Predictive@ENI WP4/INFN

Activities and progress

Alessandro Costantini, Antonio Falabella, Elisabetta Ronchieri

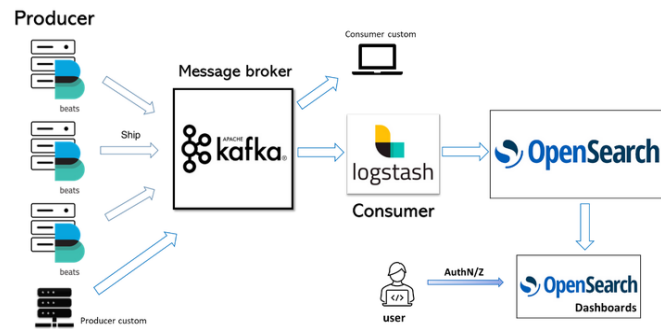INFN-CNAF

alessandro.costantini<at>cnaf.infn.it

elisabetta.ronchieri<at>cnaf.infn.it
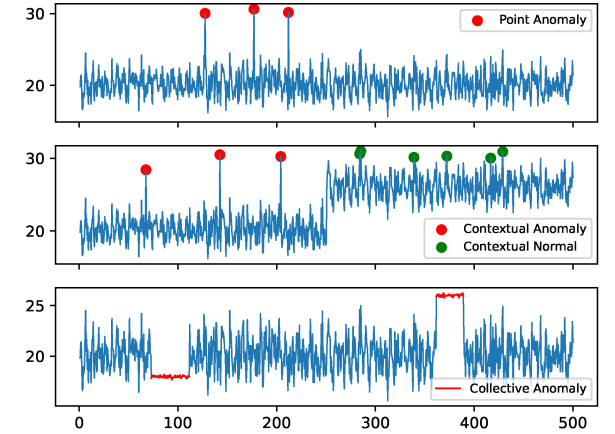
antonio.falabella<at>cnaf.infn.it
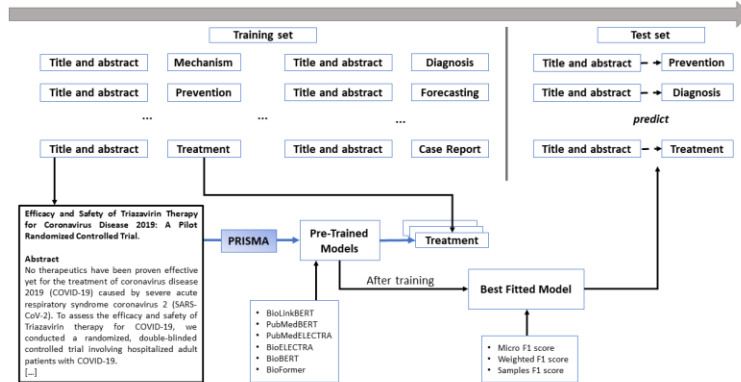
INFN CNAF AI Experiences
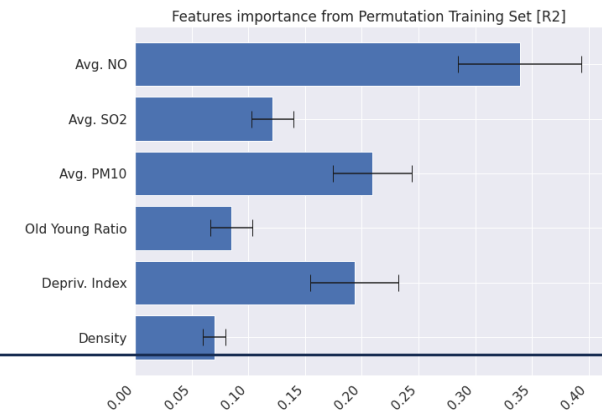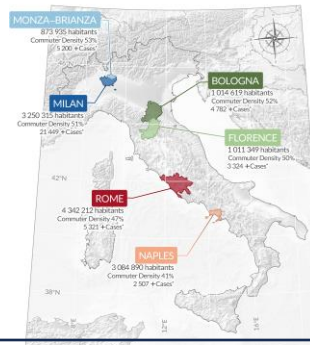
- Managing infrastructure for AI pipeline

