DISTRIBUTED MEASUREMENTS AND ANALYSIS OF THE GLOBAL INTERNET - BOINC (BERKELEY OPEN INFRASTRUCTURE FOR NETWORK COMPUTING) AND BIG DATA ENVIRONMENT

LVEE 2020 ONLINE EDITION

19.12.2020

Łukasz Świerczewski

Chairman of the Board

Co-Founder

Cyber-Complex Foundation

lswierczewski @cybercomplex.net

www.cybercomplex.net

Cyber-Complex Foundation

iThena

howfaster.net

Presentation program:

1. BOINC
   - What is BOINC? What projects do we have? What initiatives do we have? Old topic…
   - The idea of volunteer cooperation.
2. iThena – Distributed BOINC Project
3. Internet Analytics
4. Elastic Ecosystem – BIG data solutions
5. Howfaster.Net portal as result
6. Summary. Questions?

BOINC

The Berkeley Open Infrastructure for Network Computing (BOINC) is an open-source middleware system for volunteer computing and grid computing. Originally developed to support the SETI@home project, it became generalized as a platform for other distributed applications in areas as diverse as mathematics, linguistics, medicine, molecular biology, climatology, environmental science, and astrophysics, among others. BOINC aims to enable researchers to tap into the enormous processing resources of multiple personal computers around the world.

Source: https://en.wikipedia.org/wiki/Berkeley_Open_Infrastructure_for_Network_Computing

BOINC project website: https://boinc.berkeley.edu/
BOINC License: GNU Lesser General Public License

BOINC

LVEE and FOSS Lviv presentations (historical - years: 2012-2015):

1. Berkeley Open Infrastructure for Network Computing - an open distributed computing system
2. International system of nuclear gamma-ray detection using dedicated hardware sensors - Radioactive@Home project
3. Simulation of Grover's algorithm on parallel computers with shared memory and using the Olib library
4. OProject@Home - distributed computing
5. Berkeley Open Infrastructure for Network Computing — an open distributed computing system
6. Steganography - coding and intercepting the information from encoded pictures in the absence of any initial information
7. BOINC - Not only calculations

BOINC - Community

BOINC Stats: https://www.boincstats.com

Free-DC Stats: https://stats.free-dc.org/stats.php?page=index

Many, many open projects

List of BOINC projects on the official UC Berkeley website:
https://boinc.berkeley.edu/projects.php

And many other projects that are not on the official list…

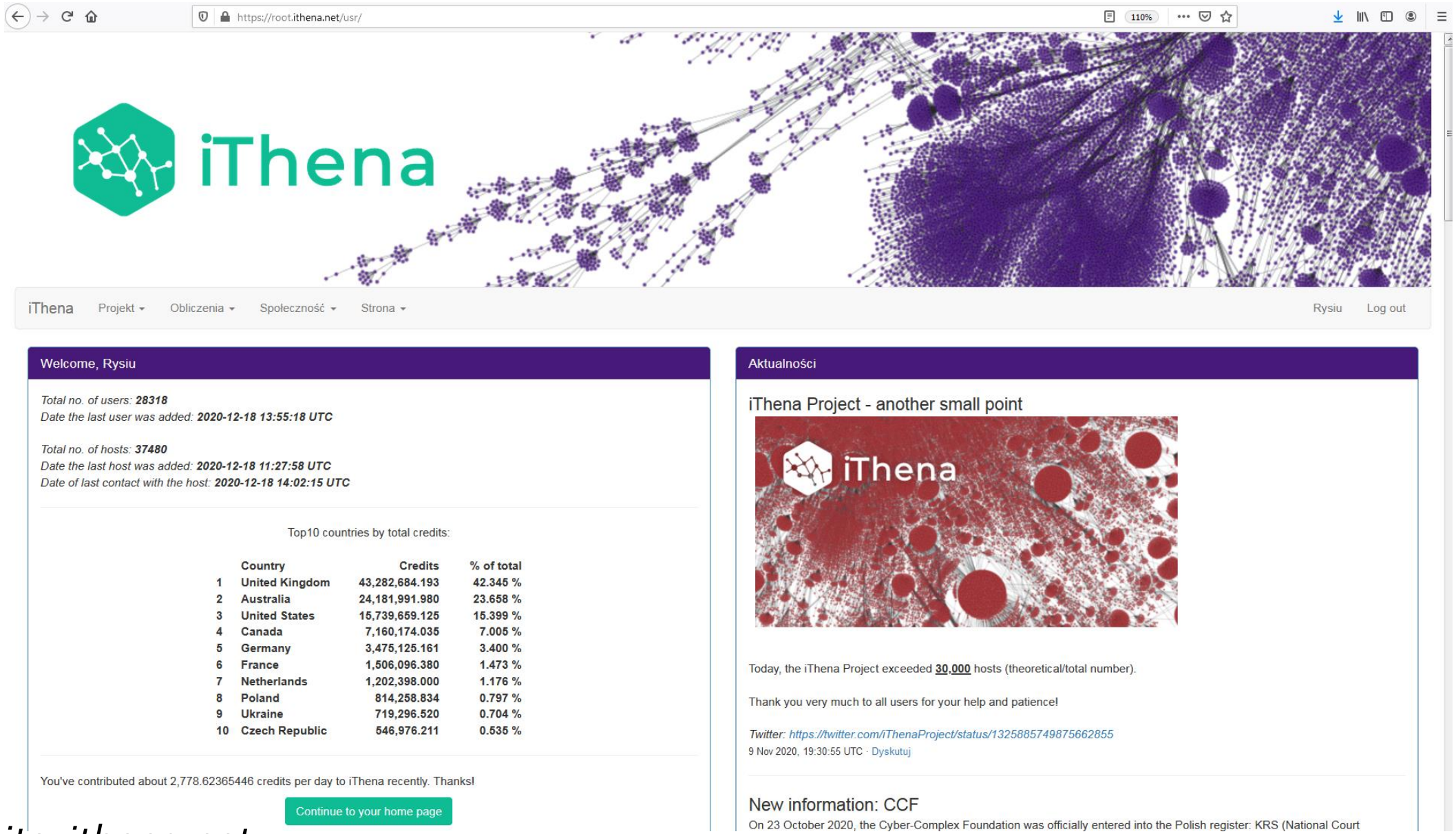
iThena Project - part of the global BOINC platform

The iThena distributed project concerns experimental mapping of network structures included in the Internet. The project is in closed beta phase. Currently, the only application available in the project (iThena CNode) performs a sequence of traceroute procedures from client computers. The resulting data is sent back to the server and submitted to the main database, where it can be further analyzed.

