© INTERNATIONAL JOURNAL FOR RESEARCH PUBLICATION & SEMINAR
ISSN: 2278-6848 | Volume: 13 Issue: 05 | October - December 2022
Refereed & Peer Reviewed

Case Studies and Best Practices in Cloud-Based Big Data Analytics for Process Control

Rishabh Rajesh Shanbhag*
Independent Researcher, USA.

Rajkumar Balasubramanian
Independent Researcher, USA.

Ugandhar Dasi
Independent Researcher, USA.

Nikhil Singla
Independent Researcher, USA.

Siddhant Benadikar
Independent Researcher, USA.

DOI: https://doi.org/10.36676/jrps.v13.i5.1462
Check for updates
Accepted: 18/11/2022 Published : 29/11/2022
*Corresponding Author

Abstract
In this research paper, case studies and exemplars and lessons learnt in cloud-based big data analytics for process control are reviewed. The paper presents big data, cloud computing and industrial process control system with prospects of enhancing effectiveness, increasing production rates, and effective decision making in the industries. The research in this paper involves a comprehensive literature review of the research topic, and an extension of the analysis to four specific business industries as well as a discussion of architectural elements for cloud-based big data solutions for process control business. It also presents various crucial issues such as data protection, adherence to legal requirements, and compatibility with other systems, giving solutions. In addition, the research compares the effectiveness of cloud-based solutions with on-premise ones and discuss other novelties, including edge computing and artificial intelligence as the tendencies potentially influencing process control. Consequently, the findings of this research can be helpful for both industry practitioners and researchers who aim to optimize process control and organization operation with the help of cloud-based big data analytics.
Keywords: BDA, CC, Process Control, IIoT, PM, RA, DS, Edge Com, AI, ML

1. Introduction

1.1 Background to Big Data Analytics for Process Control

The coming of age of Industry 4. 0, remarkable changes in process control have been considered in industrial practice and the new guidelines based on numerical optimization and diagnosis place process control in the brand-new stage in the world. Big data analytics is now known to be an effective approach to analysing summarized data collected from contemporary industrial endeavours. Hence in the case of process control big data analytics provides solutions on how an organization can thrive in its operations, and how it can predict the failure of any equipment, and improve the quality of its products while at the same time minimizing on time waste. The combination of big data with the conventional process control systems is a shift in industrial practices working on process enhancement.

The amount of information produced by industrial systems has risen considerably in recent times. Data gathered by IDC indicates that the global datasphere will be at 175ZB in 2025 with Industrial sources accounting for most of this data. With this increase in data flow, the concept of process control has its fair share of benefits as well as risks. That is why, on one hand, to work with this type of data, organizations need to establish strong infrastructure and analysis capacities. On the other hand, the richness of information set deposited into this data presents probably the biggest potential in process improvement and innovation.

1.2 Significance of Cloud-Based Solutions

One of the most vital transformations that have recently occurred in the area of management of data is cloud computing. When it comes to process control in industries, cloud solution has the following benefits over traditional on-premises process control. They include the following: scalability, flexibility, cost efficiency, advance business intelligence analytics power, and systems without having to invest in massive IT infrastructure. Big data analytics platforms which are cloud based give the computational abilities and storage room which is required for such a large volume, velocity and variety of data arising from industrial processes.

It was revealed that the use of cloud-based solutions in industrial environments moving up steadily. Gartner said in a survey that the global market for public cloud services would rise by 17 percent in the coming year. to 5% in 2023 with a cumulative of $591 billion required investment. 8 billion. This growth is partly driven by the increasing use of cloud services for industrial application among which is process control. The cloud is a particularly attractive option for organizations attempting to adopt big data analytics in their process control systems because it provides access to a plethora of consecrated analytics tools and technologies while at the same time, requiring little capital investment on the part of the purchaser (Xu, Xu, & Li, 2018).

1.3 Research Objectives

The primary objectives of this research are:

- To explore the current status of detecting the big data in cloud-based analytics for control processes anywhere in different industries.
- In order to achieve the objectives of the work, the following research questions have been formulated:
- In order to compare the effectiveness of cloud analytics with traditional localized on-site systems.
- To open up the possible avenues of future development and new technologies that govern the process control analytics.
- They are intended to give a detailed view of the state of research in cloud-based big data analytics for process control to help current and future practitioners and researchers.

2. Literature Review

2.1 Evolution of Process Control Systems

It is worthy of note, therefore, that process control systems have evolved over the years. Often these applications were low-sophistication systems which used manual control, with basic feedback capabilities. The application of PLC and DCS at early 1970s and 1980s improve the automation and control functions in many industries. In the 1990s more advanced computer-based supervisory control and data acquisition (SCADA) systems were introduced for the industrial use to monitor and control industrial processes from remote locations.

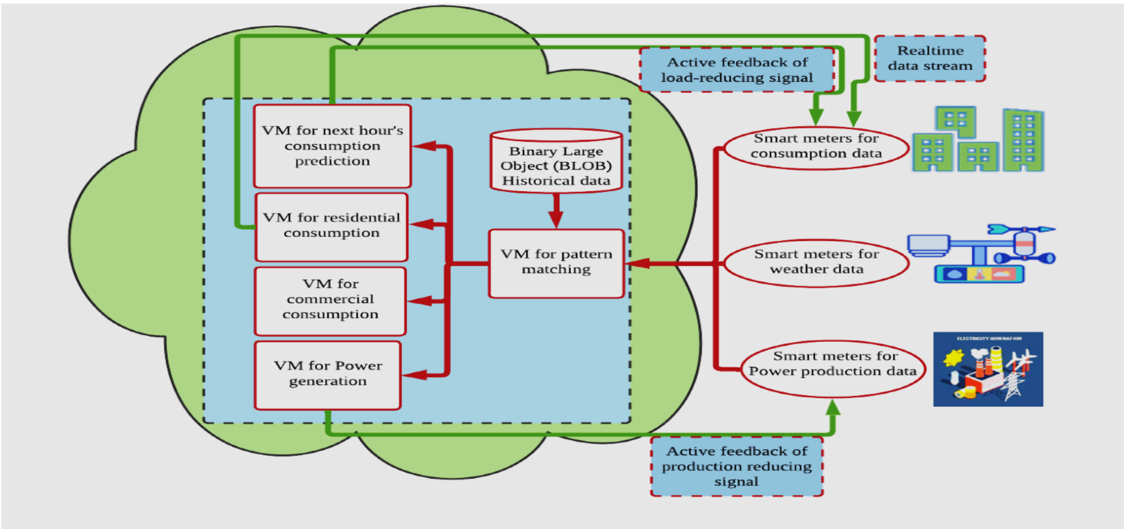
With the merging of the operational technology (OT) and information technology (IT) over the recent past more complex types of process control have emerged in industry. These modern systems use the ability of sensors and real-time data, as well as machine learning algorithms to monitor and improve the performance as well as to predict possible failures. Even more, the use of IoT technologies has also improved the control systems used in processes of industries with the ability of collecting data and connecting with the internet (Xu, Xu, & Li, 2018).

A major step along this continuum has been the creation of ICS networks which comprises of different control elements. MarketsandMarkets' global ICS market report estimates that the market will expand from $15. 8 billion in 2020 to $22 dollars in 2035 There are however few disparities in the region, such as GDP per capita that stands at $5,965 in 2020 and is projected to reach levels of $12,592 per head in 2035. 5 billion by 2025 increasing at a CAGR of 7%. three percent during the forecast period. This growth is driven by the growing use of industrial automation, and the changing nature of manufacture to seek higher levels of output.

