

Analytical Framework to Build Predictive and Optimization Functions From Manufacturing Industry Sensor Data Using Cross Sectional Sharing

Bhaskaran R (✉ rbhaskaranpsna@gmail.com)

PSNA College of Engineering and Technology

K Kalaiselvi

SRM Institute of Science and Technology

D. Murali

CMR College of Engineering and Technology

R. Venkateswara Reddy

CMR College of Engineering and Technology

G.Naga Rama Devi

CMR Institute of Technology

Sivakumar Ponnusamy

SRM Institute of Science and Technology

Research Article

Keywords: Analytical platform, Big data storage, Cross sectorial data lab platform, Operational platform, Semantic modelling framework.

Posted Date: April 13th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-410664/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Nowadays in industry sensor data are used. This needs to be shared in many areas for making the prediction. Also, it needs to be optimized for making the things to do automatically. This paper proposes a novel analytical framework to build predictive and optimization functions from manufacturing industry sensor data using cross sectional sharing which combines all different types of operation in a cross-sectional lab, which is a cooperative site in which huge quantities of data from numerous sites are composed as well as managed in a terrific way. The predictions and the optimization are made possible and store the same using the big data storage. Big Data Storage as well as Analytics Platform; Development Tools; Modelling Tools for Imitation Concepts as well as Power Framework are carried over in cross sectional lab. This is making the relations ship entities using Relational Data Base Management Systems (RDBMS). Various apache versions are used for the implementation of this which acts in a cloud platform. In the case study, the mean and variance were calculated and plotted.

I. Introduction

The Internet of Things transforms industries as well as businesses. The chief areas of Internet of Things funds comprise manufacturing activities, transportation, intelligent grid technology, smart buildings as well as Internet of Things consumer goods, smart home automation and sales. Based on the usage cases identified, we have separated the data structure for the analysis of the two main items as shown in Figure 1.

To confirm and prove the fallouts, two real-world sites were used inside the project: A French aluminium plant as well as a plastic factory in Portugal. Two cases are identified of use of equally domains. In the aluminium sector, the focus is on the production of anodes used to extract aluminium by electrolysis. Speculative care is the key objective was to forestall deterioration, to highpoint mechanical, process deviations distressing the final excellence of the green anode. Next is a quality control anode, with the aim of identifying malignant anodes with a great degree of self-assurance is removed and transmit this to electrolysis site.

In case of plastic, from coffee-making industry, this is mass-produced with slight difference as well as little quality specification. It is significant to yield the right width as well as length of the coffee pills. Also, to ensure that the holes under the pills are properly formed. The automotive industry performs the recycling process with the production part. Real-Time Operating Platform used during working hours, can be used by store staff. The section connects to existing complex systems used in the production facility. Relevant information from production site is promoted to the Cross-Sectorial Data Lab. A cooperative site consist of great quantities of data from numerous locations are composed as well as treated in a terrific way. It is intended for use by data scientists or a world process manager. Contains great data storage as well as processing. It comprises tools for improving the performance of forecasting tasks, simulations for evaluating tasks in testing as well as distribution in production environment, as well as a semantic framework that offers common language among data scientists as well as field specialists. Failure

Prediction and Detection are carried out in three ways which are explained below. One type of time-based model is called a regression model. This model collects a certain amount of historical data and uses that data to look beyond any machine failures, creating a model based on the failure time in relation to the current repair process.

It will then make a date or period of failure. Perhaps the most common model, other models of failure forecasts look at so-called “undesirable” or inconsistent behaviours and use those behaviours to predict failure. For example, suppose the centre has a robotic arm that normally moves to Point A, picks up half, and then moves part to Point B. However, when the machine is close to failure, the machine arm begins to take longer to Point B, freezing in place with its arm.

We can point to these misconducts and understand it as a sign of failure, using it to determine how close an asset is to failure. Types of predictability of survival fail the question: "How does the risk of asset failure change over time if we look at the frequency of X symptoms?" "If an area tracks many different parameters in relation to an asset (such as temperature, vibration, and noise), the survival failure prediction model will use all these factors to estimate how the failure rate changes. So while this model may not tell you exactly when the asset will fail, it will tell you that the risk of failure increases or decreases depending on those factors. From there, you can monitor those restrictions and check for equipment failures.

Guessing corrections use contextual-based corrections. Manipulating the enduring service life will permit conservation to be scheduled for normal operating hours. And forecasting care includes the detection of errors. Only condition-based and speculative remedies provide the potential to reduce oil spills and increase the life span of mechanical mechanisms. It is promising to control both how long you are ready to use and how long you can stay in this condition. This compares with other “traditional” methods in which grease can be discarded because the system failed, or because the grease was used for a certain period of service.

ii. Literature Review

In [1] defines the construction of a Big Data platform at the crossroads of the process sector domain. The goal was to create an easy-to-analyze platform that would support data collection, storage as well as processing in a wide range of industries. This proposed platform interacted with the actual plant circumstances includes the utilisation of collected data to create predictive activities to increase production processes. Analytical stage would cover the enhancement space for the construction of these works, as well as the simulation environment to test models [2]. The platform is shared between wide ranges of locations from different sectors of the industry. Sharing of categories will allow for the transmission of information transversely diverse domains. At enhancement, user-focused method was chosen to collect the needs from various shareholders to configuration development of building models from a variety of perspectives, from content to postings. The craftsmanship used was examined in two

areas of process. The examination process is performed in both aluminium production and the plastic moulding industry.

In [3], the manufacturing sector is below continuous pressure to add profits to a growing global modest market where diversity is untied to the products produced or the technology used but to the operation of business processes. Business analytics provide the chance to use hidden information as well as value inside business information systems to transform innovation, improve sales and production management, precise sales goal and sales work, and build along with management of profit. In the current task, the approach to gain the profit using business statistics was explained by the suggestion that focus should focus on three more challenging barriers. First, “measuring” the collection, compilation and packing of data is talented. After that, the transcendent “emergence of organizational culture” and data-driven decision-making is desired to make a complete business analytics environment to yield results as well as references that take place. Next, this should guide efforts to “create a business model” to address new value building, as well as to capture and protect market leadership.