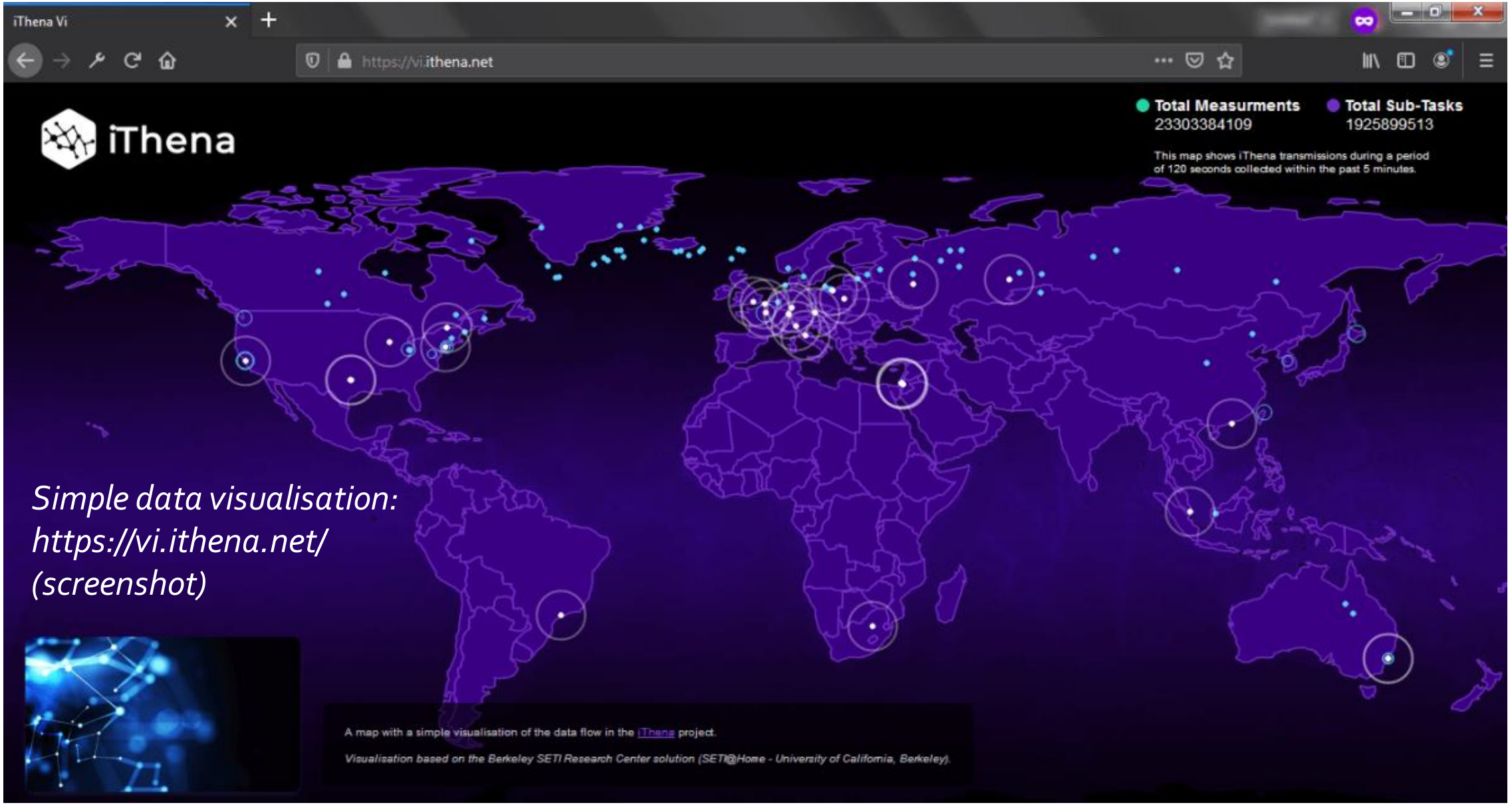
Total no. of users: **28398** (date: 19.12.2020)
Total no. of hosts: **37495** (date: 19.12.2020)

Project website: **https://ithena.net**
Simple data visualisation: **https://vi.ithena.net/**