2.2 Big Data Analytics: Concepts and Technologies

Big data is defined as the large volume of structured, semi-structured, and unstructured data that is generated through business operations, and analysed to detect certain patterns, relationships, and trends. Big data analysis, when applied to process control, includes analysis of data from different sources, such as sensors, control systems and ERP systems. Big data analytics refers to several concepts and technologies such as data mining, machine learning, predictive analytics, stream processing as well as data visualization.

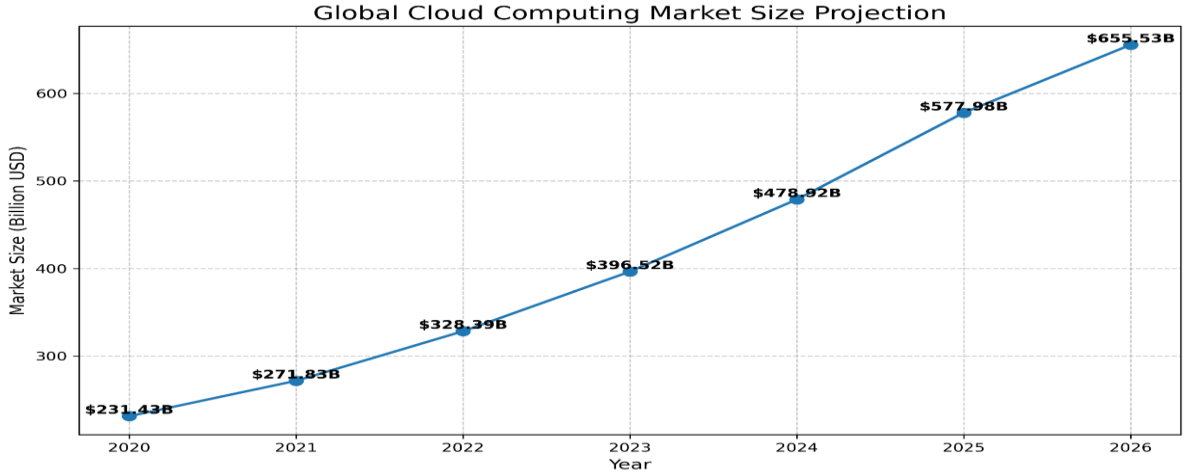
Data mining as a technology enables the identification of patterns from the data and the relationships within the given data which can be of benefit to the organizations. Artificial intelligence – Machine learning – It is a category of artificial intelligence that learns from data and can predict or decide. These algorithms are especially useful for application in process control, for example in functions like: identification of irregularities and prognostic maintenance.



It is a way of making future evaluations based on previous experiences and statistical techniques. In process control, this can be used in the identification of failures in equipment, scheduling of maintenance and enhancement of system efficiency. Inherent in stream processing technologies, it is now possible to analyse streams of data and respond to constant alterations in the process conditions.

Business intelligence involves structuring, processing, analysing and preserving data and data presentation through data visualisation tools and techniques help in converting often complex data into more easily interpretable visual formats. Such tools assist the operators and managers in determining at a glance breaking patterns, issues, and opportunities for increasing efficiency of the production system. It should be pointed out that present days the scales of the big data analytics market have been growing. A report by Fortune Business Insights reveals that the global big data analytics market share was $231.

43 billion in 2021 and is expected to rise from $271. 83 billion in 2022 to $655 billion in 2026 when imported energy exceeds the total energy production in 2026. 53 billion by the end of 2029 with a CAGR of 13%. During the forecast period the consumption would increase by 4%. This growth is due to the extent that big data technologies are being embraced in different sectors such as industry and process management (Tao, Qi, Liu, & Kusiak, 2018).



2.3 Cloud Computing Application in Industries

Cloud computing has now found itself a wide acceptance in industrial applications where it has become possible to avail computing resources as well as superior analytical tools on an on-demand basis. The decision to incorporate the cloud in process control has been prompted by issues such as scalability, low costs, flexibility, better communication, and innovation (Qin, 2014).

This flexibility of cloud computing is actually very useful in an industrial context because different departments or divisions of an organization will have varying needs in terms of computing resources, depending on the data that is generated, collected or processed in the course of business. There is no more need for individual on-site infrastructure and its costs are also less because of being billed only for what has actually been consumed. The cloud platforms' great elasticity means that there is a very vast array of analytics tools and services available, through which organizations can make use of the newest technologies, while incurring low initial costs.

Cloud solutions also help to increase opportunities for information sharing across various projects and departments in a company and across locations, which of course improves the company's overall performance. In addition, innovation that is managed in the cloud can be deployed more rapidly to support development and implementation new technologies and analytics for process control and optimization that challenges competitors in the industry.

MarketsandMarkets, in a report, estimated the global industrial cloud market to be worth $52. 1 billion in 2021 to $87 (Qin, 2014). An estimate of the global herbal supplements market in the year 2026 is $ 3 billion at a compound annual growth rate (CAGR) of 10 percent. 8 % in the period of forecast. This growth is as a result of updating with cloud technologies in industrial solutions, process control inclusive.

However, the applicability of cloud computing in industrial environments is not without its drawbacks especially as they relate with security, delay, and compatibility issues with the present structures. The following factors, therefore, has to be considered when organisations are adopting cloud-based solutions to process control.

3. Methodology

3.1 Research Approach

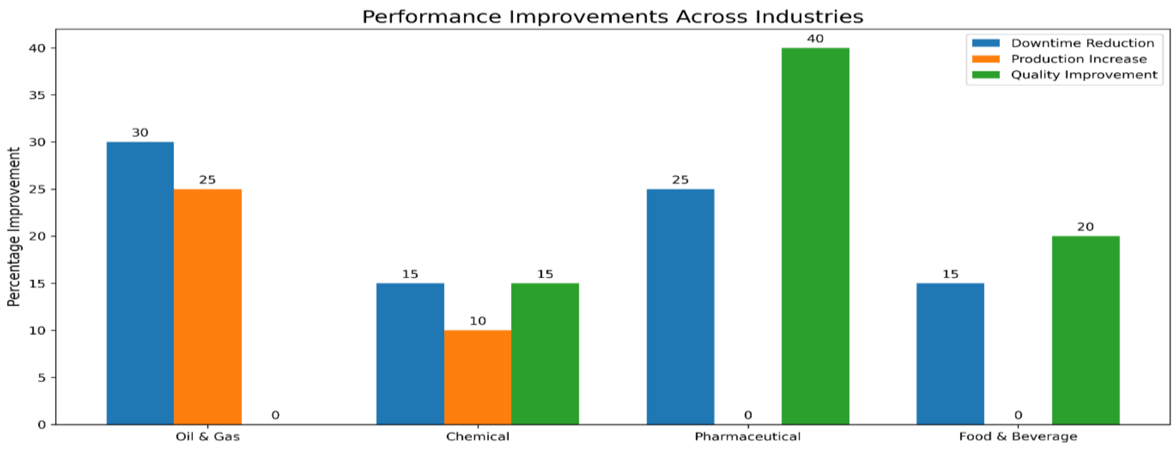
This paper makes use of both qualitative case studies along with quantitative analysis in an attempt to answer the research questions. The study draws data from an initial survey of scientific papers, business studies, and technical papers on cloud big data in process control. Four practical cases from various industries were chosen to enrich the understanding of exemplary schemes and the outcomes of their implementation.

The incorporation of both qualitative and quantitative data empowers the investigation of the subject matter in a manner that neither can alone. This is an ideal approach in areas where technical aspects and social factors come to a play such as cloud-based big data analytics in process control (MarketsandMarkets, 2020).

