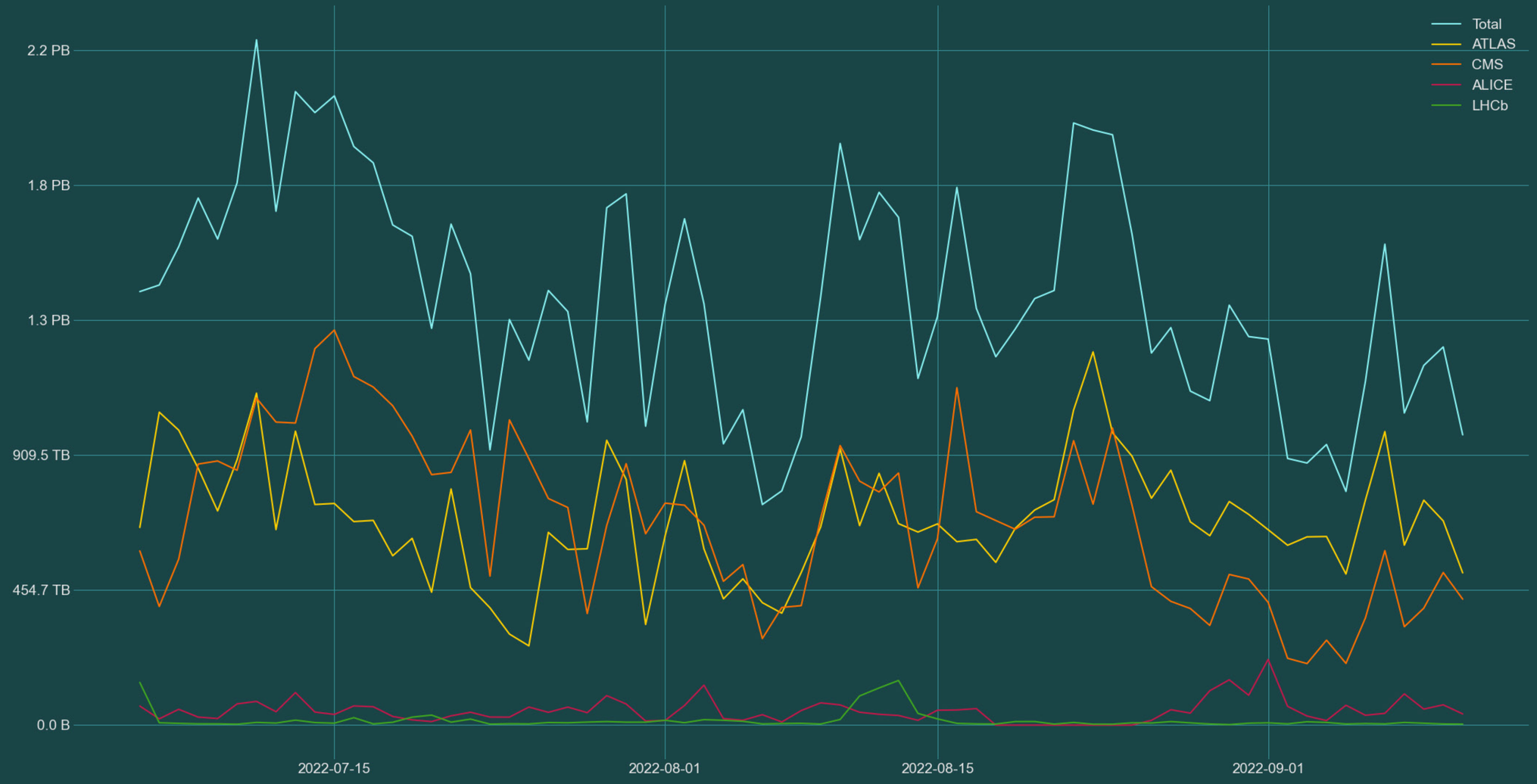
Data aggregation

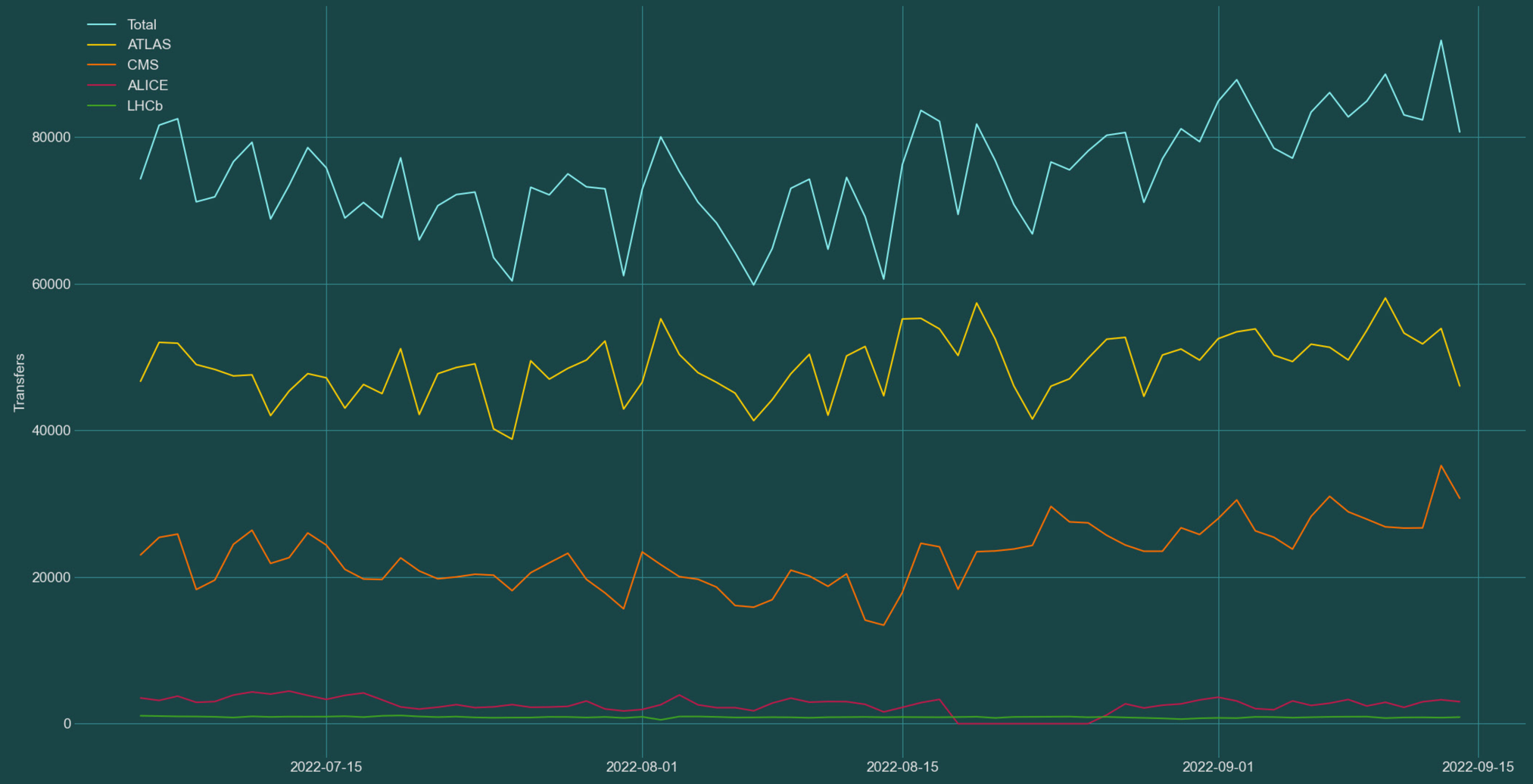
- At any given time the APIs for data transfers and job statistics offer aggregated data in a 1-hour window.
- The transfer API provides aggregated transfers for each experiment and source-destination pair. This means that for each source and destination pair there may be up to 8 entries in the response, depending on the number of VOs using the link and the state of the transfers (failed or successful).
- The APIs that expose the topology of Sites are always in sync with CRIC.