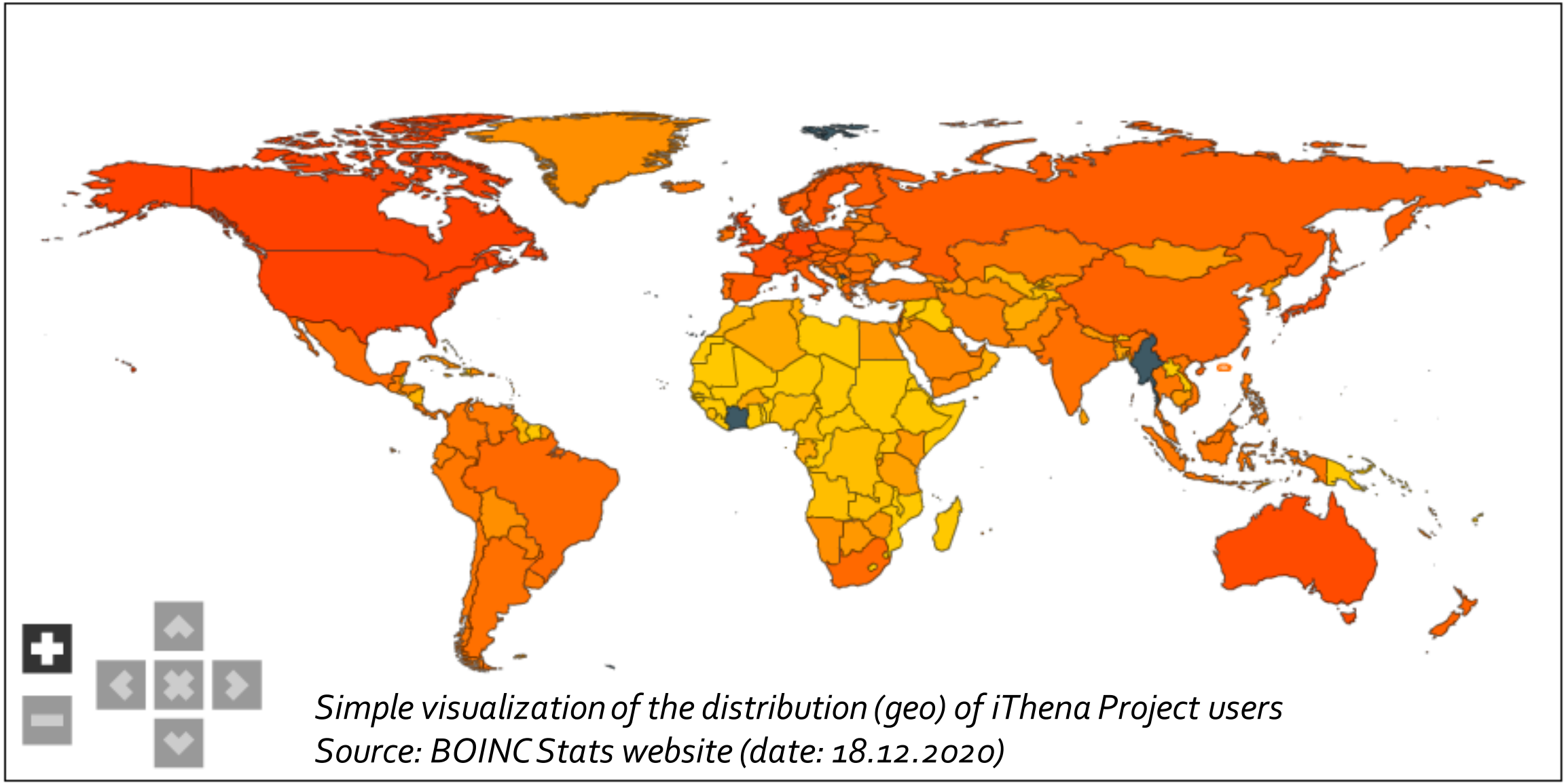
iThena page on Everipedia: **https://everipedia.org/wiki/lang_en/ithena**



Website: ithena.net



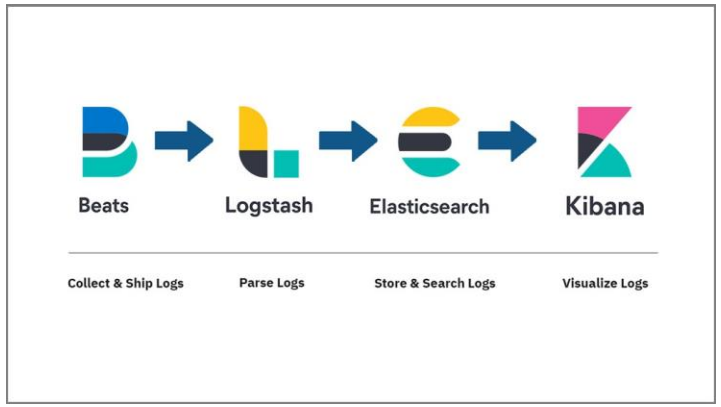
Simple data visualisation:
https://vi.ithena.net/
(screenshot)



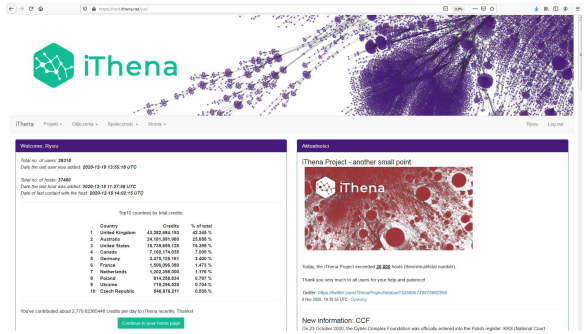
Simple visualization of the distribution (geo) of iThena Project users
Source: BOINC Stats website (date: 18.12.2020)

