International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)
doi : https://doi.org/10.32628/IJSRSET23103132

Optimizing Telemetry Data Processing Pipelines for Large-Scale Gaming Platforms

Balachandar Paulraj
Independent Researcher, USA

Article Info
Volume 7 Issue 5
Page Number : 401-410

Publication Issue :
September-October-2020

Article History
Accepted : 01 Oct 2020
Published : 12 Oct 2020

ABSTRACT
This article is devoted to the peculiarities of telemetry data processing pipelines optimization for the platforms of massively multiplayer gaming. Since the amount, velocity, and variety of gameplay data continue to increase, real-time data handling has to be optimised for the sake of system performance and player experience. Based on MPI, Apache Spark, machine learning models, the work identifies approaches for predictive analytics and real-time data processing. That it examines how cloud environments are addressing fault tolerance and proposed different ways of collecting, processing and deploying models. AI and edge computing's future advancements are also expected to address problems with data privacy, delay, and expandability.

Keywords :- Telemetry data processing, large-scale gaming, real-time analytics, MPI, Apache Spark, machine learning, fault tolerance, cloud computing, data streaming, predictive analytics etc.

I. INTRODUCTION

The massive amount, rate, and type of data generated during gameplay requires the enhancement of telemetry data processing streams for extensive gaming platforms. Steady advancements in the processing methods are important for handling real-time data depending on the complexity of the gaming environments necessary for enhancing the system's performance and the players' experience. MPI, Apache Spark and machine learning models enables gaming platforms to predict future system usage, process telemetry data fast and adapt to players' actions. Therefore, while describing the methods of telemetry data processing and the developments to discuss their impact on operational performance and real-time decision support in this paper.

II. LITERATURE REVIEW

Near-Real-Time Processing with Spark-MPI

According to Malitsky et al., 2017., The volume, velocity, and variety of data have risen rapidly; therefore there is a need for near real-time systems. New formations of data management and data processing structures and computational technologies address the emerging tasks in managing large volumes of data in scientific and experimental practices. These problems are prevented by the Spark-MPI technology which opens up great

potential to implement efficient computations. Sparked-MPI combines MPI's high-performance computing framework with the data-intensive processing capabilities of Apache Spark and delivers a solid solution for processing vast volumes of data in near real-time. Apache Spark enabled a structure called Resilient Distributed Datasets (RDDs) where high level algorithms are separated from data sources and it revolutionised the processing of data. About complex and distributed processing including machine learning and graph processing applications and SQL queries this middleware achieved scalability and fault tolerance.

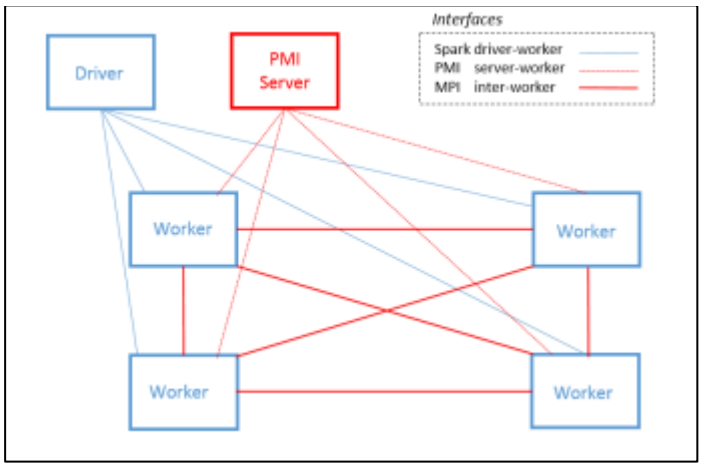

Figure 1: The Spark-MPI approach

(Source: Malitsky et al., 2017)

This is elaborated by Spark-MPI that gets the system design from MPI and the process management to improve the capability of the system to cater complex compute intensive work (Malitsky et al., 2017). It has been established that this integrated approach fits well in several domains of study. For example, scientific number research on Spark by Zaharia et al (2016) demonstrated the effectiveness of data centric operations under its performance improvement. According to the studies carried out by Gropp et al. (2014), MPI have been used in high-performance computing in MPI. MPI and Spark's RDD middleware allow dealing with the volumetric and velocity demands and as such it is adequate for real-time including online tomographic and ptychographic reconstruction. Therefore, integration of Spark and MPI develops a strong foundation for the in-near real-time data processing that meets the requirements of today's research facilities and enhances the possibilities for the Data Intensive Applications further.