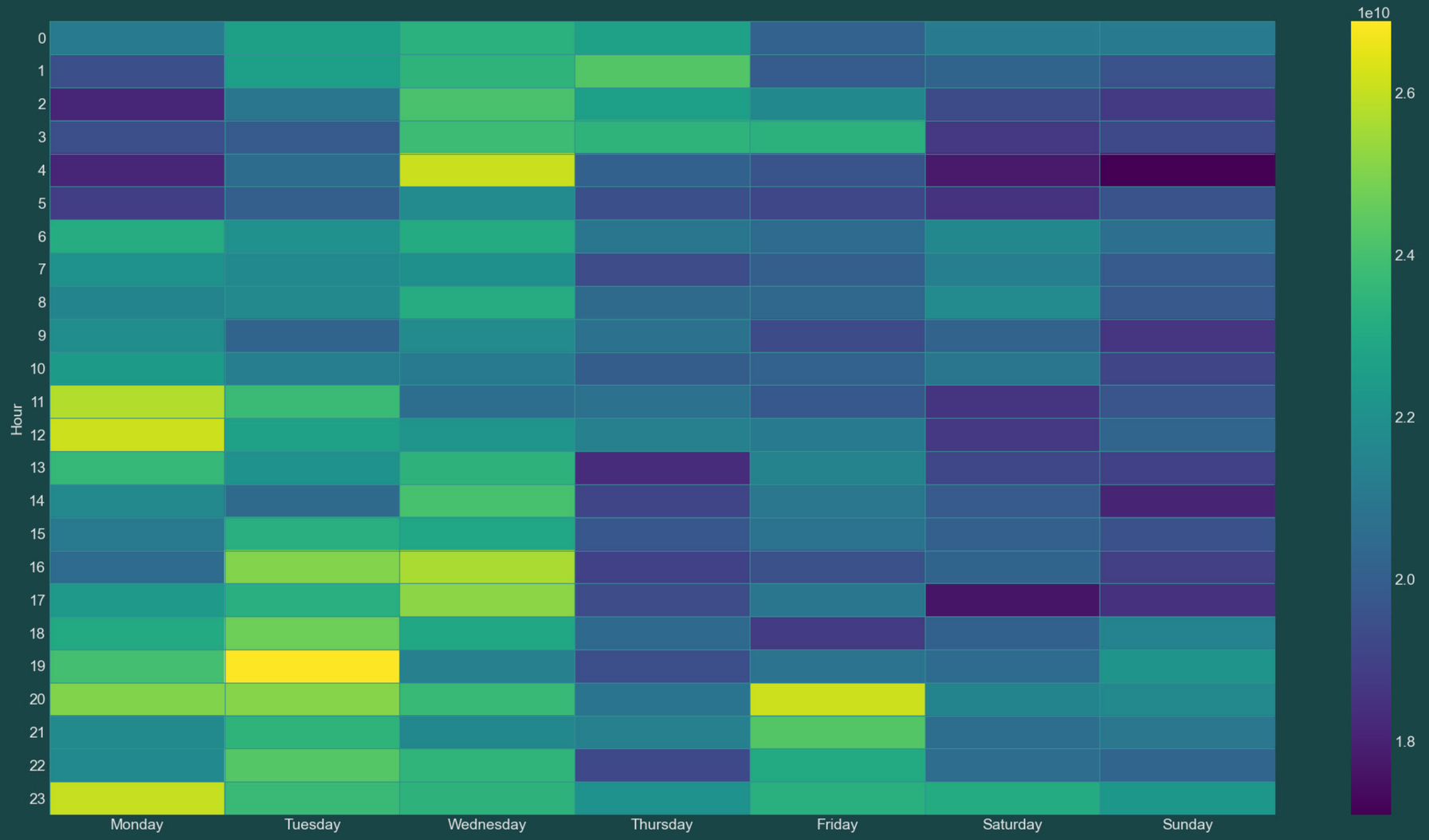

EXPERIMENTS SELECTION
TIERS SELECTION
ENABLED
DATES SELECTION
ACTIVE
SELECT DATES
TIME SCALE
COUNTRIES SELECTION
LIST
NO COUNTRY SELECTED
Afghanistan
Aland
Albania
Algeria
American Samoa
Andorra
Angola
GLOBE VIEW
MAP VIEW


success
fail
80000
70000
60000
50000
40000
30000
20000
10000
Transfers
03-07-2022-00
05-07-2022-00
07-07-2022-00
09-07-2022-00
11-07-2022-00