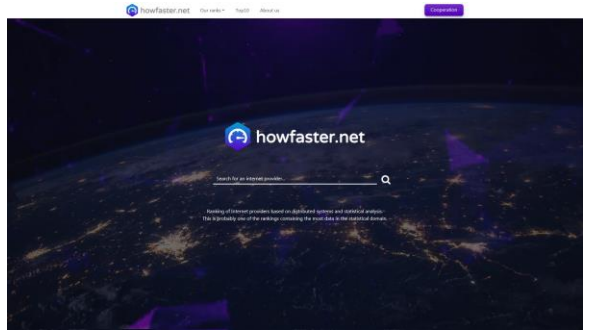
Project - simple flow schema

INPUT



OUTPUT

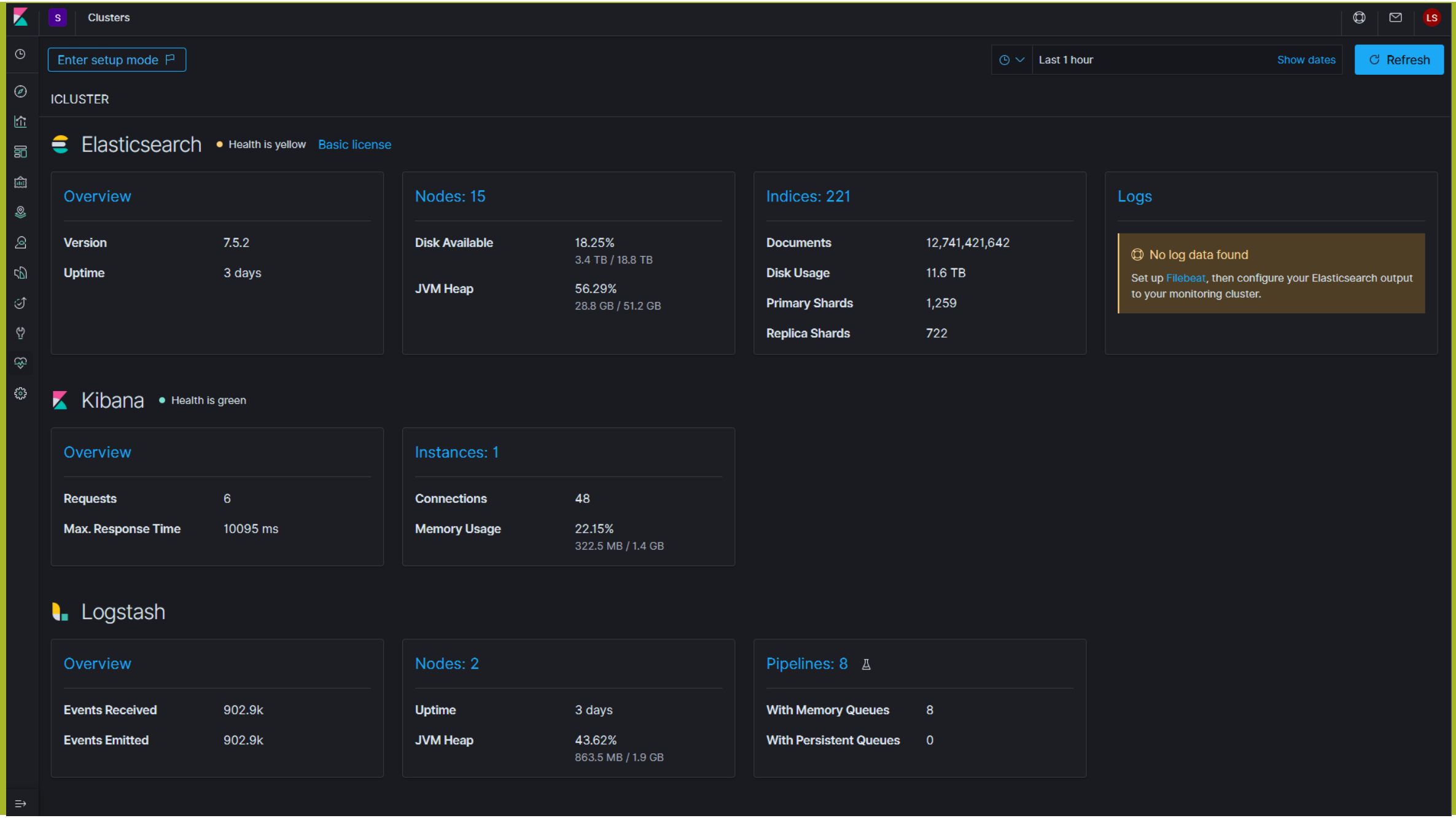




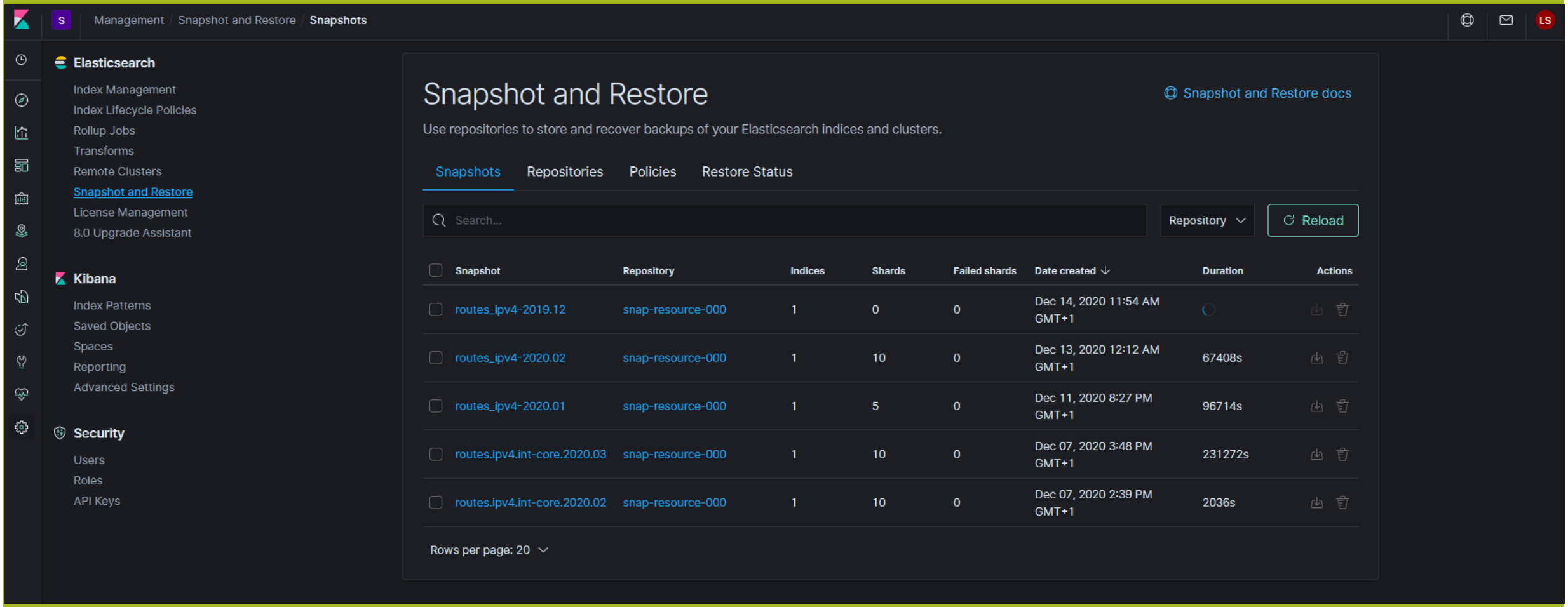
Elasticsearch and Elastic Stack - BIG Data environment