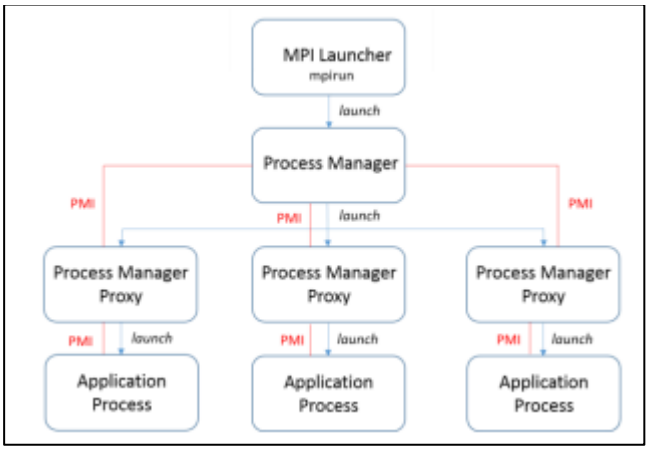

Figure 2: Hydra process manager

(Source: Malitsky et al., 2017)

Fault Tolerance in Cloud-Based Telemetry Systems

According to Hasan and Goraya., 2018., One of the key challenges in cloud computing is fault tolerance that directly impacts the reliability of data-centric applications like telemetry analysis for MMoG platforms. Continual functioning despite imperfections is especially important in cloud computing because they handle vast amounts of data and complex calculations. This study establishes that though there is a difference in the mechanism applied to the fault tolerance based on the type and class of failure, there is a unity of approach in the different situations. As it has been demonstrated above, how both checkpoint-restart and replication have contributed to preventing or reducing what can be classified as crash faults, common in distributed and dynamic systems, has been discussed. For instance, in the checkpoint-restart technique that is described by Zhao et al (2017), an application writes some notes to its state periodically, so it can restart easily and with many minimum losses. Morrow et al (2018) also points out that another widely known method of increasing availability is replication where the data is copied to the necessary nodes to so that operation can continue when a node is no longer operational (Hasan and Goraya, 2018). Current questionnaires depict how fault tolerance frameworks are developed and how they can be utilized in disparate cloud based condition. For instance, Jiang et al. 's (2019) paper offers a comparative study of fault tolerance frameworks where the authors discuss how such techniques respond to some issues in the cloud setting. Furthermore, Singh et al. (2020) integrated all of these frameworks in quantitative data to show how these frameworks are effective and feasible quantitatively.

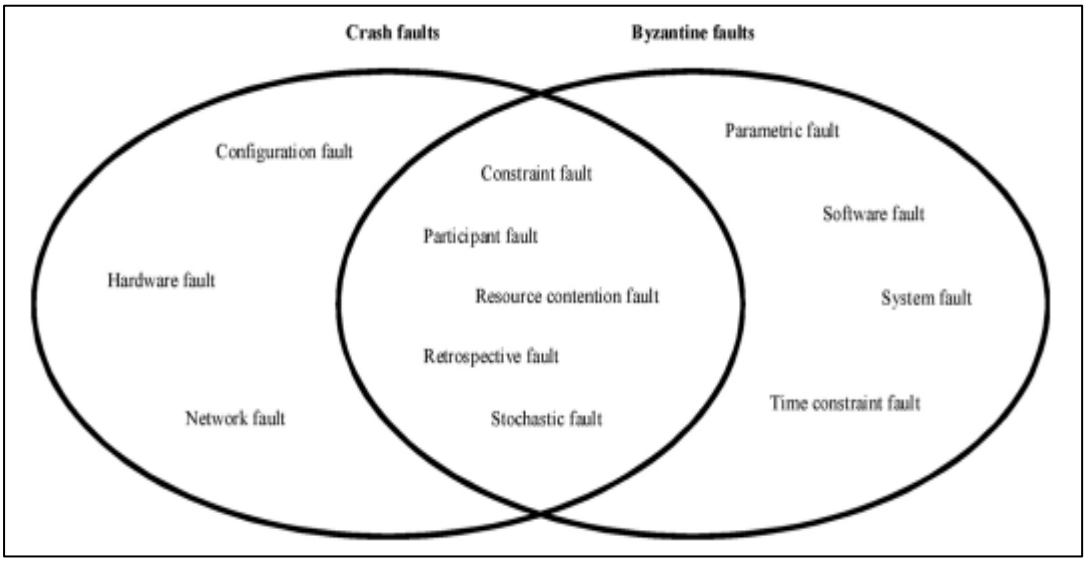

Figure 3: Fault categorization in cloud

(Source: Hasan and Goraya, ., 2018)

These strategies of fault tolerance are essentials where it concerns processing telemetry data in the game platforms. They guarantee that telemetry data pipelines are provided and accessible in case other systems' problems or data loss happen. It is also hope for the gaming platform to make further use of powerful fault tolerant mechanism so that the running speed together with the user experience will also be enhance and the services provided will be even more reliable.