In [4] Big Data (BD), which can validate valuable data on improved decision-making processes, has recently attracted a lot of attention from both academics as well as practitioners. Though, there are diverse types of analytical apps. Consequently, before quickly using and purchasing expensive BD tools, it is necessary for administrations to first understand the location of the BDA [5]. Because of the importance of BD as well as BDA, it provides a modern review that brings a comprehensive overview of the BD drawbacks and BDA schemes used to provide the better understanding of others is accompanied by a sound investment activity.

Recent developments in the manufacturing industry made the systematic positioning of Cyber-Physical Systems (CPS) [6], where data from associated methods is carefully observed as well as coordinated among the virtual factory industry as well as the cyber counting space. In addition, by using progressive knowledge analytics, interacted devices are liable to do collaboratively and robustly. Here it is proposed to build a 5-level coalition as a guide to implementing CPS.

In [7], the developing healthcare industry creates huge amount of valuable information about patient statistics, usage ideas, payments, and insurance reporting - which attracts devotion of doctors as well as scientists similar. Number of peer-examined articles had discussed the diverse measurements of a data mining application [8].

iii. Proposed System

The proposed construction was intended in the form of MONSOON, a SPIRE abbreviated as Sustainable Process Industry through Resource and Energy Efficiency research project purposes to progress infrastructure to provision process industries. The aim is to create a data-driven approach which supports the efficiency skills through model-based guessing panels in production processes.

The parts are associated by quantified connectors. The elements as well as associated areas are explained in the Functional View section.

Information View is validated based on process models as well as mantic-mantic-data models prepared in the Semantic Modelling Framework. The purpose is to deliver a common language of communication among professional in a particular domain and shareholders and data scientists [10]. The data scientists require in-depth knowledge about business statics as well as modelling products obtained from stakeholders as well as professional domain. The shareholders and domain specialist should have the better understanding of data analysis process. The Semantic Modelling Framework incorporates business management system includes the data mining methods along with the collaboration standards namely the Standardized Cross-Industry Process in Data Mining [9, 11], Language Guessing Model [13], and portable mathematical format (PFA) [15].

The concepts of Production Processes and Processes signify the decay of the entire production procedure into production steps [12]. The Product Component signifies the logical collection of Related Performance Indicators, Data Objects as well as Resources essential performing the production step. The Production Process model is quantified using a workflow diagram.

Operational resources and concepts of portable assets define sites, locations, production units, task cells, production lines, processing cells related to a particular phase of the process. Each can be described as a unique mechanical feature with unique features and capabilities [14]. In addition, the collections of apparatus with comparable features as well as objectives are defined by resource classes. It is considered to imagine, evaluate as well as achieve performance or the impression of precise activities. They are related to develop stages in the entire Production Process. KPIs are categorized into KPIs based on the types in ISO [16].

Without this distinction, the causal relationship among KPIs can be demonstrated by KPI Tasks. KPI activity inputs can also be Test metrics, which measure the performance of predictive tasks. Using the KPI Task Format, it is possible to incorporate the speculative posting function into the production process, where performance can be assessed in a variety of ways.

Data objects are simulated at two stages of output: logical as well as physical. The data object can have an input role or an output role. Output data items can be designated as an adjective of the guesswork operation. A single object data able to play the characters liable based on the data analysis. Data items type define items in terms of their value and defines continuous, ordinal, imaginary, local or sequence of data items.

Physical Data Elements Links Logical Data Elements with real-time data representation in archives or prearranged files. Numerous Physical Data Elements can be organized into a single record as well as defined by the Physical Schema.

Like Data attributes, Guessing Functions are modelled practically as well as mental levels. The Concept Tasks Function agrees which data items are included in the guesswork function besides which is the resulting data elements. In addition, Physical Predictive Functions specifies particulars around training data as well as algorithm used to construct exact guesswork as well as test statistics, describes the presentation of the prediction task as tested on the verification of the performance data.

The Platform Operation Platform framework is shown in Figure 2.

The key function is data integration, intervention, and truncing. It offers an advanced layer of integration among Big Data based processing as well as analytics platform, as well as plant environment.

VPIRA has internal construction, containing connectivity resources for numerous data sources as well as purposes. The connectors permit the mixing of data from various systems embedded in the plant environment. Integrated data is transmitted within a data flow engine.

Data mobility is again organized in a dynamic way, joining numerous sources to many targets, overwhelming any data differences. Data transfer from source to connector is dynamic liable on the kind of data or the content itself.

Run-time Container permits ingress of forecasting tasks from Data Lab platform. Predictive tasks are sent from Data Lab in an independent platform format (see Functions Repository description), translated by an internal scoring engine. The forecasting functions format too permits interaction with several data analysis tools. In addition to scoring goals, the engine similarly achieves all the necessary functions in the preliminary processing of raw data for the performance of a particular prediction function and performs the prediction process. Together batch as well as online goals is reinforced.

The structure of Operation Data Visualization delivers interface for a web customer which includes the management of numerous contemporaneous visualizations of functional data as well as predictive monitoring operations are used. Visual attribute is compiled through the adaptation of Virtual Process Industries Resource that may comprise performance of elemental data from the plant area activities performed in Run-time Container. Virtual Industries Resource Adapter distributes data displayed in accordance with the REST web interface providing the Operational Data Visualization Framework. Visual data is improved with innovative practice indicators using a variety of practice analysis methodologies. Data Visibility structure delivers an API to use a new practical indicator.

The structure of the Cross-Sectorial Data Lab platform is displayed in Figure 3.

The Big Data Storage as well as Analytics platform delivers real-time goods as well as performance and bulk performance and real-time Big Data. It delivers key integrations between the platform and the Data Platform as well as integration programs to implement data mining processes. The internal structure of Big Data Storage as well as Analytics Platform is given in Figure 4.

Development Tools deliver a great cooperative as well as collaborative demonstration to generate as well as collaborate with data flow operations operating on the Data Lab platform. The interface is combined with a type of analytical notebooks in which various portions of the examination are rationally organized as well as accessible in a single text. These manuals contain editors of data processing codes as well as SQL queries, as well as communicating table or graphical performances of used data.