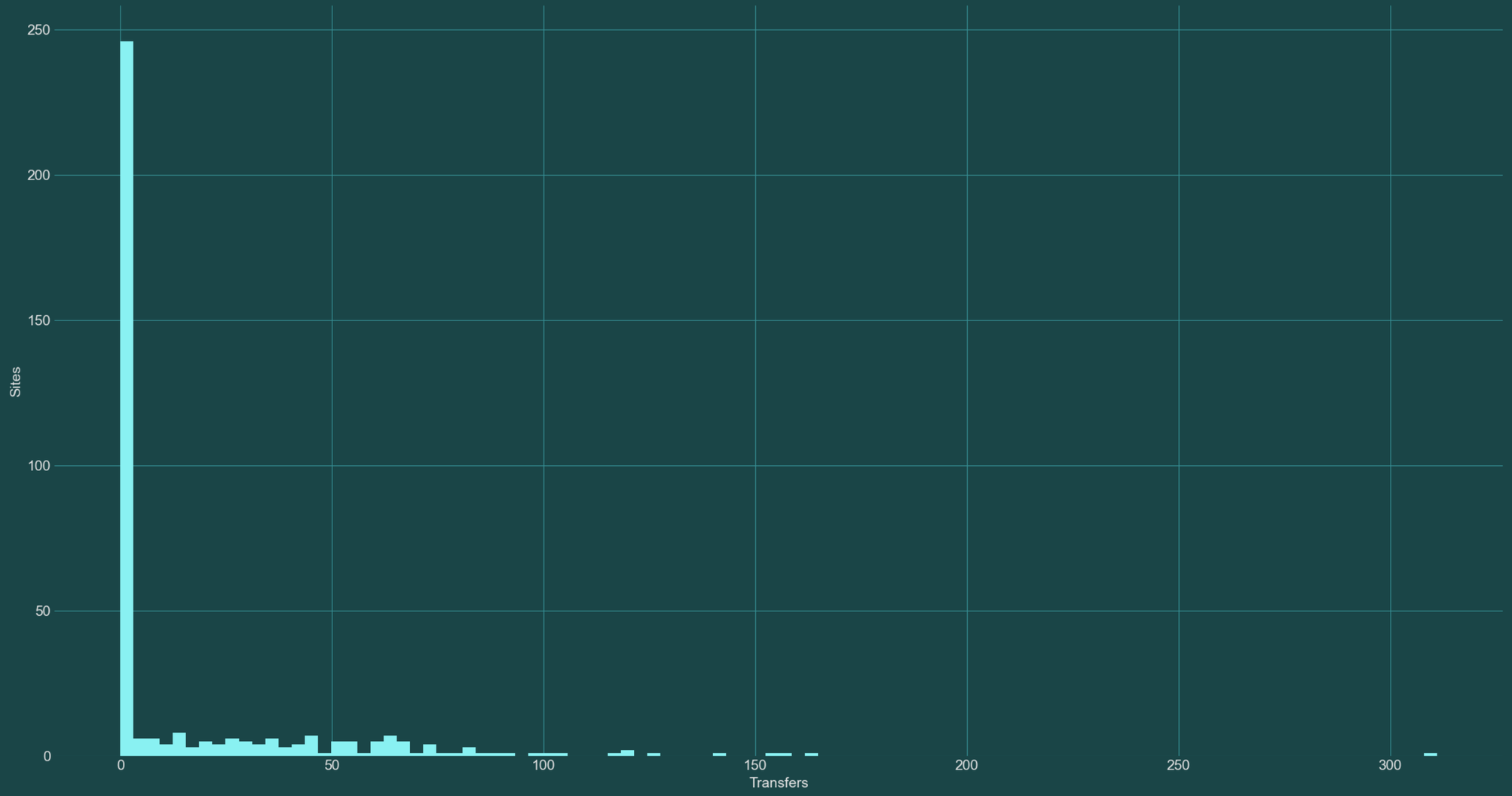

success
fail
80000
60000
40000
20000
Transfers
15-07-2022-00
01-08-2022-00
15-08-2022-00
01-09-2022-00
15-09-2022-00


Total
ATLAS
CMS
ALICE
LHCb
2.2 PB
1.8 PB
1.3 PB
909.5 TB
454.7 TB
0.0 B
2022-07-15
2022-08-01
2022-08-15
2022-09-01


Total
ATLAS
CMS
ALICE
LHCb
Transfers
80000
60000
40000
20000
0
2022-07-15
2022-08-01
2022-08-15
2022-09-01
2022-09-15


Hour
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Sunday
1e10
2.6
2.4
2.2
2.0
1.8


200
150
100
50
0
Sites
0
2000
4000
6000
8000
10000
Jobs


250
200
150
100
50
0
Sites
0
50
100
150
200
250
300
Transfers

Transfers without:

- Destination and origin: 0
- Destination: 1475
- Origin: 421

On a sample of 5408237 transfers
from 5/07/2022 to 14/09/2022

Do these Sites exist?

- Total individual sites: 371
- Sites without:
  - Transfers: 201
  - Jobs: 200
  - Jobs or transfers: 175
- Sites informations missing:
  - cpu_capacity: 261
  - cores: 259
  - core_power: 261
  - geolocation: 74

Stale sites will need to be filtered out.

Lessons Learned

- Conveying the necessary computing resources
- Grasping the large amount of data generated
- Identifying the increase of data production
- Visualizing the great work put into advancing physics
- A great tool for communicating with the public
- Celebrating the significance of collaborations

Future Work

- Visual representation of the LHC and experiments
- Representing data flow from collisions to the Data Center
- Simulating collision events
- PDF reader

Thank you.