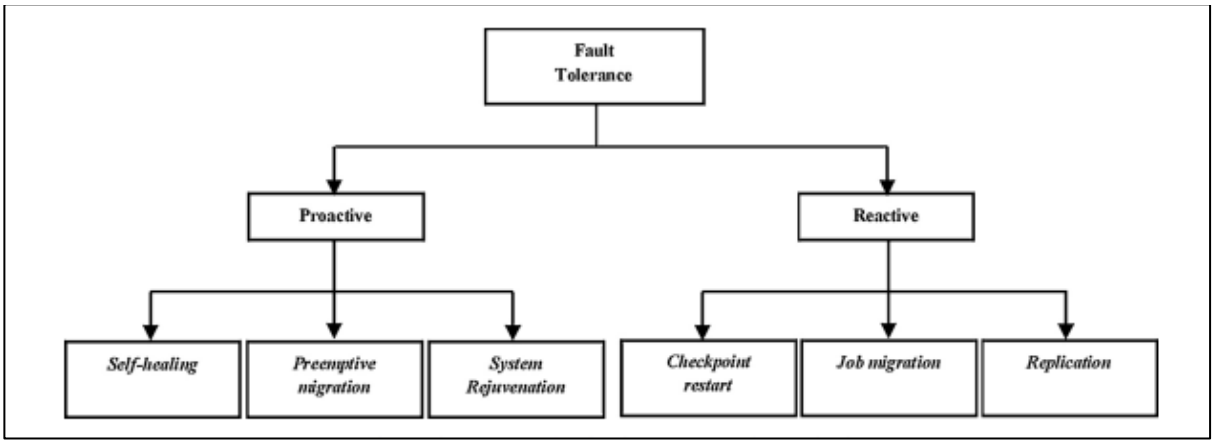

Figure 4: Classification of fault tolerance approaches in cloud

(Source: Hasan and Goraya, ., 2018)

Methods

Data collection and data processing

Data acquisition is the first stage in the improvement of telemetry data processing paths of large scale gaming facilities. This is an important aspect of information-gathering which has to be done from a variety of sources within the gaming environment and is also not limited to the past but may also focus on the present. It is concerned with game events, the network, or statistics, players and their activities including systems performance measurements. There are telemetry data that have to be gathered, and it is possible only through employing more sophisticated approaches to instrumenting game clients and servers. Most of the time, light-weight agents or middleware that do not interfere with the real-time of the game but provide correct information will be used in helping with data capture (Soikkeli, 2019). When gathered, data is worked in stages while attempting to sustain the system's throughput and acquire usable information. Ingestion real-time that activates the processing pipeline it is usually done by utilizing amazon kinesis or Apache Kafka. Afterwards, the information is either in batch or real-time to meet the needs of the application as suggested by Singh (2016). Appending flexibility together with the ability of immediate response and changes Real time processing suggests data processing with its reception with the help of Apache Flink or Spark Streaming. Batch processing concerns a large volume of previous records to analyze them or monitor trends using Apache Spark or Hadoop applications. Therefore, using this approach telemetry data is analyzed and applied to improve the efficiency of the gaming process and entertainment experience.