Semantic Modelling Tools enable customer interaction in design as well as distribution of semantic models quantified in the structure of Semantic Modelling. Moreover, Semantic Modelling Tools provides a web service template to retain the query semantic models in a machine-readable form using semantic interaction principles. The web service interface provides a practical information presented in the semantic models have an improved efficiency in production processes of Simulation and Resource Optimization structure.

The key purpose of the Simulation as well as Resource Utilization Framework is to provision the authentication as well as distribution of speculative work to maximize the overall KPIs definite by the production process. Verification is formed on Accurate determination of predictive activity measurement and variable calculation.

The Simulation as well as structure of Resource Optimization incorporates semantic models providing Semantic Modelling Tools that include evidence regarding the production process, KPIs and relationships among KPIs as well as the function of the prediction function. Authentication composition are performed utilising Data Analytics API as well as verification tasks are performed on the Distributed Data Processing Framework. Predictability tasks are fixed in the workspace.

Iv. Implementation

Nowadays, a large amount of several Big Data processing technologies are developing, frequently through scattered operations as well as inactive structures. For software developers, the selection of start-up technologies is an interesting task that necessitates a lot of deliberation of various start up details as well as compliance issues. Post view provides a consistent map of existing and emerging technologies, as well as practical features described in Function View. Implementation of the projected construction maps is potted.

Figure 5 introduces deployment view along with node types. The platform type introduced, first verified as well as assessed in both domains. The chief objective was to organize on-site crop processing platforms to gather as well as transmit data to the Cross-sectorial Data Lab. The key purpose of the experiment is to examine organization to hand over of all data.

The first type platform was introduced. The implementation used Apache Nifi (unparalleled data transfer framework, transformation) as the leading data integration technology on the Platform Operation Platform. Nifi is used to collect unique data from different process locations and reserve data used on the Data Lab platform. Performance details are displayed using Grafana, visual analytics and monitoring

framework. The main components of the Cross-Sectorial Data Lab platform trusted on the standard Apache Hadoop (hadoop.apache.org) infrastructure, that comprises a dispersed file system (Hadoop Distributed File System, HDFS) and well as resource manager. Apache Spark also delivers provision for SQL data analytics. Support for compatible GPU / CPU integration depend on the DL4J component, an open repository, distributed in-depth library, the Camera, and advanced neural network API. The platforms are linked using the Apache Kafka messaging system for a contemporaneous communication and the visual interface of the HDFS web service improves batch data. Access to the Cross-Sectorial Data Lab platform is protected with the support of Apache Knox security gateway. Cluster is governed using Apache Ambari management software, and provide the developmental support regarding fo recasting functions, Apache Zeppelin is accepted based on tool component Enhancement which is depicted in figure 6.

To facilitate site placement, all mechanisms of the Platform Operation Stage are connected on a single Site server linked to the site infrastructure. This flask can be distributed to the collection for distribution as well as trustworthiness. The Cross-Sectorial Data Lab platform is used as a cloud collection. Sango way Security was the only Internet-connected server, with individual Data Lab components, that are connected to a private network. Master has utilized key handling, observation, organization, and data Lab collection services such as Apache Ambari server, HDFS Name Node as well as YARN resource manager. The data is reserved as well as processed on the Worker nodes, performing enforcement services like HDFS Data Node as well as Spark employees.

In figure 7 the mean and variance was calculated for each mapping. Based on which the graph was plotted in figure 8 and 9.

Based on the events the device guide, value and timestamp were plotted in figure 10 by keeping 0.2 as the threshold value.

V. Conclusion

The implementation has used Apache Niffor data integration technology on the Platform Operation Platform. Performance details are displayed using Grafana which analyses visually and monitors with a framework. Component mapping is done using standard Apache Hadoop infrastructure. Apache Spark distributed the data. SQL data analytics is carried out by Apache Spark. The platforms are combined using the Apache Kafka messaging system. Cross-Sectorial Data Lab platform is protected with the Apache Knox security gateway. Cluster is achieved using Apache Ambari as well as Apache Zeppelin matures the tool component. The Cross-Sectorial Data Lab platform is used as a cloud collection includes three different node types. Sango way protection is the only Internet-linked server, with individual Data Lab components, that are connected to a privacy network. Master has utilized key management, monitoring, configuration as well as data Lab collection services like Apache Ambari server, HDFS Name Node as well as YARN resource manager. The data is kept as well as treated on the Worker nodes, performing execution services like HDFS Data Node as well as Spark employees.

