SUMMER INTERNSHIP AT CNAF: TOWARDS PREDICTIVE MAINTENANCE AT THE INFN-CNAF COMPUTING CENTRE

PRESENTED BY LUCA GIOMMI

PHD STUDENT IN DATA SCIENCE AND COMPUTATION (XXXIV CYCLE, YEAR 1)

INFN AND UNIVERSITY OF BOLOGNA

OBJECTIVE(S)

1. CNAF goal (long term): predictive maintenance system at the T1
2. *My goal (2 months): contribute to CNAF goal focusing on* log files from the T1 StoRM service
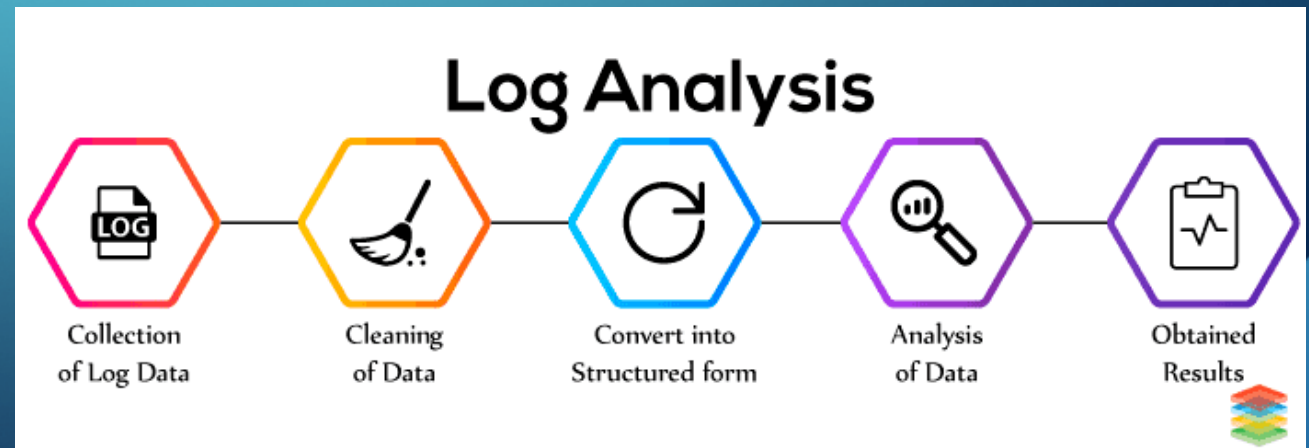
WHAT IS PREDICTIVE MAINTENANCE?

Predictive maintenance systems allow advance detection of pending failures and enable timely pre-failure interventions, thanks to prediction tools based on historical data, *ad hoc* defined *health factors*, statistical inference methods, and engineering approaches.
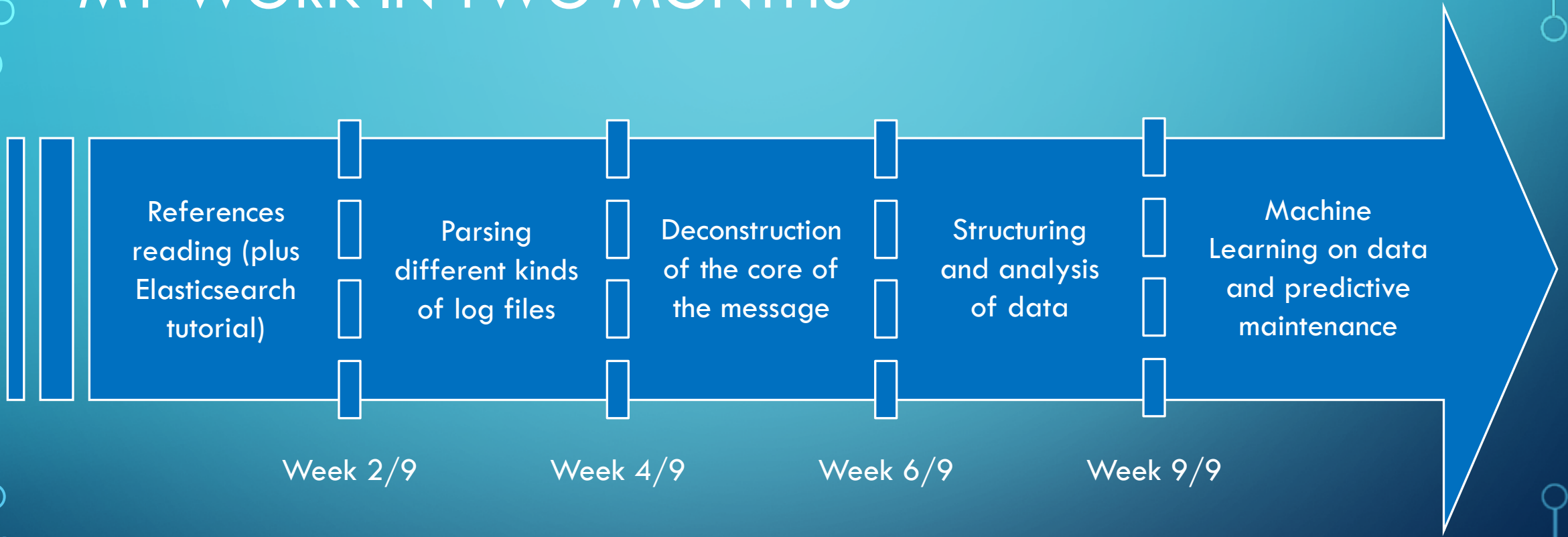
WHAT IS LOG ANALYSIS?