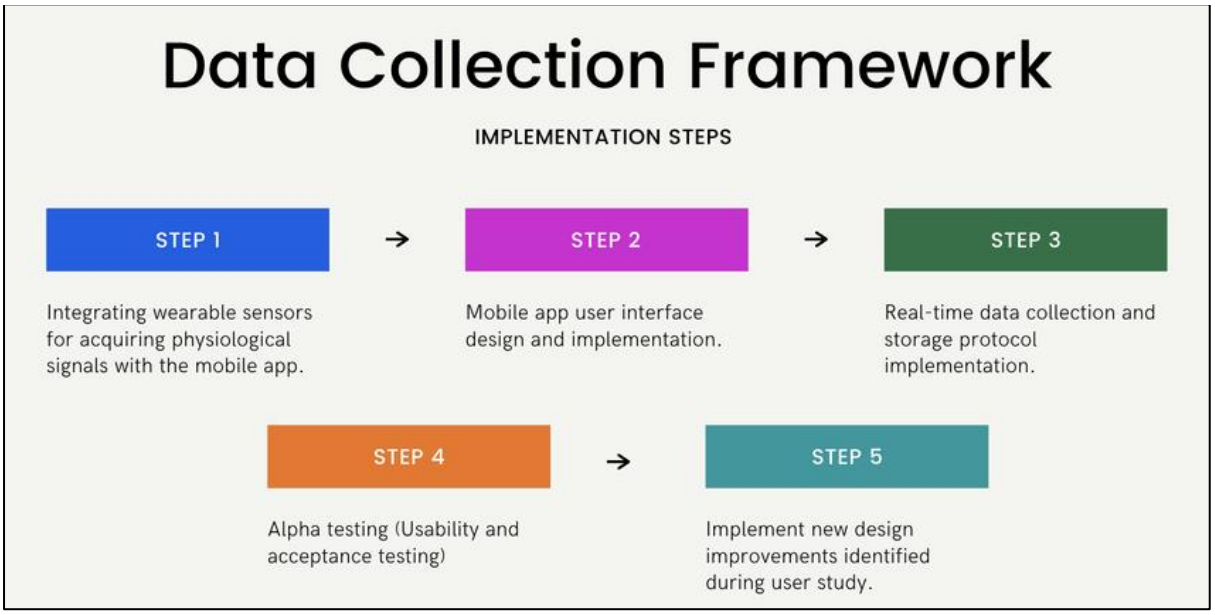

Figure 5: Data collection framework

(Source: https://www.researchgate.net/)

Designing of Machine Learning Models

The machine learning models have to be designed with the help of the few key stages to design the predictive systems which shall work. For this reason, a necessary step before applying the method is to gather and preprocess the relevant data in such a way that only clean and homogeneous data enters the subsequent phases. This comprises feature selection, normalization, and data cleaning to increase the effectiveness of the designed models (Patel et al, 2019). Next what type of problem is solved then the right algorithms are used such as clustering, regression, classification or others. Subsequently, a portion of the data is for training the model and their values are further optimized for optimal performance of the model. Hence, the cross-validation is one of the validation procedures aimed at establishing the suitability of a given model in other conditions as well. Lastly, to make it as fit as possible for real-time data conditions, the model is trained using sample data and improved based on performance metrics and user feedback.

Implementation and Deployment

Implementation and deployment of machine learning models involves putting into practice the trained model in a production model in a way that it can analyze real time data or make real time prediction. One of the activities that fall under this step involves coding of the model into a suitable framework or the appropriate application in order to make it compatible with current systems (Galati, 2017). In deployment the model is taken to a real-world setting and rather than having the option of running through test cases on input data the model is exposed to real data input and is expected to deliver real results output. This is because after deploying the particular model, there is always the need for monitoring and maintenance in order to ensure that the model is stable and running optimally. This involve in handling update, monitoring the analytics and handle any issue that arise. Successful deployment ensures the model operates correctly and provides meaningful data while maintaining its effectiveness in the long-term.