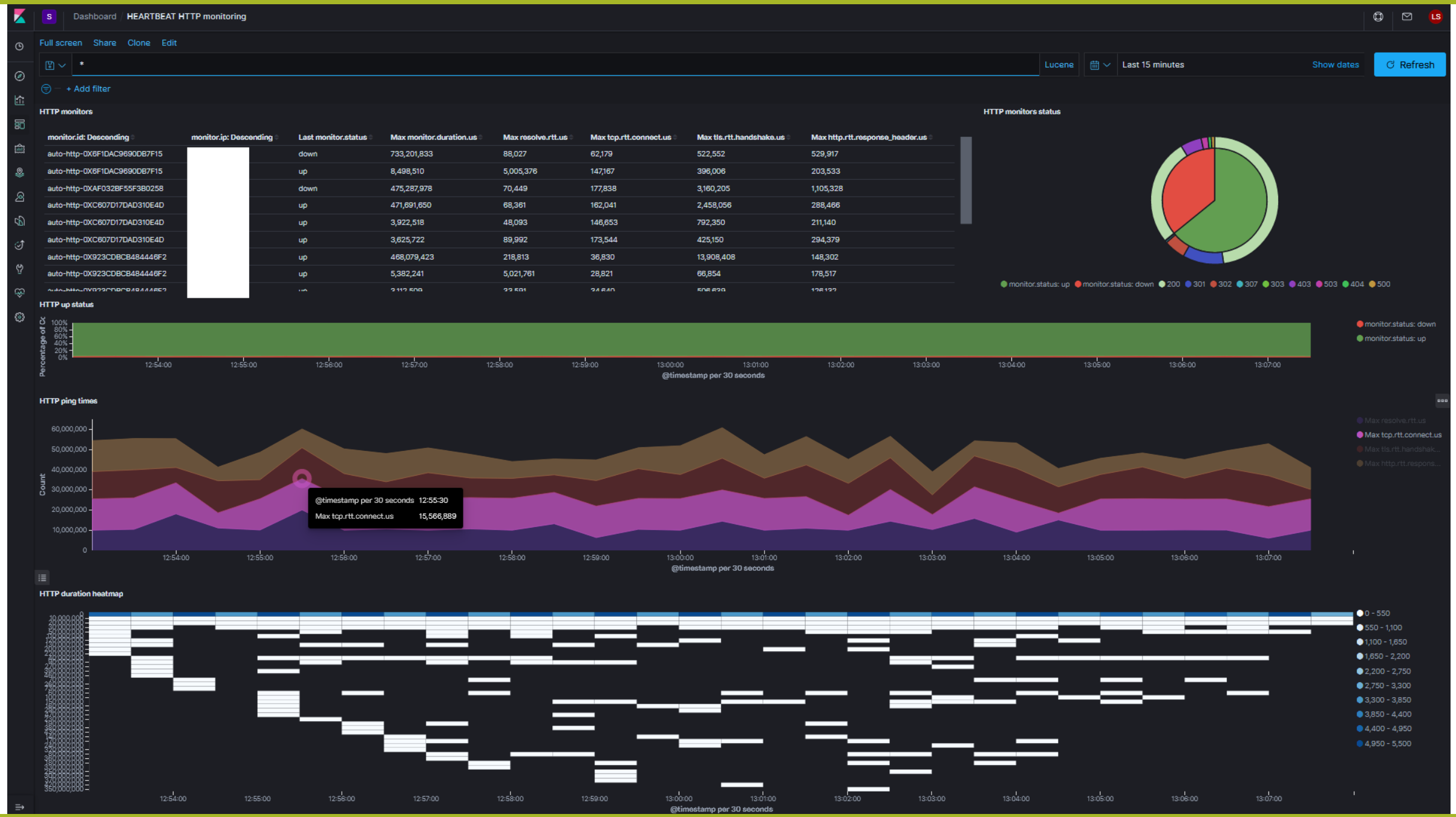
Elasticsearch is a search engine based on the Lucene library. It provides a distributed, multitenant-capable full-text search engine with an HTTP web interface and schema-free JSON documents. Elasticsearch is developed in Java. Following an open-core business model, parts of the software are licensed under various open-source licenses (mostly the Apache License), while other parts fall under the proprietary (source-available) Elastic License. Official clients are available in Java, .NET (C#), PHP, Python, Apache Groovy, Ruby and many other languages. According to the DB-Engines ranking, Elasticsearch is the most popular enterprise search engine followed by Apache Solr, also based on Lucene.