- Anomaly detection for manufacturing applications and data center failures

- Short and Long text analysis

- Interdisciplinary use cases: COVID-19 spreading

Short and Long Text Analysis

Use case: Identify text patterns and perform text classification

Activities:

- Use topic modeling

- Use LLM for medical papers concerning COVID-19

Anomalies Detection for Data Center Failures

Use case: **Identify anomalies in service log files with Natural Language Processing solutions**

Activities:

- Turn unstructure data into structure data

- Preprocess service messages

- Build log files corpora

- Build a dictionary of anomalies

- Extract the most interesting Ngram-based features

- Label each message in anomalous or not

- Cluster service messages with topic modeling techniques and unsupervised machine learning
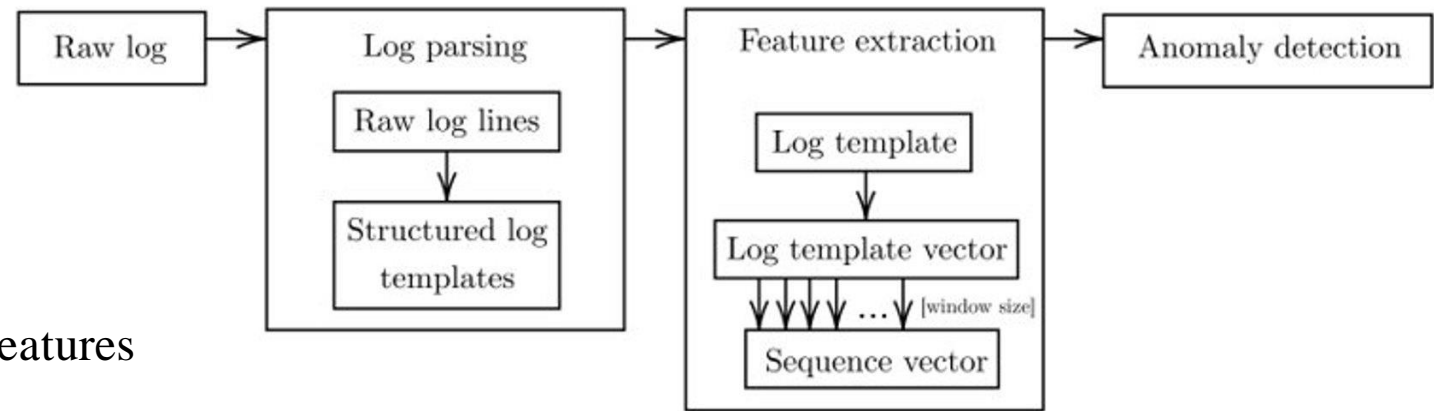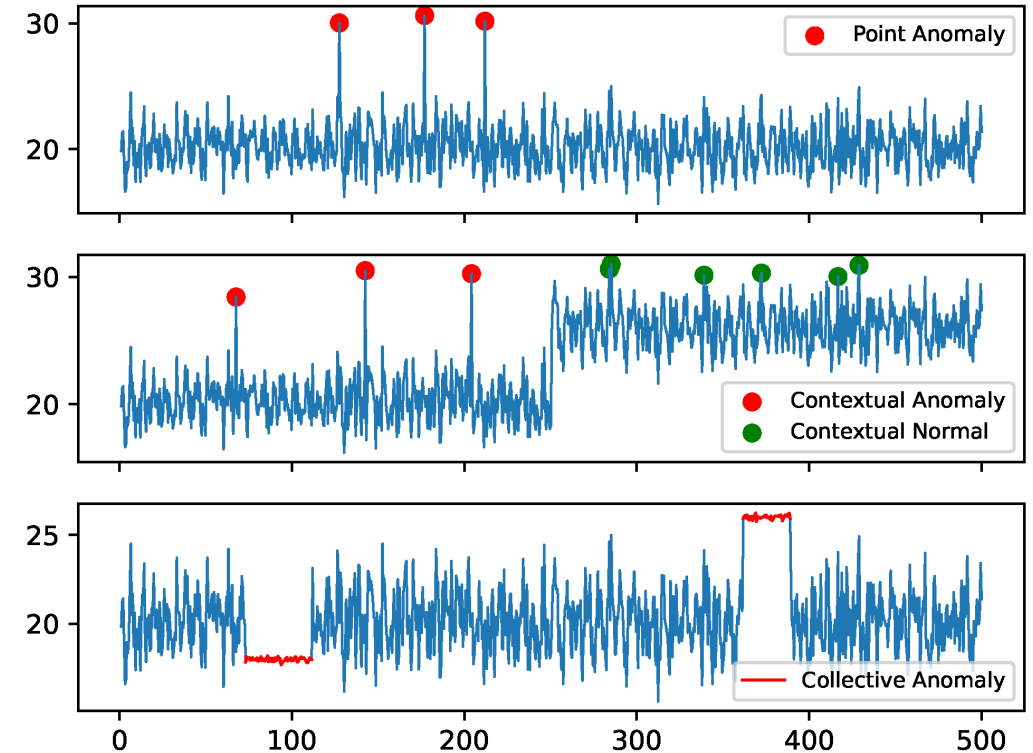
Anomalies Detection for Data Center Failures