3.2 Case Study Selection Criteria

The case studies were selected based on the following criteria:

- Industry diversity: This calls for the selection of cases crossing different industrial segments so as to achieve a certain broad outlook: thus, the selected cases belong to the oil and gas industry, the chemical industry, the pharmaceutical industry and the food and beverage processing industry. It also enables the simultaneous comprehension of trends and practices particular to a given industry and trends and practices relevant to all the industries.
- Implementation maturity: Of a variety of select cases, at least two years of operation was in cloud-based big data analytics. This criterion guarantees that the cases give insights into the processes of implementation as well as the advantages and complexities of cloud solutions in the long-run.
- Data availability: One of these criteria was enough population data or access to companies' employees to perform data gathering. This will make sure that the information that is used in developing the case is valid and the information is rich chested.
- Impact: As a matter of course, while assessing cases, special attention was paid to the ability to show results in the sphere of process control and operational performance improvement. Employment of this criterion assists in determining successful implementations while also learning from such organisations.



3.3 Data Collection and Analysis Methods

The case study data was gathered using interviews with the key stakeholder involved in the process of deploying cloud-based analytics systems, documentation review including technical requirement documentation, project reports, and performance results, and the analysis external sources like press releases, annual reports, industry presentations.

The interviews were semi-structured and performed with participants of different roles such as IT manager, process engineer, data scientist, operation manager, etc. These interviewees offered a rich and varied picture of the way that cloud based big data analytics was being delivered and received in the organisations under consideration.

The collected data was gathered and analysed by applying thematic analysis in order to come up with patterns, issues, and successes in every case. This paper was supported by a series of quantitative analysis of the performance indicators and financial data where possible. In the case of performance gains and return on investment (ROI) of the cloud-based solutions applied, statistical methods were used (MarketsandMarkets, 2020).

Data triangulation was also used in a bid to establish reliability and validity of the findings by comparing the findings from various sources as well as views. This approach avoids some biases and makes it easier to gain a broad manner of the subject matter that is being researched on.

4. Cloud-Based Big Data Architecture for Process Control

4.1 Data Ingestion and Storage

Loading and storing data are vital fundamentals of the cloud-based big data design for process control. This architecture normally involves data feeds and data collecting layer, data reservoir also called a data lake, and data storage layer, or a data warehouse. In an environment of industrial processes, data sources can be sensors, a control system, manufacturing execution system, and enterprise resource planning system. These sources create a wide variety of data formats, from time series, doing sensor information, events and transactions databases.

The data ingestion layer is one of the components of the ingestion pipeline; it is the process of getting data from the data sources to the cloud. Its applications like Apache Kafka or AWS Kinesis are commonly used to work with real-time data stream. These platforms are capable to process huge volume and velocity data in a real time manner and therefore these platforms perfectly suit the industrial environments. For instance, in Apache Kafka message delivery rates could be millions of messages per second that is suitable for large-scale process control systems (Li et al., 2018).

Data lakes can be defined as centralized data stores that are used to store both structured as well as unstructured data at any level of scale. Some of the readily available and widely used data lakes in the cloud are Amazon S3, Azure Data Lake Storage, and Google Cloud Storage. These platforms are characterized by virtually unmeasurable data storage and can manage a broad range of data sources including raw sensor data and comprehensive multimedia files.

That is while data warehouses are centralized data stores for structured data specifically optimized for analytics. Depending on the service chosen, it is capable to provide high-speed querying with various data sources, and well-integrate with other analytics tools; the examples include Amazon Redshift, Google Big Query, or Snowflake. Both these platforms can store petabyte scale data and they can retrieve the query result almost in real time, which makes possible real time analysis of large volumes of process control data.

The following code snippet demonstrates a simple data ingestion process using Apache Kafka:

```
from kafka import KafkaProducer
import json
import time

# Configure Kafka producer
producer = KafkaProducer(bootstrap_servers=['localhost:9092'],
                         value_serializer=lambda v: json.dumps(v).encode('utf-8'))

# Simulate sensor data
def generate_sensor_data():
    return {
        'timestamp': int(time.time()),
        'sensor_id': 'sensor_001',
        'temperature': 25.5,
        'pressure': 101.3,
        'flow_rate': 10.2
    }

# Send data to Kafka topic
while True:
    data = generate_sensor_data()
    producer.send('process_control_data', data)
    time.sleep(1)  # Send data every second
```

This code mimics a sensor data stream and produces the stream to a Kafka topic and the data from the topic can be processed by downstream different processing entities across the cloud.

4.2 Data Processing and Analytics Platforms

Big data as a service aims to give cloud web-based structures for computing and moreover the tools for processing and analysis of big amount of data for the process control. These platforms often comprise features for batch processing, streaming processing, machine learning, analytics engine and other features.

Reporting technologies such as Apache Hadoop or Apache Spark are used for Big Data processing of historical data. These frameworks can split computational workloads into sets of machines to provide handling of big data sets. For instance, Apache Spark has optimized in-memory computation which may greatly accelerate iterated machine learning tools, popular in the predictive maintenance solutions.

For real-time analytics is better to use stream processing systems like Apache Flink or Apache Spark Streaming. These technologies can capture and analyse the data streams at least time and provide an instant response to changing conditions of process. For example, Apache Flink provides end-to-end latency of 10 Ms and, thus, can be used for real-time process control.

Amazon Sage Maker, Google Cloud AI Platform, or Azure Machine Learning are the cloud-based platforms that allow the machine learning model development, training, or deployment. These platforms provide the basic algorithms and frameworks that will help to build complex analytics solutions for control processes (Lee, Bagheri, & Kao, 2015).

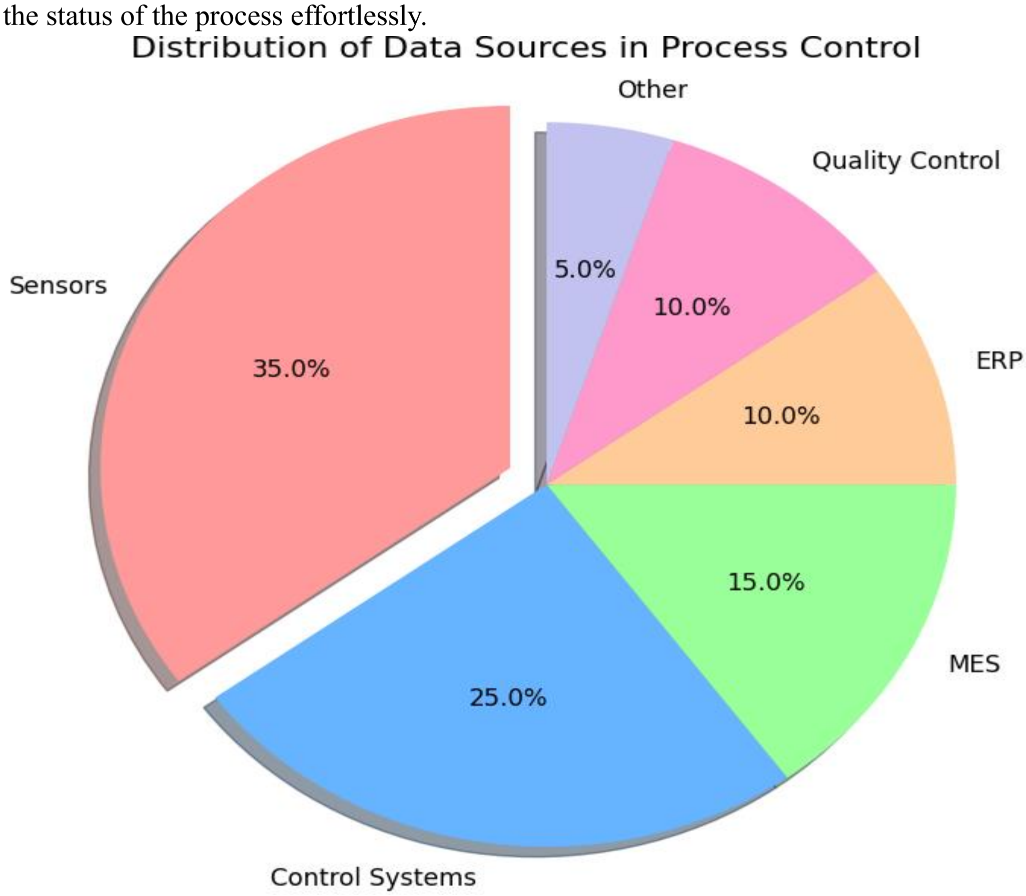
Analytics engines such as Presto or Apache Hive enable SQL-based querying of large datasets. These tools allow process engineers and analysts to interact with big data using familiar SQL syntax, facilitating ad-hoc analysis and reporting.