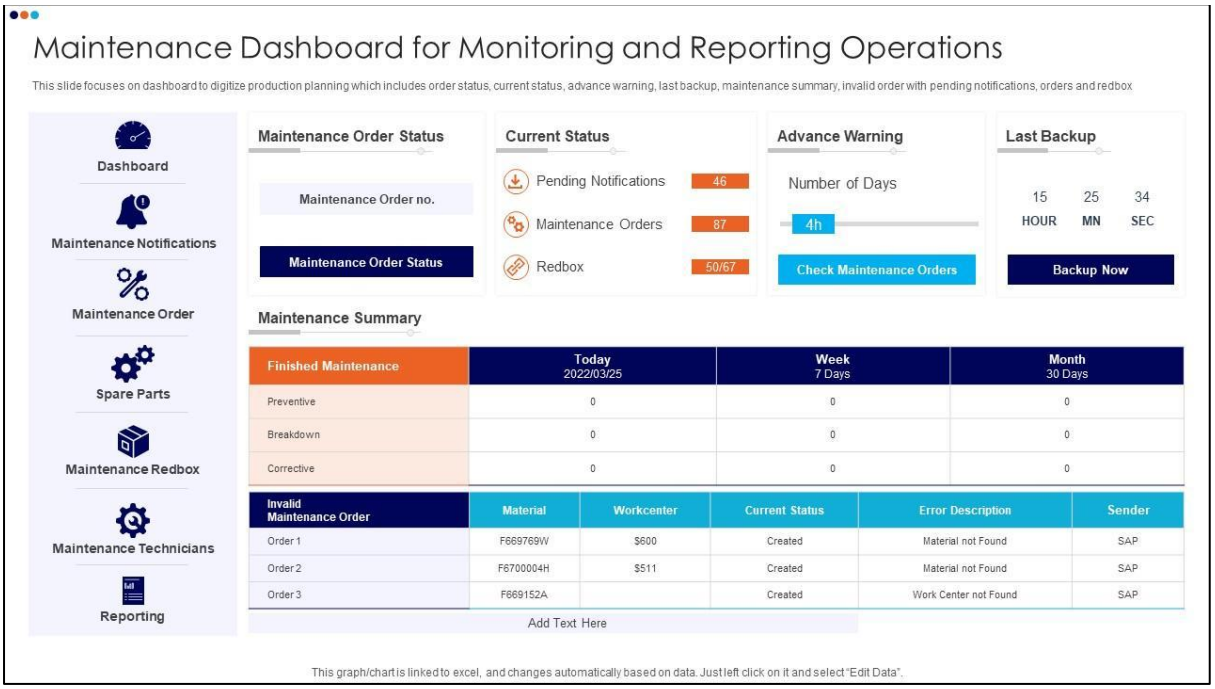

Figure 6: Monitoring and Maintenance

(Source: https://www.slideteam.net)

III. Result

Predictive Analytics in Sales and Demand

Another application of telemetry in the processing of huge gaming systems includes the use of machine learning and data analytics to predict system requirements and player behavior. it is possible to predict peak loads, potential failures in the system, and players' actions based on analyzing the available information on historical data on gameplay, characteristics of the system, and usage patterns (Basak et al, 2017). These insights help in making timely changes to how data is processed based on the current traffic, by adjusting the flow to prevent bottlenecks or investing in more resources to accommodate the increase in traffic. Gaming platforms can ensure a faster response to real-time situation and better performance by allocating resources depending on expected workload.

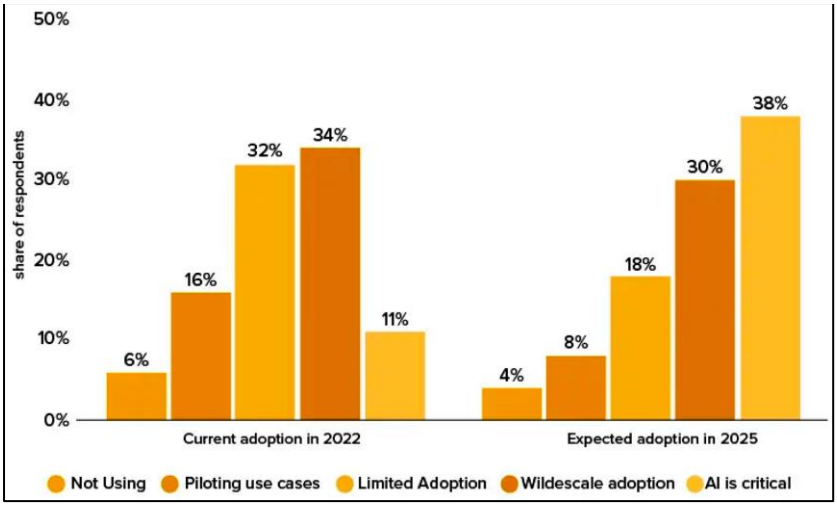

Figure 7: Predictive Analytics in Gaming

(Source: https://www.mdpi.com/)

Innovation Strategies for Inventory Management and Replenishment

The major aspects of large-scale gaming platforms' telemetry data processing pipeline innovation include data management and analysis. One of the methods is the ability to utilize real-time data streaming technologies, such as Apache Kafka to manage large volumes of telemetry data and interpret such data. So for the purpose of predicting the gaming trends and improving the functionality of the system advanced analytics and machine learning algorithms are used (Dixon, 2017). Automation technologies reduce the time or delay in mapping data while at the same time increasing the quantity or rate at which they are ingested. Integration of the information through the use of predictive analytics further helps in tackling certain performance issues before they are felt during gameplay. Other benefits of collaborative methods also include scalability and responsiveness, which are enhanced with integration of accessible cloud services. Taken together these advances make for systems that are more stable, and user experiences which are more optimal as well as processes that are more efficient in terms of data processing.

Redesigning the Lines of Logistics and Supply

The improvement of the flow and processing of the data to make them run faster and more efficiently is part of the general redesign of the supply and logistic chains in the context of telemetry data pipeline for massive gaming platforms. To ensure the effectiveness in managing the telemetry data, this procedure requires redesigning input, storage and processing of data. Through the application of the distributed computing frameworks and real time data streaming platforms the system is in a position to manage the large data generated by the gaming activities (Veith, 2019). These components can be rearranged by gaming systems to achieve greater data processing speed, reduce system contention, and enhance the user experience in general. Through this method, efficiency and player satisfaction is improved as telemetry data is processed and analyzed faster thus providing real time feedback and adjusting to changing gameplay scenarios.