Use case: **Identify anomalies in monitoring physical machine metric data**

Activities:

- Explore available data with time series, considering thresholds whenever possible

- Correlate variables to exclude redundant ones

- Identify anomaly slots

- Use Bayesian Optimization approach – a strategy for global optimization of expensive-to-evaluate functions - to predictive maintenance on imbalance data

- Use the JumpStarter solution - a multivariate time series anomaly detection approach – to compute anomaly score and label the various observations

Anomalies Detection for Data Center Failures

Use case: Identify anomaly pattern in heterogeneous data covering service log files and machine metrics

Activities:
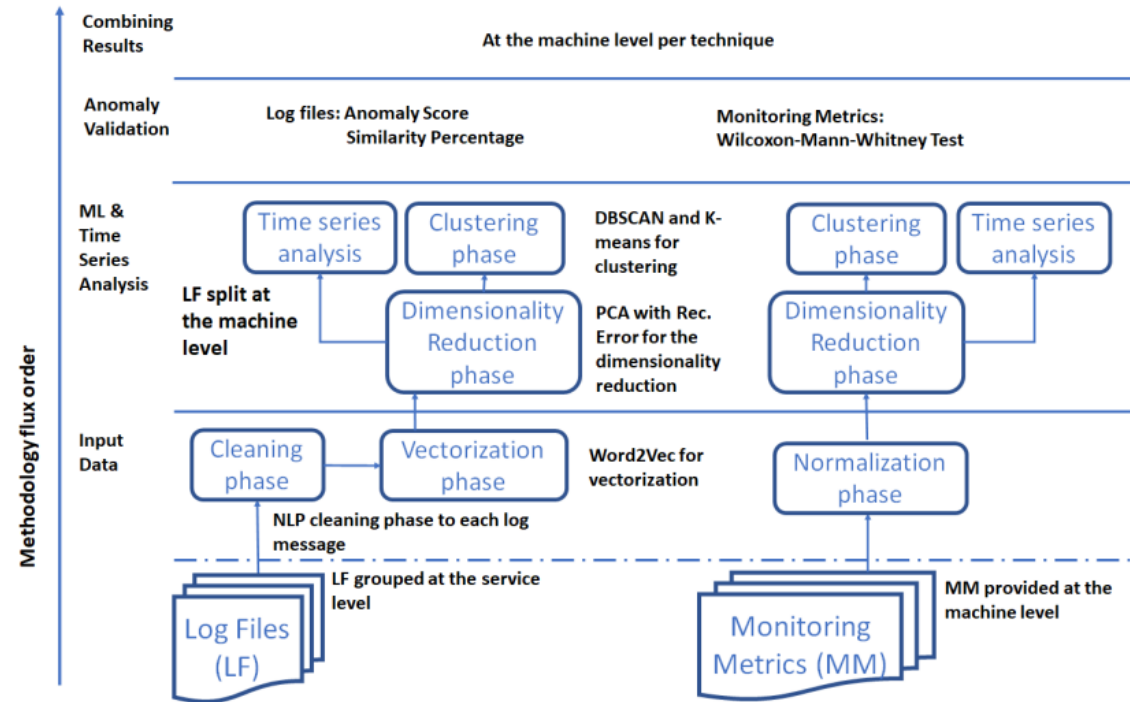
Bottom-up Anomaly Detection Approach with Log Files and Monitoring Metrics.