References

1. Moghaddam, M., & Nof, S. Y. (2016). Real-time optimization and control mechanisms for collaborative demand and capacity sharing. *International Journal of Production Economics*, 171, 495–506.
2. Yamila, M., Omar, M., Minoufekar, P., & Plapper, Business analytics in manufacturing: Current trends, challenges and pathway to market leadership, *Operations Research Perspectives*, Volume 6, 2019, 100127, ISSN 2214–7160, <https://doi.org/10.1016/j.orp.2019.100127>.
3. Shearer, C. (2000). The CRISP-DM model: The new blueprint for data mining. *Journal of Data Warehousing*, 5(4), 13–22.
4. Uthayasankar Sivarajah, M. M., Kamal, Z., Irani, V., & Weerakkody, Critical analysis of Big Data challenges and analytical methods, *Journal of Business Research*, 70, 2017, Pages 263–286, ISSN 0148–2963, <https://doi.org/10.1016/j.jbusres.2016.08.001>.
5. Lee, J., Bagheri, B., & Kao, H.-A. (2014). A Cyber-Physical Systems architecture for Industry 4.0-based manufacturing systems. *SME Manufacturing Letters*. 3. 10.1016/j.mfglet.2014.12.001.
6. Islam, MdSaiful, et al. "A Systematic Review on Healthcare Analytics: Application and Theoretical Perspective of Data Mining." *Healthcare (Basel, Switzerland)* vol. 6,2 54. 23 May. 2018, doi:10.3390/healthcare6020054.
7. Shen, C.-P., Jigjidsuren, C., Dorjgochoo, S., Chen, C.-H., Chen, W.-H., Hsu, C.-K., Muller, R., Robson, B., Apte, C., Weiss, S., et al. (2012). A data-mining framework for transnational healthcare system. *Journal of Medical Systems*, 36, 2565–2575.
8. Masolo, C., Borgo, S., Gangemi, A., Guarino, N., Oltramari, A., Schneider, L. DOLCE: A descriptive ontology for linguistic and cognitive engineering. *WonderWebProj. Deliv. D17 V2*, Tech. Rep. 2002, 1, 2–3.
9. Lemaignan, S., Siadat, A., Dantan, J. Y., Semenenko, A. MASON: A proposal for an ontology of manufacturing domain. In *Proceedings of the DIS 2006: IEEE Workshop on Distributed Intelligent Systems-Collective Intelligence and Its Applications*, Prague, Czech Republic, 15–16 June 2006; Volume 2006, pp. 195–200.
10. Keet, C. M., Ławrynowicz, A., D'Amato, C., Kalousis, A., Nguyen, P., Palma, R., Stevens, R., & Hilario, M. (2015). The Data Mining OPTimization Ontology. *J. Web Semant*, 32, 43–53.
11. Panov, P., Dzeroski, S., Soldatova, L. N., *OntoDM: An ontology of data mining*. In: *ICDM Workshops*. IEEE Computer Society; 2008, p.752–760.
12. Panov, P., Soldatova, L., Džeroski, S. *OntoDM-KDD: Ontology for representing the knowledge discovery process*. In *Proceedings of the Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Paphos, Cyprus, 28–30 November 2012; 2013; Volume 8140, pp. 126–140.
13. Bernstein, A., Provost, F., & Hill, S. (2005). Toward intelligent assistance for a data mining process: An ontology-based approach for cost-sensitive classification. *IEEE Trans. on Knowl. and Data Eng.*,

14. Vanschoren, J., Soldatova, L. N., Exposé: An ontology for data mining experiments. In Proceedings of the SoKD 2010—Third Generation Data Mining Workshop at ECML PKDD, Barcelona, Spain, 20–24 September 2010; 2010; pp. 31–46.
15. Patterson, E., Baldini, I., Mojsilović, A., Varshney, K. R. Semantic Representation of Data Science Programs. In Proceedings of the Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, Stockholm, Sweden, 13–18 July 2018; International Joint Conferences on Artificial Intelligence Organization: Stockholm, Sweden, 2018; pp. 5847–5849.

Figures

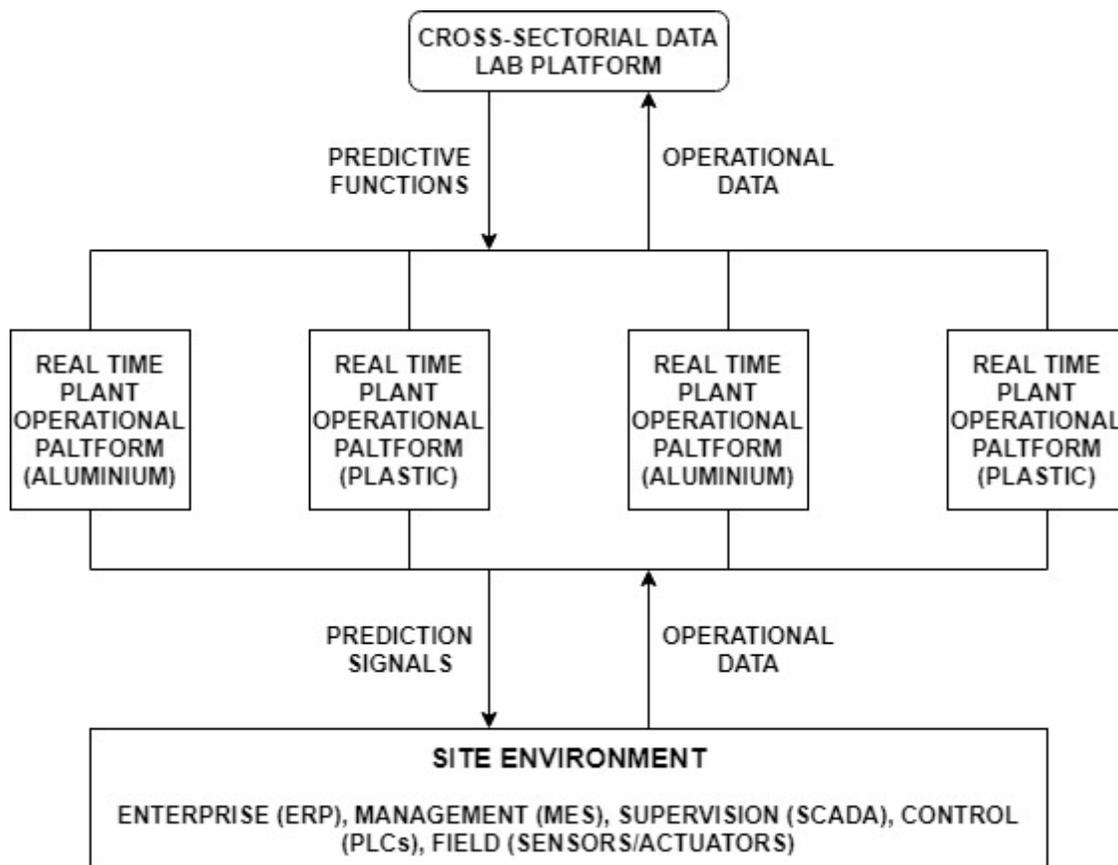


Figure 1

An overview of high-end buildings.