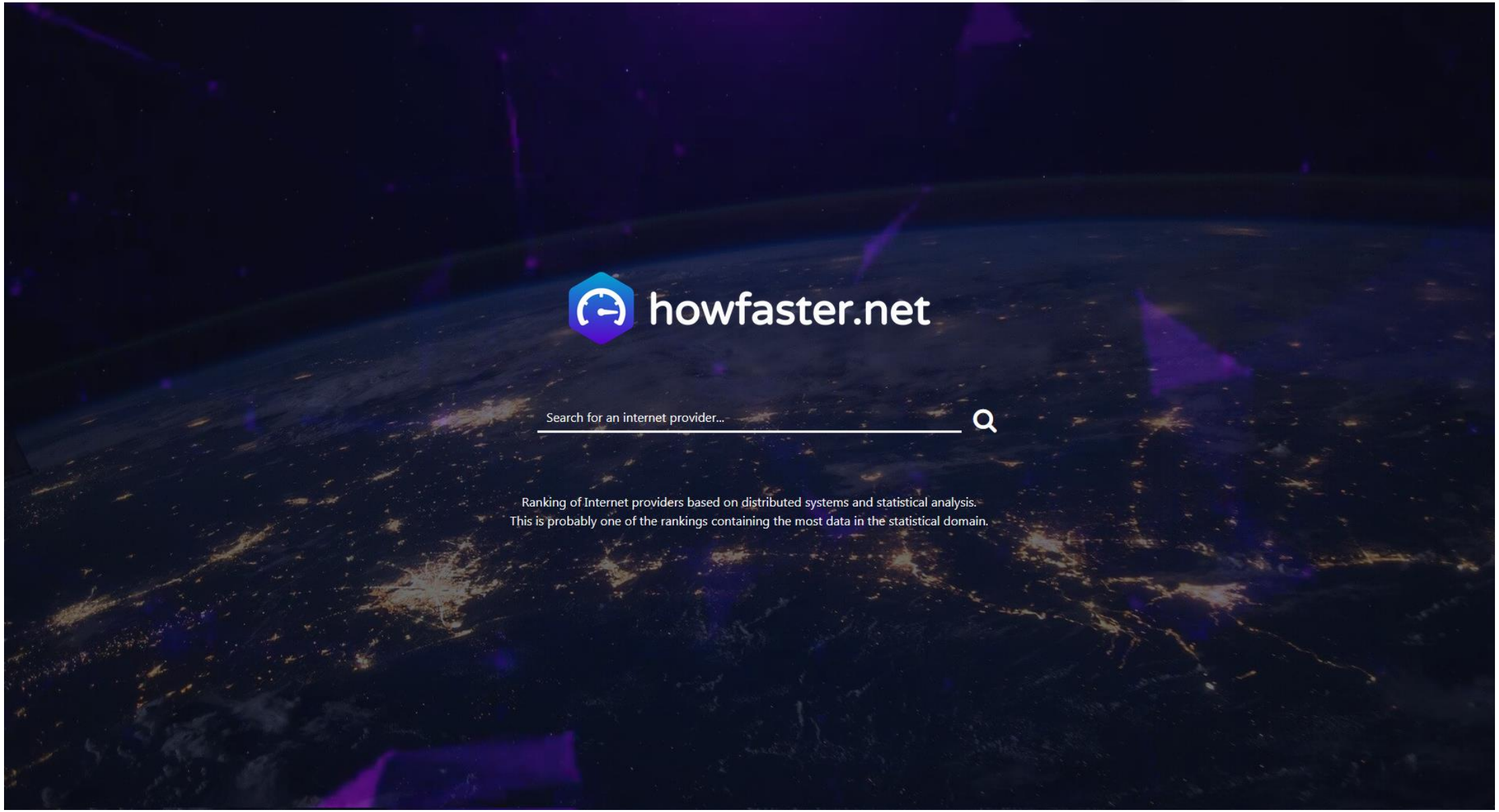
Source: https://en.wikipedia.org/wiki/Elasticsearch
Official Elastic website: https://www.elastic.co


Clusters
Enter setup mode
Last 1 hour
Show dates
Refresh
ICLUSTER
Elasticsearch
Health is yellow
Basic license
Overview
Version 7.5.2
Uptime 3 days
Nodes: 15
Disk Available 18.25%
3.4 TB / 18.8 TB
JVM Heap 56.29%
28.8 GB / 51.2 GB
Indices: 221
Documents 12,741,421,642
Disk Usage 11.6 TB
Primary Shards 1,259
Replica Shards 722
Logs
No log data found
Set up Filebeat, then configure your Elasticsearch output to your monitoring cluster.
Kibana
Health is green
Overview
Requests 6
Max. Response Time 10095 ms
Instances: 1
Connections 48
Memory Usage 22.15%
322.5 MB / 1.4 GB
Logstash
Overview
Events Received 902.9k
Events Emitted 902.9k
Nodes: 2
Uptime 3 days
JVM Heap 43.62%
863.5 MB / 1.9 GB
Pipelines: 8
With Memory Queues 8
With Persistent Queues 0