Anomalies Detection for Data Center Failures

Use case: **Identify anomalies in monitoring electrical plant and cooling system data**
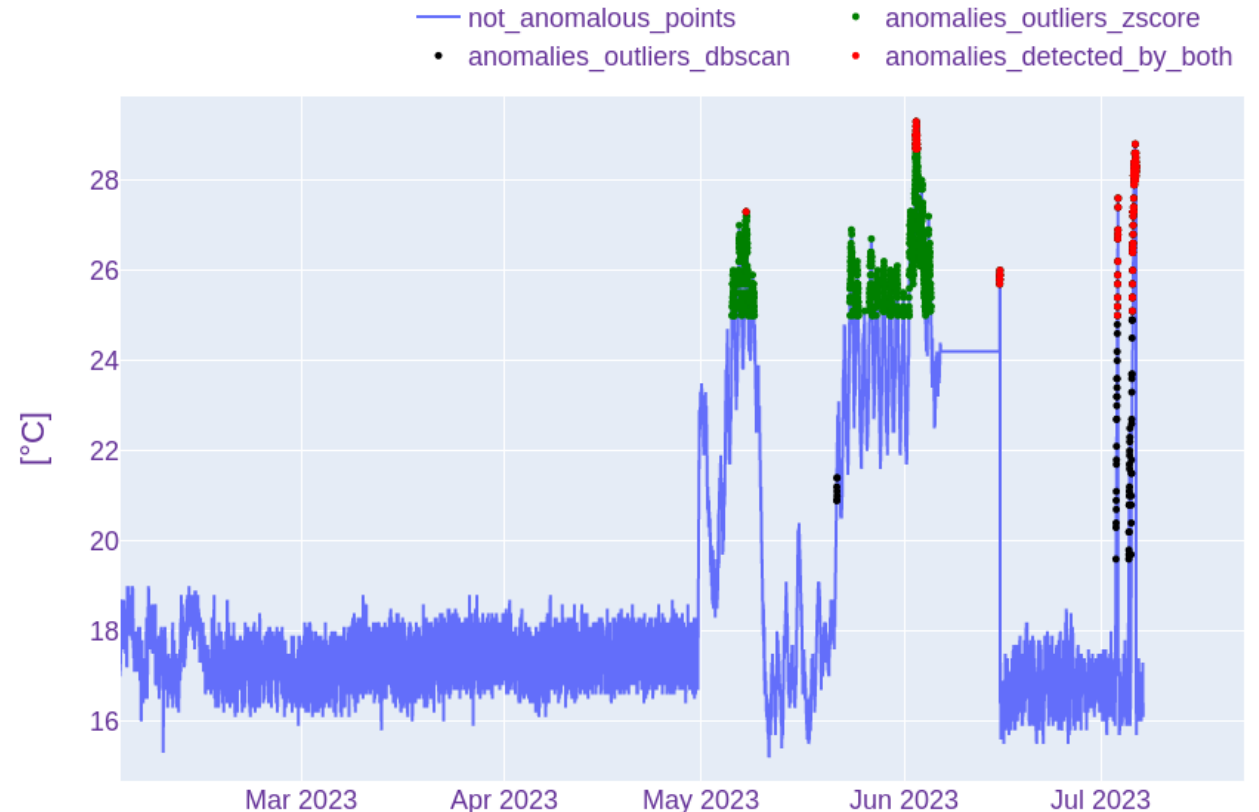
Activities:

- Preprocess data

- Reduce data dimension with PCA

- Identify anomalies with DBSCAN

- Identify anomalies with z-score

- Investigate Graph Neural Network with multivariate time series

Infrastructure Management and Data Aggregation

Use cases: Support data segregation

Activities:

