

Real-Time data analytics of a Cyber Physical System: Dewatering Machine



Technologies for Big Data Management

Students

Damiano Buzzo
Francesco Pizzuto
Eduardo Tiano

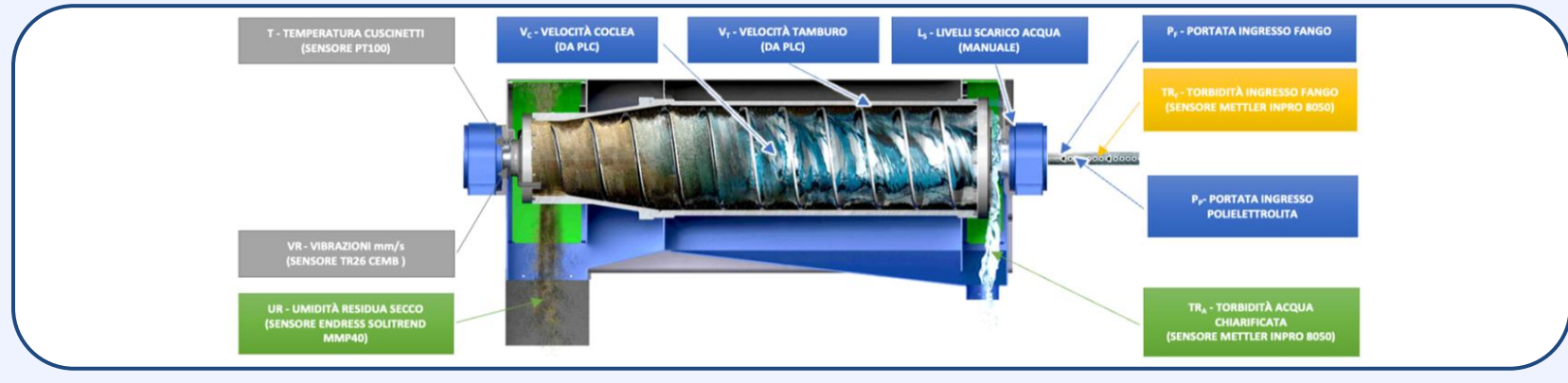
Prof. Massimo
Callisto De Donato

Scope of the project



This project explores **Kibana** and **Elasticsearch** as essential tools for **data analysis, visualisation, and real-time monitoring** within a Big Data architecture.

Our implementation focusses on a **sensor data analysis pipeline** for a specific industrial use case: **Dewatering Machines**.



Objectives



- **Implement a data analysis pipeline** for sensor data processing.
- **Perform data exploration** using a **search engine (Elasticsearch)** and a **reporting tool (Kibana)**.
- **Enable geospatial visualisation** to track sensor data on maps.
- **Develop monitoring dashboards**

Objectives



- **Develop alerting mechanisms** for real-time insights.
- **Integrate AI-driven analytics** to enhance decision-making.
- **Investigate API-based integration** in operational environments (**C#**).
- **Identify challenges and limitations** encountered in the implementation.

Technologies



- Lightweight messaging protocol for IoT and real-time applications.
- Publish-subscribe model
- Efficient and ideal for low-bandwidth networks.



- Distributed event streaming platform for high-throughput, real-time data processing.
- Publish-subscribe model.
- Used for log aggregation, messaging, and real-time analytics.

Technologies



elasticsearch

- Distributed search and analytics engine
- Indexes data for quick searching and allows full-text queries.
- Ideal for logging, monitoring, and search