Computers, networks, and other IT systems generate records called logs that document system activities. The logs comprise of several messages that are chronologically arranged and stored on a disk. Log analysis is the evaluation of these records and is used by organizations to help mitigate a variety of risks. The steps for the processing of Log Analysis are:
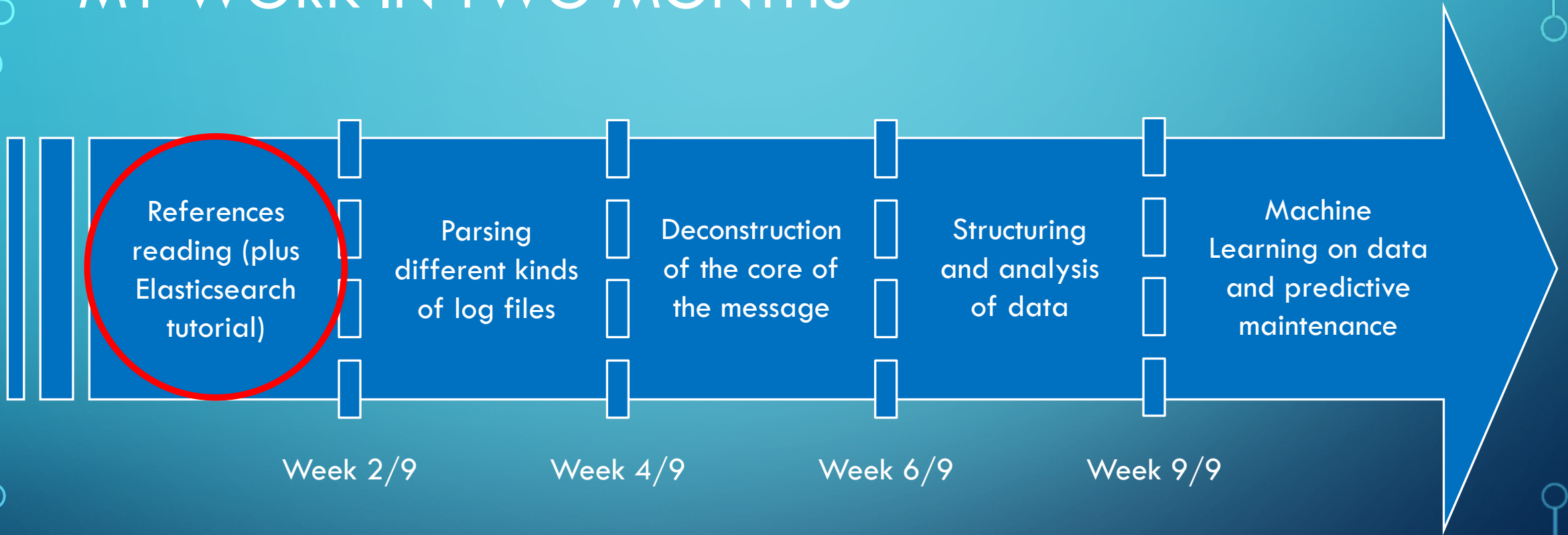
- Collection and Cleaning of data
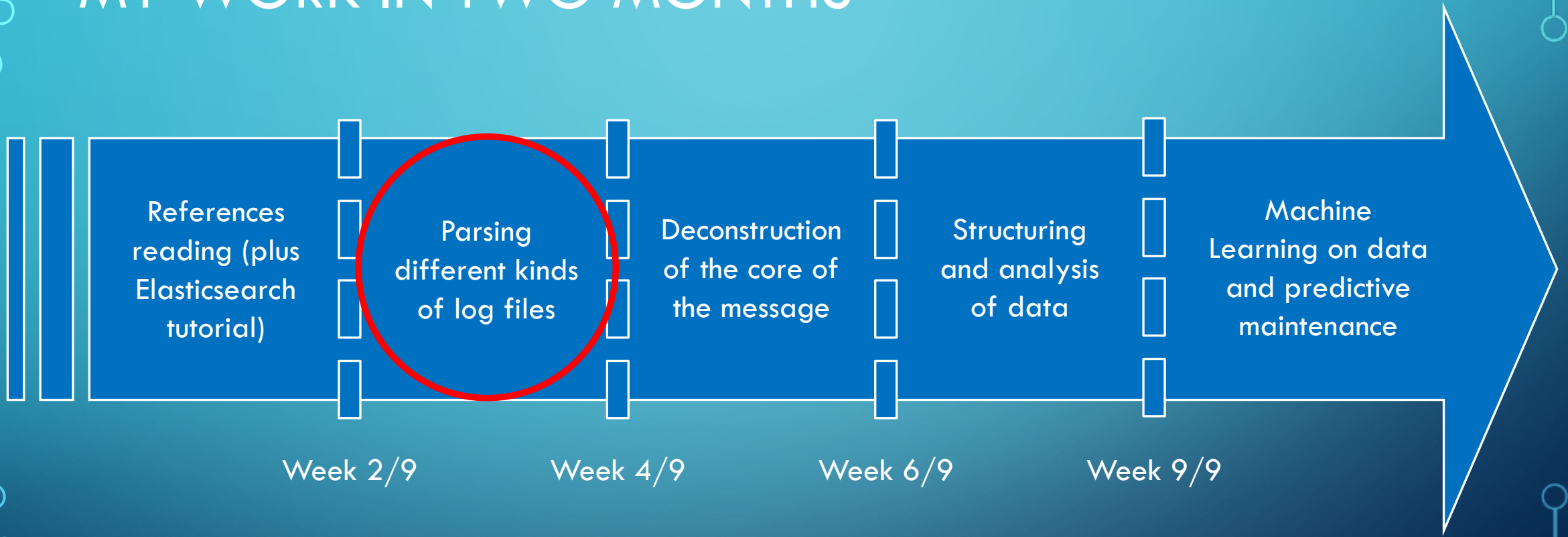
- Structuring of Data