- Kafka **topics** enforce **data segregation**

- **Data management**
  - Replication
  - Retention
  - Partition

- Data from producer to Kafka cluster **encrypted using SSL**

- **Specific ACL** for producers and consumers

- **OIDC AutheN/AuthZ** to get access to the data
  - Data consumed by single consumers/consumer groups
  - Multitenancy support enabled for Groups/Users
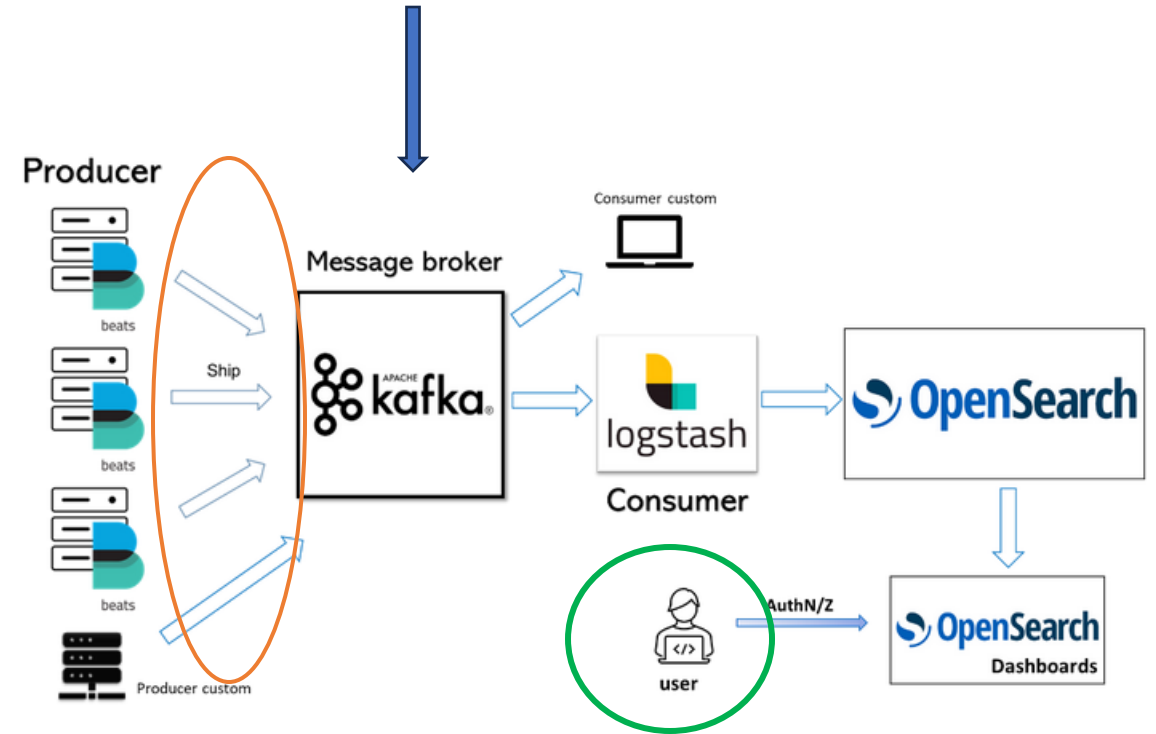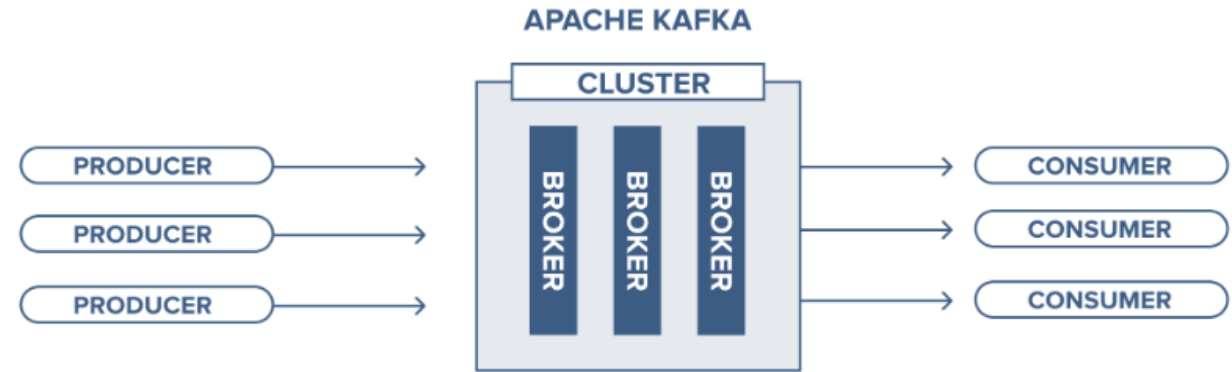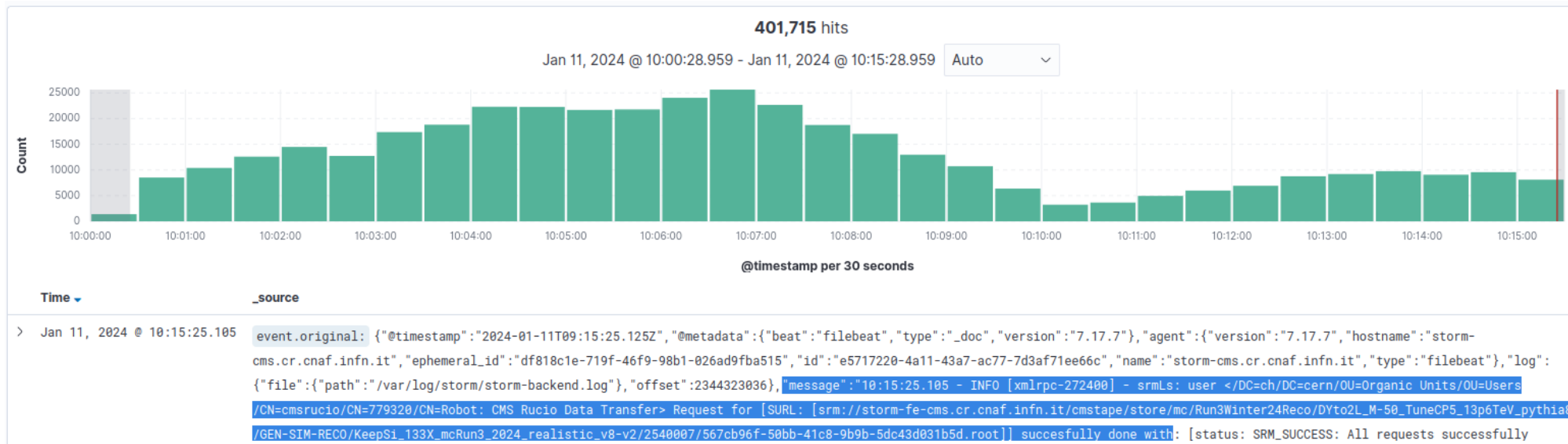
Infrastructure Management and Data Aggregation