kibana

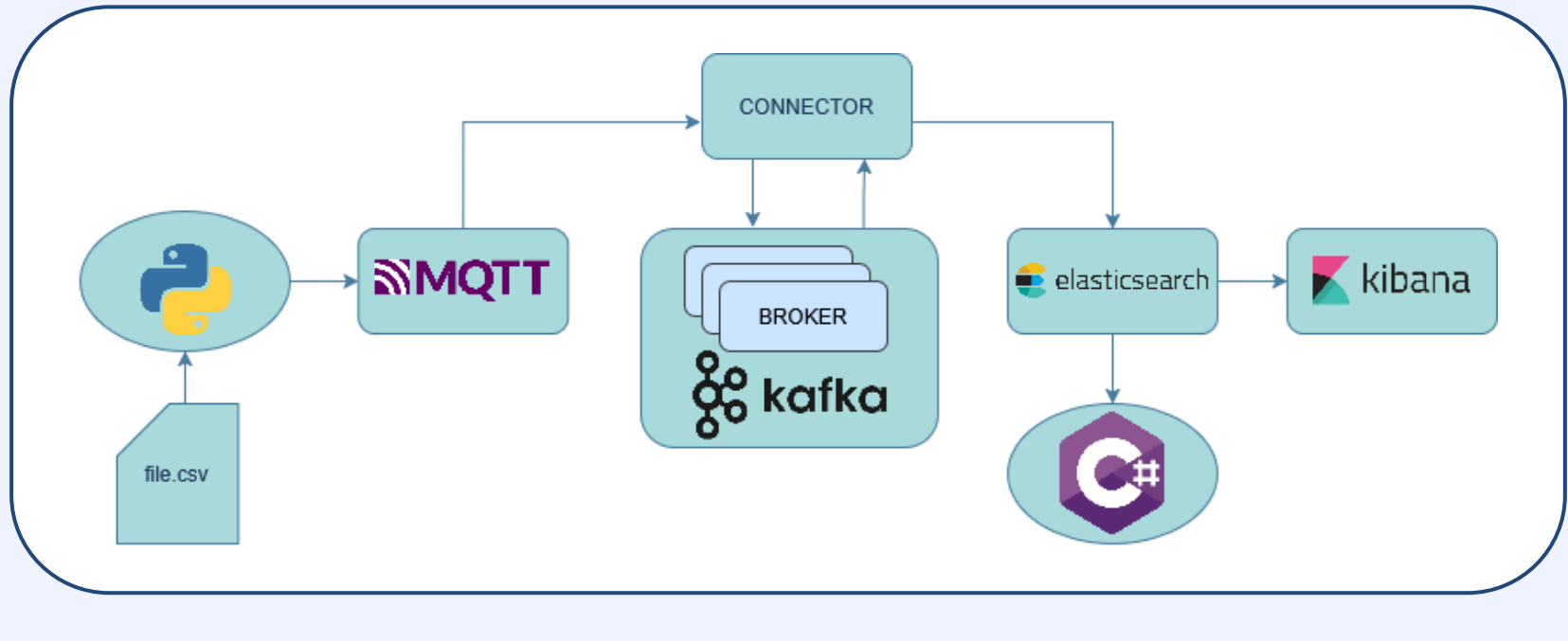
- Visualization tool for Elasticsearch data.
- Interactive dashboards and real-time analytics.
- Enables data exploration and effective log monitoring.

Technologies



- Provides containerisation.
- Simplifies deployment, scaling, and management.
- Ensures consistency across different environments.
- Permits rapid setup of distributed systems.
- Supports multiple-container management.

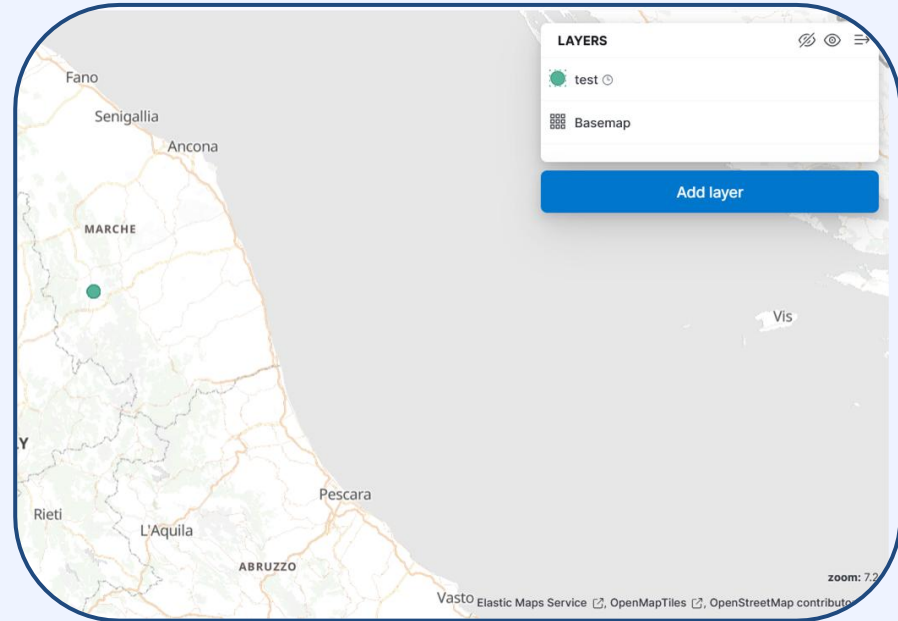
Architecture



Maps

Data in the maps can be visualized as:

- Integrated from **different sources**
- **Realistic** and **detailed** map
- **Animated** to **visualise it during time**



Dashboards



Dashboarding tools provide **easy-to-interpret** and **effective insights** on data.