IV. Discussion

As such, machine learning models, live analytics and pipeline optimization are necessary in telemetry data processing involving large-scale gaming platforms to enhance system efficiency and players' satisfaction. That is why it is crucial to design this SOG and implement such models to address a large volume, high speed, and high variety of telemetry data in gaming contexts (Callegaro et al, 2019). It is a fact that to increase the reliability of the forecasts and get the necessary timely information, it is necessary to solve the problems of constructing effective models, increasing their performance, and data preparation. Lowering the latency and improving the response time of the gameplay is possible only by using reliable models that provide the possibility of real-time calculations (Christidis et al, 2019). In addition, constant examinations and modifications of these models help to address shifts in users' behavior and gaming characteristics, maintain the performance and reliability of the system. This helps to ensure that the telemetry pipelines are always optimized through this approach, thereby enhancing the whole affair in gaming.

V. Future Directions

Other experimental and more complex machine learning models and AI-based analytics will likely be integrated into telemetry data processing pipeline improvements for mass-scale gaming systems in future developments (Ruffy Varga, 2019). It will mainly center around reducing communication overhead in distributed architectures and utilizing edge computing for less latency. Advancements in analytical tool can improve the performance of the game because player and systemic behavior can now be predicted in detail. Also, as pipelines process more private or personal information, the evolution of data protection and security

will continue. This has made it necessary for developers and data scientists to design fast, flexible architectures that could easily adapt to any prevailing change in the game technology field.

VI. Conclusion

In conclusion, complex and high volume telemetry data processing is crucial for large scale gaming platforms, whose telemetry data processing pipelines should be fine-tuned in order to handle the streaming. By incorporating advanced technology and application of machine learning models, platforms will be able to enhance the players' game sessions, enhance the systems performance and also carry out real-time data analytics. Implementing advanced approaches such as advanced data streaming and analytics ensures that processing pipelines are receptiveness to changing conditions preferably in gaming and efficient. In the future, further developments in these procedures will be oriented at meeting the needs of the changing business in the gaming sector, including the questions of scalability, real-time reaction, and data security.

VII. References

[1]. Malitsky, N., Chaudhary, A., Jourdain, S., Cowan, M., O'Leary, P., Hanwell, M. and Van Dam, K.K., 2017, August. Building near-real-time processing pipelines with the Spark-MPI platform. In 2017 New York Scientific Data Summit (NYSDS) (pp. 1-8). IEEE.

[2]. Hasan, M. and Goraya, M.S., 2018. Fault tolerance in cloud computing environment: A systematic survey. Computers in Industry, 99, pp.156-172.

[3]. Soikkeli, E., 2019. Scaling out Big Data Distributed Pricing in Gaming Industry (Master's thesis).

[4]. Singh, M.P., 2016. Distributed approach to analyze physiological time series signals in medical telemetry. Computer Science, 74, p.13.

[5]. Patel, H., Jindal, A. and Szyperski, C., 2019, November. Big data processing at microsoft: Hyper scale, massive complexity, and minimal cost. In Proceedings of the ACM Symposium on Cloud Computing (pp. 490-490).

[6]. Galati, T., 2017. Exploring game analytics solutions for data-driven user research in indie studios (Doctoral dissertation).

[7]. Basak, A., Venkataraman, K., Murphy, R. and Singh, M., 2017. Stream Analytics with Microsoft Azure: Real-time data processing for quick insights using Azure Stream Analytics. Packt Publishing Ltd.

[8]. Dixon, J., 2017. Monitoring with Graphite: Tracking Dynamic Host and Application Metrics at Scale. " O'Reilly Media, Inc.".

[9]. Veith, A.D.S., 2019. Quality of Service Aware Mechanisms for (Re) Configuring Data Stream Processing Applications on Highly Distributed Infrastructure (Doctoral dissertation, Université Rennes 1).

[10]. Callegaro, D., Baidya, S. and Levorato, M., 2019, August. A measurement study on edge computing for autonomous UAVs. In Proceedings of the ACM SIGCOMM 2019 Workshop on Mobile AirGround Edge Computing, Systems, Networks, and Applications (pp. 29-35).

[11]. Ruffy Varga, F., 2019. Data-driven data center traffic control (Doctoral dissertation, University of British Columbia).

[12]. Christidis, A., Davies, R. and Moschoyiannis, S., 2019, November. Serving machine learning workloads in resource constrained environments: A serverless deployment example. In 2019 IEEE 12th Conference on Service-Oriented Computing and Applications (SOCA) (pp. 55-63). IEEE.