- Analysis of Data

Log Analysis

Collection of Log Data — Cleaning of Data — Convert into Structured form — Analysis of Data — Obtained Results

MY WORK IN TWO MONTHS

References reading (plus Elasticsearch tutorial)	Parsing different kinds of log files	Deconstruction of the core of the message	Structuring and analysis of data	Machine Learning on data and predictive maintenance
	Week 2/9	Week 4/9	Week 6/9	Week 9/9

4

MY WORK IN TWO MONTHS

References reading (plus Elasticsearch tutorial)

Parsing different kinds of log files

Deconstruction of the core of the message

Structuring and analysis of data

Machine Learning on data and predictive maintenance

Week 2/9   Week 4/9   Week 6/9   Week 9/9

SOME REFERENCES

- G. A. Susto, A. Schirru, S. Pampuri, S. McLoone and A. Beghi, "Machine Learning for Predictive Maintenance: A Multiple Classifier Approach," in *IEEE Transactions on Industrial Informatics*, vol. 11, no. 3, pp. 812-820, June 2015.

- T. Kimura, A. Watanabe, T. Toyono and K. Ishibashi, "Proactive failure detection learning generation patterns of large-scale network logs," *2015 11th International Conference on Network and Service Management (CNSM)*, Barcelona, 2015, pp. 8-14.

- P. Bambharolia, P. Bhavsar, V. Prasad, "Failure Prediction And Detection In Cloud Datacenters", IJSTR, vol. 06, 2017.

- G. Wang, L. Zhang and W. Xu, "What Can We Learn from Four Years of Data Center Hardware Failures?," *2017 47th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*, Denver, CO, 2017, pp. 25-36.

- https://digitalguardian.com/blog/what-log-analysis-use-cases-best-practices-and-more

- https://www.xenonstack.com/blog/data-science/log-analytics-log-mining-anomaly-detection/

Example of "StoRM Frontend" log messages

```
09/06 03:07:20.907 Thread 380 -  INFO [aaec5ef5-09fc-4942-bceb-99cc9a2c3c54]: process_request :
Connection from 2001:1458:301:bb::100:e7
09/06 03:07:20.949 Thread 20 -  INFO [ed2c88f5-9cfd-4164-93e9-c3b82c61916a]: Request 'PTG status' from
Client IP=':::ffff:192.12.15.94' Client DN='/DC=ch/DC=cern/OU=Organic
Units/OU=Users/CN=ddmadmin/CN=531497/CN=Robot: ATLAS Data Management' # Requested token 'a9f73e94-72c6-
44cb-adfb-af9122927542'
```

Example of "StoRM Backend" log messages

```
00:00:00.018 - INFO [qtp663846109-31570] - srmPutDone: user</DC=ch/DC=cern/OU=Organic
Units/OU=Users/CN=atlpilo2/CN=531497/CN=Robot: ATLAS Pilot2> Request for [token: e388fadd-3e1b-440f-a889-
1b2506dae8aa] for [SURL: [srm://storm-
fe.cr.cnaf.infn.it:8444/atlas/atlasdatadisk/rucio/mc15_13TeV/d1/43/log.15310182._027646.job.log.tgz.1]]
succesfully done with  [status: SRM_SUCCESS: All file requests are successfully completed]
00:00:00.146 - WARN [qtp535195512-32019] - Invalid token provided - request: (GET
/recalltable/cardinality/tasks/readyTakeOver)@2003695725 org.eclipse.jetty.server.Request@776df86d
```

Example of "Messages" log messages

```
Sep 16 03:24:18 storm-fe-atlas-07 mmfs: [N] Connecting to 131.154.195.122 ui03-cms <c0n860>
Sep 16 03:53:51 storm-fe-atlas-07 puppet-agent[11807]: Finished catalog run in 28.71 seconds
```