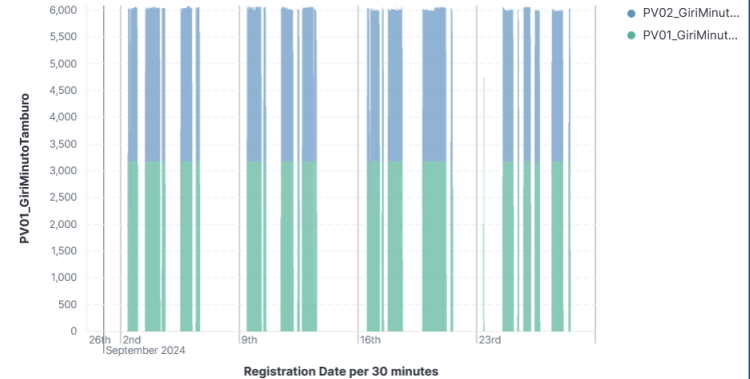
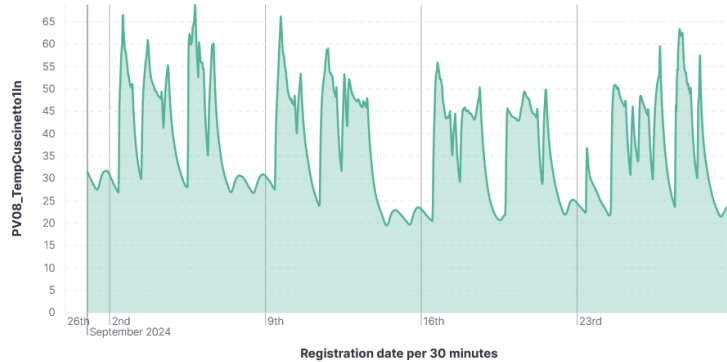
We developed two dashboards:

- **Machine Status Monitoring:** insights about some relevant data.
- **Process Efficiency:** last power consumption values.

Machine Status Monitoring



Area diagram for PV08_TempCuscinetto1In



Machine Status Monitoring

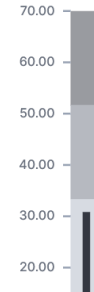


Vibrations



Gauge diagram for PV08_TempCuscinetto1Out

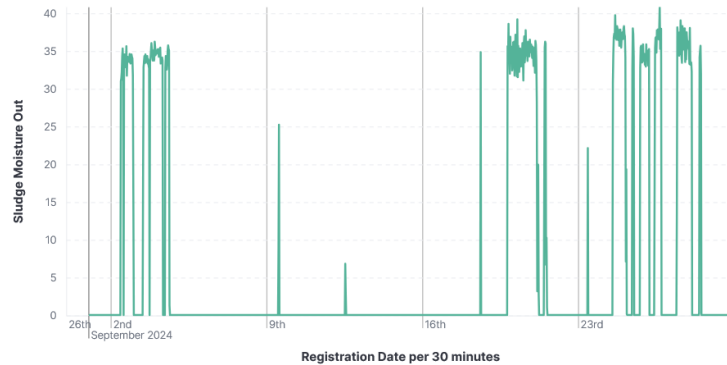
Median of temperature



Process Efficiency



Sludge Moisture Out



Comparison Table

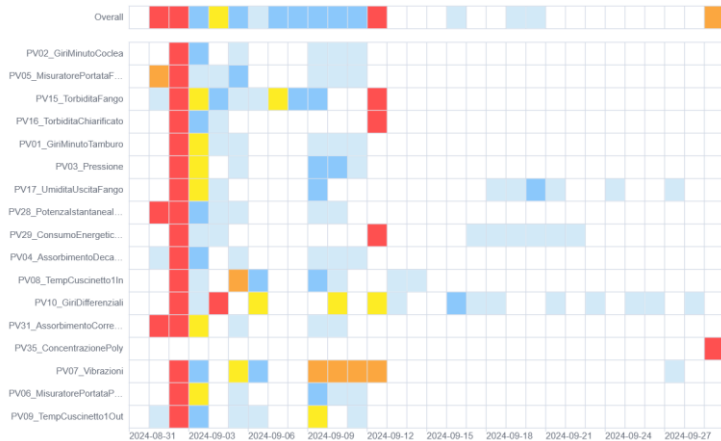
Registration Date per 3	Average Sludge Flow M	Average Poly Flow Met	Average Resulting Dryr
01:00	0.004	0	0.1
01:30	0.006	0	0.1
02:00	0.006	0	0.1
02:30	0.006	0	0.1
03:00	0.006	0	0.1
03:30	0.005	0	0.1
04:00	0.003	0	0.1
04:30	0.006	0	0.1
05:00	0.006	0	0.1