[13]. Santhosh Palavesh. (2019). The Role of Open Innovation and Crowdsourcing in Generating New Business Ideas and Concepts. International Journal for Research Publication and Seminar, 10(4), 137–147. https://doi.org/10.36676/jrps.v10.i4.1456

[14]. Vijaya Venkata Sri Rama Bhaskar, Akhil Mittal, Santosh Palavesh, Krishnateja Shiva, Pradeep Etikani. (2020). Regulating AI in Fintech: Balancing Innovation with Consumer Protection. European Economic Letters (EEL), 10(1). https://doi.org/10.52783/eel.v10i1.1810

[15]. Challa, S. S. S. (2020). Assessing the regulatory implications of personalized medicine and the use of biomarkers in drug development and approval. European Chemical Bulletin, 9(4), 134-146.D.O.I10.53555/ecb.v9:i4.17671

[16]. Challa, S. S. S., Tilala, M., Chawda, A. D., & Benke, A. P. (2019). Investigating the use of natural language processing (NLP) techniques in automating the extraction of regulatory requirements from unstructured data sources. Annals of Pharma Research, 7(5), 380-387.

[17]. Challa, S. S. S., Chawda, A. D., Benke, A. P., & Tilala, M. (2020). Evaluating the use of machine learning algorithms in predicting drug-drug interactions and adverse events during the drug development process. NeuroQuantology, 18(12), 176-186. https://doi.org/10.48047/nq.2020.18.12.NQ20252

[18]. Challa, S. S., Tilala, M., Chawda, A. D., & Benke, A. P. (2019). Investigating the use of natural language processing (NLP) techniques in automating the extraction of regulatory requirements from unstructured data sources. Annals of PharmaResearch, 7(5), 380-387.

[19]. Dr. Saloni Sharma, & Ritesh Chaturvedi. (2017). Blockchain Technology in Healthcare Billing: Enhancing Transparency and Security. International Journal for Research Publication and Seminar, 10(2), 106–117. Retrieved from https://jrps.shodhsagar.com/index.php/j/article/view/1475

[20]. Saloni Sharma. (2020). AI-Driven Predictive Modelling for Early Disease Detection and Prevention. International Journal on Recent and Innovation Trends in Computing and Communication, 8(12), 27–36. Retrieved from https://www.ijritcc.org/index.php/ijritcc/article/view/11046

[21]. 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

[22]. Bhaskar, V. V. S. R., Etikani, P., Shiva, K., Choppadandi, A., & Dave, A. (2019). Building explainable AI systems with federated learning on the cloud. Journal of Cloud Computing and Artificial Intelligence, 16(1), 1–14.

[23]. Vijaya Venkata Sri Rama Bhaskar, Akhil Mittal, Santosh Palavesh, Krishnateja Shiva, Pradeep Etikani. (2020). Regulating AI in Fintech: Balancing Innovation with Consumer Protection. European Economic Letters (EEL), 10(1). https://doi.org/10.52783/eel.v10i1.1810

[24]. Dave, A., Etikani, P., Bhaskar, V. V. S. R., & Shiva, K. (2020). Biometric authentication for secure mobile payments. Journal of Mobile Technology and Security, 41(3), 245-259.

[25]. Prasad, N., Narukulla, N., Hajari, V. R., Paripati, L., & Shah, J. (2020). AI-driven data governance framework for cloud-based data analytics. Volume 17, (2), 1551-1561.

[26]. 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

[27]. Secure Federated Learning Framework for Distributed Ai Model Training in Cloud Environments. (2019). International Journal of Open Publication and Exploration, ISSN: 3006-2853, 7(1), 31-39. https://ijope.com/index.php/home/article/view/145

[28]. Challa, S. S. S., Tilala, M., Chawda, A. D., & Benke, A. P. (2019). Investigating the use of natural language processing (NLP) techniques in automating the extraction of regulatory requirements from unstructured data sources. Annals of Pharma Research, 7(5),

[29]. Bhavesh Kataria, Jethva Harikrishna, "Performance Comparison of AODV/DSR On-Demand Routing Protocols for Ad Hoc Networks", International Journal of Scientific Research in Science and Technology, Print ISSN : 2395-6011, Online ISSN : 2395-602X, Volume 1, Issue 1, pp.20-30, March-April-2015. Available at : https://doi.org/10.32628/ijsrst15117

[30]. Bhavesh Kataria, "XML Enabling Homogeneous and Platform Independent Data Exchange in Agricultural Information Systems, International Journal of Scientific Research in Science, Engineering and Technology, Print ISSN : 2395-1990, Online ISSN : 2394-4099, Volume 1, Issue 2, pp.129-133, March-April-2015. Available at : https://doi.org/10.32628/ijsrset152239

[31]. Bhavesh Kataria, Characterization and Identification of Rice Grains Through Digital Image Analysis in International Journal – Sanshodhan, ISSN 0975- 4245, December, 2011 (Print)

[32]. Tilala, M., & Chawda, A. D. (2020). Evaluation of compliance requirements for annual reports in pharmaceutical industries. NeuroQuantology, 18(11), 27.