Example of "Monitoring" log messages

```
09/06 03:08:36 : [#   887 lifetime=14:47:00] S [OK:1404529,F:44400,E:0,m:0.000,M:451.458,Avg:0.123] A
[OK:193461,F:0,E:0,m:0.005,M:4.810,Avg:0.027] Last:(S [OK:1632,F:33,E:0,m:0.001,M:2.635] A
[OK:162,F:0,E:0,m:0.009,M:1.109])
```
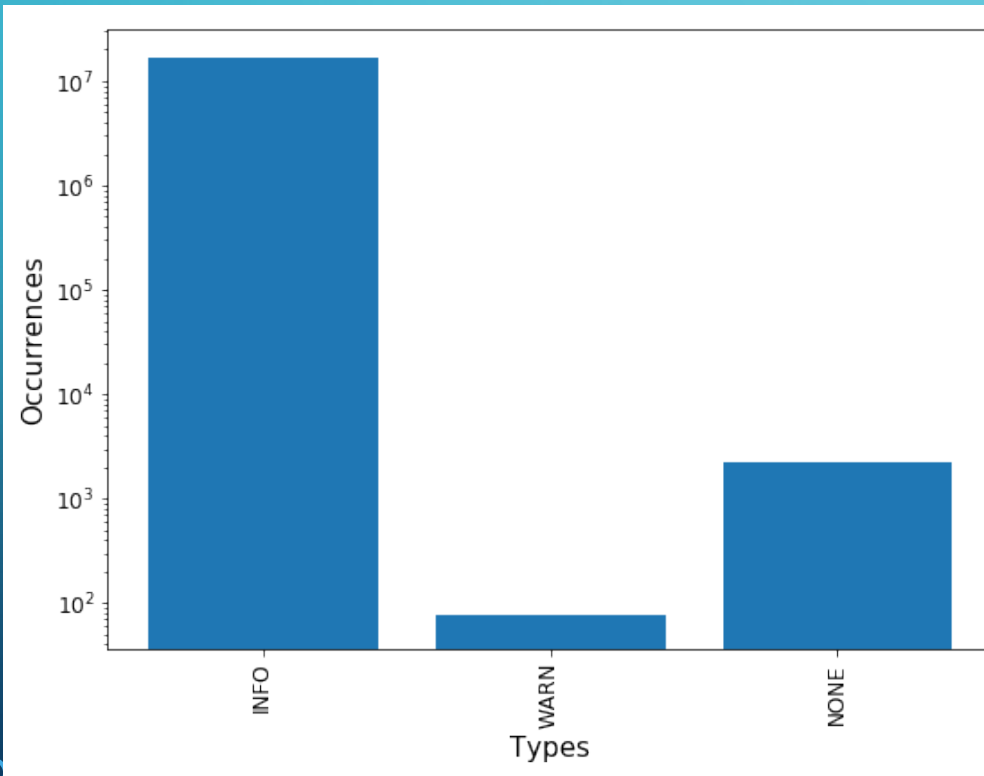
```
08/31 03:21:41.374 Thread 54 -   INFO [c42e140a-ed4f-4c2f-9a8c-ae96f53e8627]: process_request :
Connection from ::ffff:131.154.198.110
```
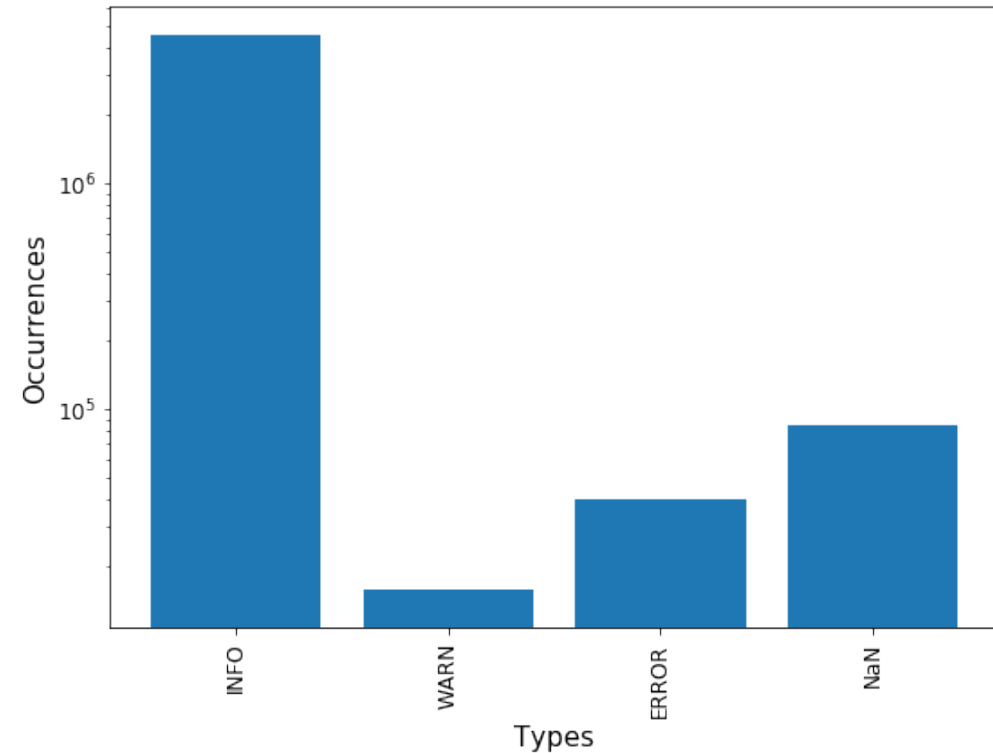
Parsing

```
timestamp,datetime,thread,type,request-ID,message
1535685701.374,2018-08-31 03:21:41.374000,54,INFO,c42e140a-ed4f-4c2f-9a8c-
ae96f53e8627,process_request : Connection from ::ffff:131.154.198.110
```
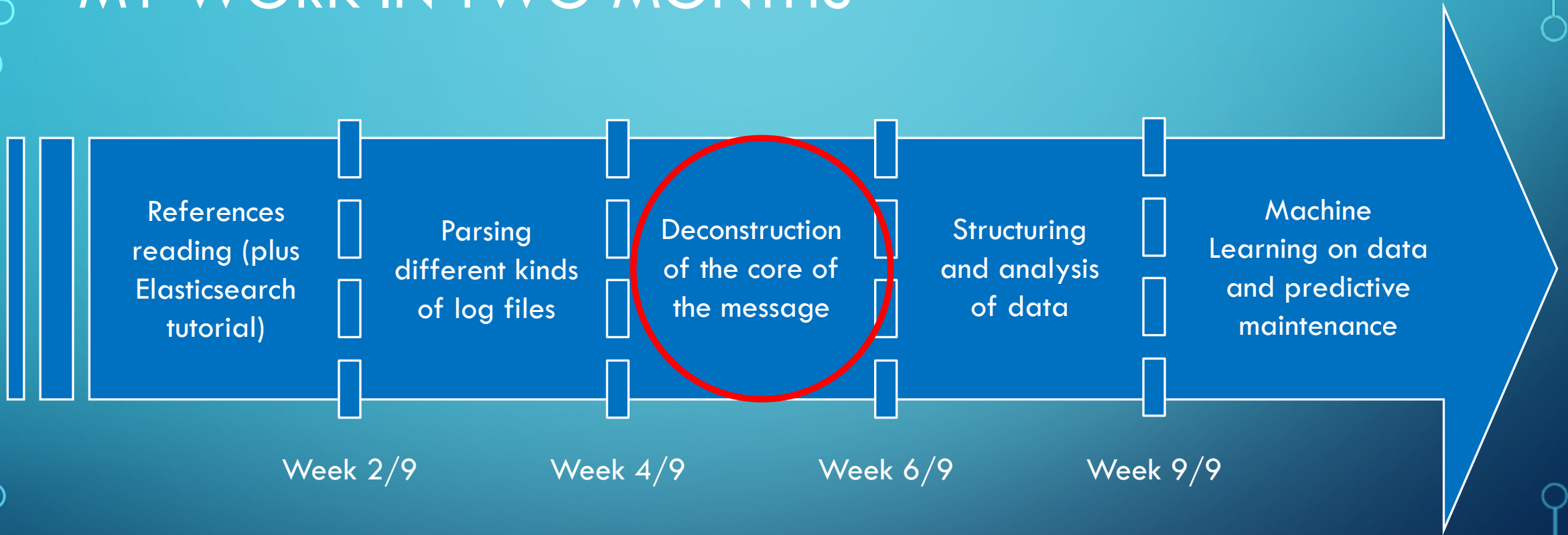