Use case: Support for log analysis

Activities:

- *Log data aggregation*

System/Service Logs

401,715 hits

Jan 11, 2024 @ 10:00:28.959 - Jan 11, 2024 @ 10:15:28.959   [Auto ▾]

@timestamp per 30 seconds

| Time ▾	_source
Jan 11, 2024 @ 10:15:25.105	event.original: {"@timestamp":"2024-01-11T09:15:25.125Z","@metadata":{"beat":"filebeat","type":"_doc","version":"7.17.7"},"agent":{"version":"7.17.7","hostname":"storm-cms.cr.cnaf.infn.it","ephemeral_id":"df818c1e-719f-46f9-98b1-026ad9fba515","id":"e5717220-4a11-43a7-ac77-7d3af71ee66c","name":"storm-cms.cr.cnaf.infn.it","type":"filebeat"},"log":{"file":{"path":"/var/log/storm/storm-backend.log"},"offset":2344323036},"message":"10:15:25.105 - INFO [xmlrpc-272400] - srmLs: user </DC=ch/DC=cern/OU=Organic Units/OU=Users/CN=cmsrucio/CN=779320/CN=Robot: CMS Rucio Data Transfer> Request for [SURL: [srm://storm-fe-cms.cr.cnaf.infn.it/cmstape/store/mc/Run3Winter24Reco/DYto2L_M-50_TuneCP5_13p6TeV_pythia8/GEN-SIM-RECO/KeepSi_133X_mcRun3_2024_realistic_v8-v2/2540007/567cb96f-50bb-41c8-9b9b-5dc43d031b5d.root]] successfully done with: [status: SRM_SUCCESS: All requests successfully