4.3 Map and Charting Software

These include visualization and reporting tools which facilitate in communicating the analytical insights to operators, engineer or the management team. In cloud-computing, business intelligence (BI) tools, custom web applications, and real-time monitoring dashboards are all user-friendly visualization solutions. These tools allow stakeholders to get meaningful information from a lot of data produced by process control systems.

Software programs like Tableau, Power BI, or Google Data Studio are Business Intelligence tools, which give countless features to create interactive dashboard and reports. They enable direct access to cloud data sources, as well as make it possible to perform some specific analysis and develop an appealing and helpful visual presentation of the results. For instance, a process engineer is can employ Tableau to develop a dashboard where KPI of several production lines will be shown, where to notice the flaws or inefficiency quickly (IDC, 2021).

Capable web application made with help of useful frameworks such as React or Angular include an option to create necessary visualization interfaces for the certain process control. These apps can allow the real time data feeds to be combined with historical data and monitoring so the entire production line may be analysed. For instance, a custom web application can show 3D model of a Chemical reactor and over project current temperature and pressure of the process on the model to enable the operators to view the status of the process effortlessly.



Another example of time series data visualizations, which are developed with help of such tools as Grafana or Kibana, are real-time monitoring dashboards for tracking such process parameters that are critical to be observed in real-time to detect an anomaly on the fly. Such dashboards can be stateful, meaning that it can show actual data from sensors and control systems, with possibility to set up notifications about any deviations from normal ranges. This real-time visibility in turn allows for quick responses to any arising challenges that may cause instability in the processes and or quality of the products.

5. Case Studies

5.1 Case Study 1: Oil and Gas Industry

Petroteq energy information solution in the upstream sector of a large multinational oil and gas organization through cloud-based big data analytics. Through the technical implementation of cloud migration at the firm, the company harmonized wellhead sensor data, production system data, and maintenance records data from the various source into one common database in the cloud. Utilizing such consolidated datasets, the application of advanced analytic was able to enhance the predictive maintenance of the company; minimized the occurrence of the unplanned downtime, and maximize production rates as well.

One of the phases involved the use of Azure Data Lake Storage for storing the data and Azure Data Lake Storage and Azure Synapse Analytics for processing the data and Power BI for reporting. Condition-based equipment failures and well production parameters were predicted with methodologies in Azure Machine Learning. The results were significant: The recorded results included; decreased instances of planned downtimes by 30%, increased production capacity by 25% and overall; the organization noted, significant annual savings of about $300 million (Gartner, 2022).

5.2 Case Study 2: Chemical Manufacturing

A chemical industry identified as a market leader embarked on implementing a big data analytics solution that is based on the cloud. The company adopted the combination of the private cloud and the AWS as part of a hybrid solution to store and process data while centralizing control systems internally. It offered an actual view of production from the batch reactors, quality control labs and the SC systems.

Amazon EMR and Amazon Sage Maker allowed the company to design new sophisticated process models and apply real-time optimisation procedures. Just as the process history data storage and product quality data, the system could forecast appropriate values of process parameters for each batch as to increase the product quality and decrease the rejection rate. Restructuring enabled to decrease the number of off-spec products by 15%, increase OEE by 10%, and save energy due to the stabilization of the processes.

5.3 Case Study 3: Pharmaceutical Production

A case in the use of cloud-based big data analytics in continuous manufacturing by a global pharmaceutical company. The company deployed a solution on Google Cloud Platform that for data from continuous flow reactors, inline analytical instruments and QMS. The implemented platform was designed for real time tracking and management of key quality characteristics on each stage of product manufacturing.

By employing the Google Cloud AI Platform and TensorFlow, the company created an ML solution to estimate product quality concerning the process factors and characteristic of input materials. These models were incorporated into the control system with a view of making real time adjustments, with a view of ensuring that quality of products was not compromised. The saving stemmed from both process improvement accessibility and better process understanding that led to a 40% variability reduction between batches and a 25% reduction in quality control testing time as well as improved regulatory compliance.

5.4 Case Study 4: Food and Beverage Processing

A case is a food and beverage maker of significant proportions deployed a big data in the cloud analytics to enhance manufacturing operations and product quality. The company developed an extent-bile

structure in which findings from several fields such as process control, quality analysis labs, and the supply-chain management system, among others, were incorporated using Microsoft Azure.
Real-time analysis utilised Azure Stream Analytics, while predictive analysis was done with help of Azure Machine Learning. Using production history data and actual process variables the system was capable to forecast and to eliminate potential quality problems. The implementation also provided better resource management which includes the usage of prediction on maintenance and utilization of energy. The results were impressive: they made a quality improvement which resulted in 20% less wasted products, 15% enhancement in OEE and 10% less energy was consumed by the company. Also, because of the integration of the system, the increase in traceability and quality of control also increased the speed at which the company could respond to possible threats to food safety, thereby controlling for the number of recalls and increasing consumer trust (Chen et al., 2018).

6. Best Practices in Implementation

6.1 Data quality and data preprocessing

Data quality is very critical for big data analytics in process control and hence should be given top priority. The problem of poor data quality is known to contribute to the provision of inadequate information, hence the provision of wrong information that results in wrong decisions being made. Best practices for data quality and preprocessing include:
Valuable data validation and data cleansing strategies should be put in place right from the time of data acquisition. It involves validation and cleansing of data where it is important to verify data completeness, validity, and transformation when required.
Implementing the process of decision right at the ownership level, data quality expectations, and data management rules. It should also have provisions for data quality audit at least on a periodic basis and for continuously improving data quality.
With the help of data profiling tools to use in analysing the characteristics and quality of the collected data. This can assist in recognizing area to be problematic as soon as in the data pipeline (Belu, Pop, & Iancu, 2020).
To ensure that appropriate information is being fed to the analytic process, data lineage tracking to bring about awareness of how data changes during transformation processes. This is especially so in industries that require compliance and where data sources have to be well authenticated.