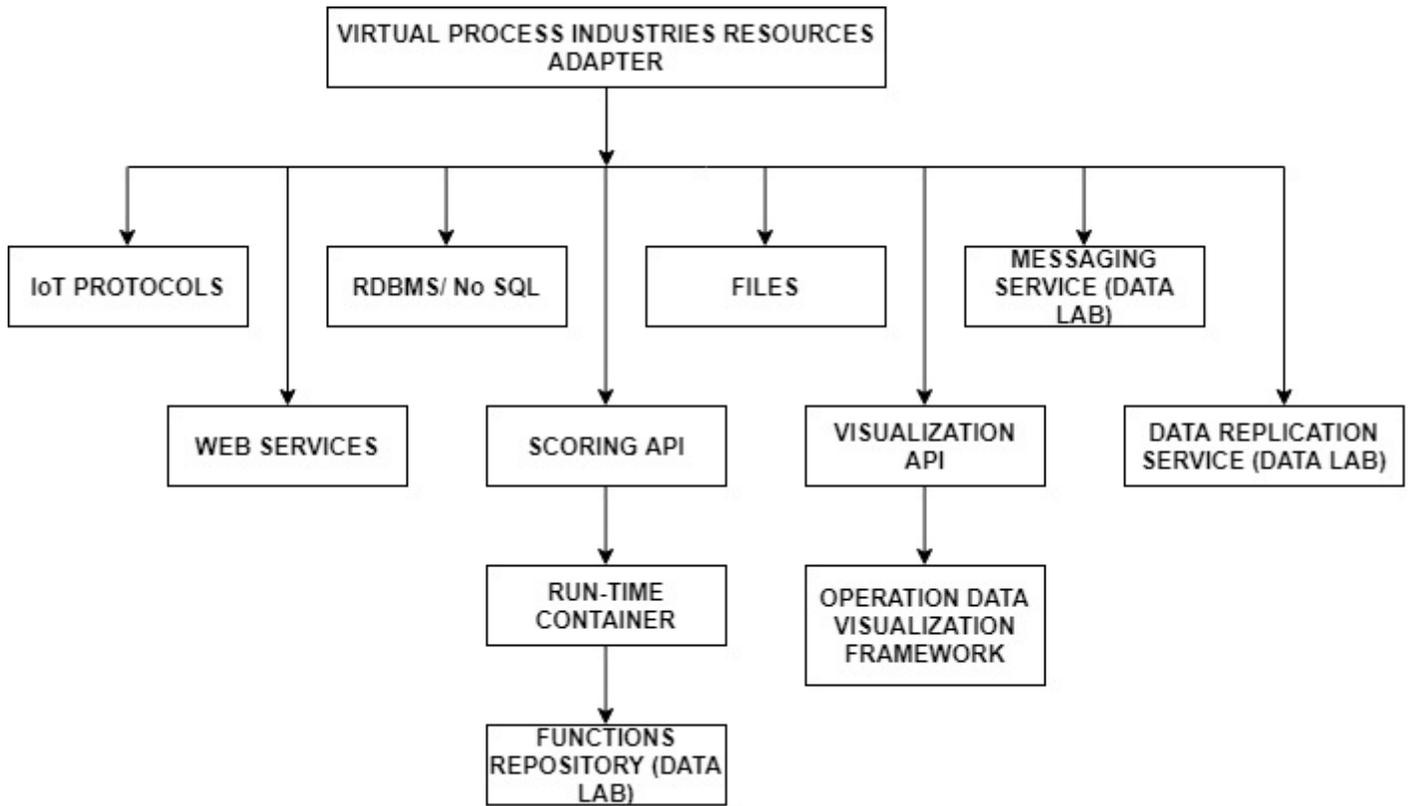


Figure 2

Operational Platform framework.