StoRM-FE 31/08/2018

StoRM-BE 07/09/2018

MY WORK IN TWO MONTHS

References reading (plus Elasticsearch tutorial)

Parsing different kinds of log files

Deconstruction of the core of the message

Structuring and analysis of data

Machine Learning on data and predictive maintenance

Week 2/9

Week 4/9

Week 6/9

Week 9/9

```
08/31 03:21:41.374 Thread 54 -  INFO [c42e140a-ed4f-4c2f-9a8c-ae96f53e8627]: process_request :
Connection from ::ffff:131.154.198.110
```

	timestamp	datetime	thread	type	request-ID	message
0	1.535686e+09	2018-08-31 03:21:41.374000	54	INFO	c42e140a-ed4f-4c2f-9a8c-ae96f53e8627	process_request : Connection from ::ffff:131.1...

DN	Request	ip	num_surl	result	surl	token_requested
NaN	Connection	::ffff:131.154.198.110	NaN	NaN	NaN	NaN

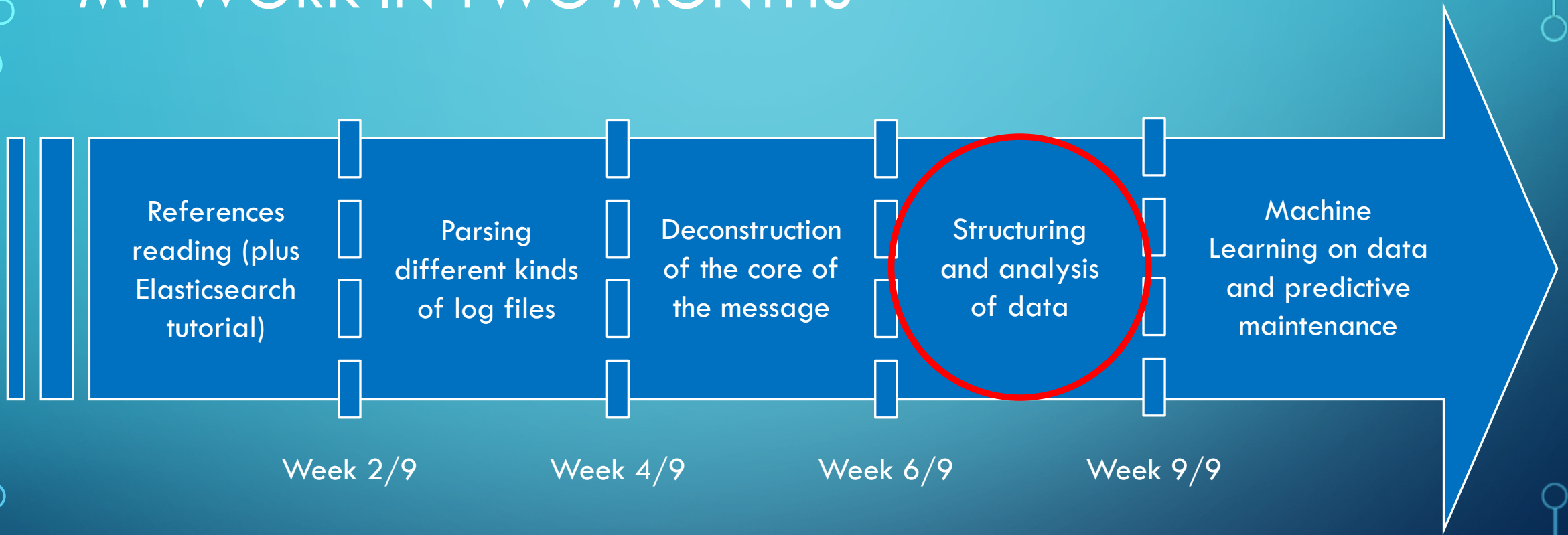```