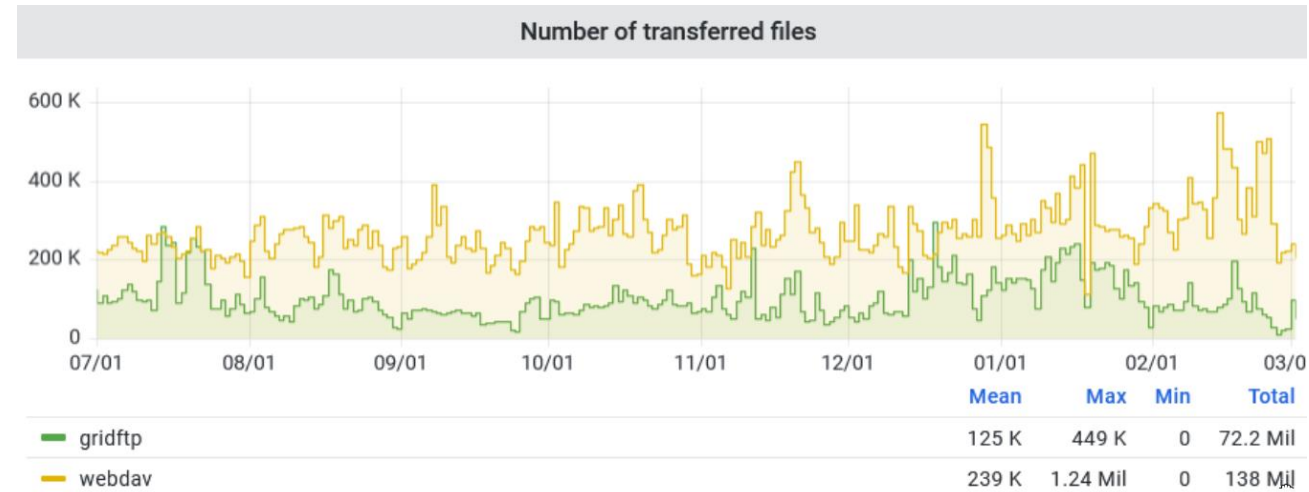# Infrastructure Management and Data Aggregation

Use case: Support log analysis

Activities:

- *Aggregation using OpenSearch Amazon Utility*
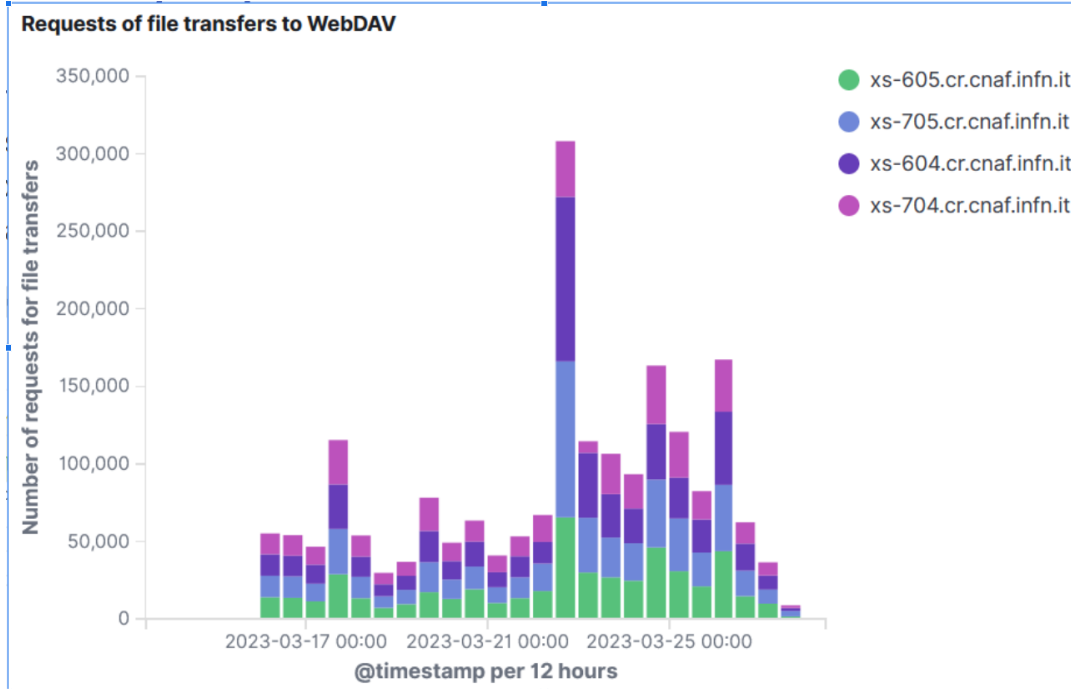
- *Aggregation using Other Utility (Grafana)*

Made Data available for analysis (AD, other AI models)