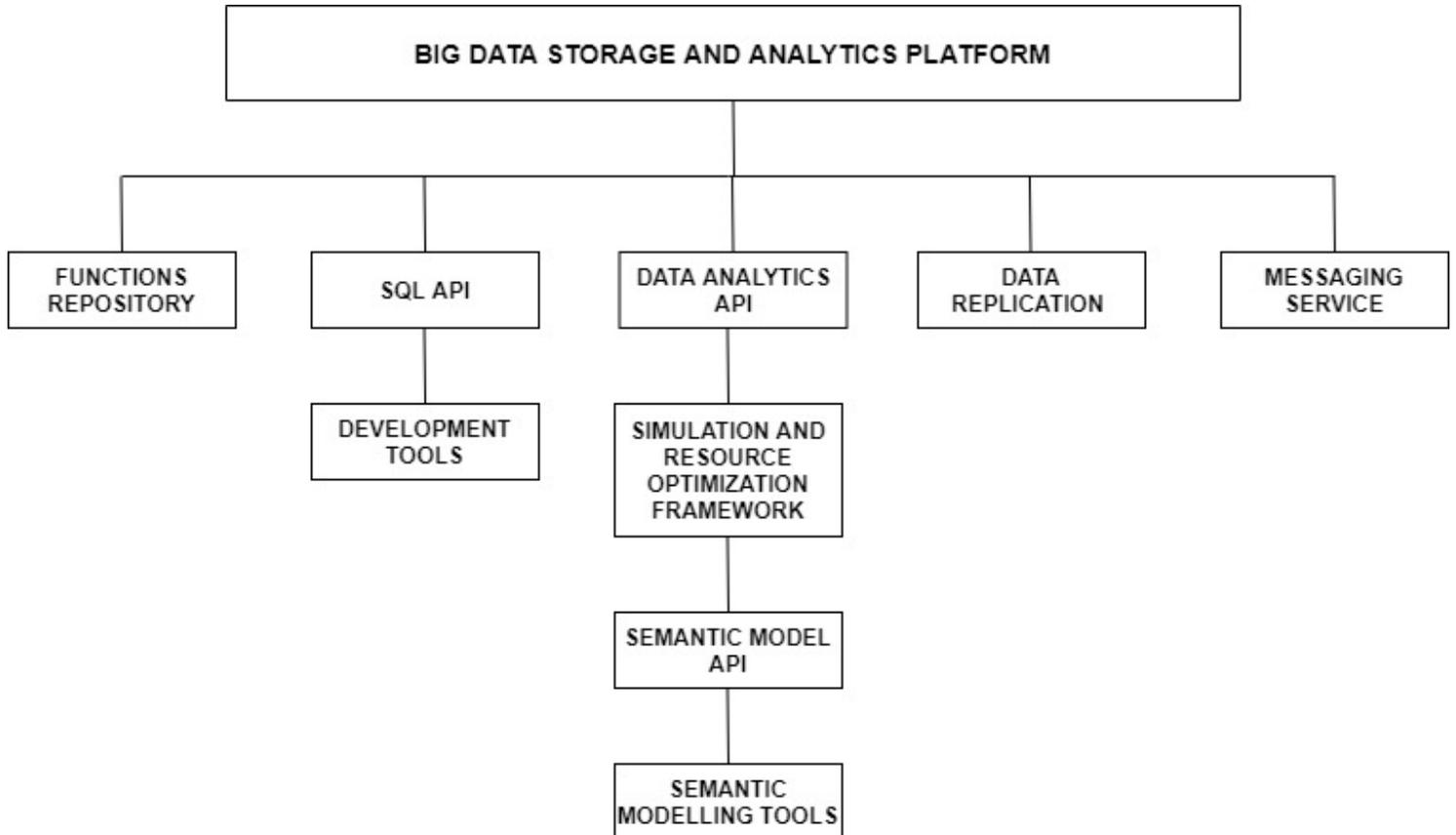


Figure 3

The Forum for cross-sectoral data.

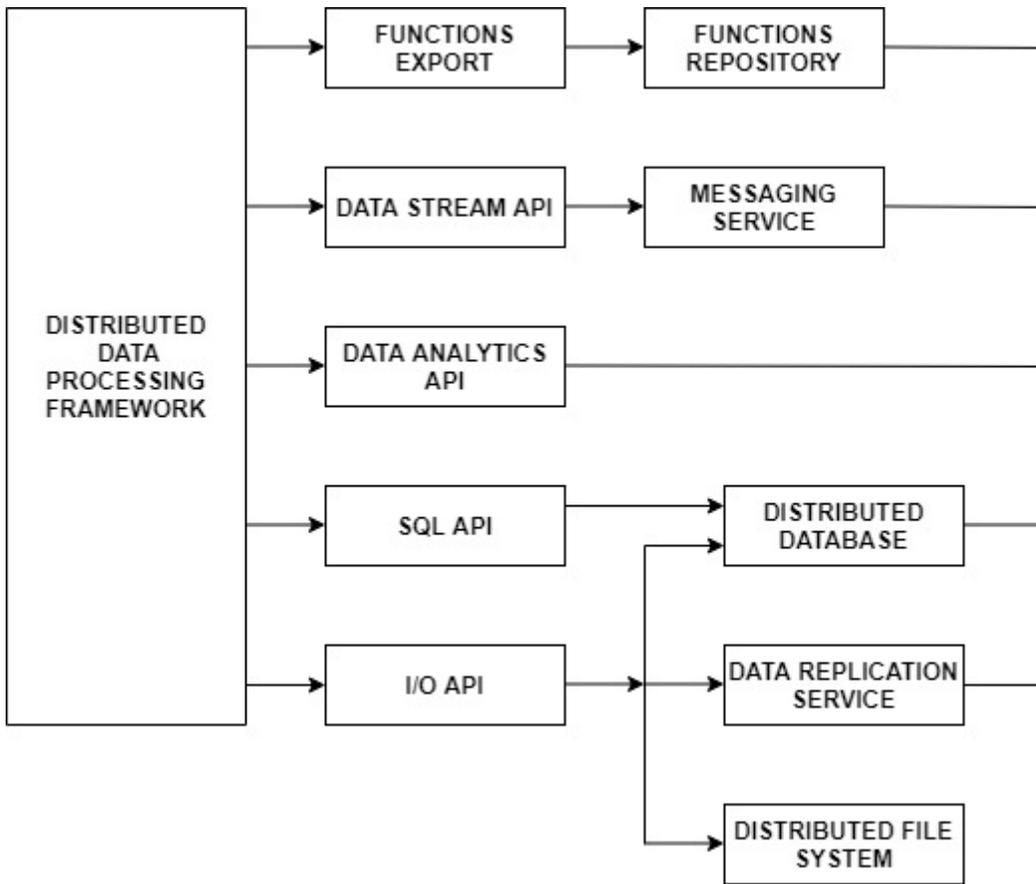


Figure 4

Big Data Storage and Analytics Platform Architecture.