08/31 03:23:02.650 Thread 4 - INFO [1ebfffbd-9575-4842-b2fb-5e483e5d7975]: Request 'PTG' from Client
IP='::ffff:131.154.198.111' Client DN='/DC=ch/DC=cern/OU=Organic Units/OU=Users/CN=atlpilo2/CN=531497/CN=Robot:
ATLAS Pilot2' # Requested '1' SURL(s): 'srm://storm-
fe.cr.cnaf.infn.it/atlas/atlasdatadisk/rucio/mc15_13TeV/99/c1/TXT.13035170._004547.tar.gz.1'
```

| timestamp | datetime | thread | type | request-ID | message |
|---|---|---|---|---|---|
| 1.535686e+09 | 2018-08-31 03:23:02.650000 | 4 | INFO | 1ebfffbd-9575-4842-b2fb-5e483e5d7975 | Request 'PTG' from Client IP='::ffff:131.154.1... |

| DN | Request | ip | num_surl | result | surl | token_requested |
|---|---|---|---|---|---|---|
| /DC=ch/DC=cern/OU=Organic Units/OU=Users/CN=at... | PTG | ::ffff:131.154.198.111 | 1.0 | NaN | ['srm://storm-fe.cr.cnaf.infn.it/atlas/atlasda... | NaN |

