A Practical Study on Optimization of Big Data Streaming and Data Analytics Infrastructure for Efficient AI-Based Processing

Onur Izmitlioglu, Mujdat Soyturk - VeNIT Research Lab - Marmara University

VeNIT Lab
www.venit.org

MARMARA UNIVERSITY 1883

General Description

Description

Reliable quality control systems are the focus of Industry 4.0 to improve production qualities and decrease service failure rate. With the proposed data streaming and analytics infrastructure, it is aimed to enhance the Industry 4.0 applications with beyond state-of-the-art AI-based framework for improving reliability tests at edge and local cloud.

Challenge

- Creating an AI-based quality control system for Industry 4.0
- Enabling the estimation of remaining useful life of the industrial products while in use
- Eliminating unreliable & relatively slow process of human intervention to test setup with an automated approach

State of the art

- Many Industry 4.0 applications focus on calculating measurements using various functions and comparing the calculations with thresholds determined by R&D departments.
- Most tests include manual inspection of data by human operators.
- This results in gaining limited information regarding the root cause of failure since the tests are conducted independently.

Innovation (beyond state of art)

- Establishing a holistic view of the process by capturing multiple signals from sensors.
- Implementation of a big data streaming platform to transfer, process and store streams of sensory data to enable predictive maintenance.
- Simultaneous collection and transmission of sensory data to the AI-based fault detection models.
- Analyzing measurements using AI-based models for enhanced reliability mechanisms.

Expected results

- Correlating multiple sensory data for failure detection.
- Root cause analysis of correlated data including aging and items in use.
- Foundation for real-time health status monitoring.
- Scalable data collection infrastructure for multiple test vehicles considering the future requirements and needs of the manufacturing plant.

Measurable targets

- End-to-end latency within data streaming platform
- Sensitivity and specificity of the developed

AI-based algorithms

Prototype in Action

Hardware Specifications

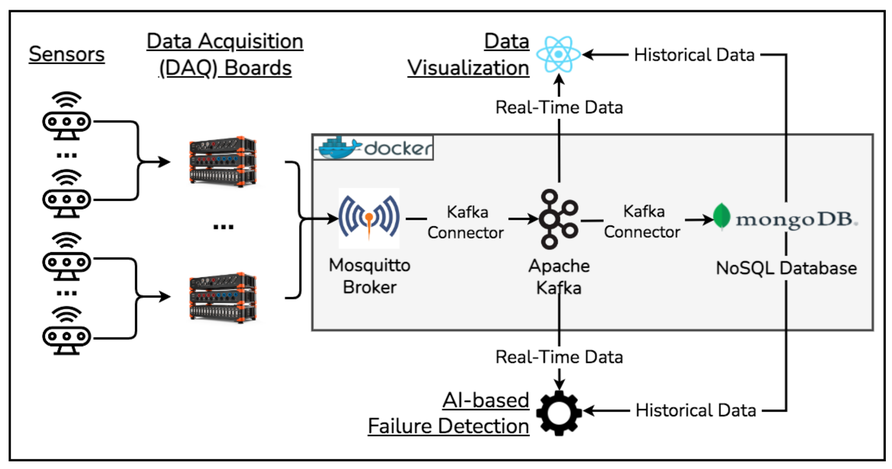
Many Industry 4.0 applications which utilizes multiple sensors as data source would benefit from the proposed infrastructure. In our experiments, from the production line, 3 channels of current, 3 channels of voltage, 1 channel vibration, 1 channel sound and 1 channel torque data are being collected. DAQ (Data Acquisition) card collects the signals with at least 200 kS/s sampling rate and at least 16-bit resolution.

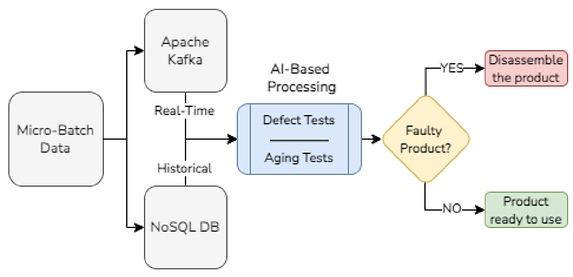
Prototype for Data Streaming Platform

The multi-sensory data streaming infrastructure is designed and implemented with a Kafka-centric approach, which enables the transfer of event streaming data collected from the MQTT broker into multiple sink components, e.g., NoSQL database and AI-based models, simultaneously.