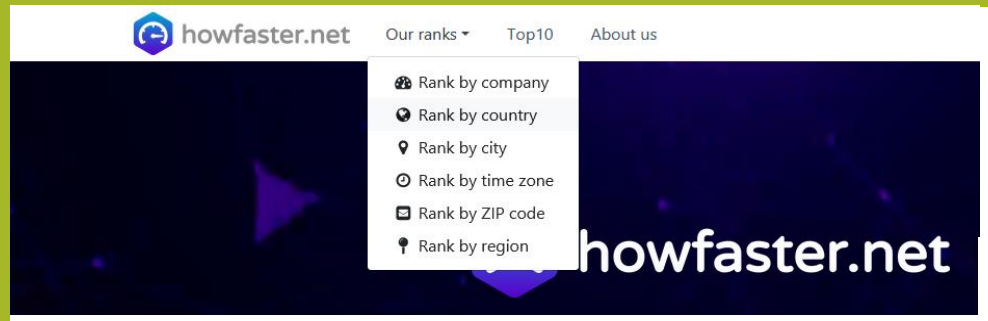

S
Clusters / ICLUSTER / Elasticsearch
LS
Last 1 hour
Show dates
Refresh
Overview
Nodes
Indices
Status
Yellow
Nodes
15
Indices
221
Memory
28.7 GB / 51.2 GB
Total shards
2537
Unassigned shards
556
Documents
12,741,544,660
Data
11.6 TB
Search Rate (/s)
2,500 /s
2,000 /s
1,500 /s
1,000 /s
11:50 11:55 12:00 12:05 12:10 12:15 12:20 12:25 12:30 12:35 12:40 12:45
Total Shards 1,869.63 /s
Search Latency (ms)
20 ms
18 ms
16 ms
14 ms
12 ms
10 ms
11:50 11:55 12:00 12:05 12:10 12:15 12:20 12:25 12:30 12:35 12:40 12:45
Search Latency 13.08 ms
Indexing Rate (/s)
1,250 /s
1,000 /s
750 /s
500 /s
250 /s
0 /s
11:50 11:55 12:00 12:05 12:10 12:15 12:20 12:25 12:30 12:35 12:40 12:45
Total Shards 925.1 /s
Primary Shards 617.97 /s
Indexing Latency (ms)
4 ms
3.5 ms
3 ms
2.5 ms
2 ms
1.5 ms
1 ms
11:50 11:55 12:00 12:05 12:10 12:15 12:20 12:25 12:30 12:35 12:40 12:45
Indexing Latency 1.33 ms


Management / Snapshot and Restore / Snapshots
LS
Elasticsearch
Index Management
Index Lifecycle Policies
Rollup Jobs
Transforms
Remote Clusters
Snapshot and Restore
License Management
8.0 Upgrade Assistant
Kibana
Index Patterns
Saved Objects
Spaces
Reporting
Advanced Settings
Security
Users
Roles
API Keys
Snapshot and Restore
Snapshot and Restore docs
Use repositories to store and recover backups of your Elasticsearch indices and clusters.
Snapshots
Repositories
Policies
Restore Status
Search...
Repository
Reload
Snapshot
Repository
Indices
Shards
Failed shards
Date created
Duration
Actions
routes_ipv4-2019.12 snap-resource-000 1 0 0 Dec 14, 2020 11:54 AM GMT+1
routes_ipv4-2020.02 snap-resource-000 1 10 0 Dec 13, 2020 12:12 AM GMT+1 67408s
routes_ipv4-2020.01 snap-resource-000 1 5 0 Dec 11, 2020 8:27 PM GMT+1 96714s
routes.ipv4.int-core.2020.03 snap-resource-000 1 10 0 Dec 07, 2020 3:48 PM GMT+1 231272s
routes.ipv4.int-core.2020.02 snap-resource-000 1 10 0 Dec 07, 2020 2:39 PM GMT+1 2036s
Rows per page: 20


Dashboard / HEARTBEAT HTTP monitoring
Full screen Share Clone Edit
Lucene
Last 15 minutes
Show dates
Refresh
+ Add filter
HTTP monitors
monitor.id: Descending
monitor.ip: Descending
Last monitor.status
Max monitor.duration.us
Max resolve.rtt.us
Max tcp.rtt.connect.us
Max tls.rtt.handshake.us
Max http.rtt.response_header.us
auto-http-0X6F1DAC9690DB7F15 down 733,201,833 88,027 62,179 522,552 529,917
auto-http-0X6F1DAC9690DB7F15 up 8,498,510 5,005,376 147,167 396,006 203,533
auto-http-0XAF032BF55F3B0258 down 475,287,978 70,449 177,838 3,160,205 1,105,328
auto-http-0XC607D17DAD310E4D up 471,691,650 68,361 162,041 2,458,056 288,466
auto-http-0XC607D17DAD310E4D up 3,922,518 48,093 146,653 792,350 211,140
auto-http-0XC607D17DAD310E4D up 3,625,722 89,992 173,544 425,150 294,379
auto-http-0X923CDBCB484446F2 up 468,079,423 218,813 36,830 13,908,408 148,302
auto-http-0X923CDBCB484446F2 up 5,382,241 5,021,761 28,821 66,854 178,517
HTTP monitors status
monitor.status: up
monitor.status: down
200
301
302
307
303
403
503
404
500
HTTP up status
Percentage of Cc
100%
80%
60%
40%
20%
0%
12:54:00
12:55:00
12:56:00
12:57:00
12:58:00
12:59:00
13:00:00
13:01:00
13:02:00
13:03:00
13:04:00
13:05:00
13:06:00
13:07:00
@timestamp per 30 seconds
monitor.status: down
monitor.status: up
HTTP ping times
Count
60,000,000
50,000,000
40,000,000
30,000,000
20,000,000
10,000,000
0
@timestamp per 30 seconds 12:55:30
Max tcp.rtt.connect.us 15,566,889
@timestamp per 30 seconds
Max resolve.rtt.us
Max tcp.rtt.connect.us
HTTP duration heatmap
0 - 550
550 - 1,100
1,100 - 1,650
1,650 - 2,200
2,200 - 2,750
2,750 - 3,300
3,300 - 3,850
3,850 - 4,400
4,400 - 4,950
4,950 - 5,500
@timestamp per 30 seconds


howfaster.net
Search for an internet provider...
Ranking of Internet providers based on distributed systems and statistical analysis.
This is probably one of the rankings containing the most data in the statistical domain.