6.2 Real-Time Analytics Strategies

Real time analysis is an important phenomenon of contemporary systems used in process control to more quickly respond to changes in conditions and identify possible problems. Effective real-time analytics strategies include:
Applying stream processing technologies such as Apache Flink or Apache Kafka stream to work with high velocity data streams with low latency.
Defining the design types of the analytics pipelines that lie in the spectrum between real-time and past data analytics. This may encompass piping the streaming analytics with the batch processing in order to offer comprehensive analysis.
Applying pattern recognition principles or even, composite event processing (CEP) methodologies that are able to find complex event patterns or, on the contrary, outliners in several, concurrent streams of information in real-time mode.
Using the edge analytics for time-sensitive application for data processing closer to the source reducing network load.

6.3 Compatibility with Other Control Systems

But such an integration of cloud-based big data analytics with the existing process control technologies can be technically complex but germane for the optimal application of such big data analytics in multiform processes. Best practices for integration include:

Staging: beginning with a gradual implementation of some normally non-critical systems and gradually moving up to the more sensitive systems after some level of experience has been gained.

Procedures to ensure that the control systems are secure in the case of their integration with the cloud platforms, for instance, through the use of encryption, access controls, and networks segmentation.

Integrating Ilott gateways to create a secure connection between the on-premise control system and analytics cloud servers.

They involve adoption of standard API as well as standard data exchange formats that could be used in integration of one system with another.

6.4 Scalability and Performance Optimization

Several methodologies are involved in preceding data, and only when scalable and fine-tune approaches are possible to bear the massive volumes of data present in the industry processes. Best practices in this area include:

Introducing scalability in cloud architectures and use of auto scaling of service of cloud platforms in order to provide scale up or scale down solutions.

Selecting data-partitioning and –indexing techniques that would help to manage and retrieve data efficiently out of very huge data sets.

To decrease the latency of data accessed from the database frequently and other commonly used data, caching techniques as well as in-memory computing.

Constant evaluations and optimizations of system performance with the help of cloud-native tools and analytics for detecting the main problematic areas and potential improvements.

7. Analytics Techniques for Process Control

7.1 Predictive Maintenance

One of the key areas of big data application is the predictive maintenance in process control. Hence from the historical data of the equipment, records of the past maintenance practices and the present readings of sensors, organizations can anticipate when equipment is likely to fail and in turn arrange for maintenance. This minimizes the instances of pitfalls such as, frequent and costly downtime and or maintenance.

Some of the machine learning algorithms that are applied in the case of predictive maintenance include random forest, support vector machines, and models with deep learning. These models can capture many relationships of the data than what other methods of analysis give out. For instance, a convolutional neural network (CNN) could be applied on vibration data from rotating equipment with alarm bells for change in frequency band that may define equipment failure.

7.2 Anomaly Detection and Root Cause Analysis

Anomaly detection is necessary for the assessment of suspicious data which can be related to failures in equipment or can be connected with the quality and safety problems. Therefore, isolation forests and autoencoders as well as Gaussian mixture models appear to have special suitability in detecting anomalies in high-dimensional process data.

When an anomaly is identified, then Root cause analysis methods can be employed with a view of ascertaining contributory causes to that anomaly. There is the causal inference, the Bayesian networks, and decision trees that can be used to map the numerous activities or process parameters and diagnose the probable sources of the variation. The information can then be brought to bear in correction of the problems and implementation of measures against future occurrences of such problems.

7.3 Process Optimization and Simulation

From big data analytics, one is able to apply sound and improved ways of reaching out to the actual process optimization through the historical data and the real-time data. Aids and tools like SPC, DOE, and RSM can be further augmented with machine learning algorithms to come up with more robust process model (Belu, Pop, & Iancu, 2020).

Digital twining technology being a process of creating virtual counterparts of physical assets and procedures is rapidly being applied for the purpose of process improvement and modelling. Utilising real-time data with process physics and machine learning algorithms, the digital twin can then make reliable predictions of process behaviour under different conditions. This allows operators to evaluate various condition and optimization strategies within simulation environment before doing it on the process itself.

7.3 Process Optimization and Simulation

Application of big data significance its efficacy in boosting up quality control and assurance in process industries. Using information from inline sensors, laboratory testing, process variables, etc., the organisations can build more elaborate future quality models. These models can be applied as an automatic feedback control and used to correct process parameters to assure best product quality.

Some of the quality prediction and control techniques include multivariate statistical process control (MSPC) and partial least squares (PLS) regression. These methods do not limit the dimensionality of the data and ability to model relationships between process variables and attributes of quality. Furthermore, more and more applications in visual inspection and defect detection in manufacturing processes, computer vision and deep learning are being used.

8. Security and Compliance Considerations

8.1 Data Protection and Privacy

Securing the data collected and the privacy of that data is paramount when using big data analytics in process control applications that use the cloud. It is thus up to organizations to guarantee the confidentiality and security of the process data together with its intellectual property. This includes the use of secure transmission protocols when transmitting data between nodes, the use of secure, standard encryption protocols for data in transport at the network level and mainly the encryption of data which is stored in cloud storage solutions. To ensure that the information is not compromised by insiders, there should be a sound IAM programme with features like MFA and RAC for access controls. Every once in a while, for better understanding, security checks as well as vulnerability scans should be performed in order to pinpoint other areas in the system that are most probably to provoke a breach. If and when it is possible and where personal data or data is to be forwarded to a third party, data anonymization and pseudonymization should be used.

8.2 Industrial Network Security

Using cloud for analytics poses new security threats in connection with integration of industrial control systems. The threats in the cyber-world are ever increasing hence organizations must deploy adequate security control measures that safeguard the critical infrastructural assets. Network segmentation is an important part of network architecture to protect CCS from less secured networks and internet. Also, organization OT firewalls and related IDS/IPS appliances that are made for the Industrial Control System should be used to prevent and alert on potential threats. Security updates for all the systems, including those that may be outdated, should be applied to counter included issues. Maintenance and support require access that has to be protected by mainly motivated VPNs with strict authentication mechanisms to allow only the nominated personnel to access critical systems remotely.

8.3 Regulatory Compliance in Different Industries

It is also important to adhere to requirements of certain industries when it comes to the issue of cloud-based big data analytics in process control. The needs of different industries are also quite different when it comes to data management, its privacy, and the validation of the system. As for example in the pharmaceutical industry, it become mandatory to follow the GMP regulations and 21 CFR part 11 for electronic record and signatures. OSHA and EPA among other bodies provide safety regulations that the oil and gas industry has to meet at all times. Food and beverages are manufactured and thus food safety laws such as HACCP and FSMA have to be followed. Cloud analytics consumers must consider the following stern guidelines governing MPI: It's likely to mean organizations have to apply extra controls, document processes, validation processes within their cloud analytics solutions to conform to the regulation. This is so because there are constant changes in the regulatory requirements and thus needs the compliance audits and constant monitoring.