The NoSQL-based data storage component, MongoDB, allows the historic data storage which includes a time-series collection feature that will be used as a source for AI-enriched aging tests. The system becomes resource independent, and consistent. Infrastructure built by VeNIT Lab is designed on top of Docker containerization technology to expand the data streaming platform in a fault-tolerant and scalable fashion.

By the streaming infrastructure, the foundation for prognostics and condition monitoring using state-of-the-art frontend technologies is also provided.





Testing the Prototype

Test strategy

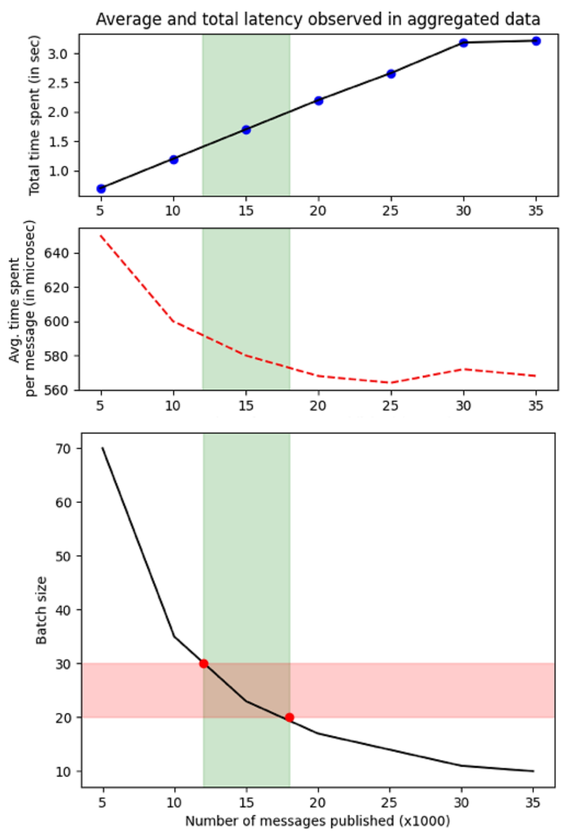
The big data streaming and data analytics infrastructure is designed and implemented by VeNIT Lab (Marmara University) to enable AI-based models detect early failures of electric motors, including multiple components that collect, transfer and store data in a plain yet influential format.

To evaluate the reliability of the proposed big data streaming platform, the quantitatively measurable target of “latency observed on transferred data within the streaming platform” has been investigated on the prototype, which is designed and implemented at Marmara University VeNIT Research lab.

Tests were conducted on a local iMac computer which resides in Marmara University VeNIT Lab with specifications:

- MacOS 10.13
- an Intel i5 processor running at 2.5 GHz,
- 8 GB of RAM

Infrastructure built by VeNIT Lab is designed on top of Docker containerization technology to expand the data streaming platform effectively, considering the outlook on future requirements and needs of the test setup. Docker allows to run and adjust multiple isolated spaces which share the same OS. Hence, the data streaming infrastructure can be deployed to various types of solutions, including edge, local cloud (fog) and cloud.



Results

The latency-throughput tradeoff during the transfer of data from data sources to AI-based failure detection models were examined and **micro-batching** technique is experimented on aggregated data points between sizes 5 to 50 per batch, to obtain the optimal batch size for data transfer.

The total latency observed during the transfer of 7-seconds synthetic sensory data, optimal batch size is devised for the particular use case. Configurations of component-specific parameters would be the final step to realize the benefits of the data streaming platform.

*This study is supported by **InSecTT** and **iRel40** projects.*

***InSecTT** has received funding from the ECSEL Joint Undertaking (JU) under grant agreement No 876038. The JU receives support from the European Union’s Horizon 2020 research and innovation programme and Austria, Sweden, Spain, Italy, France, Portugal, Ireland, Finland, Slovenia, Poland, Netherlands, Turkey.*

***iRel40** is a European co-funded innovation project that has been granted by the ECSEL Joint Undertaking (JU) under grant agreement No 876659. The funding of the project comes from the Horizon 2020 research programme and participating countries. National funding is provided by Germany, including the Free States of Saxony and Thuringia, Austria, Belgium, Finland, France, Italy, the Netherlands, Slovakia, Spain, Sweden, and Turkey.”*