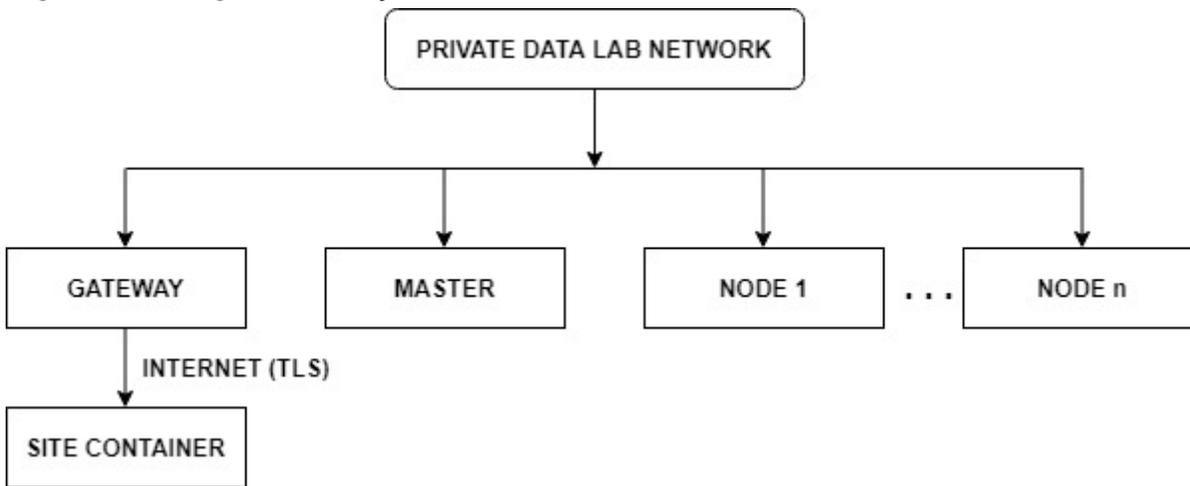


Figure 5

Deployment view of principal nodes.

```


snappy-demo2,130463295225700443,xgatt/value, {\ device\ : \ 84...
Took 20 seconds

val fanEvents = fanData.map {
  line =>
    val elements = line.split(",")

    val dt = org.joda.time.DateTime.parse(elements(6).replace(" ", "T"))
    val v = (":[^\\"]\"([0-9A-F]*)\".r findFirstMatchIn elements(5)).flatMap{ m => Some(java.lang.Integer.parseInt(m.group(1), 16)) }.getOrElse(0)

    val x = (v & 0xFF).toByte / 64.0
    val y = ((v >> 8) & 0xFF).toByte / 64.0
    val z = (v >> 16).toByte * -1.0 / 64.0

    val dist = scala.math.sqrt(x*x + y*y + z*z)

    (dt, dist)
  }.cache()
fanEvents.take(100).mkString("\n")

fanEvents: org.apache.spark.rdd.RDD[(org.joda.time.DateTime, Double)] = MapPartitionsRDD[18] at map at <console>:29
res12: String =
(2015-03-24T22:38:42.000Z,0.2390165396996618)
(2015-03-24T22:38:42.000Z,0.25242960033442985)
(2015-03-24T22:38:42.000Z,0.24506854908372067)
(2015-03-24T22:38:42.000Z,0.24655834121967968)
(2015-03-24T22:38:42.000Z,0.25242960033442985)
(2015-03-24T22:38:42.000Z,0.24506854908372067)
(2015-03-24T22:38:42.000Z,0.25242960033442985)
(2015-03-24T22:38:42.000Z,0.2312288841710741)
(2015-03-24T22:38:42.000Z,0.25242960033442985)
(2015-03-24T22:38:42.000Z,0.24506854908372067)
(2015-03-24T22:38:43.000Z,0.25194555463432966)
(2015-03-24T22:38:43.000Z,0.24506854908372067)
(2015-03-24T22:38:43.000Z,0.25242960033442985)
(2015-03-24T22:38:43.000Z,0.25242960033442985)
(2015-03-24T22:38:43.000Z,0.25242960033442985)
(2015-03-24T22:38:43.000Z,0.2538762001448738)
(2015-03-24T22:38:43.000Z,0.253...

Took 1 seconds

```

Figure 6

Tool component development.

```



val windowSizeMillis = 5000
val eventsStat = fanEvents.groupBy(t => t._1.getMillis / windowSizeMillis).mapValues(v => breeze.stats.meanAndVariance(v.map(...2))).sortByKey().cache()
eventsStat.take(100).mkString("\n")

windowSizeMillis: Int = 5000
eventsStat: org.apache.spark.rdd.RDD[(Long, breeze.stats.MeanAndVariance)] = ShuffledRDD[24] at sortByKey at <console>:33
res14: String =
(285447344,MeanAndVariance(0.2496881940178851,2.6480104917526904E-5,30))
(285447345,MeanAndVariance(0.2508021056975523,1.3615143034488494E-5,50))
(285447346,MeanAndVariance(0.2502614464003203,2.0995543999580213E-5,50))
(285447347,MeanAndVariance(0.25078243724056215,2.216586015606995E-5,49))
(285447348,MeanAndVariance(0.24924589951654072,3.2448705674301964E-5,51))
(285447349,MeanAndVariance(0.2489581053043226,7.292858564185159E-5,49))
(285447350,MeanAndVariance(0.2491688251041837,3.971432010276223E-5,50))
(285447351,MeanAndVariance(0.25093823359303913,1.86573175564571E-5,50))
(285447352,MeanAndVariance(0.2508745843010505,2.135441507669427E-5,50))
(285447353,MeanAndVariance(0.25109257541413793,1.285800583332672E-5,51))
(285447354,MeanAndVariance(0.25005666493639456,2.702...

Took 3 seconds

```

```

val windowSizeMillis = 5000
val eventsStat = fanEvents.groupBy(t => t._1.getMillis / windowSizeMillis).mapValues(v => breeze.stats.meanAndVariance(v.map(...2))).sortByKey().cache()
eventsStat.take(100).mkString("\n")

```

Figure 7

Mean and Variance.