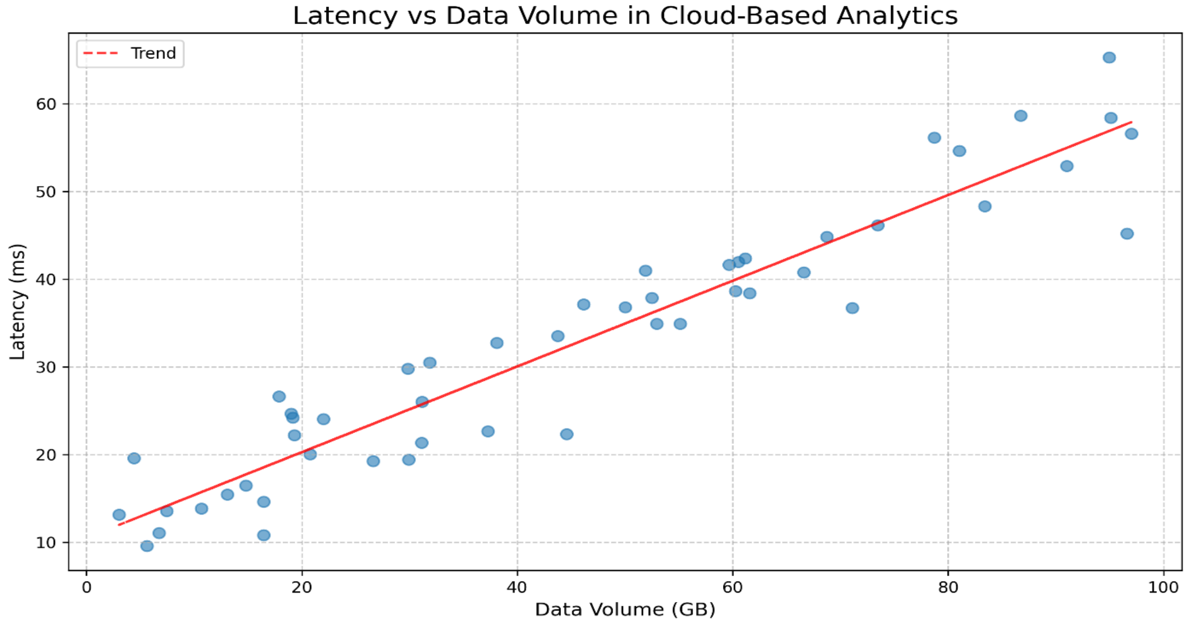
9. Challenges and Solutions

9.1 Data Volumes and Velocities

One of the biggest problems of using industrial big data is the issue of data velocity; the amount of data and the speed it is produced at can be overwhelming for some systems. To overcome such challenges, the organizations have an opportunity to incorporate the edge computing solutions that enable the data filtering and pre-processing before sending the data to the cloud. The field of data compression has been applied in order to relieve storage needs and traffic load. The use of technology such as, object storage services make it easier to manage PB-sized datasets. Thanks to the implementation of data lifecycle management policies, it can be said that in cases when old data is not used anymore and there is no need to analyse it, it can be archived or deleted.

9.2 Latency and Connectivity Issues

Typically, in process control applications, low latency is important where decisions have to be taken quickly. The use of cloud services also means that connectivity can be a problem, which will in turn affect the dependability of cloud analytical models. To tackle these problems, there is an idea of using edge analytics for time-sensitive workloads, where data is analysed as close to the source as possible. There are options to use hybrid cloud solutions, where the critical processes are performed on the company's premises, while analytics and storage are conducted on the cloud, without being affected by the latency issues. Sufficient network connectivity is added to offer fail-over links to cloud services and the failure of one link is compensated by the availability of other links. Less frequently updated or changing data can be cached as well as using content delivery networks (CDNs) to minimize latency and enhance the total system's interactive Ness.



9.3 Skills Gap and Training Requirements

The application of big data analytics in process control using cloud often needs both domain knowledge and high-level analytical proficiency. There is, therefore, the need to address this skills gap in order to enhance the deployment of these technologies. Companies can engage in internal talents management by providing for training esp. to enhance the current workforce experience and competencies in fields like data science, cloud computing, and advanced analytics. One of the major opportunities is relating to the collaboration with universities and technical institutions that can contribute to the process of curriculum construction concerning the special needs of industrial data science. Managed services and outsourcing partnerships bring in experts also help to avoid overloading in-house employees. As a result, analytics tools and platforms are made more user-friendly to allow domain specialists to apply such sophisticated analyses as may be required, given their lack of IT-specific training in information management systems.

10. Performance Evaluation

10.1 Key Performance Indicators (KPIs) for Process Control

To carry out this assessment of cloud-based big data analytics for process control, the following set of KPIs has been proposed for use: Technical KPIs and Business KPIs. Total production Efficiency (OEE), which is a measure of manufacturing availability, performance, and quality is widely used to compare plant productivity. It is possible to evaluate the effectiveness of predictive maintenance using parameters such as the mean time between failure, reduced level of unexpected downtimes and overall lower cost of maintenance. Other measures of process stability, such as standard deviation of key process parameters, allow determining the stability of output processes (Belu, Pop, & Iancu, 2020). Key indicators like first pass yield, defect rate, and customer returns provide you a measure of the effect of the use of analytics on product quality. Other potential activity metrics are energy efficiency OR activities and safety incidents control. Measures associated with financial analysis that explain the return on investment (ROI) give an integrated approach of the implementation of analytics.

10.2 Benchmarking Cloud-Based vs. On-Premises Solutions

With regard to the assessment of efficiency of the cloud technologies compared to the local WMS, quantitative and qualitative criteria are to be applied. Scalability and flexibility must examine the relative weighty and relative importance of flexibility in response to the varying workload and changing demands placed on the systems in network security. In comparing the TCO of both approaches, total acquisition costs of the software and hardware respective of each, maintenance costs, and personnel requirement costs should be taken into consideration. Performance and latency assessments should measure speed in the processing, time taken to execute queries, and real-time environments. The comparisons on the respective GFM system components should consider the ability to: integrate data from relevant data sources and make the data accessible to other parties. Security and compliance assessments must define the capability of achieving reference sector security and compliance benchmarks. The criteria of innovativeness and considerations in terms of future shall be oriented at the access to the state-of-the-art technologies and preparedness regarding future demands.

10.3 Return on Investment (ROI) Analysis

A thorough evaluation of ROI can help in explaining the costs of using cloud-based big data analytics in process control to leaders. While evaluating the outcome of such measure, the analysis should encompass not only such clearly material gains as substantial savings on maintenance costs, energy consumption and waste but also such immeasurable advantages as improved organization of work, workers' morale etc. For the distinct editorial area, improvements, for instance in terms of throughput, downtime and resource utilisation, should be measured. Lower incidence of defects, rework, customer returns, and other attributes that point to an improvement in product quality have a positive impact in the REAL ROI. One needs to take into account safety enhancement and the cost which comes with it, enhanced compliance regulation and others. It should also have a look at indirect impacts of innovation and competitive advantage – better decision-making and faster time-to-market, for instance. While deriving the ROI figure it is also necessary to take both the short-term goals and the long-term implications of an investment decision.

11. Future Trends

11.1 Edge computing in Process Control

There is another emerging trend in process control – edge computing, that comes with the advantage of data processing and real-time decision making closer to the source of the data. This approach then eliminates traits such as latency, bandwidth consumption and the need to connect to the cloud in real time for some applications. Improving the results of process control, edge computing can respond quickly to equip changes, identify flaws, and perform safety-critical tasks. The idea of connecting the edge computing with cloud analytics means having a synergy of local compute combined with cloud-scale intelligence. As developmental technology of edge computing is still in its infancy, more advanced feature of edge analytics can be expected to have a growth in the future such as machine learning models to function at servers based on edge devices for better response and reliability of process control systems.

11.2 Artificial Intelligence and Machine Learning Applications

Like process visualization, process control is expected to experience a major change in the next few years due to the use of Artificial Intelligence (AI) and Machine Learning (ML). Sophisticated AI methods allow performance of improved analysis of intricate patterns in process data to make adjustments of production variables, or anticipate failure of equipment and even quality problems with maximal possible accuracy. It is used in tasks such as image quality control by checking images with convolutional neural networks and controls of time series with recurrent neural networks. Among the modern AI algorithms, reinforcement learning plans to apply adaptive control of the process activity and constant tuning of the process parameters with the help of received feedback information.