Machine learning tool

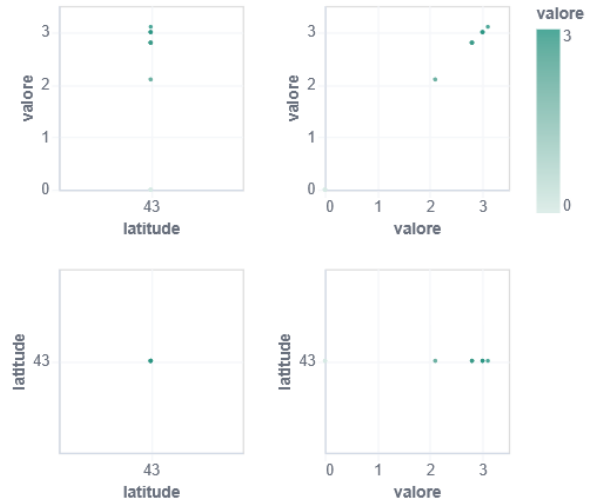


The tool can be used to solve different kinds of problems.

Anomaly Detection



Regression



API Connector



- Easy access importing

Elastic.Clients.Elasticsearch nuget

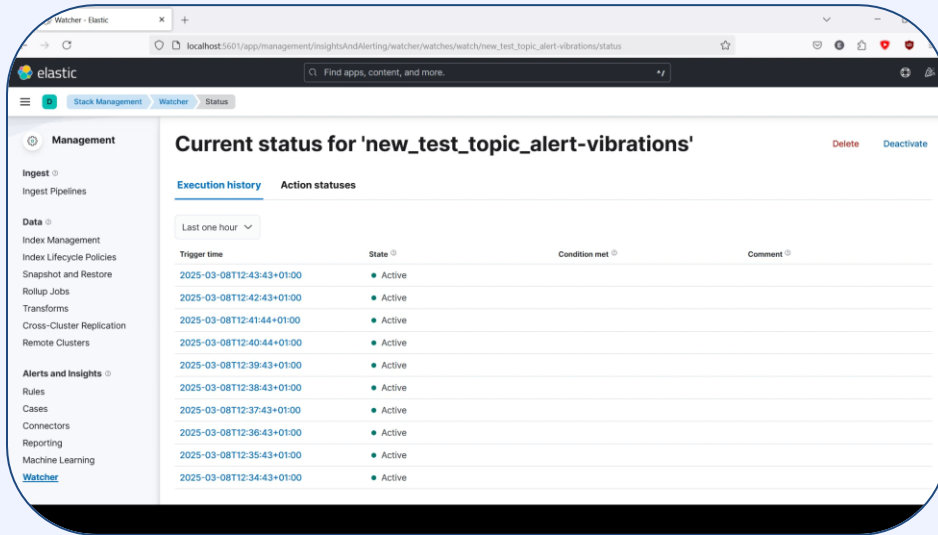
packet

- Data retrieval using Elastic client
- Query in a certain index, with certain conditions and size

```
var settings = new ConnectionSettings(new Uri(ElasticSearchUri))  
    .DefaultIndex(IndexName);  
_client = new ElasticClient(settings);
```

```
var searchResponse = await _client.SearchAsync<Data>(s => s  
    .Index(IndexName)  
    .Query(q => q.MatchAll())  
    .Size(10000));
```

Watcher



- Monitoring feature
- Periodic control over time
- Actions performed when a condition is met

Export/import of saved objects

The user can export and import saved objects (dashboards, data views, ...) to share them with others instances of Elasticsearch



Saved Objects

Refresh

Import

Export 5 objects

Manage and share your saved objects. To edit the underlying data of an object, go to its associated application.

Q













Search...

Type ▾

Tags ▾

Delete

Export ▾

<input type="checkbox"/>	Type	Title	Tags	Spaces	Last updated ▾	Actions
<input type="checkbox"/>		test			4 seconds ago	
<input type="checkbox"/>		new-test			4 seconds ago	
<input type="checkbox"/>		Process Efficiency		—	4 seconds ago	
<input type="checkbox"/>		Machine Status Monitoring		—	4 seconds ago	
<input type="checkbox"/>		Advanced Settings [8.6.0]		—	27 minutes ago	

Limitations



- Cross-reference of data. It's difficult to create data views from different sources.
- Documentation lacks clarity and examples.
- The Machine Learning feature has a 30-day trial.
- The ML tool throws exceptions even when jobs are completed.
- Kafka needs at least 3 brokers to work properly.

Limitations



- Data can't be shown with absolute values (only average/median are allowed).
- Cartesian x-axes allow no type but temporal data.
- Alert's configuration is tricky and requires complex rules.
- Email sending (action to perform) is very complex to configure.
- Roles and authentication are very complex to configure.

Possible future improvements



- Conduct a domain-specific research to enhance dashboards
- Improve alerts and notifications with other actions, such as emails
- Cloud deployment (AWS, Azure, ...)
- Real sensor integration



THANK YOU
for your attention