# MY WORK IN TWO MONTHS

References reading (plus Elasticsearch tutorial)

Parsing different kinds of log files

Deconstruction of the core of the message

Structuring and analysis of data

Machine Learning on data and predictive maintenance

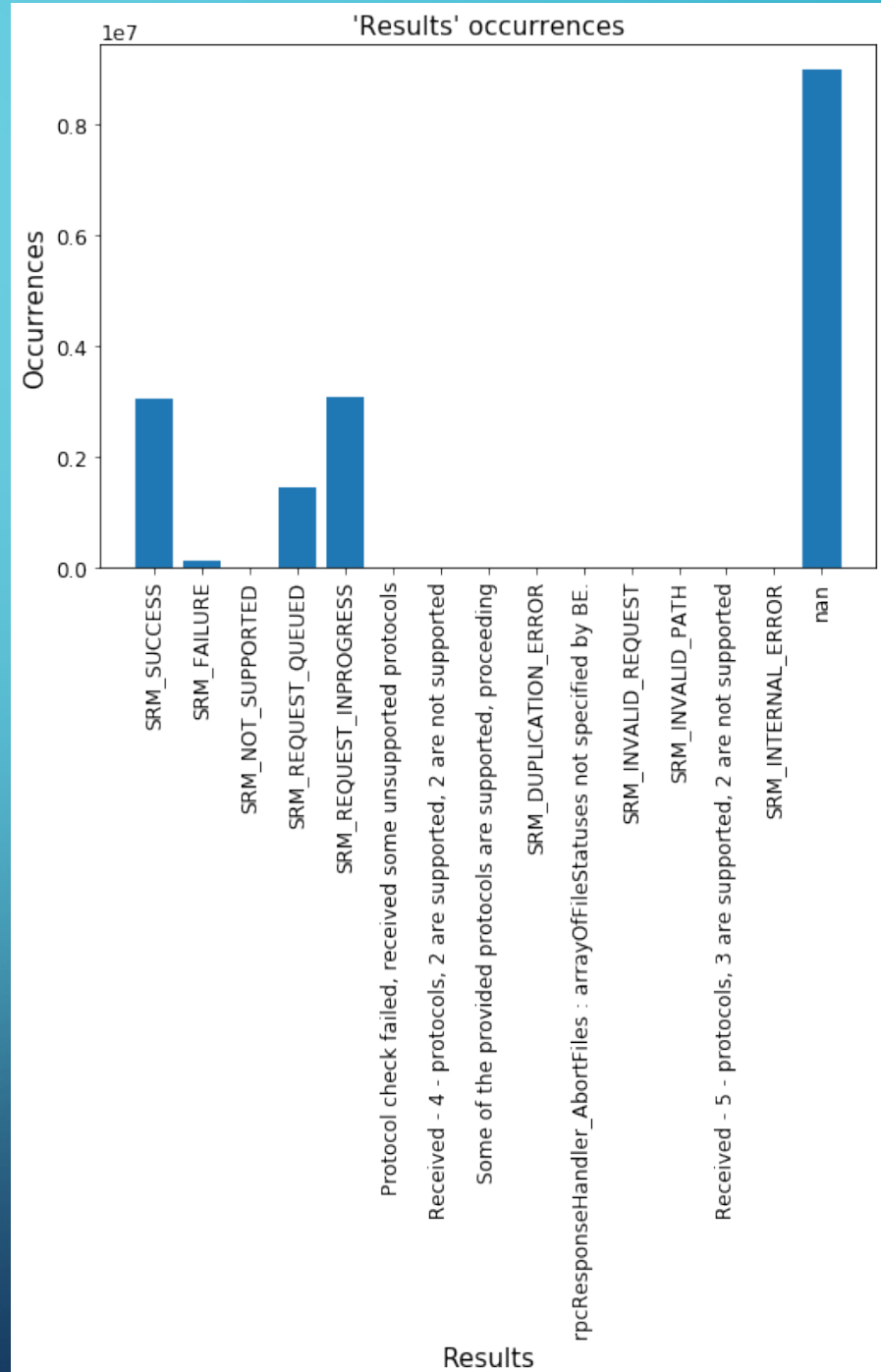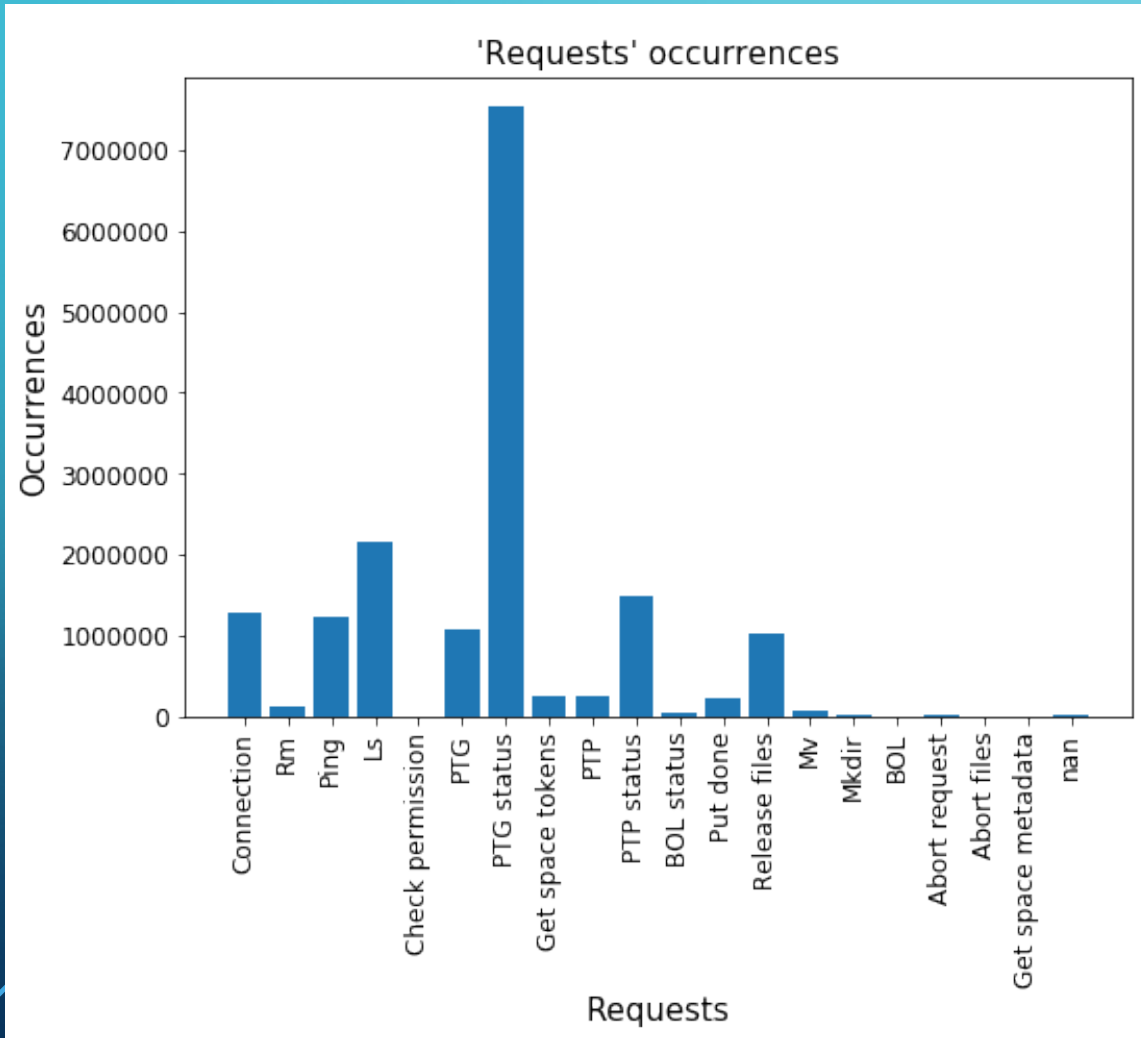Week 2/9          Week 4/9          Week 6/9          Week 9/9

# STRUCTURING AND ANALYSIS OF DATA (I.E. DATA PREPARATION FOR ML)

Steps followed:

1. Feature selection

2. One hot encoding

3. 1st-order labelling (specific for StoRM log files)
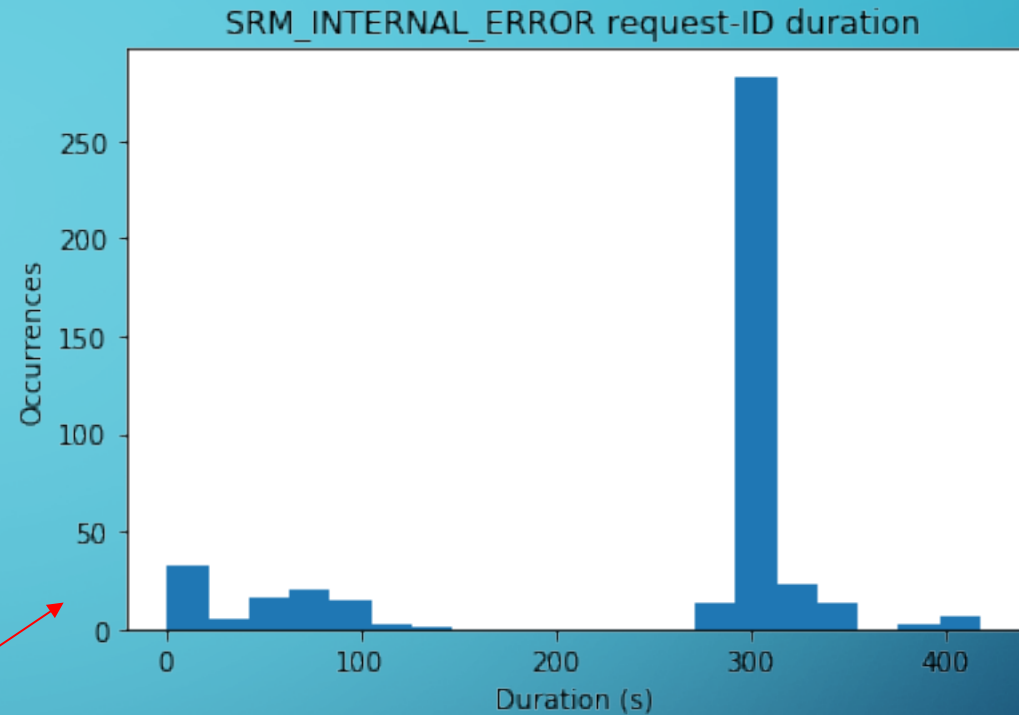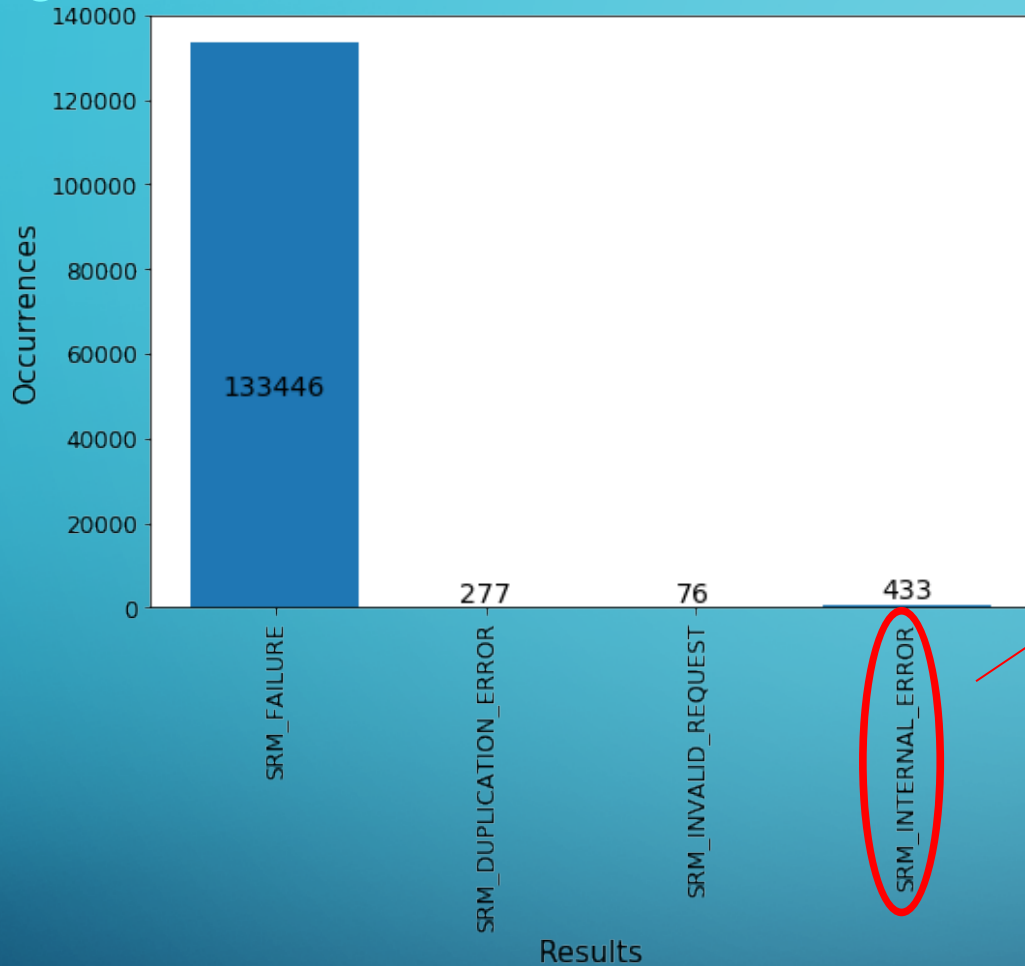
# 1. Feature selection



'Requests' occurrences

'Results' occurrences

# 2. One hot encoding

# 3. 1st-order labelling



SRM_INTERNAL_ERROR must be classified as error for ML