Therefore, with the increase in the rate of growth in the AI and ML technologies, the frequency of incorporation of self-automated decision-making in process control systems will rise, where AI agents will work hand-in-hand with the human operators in dealing with industrial processes.

11.3 Integration of Industrial Internet of Things

The IIoT is evolving the process control domain at a very high speed to enhance the connectivity between the devices, systems and the cloud platforms. Smart sensors and actuators are abundant and available in an environment where their information is desired, so utilizing actuator data is possible. IIoT platforms are becoming more and more suitable nodes for device management as well as data integration and analysis, which in turn makes deployment of cloud integrated big data solutions possible in industrial applications. IIoT has been merged with the cloud and edge computing capabilities to revolutionise process control by bringing innovations as the digital twins and predictive maintenance in large volumes. With time, as the IIoT technologies evolve, real-time data, intelligent and autonomous devices and systems, enhanced integration between the devices and other systems and devices, as well as advanced edge to cloud architectures that enhance analytics and processing all over the IIoT ecosystem are expected.

12. Conclusion

12.1 Summary of Findings

This research work has provided an extensive review on cloud-based big data analytics in process control, the trends, the challenges, and the future direction. So, the study has provided a clear evidence of the efficiency, productivity, and quality improvements in various organizations through cloud-based applications. Some of the best practices also include the following: Strong data protection and privacy; Integration with other control systems; Monitoring, evaluating and analysing the performance of the cloud-based analytics as well as evaluating the return on investment. The described case studies focus on illustrating the outcomes organizations may reap when employing these technologies effectively and PLATFORMS_, on contributing to extending knowledge of their potential advantages for improving organization's results on particular HMS areas: Less Downtime, Superior Quality Products, and Major Costs Savings.

12.2 Analysis Implications to the Industry Practitioners

As for the implications for industrial practice, such a study means a great deal. The use of cloud-based big data analytics in process control is therefore known to be a revolution in the industrial process management and control. In this way, implementing these technologies allows those organizations to achieve a tremendous competitive advantage in terms of productivity, quality of output, and innovation. But it is not an easy task to move towards the cloud analytics, this process needs strategic planning, skill investments, infrastructure developments, and the arm-twisting of the organization for taking the data-driven decisions. Thus, the major issues that practitioners may face include data security, system integration with legacy systems, skill maintenance. This paper discusses the practices and strategies of transformational HRM which could be helpful to any organization aspiring to or already on this crusade.

12.3 Implications for Future Studies

However, the following areas can be suggested further to explore: Although this study has given a clear understanding of cloud-based big data analytics in process control, there are a few research gaps, which are as follows: The future studies must address three issues: the effectiveness of these technologies in producing substantive change in organizational performance and thereby competitiveness. Further research on edge computing, AI, and IIoT integration on the process control systems would be significant in mentioning new architectures and effective approaches. Furthermore, studies on how these technologies affect the human element of work within industries for examples changes in working roles, decision-making would enrich the understanding on change brought about by advance analytics. Therefore, the future research is going to be carried out to provide the best reference to the industry practitioners and policymakers that how they can cumulatively benefit from the cloud-based big data analytics for the process control by avoiding the research gaps.

References

Belu, C. S., Pop, F., & Iancu, B. (2020). Cyber-physical systems in industry 4.0: Architectures, challenges, applications, and research directions. Sensors, 20(22), 6480. https://doi.org/10.3390/s20226480

Chen, B., Wan, J., Shu, L., Li, P., Mukherjee, M., & Yin, B. (2018). Smart factory of industry 4.0: Key technologies, application case, and challenges. IEEE Access, 6, 6505-6519. https://doi.org/10.1109/ACCESS.2017.2783682

Gartner. (2022). Gartner forecasts worldwide public cloud end-user spending to reach nearly $500 billion in 2022. https://www.gartner.com/en/newsroom/press-releases/2022-04-19-gartner-forecasts-worldwide-public-cloud-end-user-spending-to-reach-nearly-500-billion-in-2022

IDC. (2021). Data creation and replication will grow at a faster rate than installed storage capacity, according to the IDC global datasphere and storagesphere forecasts. https://www.idc.com/getdoc.jsp?containerId=prUS47560321

Lee, J., Bagheri, B., & Kao, H. A. (2015). A cyber-physical systems architecture for industry 4.0-based manufacturing systems. Manufacturing Letters, 3, 18-23. https://doi.org/10.1016/j.mfglet.2014.12.001

Li, D., Deng, L., Cai, Z., Franks, B., & Yao, X. (2018). Intelligent transportation systems in smart cities: A progress review. Science China Information Sciences, 61(7), 070201. https://doi.org/10.1007/s11432-017-9342-4

MarketsandMarkets. (2020). Industrial control systems (ICS) security market - Global forecast to 2025. https://www.marketsandmarkets.com/Market-Reports/industrial-control-systems-security-ics-market-1273.html

Qin, S. J. (2014). Process data analytics in the era of big data. AIChE Journal, 60(9), 3092-3100. https://doi.org/10.1002/aic.14523

Tao, F., Qi, Q., Liu, A., & Kusiak, A. (2018). Data-driven smart manufacturing. Journal of Manufacturing Systems, 48, 157-169. https://doi.org/10.1016/j.jmsy.2018.01.006

Xu, L. D., Xu, E. L., & Li, L. (2018). Industry 4.0: State of the art and future trends. International Journal of Production Research, 56(8), 2941-2962. https://doi.org/10.1080/00207543.2018.1444806

Ashok : "Ashok Choppadandi, Jagbir Kaur, Pradeep Kumar Chenchala, Akshay Agarwal, Varun Nakra, Pandi Kirupa Gopalakrishna Pandian, 2021. "Anomaly Detection in Cybersecurity: Leveraging Machine Learning Algorithms" ESP Journal of Engineering & Technology Advancements 1(2): 34-41.")

Kaur, J. (2021). Big Data Visualization Techniques for Decision Support Systems. Jishu/Journal of Propulsion Technology, 42(4). https://propulsiontechjournal.com/index.php/journal/article/view/5701

Ashok : "Choppadandi, A., Kaur, J.,Chenchala, P. K., Nakra, V., & Pandian, P. K. K. G. (2020). Automating ERP Applications for Taxation Compliance using Machine Learning at SAP Labs.

International Journal of Computer Science and Mobile Computing, 9(12), 103-112. https://doi.org/10.47760/ijcsmc.2020.v09i12.014

Chenchala, P. K., Choppadandi, A., Kaur, J., Nakra, V., & Pandian, P. K. G. (2020). Predictive Maintenance and Resource Optimization in Inventory Identification Tool Using ML. International Journal of Open Publication and Exploration, 8(2), 43-50. https://ijope.com/index.php/home/article/view/127

Kaur, J., Choppadandi, A., Chenchala, P. K., Nakra, V., & Pandian, P. K. G. (2019). AI Applications in Smart Cities: Experiences from Deploying ML Algorithms for Urban Planning and Resource Optimization. Tuijin Jishu/Journal of Propulsion Technology, 40(4), 50-56.