16

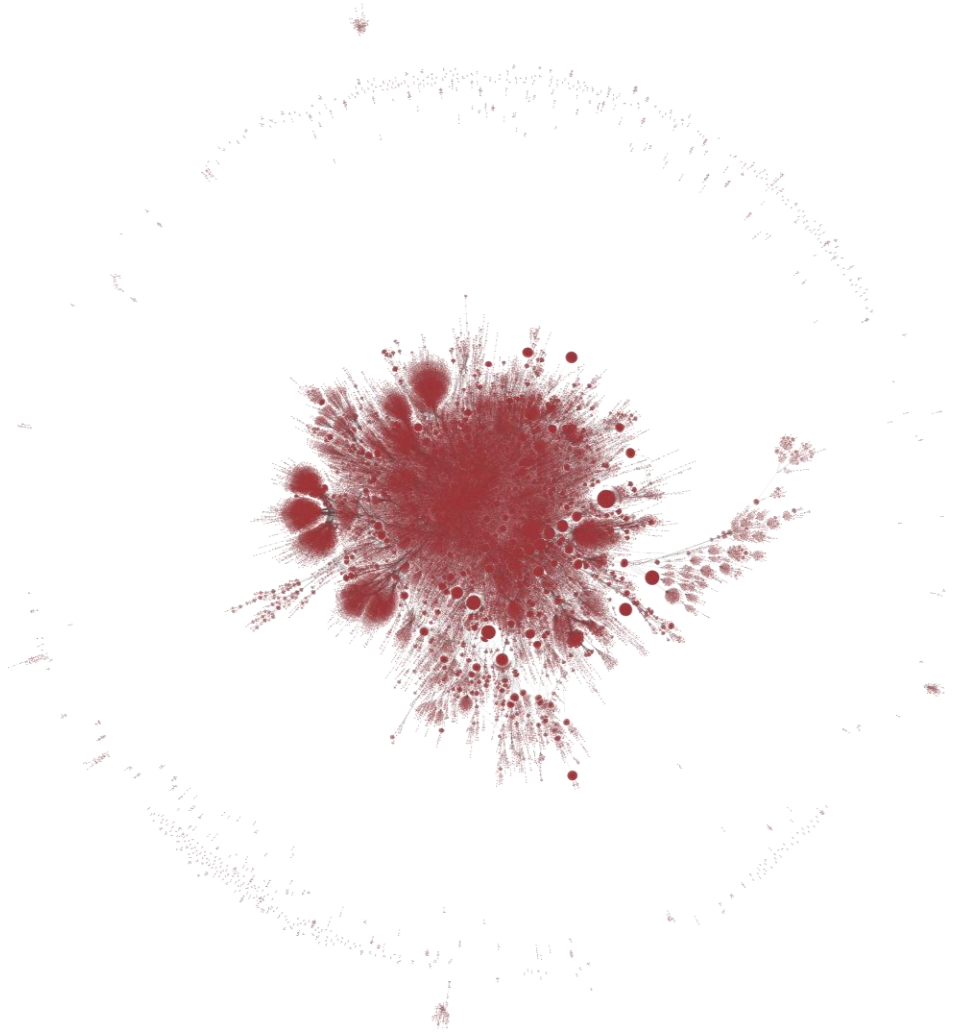


Website: **https://howfaster.net**

Number of:
AS: **6396**
Country: **238**
Regions: **3078**
City: **6307**
Measurments: **27632107**

Date: 19.12.2020 (data from the first two months of the iThena project only)

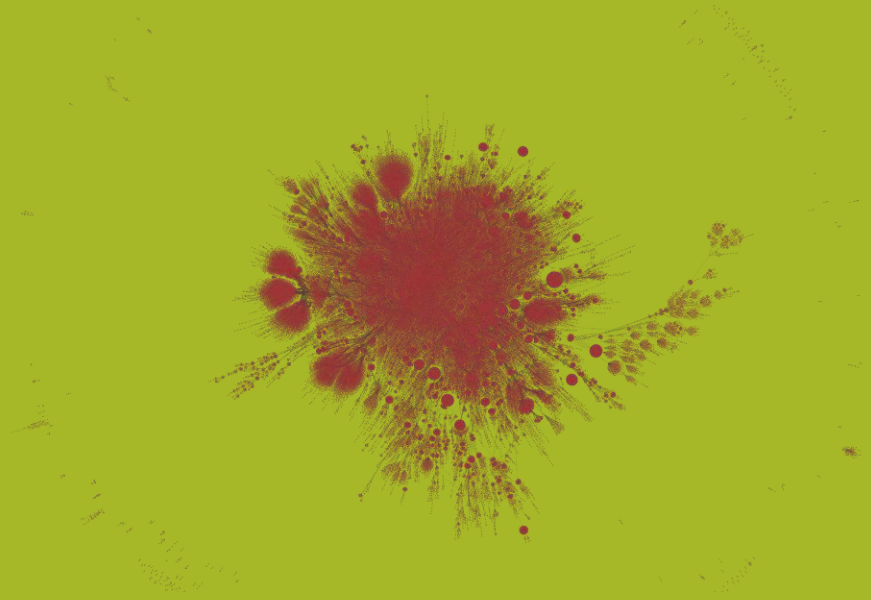
Rank		Min hops	Avg hops	Max hops	Min delay	Avg delay	Max delay	Count No.
1	Dillingen	4.79	5.04	6.38	21.4	32.67	79.78	24
2	Landen	4.76	5.61	7.43	23.3	34.11	53.71	46
3	Temple Terrace	5.98	8.27	10.97	26.38	35.02	72.03	88
4	Munchen	3.08	8.62	13.24	14.03	37.49	113.46	248
5	The Colony	3.18	5.04	11.52	12.83	39.28	207.77	851
6	Cesson-Sevigne	6.4	7.63	9.75	32.46	40.03	55.89	48
7	Havant	5.54	7.92	12.38	19.57	41.59	82.69	26
8	Leiderdorp	6.13	6.28	7.19	36.19	42.57	56.73	32
9	Zutphen	5.42	6.57	9.08	19.65	43.45	89.94	60
10	Bonn	3.2	5.12	22.19	12.23	44.36	248.92	6299
11	Lagny-sur-Marne	6.27	6.73	7.4	32.75	44.6	79.11	30
12	Merzig	5.05	5.24	6.22	30.42	45.3	106.88	37
13	Monroe	3	5.46	26.34	9.93	47.04	534.28	33829



ANY QUESTIONS?

LSWIERCZEWSKI@CYBERCOMPLEX.NET

WWW.CYBERCOMPLEX.NET



THANK YOU FOR YOUR ATTENTION!

LSWIERCZEWSKI@CYBERCOMPLEX.NET

WWW.CYBERCOMPLEX.NET