```
08/31 11:07:30.386 Thread 156 -  INFO [12dfa6bf-26ef-4066-b64e-d6ae587c16d8]: process_request : Connection from ::ffff:131.154.197.73
08/31 11:07:30.510 Thread 156 -  INFO [12dfa6bf-26ef-4066-b64e-d6ae587c16d8]: Request 'PTP status' from Client IP='::ffff:131.154.197.73'
Client DN='/DC=ch/DC=cern/OU=Organic Units/OU=Users/CN=atlpilo2/CN=531497/CN=Robot: ATLAS Pilot2' # Requested token '4d80b737-1796-4182-
a063-eb61e7e3a44d'
08/31 11:07:30.518 Thread 156 -  INFO [12dfa6bf-26ef-4066-b64e-d6ae587c16d8]: Result for request 'PTP status' is 'SRM_REQUEST_INPROGRESS'
08/31 11:07:30.524 Thread 156 -  INFO [12dfa6bf-26ef-4066-b64e-d6ae587c16d8]: ns1__srmAbortRequest : Request: Abort request. IP:
::ffff:131.154.197.73. Client DN: /DC=ch/DC=cern/OU=Organic Units/OU=Users/CN=atlpilo2/CN=531497/CN=Robot: ATLAS Pilot2. token: 4d80b737-
1796-4182-a063-eb61e7e3a44d
08/31 11:07:31.639 Thread 156 -  INFO [12dfa6bf-26ef-4066-b64e-d6ae587c16d8]: Result for request 'Abort request' is 'SRM_INTERNAL_ERROR'
```

WHAT'S NEXT?

A possible next step consists of using the evolution of a request-ID to teach ML how the request-IDs are made in order to determine in advance those that can degenerate in an error.

Request-ID_i →

	Variable_1	Variable_2	Variable_3	Variable_4	....	Variable_n
t_0	0	0	1	0		0
t_1	1	0	0	1		1
t_2	0	1	0	0		1
...						
t_m	1	1	0	0		0

n=39
m varies

If in the t_m instance, SRM_INTERNAL_ERROR=1 → Request-ID_i label=1

- Thanks for the experience at CNAF! A lot learned… and more to learn!

- I am now 1° year PhD student in DSC

- Interested in this (and more!) ML-related research and applications.
  - ➢ luca.giommi3@unibo.it