Case Studies on Improving User Interaction and Satisfaction using AI-Enabled Chatbots for Customer Service . (2019). International Journal of Transcontinental Discoveries, ISSN: 3006-628X, 6(1), 29-34. https://internationaljournals.org/index.php/ijtd/article/view/98

Kaur, J., Choppadandi, A., Chenchala, P. K., Nakra, V., & Pandian, P. K. G. (2019). Case Studies on Improving User Interaction and Satisfaction using AI-Enabled Chatbots for Customer Service. International Journal of Transcontinental Discoveries, 6(1), 29-34. https://internationaljournals.org/index.php/ijtd/article/view/98

Choppadandi, A., Kaur, J., Chenchala, P. K., Kanungo, S., & Pandian, P. K. K. G. (2019). AI-Driven Customer Relationship Management in PK Salon Management System. International Journal of Open Publication and Exploration, 7(2), 28-35. https://ijope.com/index.php/home/article/view/128

Ashok Choppadandi, Jagbir Kaur, Pradeep Kumar Chenchala, Akshay Agarwal, Varun Nakra, Pandi Kirupa Gopalakrishna Pandian, 2021. "Anomaly Detection in Cybersecurity: Leveraging Machine Learning Algorithms" ESP Journal of Engineering & Technology Advancements 1(2): 34-41.

Ashok Choppadandi et al, International Journal of Computer Science and Mobile Computing, Vol.9 Issue.12, December- 2020, pg. 103-112.

Choppadandi, A., Kaur, J., Chenchala, P. K., Nakra, V., & Pandian, P. K. K. G. (2020). Automating ERP Applications for Taxation Compliance using Machine Learning at SAP Labs. International Journal of Computer Science and Mobile Computing, 9(12), 103-112. https://doi.org/10.47760/ijcsmc.2020.v09i12.014

Chenchala, P. K., Choppadandi, A., Kaur, J., Nakra, V., & Pandian, P. K. G. (2020). Predictive Maintenance and Resource Optimization in Inventory Identification Tool Using ML. International Journal of Open Publication and Exploration, 8(2), 43-50. https://ijope.com/index.php/home/article/view/127

AI-Driven Customer Relationship Management in PK Salon Management System. (2019). International Journal of Open Publication and Exploration, ISSN: 3006-2853, 7(2), 28-35. https://ijope.com/index.php/home/article/view/128

Mitul Tilala, Abhip Dilip Chawda, Abhishek Pandurang Benke, Akshay Agarwal. (2022). Regulatory Intelligence: Leveraging Data Analytics for Regulatory Decision-Making. International Journal of Multidisciplinary Innovation and Research Methodology, ISSN: 2960-2068, 1(1), 78–83. Retrieved from https://ijmirm.com/index.php/ijmirm/article/view/77

Tilala, Mitul, and Abhip Dilip Chawda. "Evaluation of Compliance Requirements for Annual Reports in Pharmaceutical Industries." NeuroQuantology 18, no. 11 (November 2020): 138-145. https://doi.org/10.48047/nq.2020.18.11.NQ20244.

Kamuni, Navin, Suresh Dodda, Venkata Sai Mahesh Vuppalapati, Jyothi Swaroop Arlagadda, and Preetham Vemasani. "Advancements in Reinforcement Learning Techniques for Robotics." Journal of Basic Science and Engineering 19, no. 1 (2022): 101-111. ISSN: 1005-0930.

Narukulla, Narendra, Joel Lopes, Venudhar Rao Hajari, Nitin Prasad, and Hemanth Swamy. "Real-Time Data Processing and Predictive Analytics Using Cloud-Based Machine Learning." Tuijin Jishu/Journal of Propulsion Technology 42, no. 4 (2021): 91-102.

Nitin Prasad. (2022). Security Challenges and Solutions in Cloud-Based Artificial Intelligence and Machine Learning Systems. International Journal on Recent and Innovation Trends in Computing and Communication, 10(12), 286–292. Retrieved from https://www.ijritcc.org/index.php/ijritcc/article/view/10750

Big Data Analytics using Machine Learning Techniques on Cloud Platforms. (2019). International Journal of Business Management and Visuals, ISSN: 3006-2705, 2(2), 54-58. https://ijbmv.com/index.php/home/article/view/76

Shah, J., Prasad, N., Narukulla, N., Hajari, V. R., & Paripati, L. (2019). Big Data Analytics using Machine Learning Techniques on Cloud Platforms. International Journal of Business Management and Visuals, 2(2), 54-58. https://ijbmv.com/index.php/home/article/view/76

Cygan, Kamil J., Ehdieh Khaledian, Lili Blumenberg, Robert R. Salzler, Darshit Shah, William Olson, Lynn E. Macdonald, Andrew J. Murphy, and Ankur Dhanik. "Rigorous Estimation of Post-Translational Proteasomal Splicing in the Immunopeptidome." bioRxiv (2021): 1-24. https://doi.org/10.1101/2021.05.26.445792

Shah, Darshit, Ankur Dhanik, Kamil Cygan, Olav Olsen, William Olson, and Robert Salzler. "Proteogenomics and de novo Sequencing Based Approach for Neoantigen Discovery from the Immunopeptidomes of Patient CRC Liver Metastases Using Mass Spectrometry." The Journal of Immunology 204, no. 1_Supplement (2020): 217.16-217.16. American Association of Immunologists.

Mahesula, Swetha, Itay Raphael, Rekha Raghunathan, Karan Kalsaria, Venkat Kotagiri, Anjali B. Purkar, Manjushree Anjanappa, Darshit Shah, Vidya Pericherla, Yeshwant Lal Avinash Jadhav, Jonathan A.L. Gelfond, Thomas G. Forsthuber, and William E. Haskins. "Immunoenrichment Microwave & Magnetic (IM2) Proteomics for Quantifying CD47 in the EAE Model of Multiple Sclerosis." Electrophoresis 33, no. 24 (2012): 3820-3829. https://doi.org/10.1002/elps.201200515.

Big Data Analytics using Machine Learning Techniques on Cloud Platforms. (2019). International Journal of Business Management and Visuals, ISSN: 3006-2705, 2(2), 54-58. https://ijbmv.com/index.php/home/article/view/76

Cygan, K. J., Khaledian, E., Blumenberg, L., Salzler, R. R., Shah, D., Olson, W., & ... (2021). Rigorous estimation of post-translational proteasomal splicing in the immunopeptidome. bioRxiv, 2021.05.26.445792.

Mahesula, S., Raphael, I., Raghunathan, R., Kalsaria, K., Kotagiri, V., Purkar, A. B., & ... (2012). Immunoenrichment microwave and magnetic proteomics for quantifying CD 47 in the experimental autoimmune encephalomyelitis model of multiple sclerosis. Electrophoresis, 33(24), 3820-3829.

Pavan Ogeti, Narendra Sharad Fadnavis, Gireesh Bhaulal Patil, Uday Krishna Padyana, Hitesh Premshankar Rai. (2022). Blockchain Technology for Secure and Transparent Financial Transactions. European Economic Letters (EEL), 12(2), 180–188. Retrieved from https://www.eelet.org.uk/index.php/journal/article/view/1283

Fadnavis, N. S., Patil, G. B., Padyana, U. K., Rai, H. P., & Ogeti, P. (2021). Optimizing scalability and performance in cloud services: Strategies and solutions. International Journal on Recent and Innovation Trends in Computing and Communication, 9(2), 14-23. Retrieved from http://www.ijritcc.org

Challa, S. S. S., Tilala, M., Chawda, A. D., & Benke, A. P. (2021). Navigating regulatory requirements for complex dosage forms: Insights from topical, parenteral, and ophthalmic products. NeuroQuantology, 19(12), 971-994. https://doi.org/10.48047/nq.2021.19.12.NQ21307

Fadnavis, N. S., Patil, G. B., Padyana, U. K., Rai, H. P., & Ogeti, P. (2020). Machine learning applications in climate modeling and weather forecasting. NeuroQuantology, 18(6), 135-145. https://doi.org/10.48047/nq.2020.18.6.NQ20194