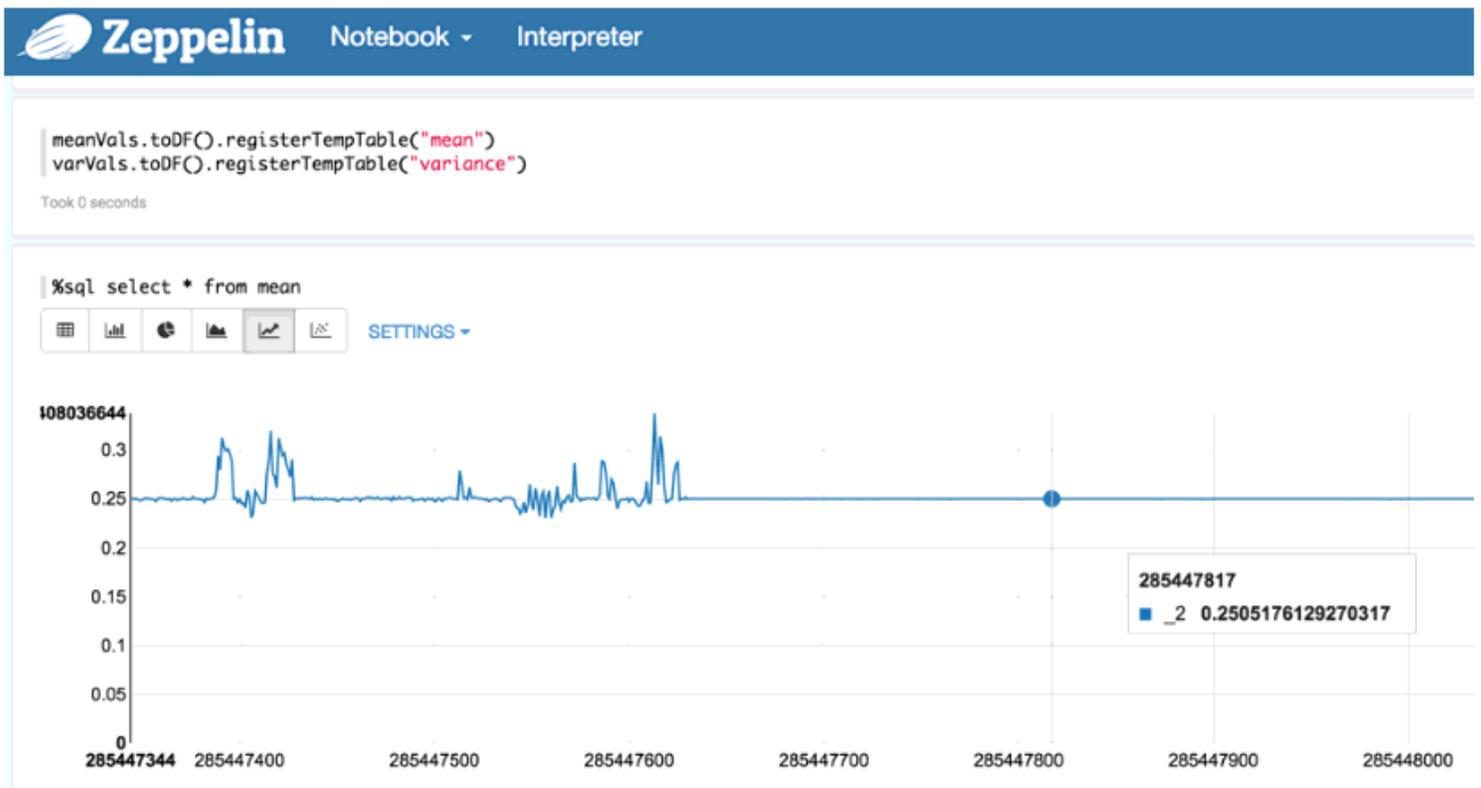


Figure 8

Mean.

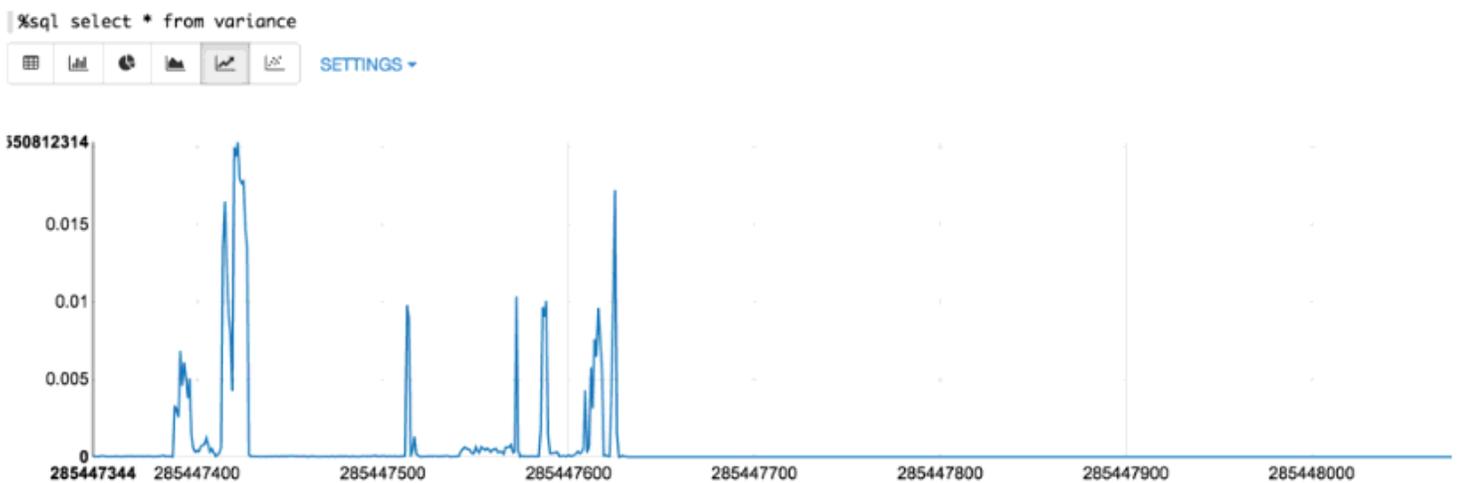


Figure 9

Variance.

```
%sql select deviceGuid, value, timestamp from notifications where value > ${threshold=0.2}
```

FINISHED ▶ ⌵ ⌵ ⌵

threshold 0.0003

SETTINGS -

All fields:
deviceGuid value timestamp

Keys
timestamp

Groups
deviceGuid

Values
value SUM

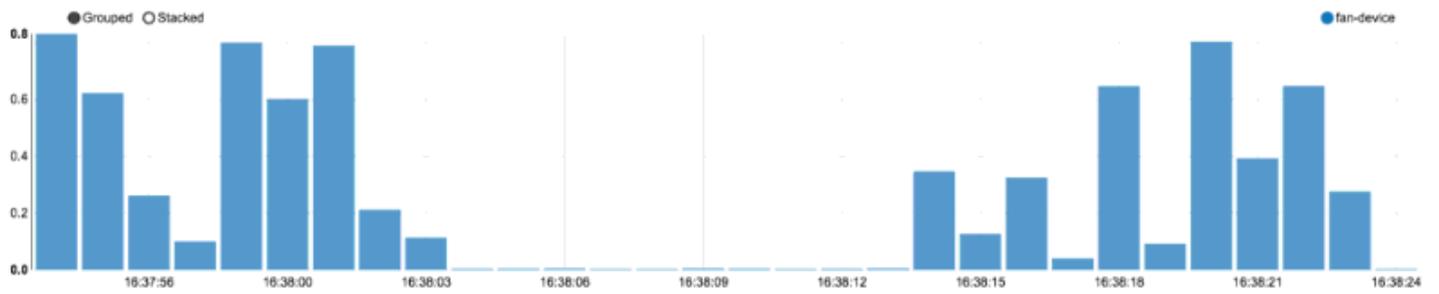


Figure 10

Parameters.