[33]. Ghavate, N. (2018). An Computer Adaptive Testing Using Rule Based. Asian Journal For Convergence In Technology (AJCT) ISSN -2350-1146, 4(I). Retrieved from http://asianssr.org/index.php/ajct/article/view/443

[34]. Shanbhag, R. R., Dasi, U., Singla, N., Balasubramanian, R., & Benadikar, S. (2020). Overview of cloud computing in the process control industry. International Journal of Computer Science and Mobile Computing, 9(10), 121-146. https://www.ijcsmc.com

[35]. https://ijisae.org/index.php/IJISAE/article/view/6761

[36]. Tripathi, A. (2020). AWS serverless messaging using SQS. IJIRAE: International Journal of Innovative Research in Advanced Engineering, 7(11), 391-393.

[37]. Tripathi, A. (2019). Serverless architecture patterns: Deep dive into event-driven, microservices, and serverless APIs. International Journal of Creative Research Thoughts (IJCRT), 7(3), 234-239. Retrieved from http://www.ijcrt.org

[38]. Bhavesh Kataria, "The Challenges of Utilizing Information Communication Technologies (ICTs) in Agriculture Extension, International Journal of Scientific Research in Science, Engineering and Technology, Print ISSN : 2395-1990, Online ISSN : 2394-4099, Volume 1, Issue 1, pp.380-384, January-February-2015. Available at : https://doi.org/10.32628/ijsrset1511103

[39]. Bhavesh Kataria, Analysis of Rice Grains Through Digital Image Processing, SCI-TECH Research (National Journal) ISSN 0974 – 9780, February, 2012 (Print)

[40]. Thakkar, D. (2020, December). Reimagining curriculum delivery for personalized learning experiences. International Journal of Education, 2(2), 7. Retrieved from https://iaeme.com/Home/article_id/IJE_02_02_003

[41]. Kanchetti, D., Munirathnam, R., & Thakkar, D. (2019). Innovations in workers compensation: XML shredding for external data integration. Journal of Contemporary Scientific Research, 3(8). ISSN (Online) 2209-0142.

[42]. Aravind Reddy Nayani, Alok Gupta, Prassanna Selvaraj, Ravi Kumar Singh, & Harsh Vaidya. (2019). Search and Recommendation Procedure with the Help of Artificial Intelligence. International Journal for Research Publication and Seminar, 10(4), 148–166. https://doi.org/10.36676/jrps.v10.i4.1503

[43]. Vaidya, H., Nayani, A. R., Gupta, A., Selvaraj, P., & Singh, R. K. (2020). Effectiveness and future trends of cloud computing platforms. Tuijin Jishu/Journal of Propulsion Technology, 41(3). Retrieved from https://www.journal-propulsiontech.com

[44]. Rinkesh Gajera , "Leveraging Procore for Improved Collaboration and Communication in Multi-Stakeholder Construction Projects", International Journal of Scientific Research in Civil Engineering (IJSRCE), ISSN : 2456-6667, Volume 3, Issue 3, pp.47-51, May-June.2019

[45]. Gudimetla, S. R., et al. (2015). Mastering Azure AD: Advanced techniques for enterprise identity management. Neuroquantology, 13(1), 158-163. https://doi.org/10.48047/nq.2015.13.1.792

[46]. Bhavesh Kataria, "Role of Information Technology in Agriculture : A Review, International Journal of Scientific Research in Science, Engineering and Technology, Print ISSN : 2395-1990, Online ISSN : 2394-4099, Volume 1, Issue 1, pp.01-03, 2014. Available at : https://doi.org/10.32628/ijsrset141115

[47]. Bhavesh Kataria, "Variant of RSA-Multi prime RSA, International Journal of Scientific Research in Science, Engineering and Technology, Print ISSN : 2395-1990, Online ISSN : 2394-4099, Volume 1, Issue 1, pp.09-11, 2014. Available at https://doi.org/10.32628/ijsrset14113Gudimetla, S. R., & et al. (2015). Beyond the barrier: Advanced strategies for firewall implementation and management. NeuroQuantology, 13(4), 558-565. https://doi.org/10.48047/nq.2015.13.4.876

[48]. Saoji, R., Nuguri, S., Shiva, K., Etikani, P., & Bhaskar, V. V. S. R. (2019). Secure federated learning framework for distributed AI model training in cloud environments. International Journal of Open Publication and Exploration (IJOPE), 7(1), 31. Available online at https://ijope.com.