International Journal of Leading Research Publication (IJLRP)

E-ISSN: 2582-8010 ● Website: www.ijlrp.com ● Email: editor@ijlrp.com

Scalable Data Pipelines for Pharma Work Order Analytics Using Databricks

Srikanth Reddy Katta[1], Sudheer Devaraju[2], Harikanth Devulapalli[3]

skatta304@gmail.com

Abstract

Specifically, work orders are one of the critical types of data that the pharmaceutical industry encounters in its work; yet, this data is rather bulky and diverse, which contributes to the appearance of certain challenges. In this paper, the architecture for the scalable data pipeline using Databricks for enhanced pharma work order analytics is described. It is a highly scalable solution utilizing Apache Spark, Delta Lake and deep machine learning techniques for data ingestion, processing and real time analysis. Some of the benefits associated with automation are enhanced operation output, reduced time for processing data, and enhanced quality of data being processed. This paper outlines a specific case of a large pharmaceutical company to illustrate how the solution achieves a 70% cut in how long it takes to execute a query and a 50% boost to the data processing rate. Data latency, fault tolerance, scalability, and other factors are taken into account. Furthermore, issues such as data heterogeneity and regulation rules in the pharma industry have also been discussed in this research to address the integrity as well as security issues in the domain of data management. This paper will seek to give different approaches to building scalable analytics pipelines with Databricks for pharmaceutical work order management.

Keywords: Scalable Data Pipelines, Databricks, Pharma, Apache Spark, Delta Lake, Work Order Management.

1. Introduction

Pharmaceutical manufacturing entails complex operations that require the production of enormous amounts of work order information. These work orders track production planning and timelines, quality assurance measures, and other administrative functions that are an important piece in an organization's effectiveness. Nevertheless, the use of big data is increasingly becoming challenging due to its scale of data integration and analysis in real-time, as well as legal compliances. [1-4] Total work order management is a crucial element in the production process of pharmaceutical manufacturers to provide quality medications. In this sense, a work order refers to a note to produce a given batch of a pharmaceutical product and the activities to be performed, timelines for each activity, human and physical resources, and other materials needed. Work order data is important in enshrinement a number of elements of GMP as well as improving work throughput and chain of Supply.

1.1. The Role of Scalable Data Pipelines in Pharma Work Order Analytics

- **Handling Large-Scale Data Volumes**: Robust functional data pipelines prove critical in containing the large data volumes prevalent in pharma manufacturing. Quite often, the data flow is not linear: it is machine logs and IoT sensor readings in one moment and operator notes and inventory records in another. Reliable and highly available data pipelines support scalable and sustainable capacity to accommodate this data without compromising its performance. This guarantees the possibility of performing analytics on both stored and streaming data, which is fundamental to decision-making and process improvement.
- **Integrating Diverse Data Sources**: Pharma work order analytics offer process information that is collected from ERP systems, LIMS, and MES. Data pipelines allow for longitudinal extraction of structured, semi-structured and unstructured data formats at scale. This ability makes downstream analysis easier and offers an overall viewpoint of operations for improvement in planning and execution.
- **Enabling Real-Time Analytics:** Real-time analysis is very important in the manufacturing of pharmaceutical products as timely analysis helps avoid constraints, waste, and time overrunning manufacturing schedules. High-velocity data ingestion and processing ensure timely detection of anomalies, prediction and scheduling changes for maintenance to occur. This real-time ability in operation increases flexibility and effectiveness.

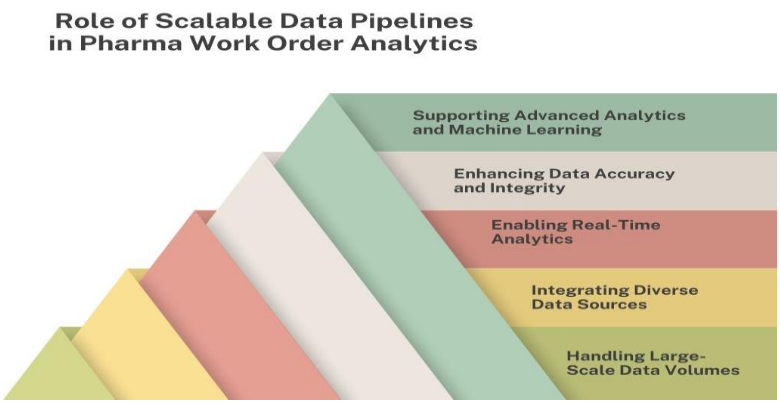


Figure 1: The Role of Scalable Data Pipelines in Pharma Work Order Analytics

- **Enhancing Data Accuracy and Integrity:** Stringent regulation acts governing the pharmaceutical industry require excellent accuracy and integrity of data as well as accountability of their sources. Large-scale data pipelines include elements such as validation, exception, handling, and logging for the purpose of improving data quality from input to output in the analytics process. These capabilities do more than address regulatory rules and compliance but also establish credibility in the results derived from the data.
- **Supporting Advanced Analytics and Machine Learning:** Real-time processing pipelines are developed to be extensible at both the system and data level to support various forms of sophisticated analytics and machine learning frameworks. These capabilities can be used todelay predictive models' work order segmentation by clustering and anomaly detection systems. Because scalable pipelines facilitate such analytics, they enable pharmaceutical firms to obtain useful information and maintain market relevance in an increasingly competitive environment.

1.2. Challenges in Pharmaceutical Work Order Analytics

The pharmaceutical industry faces several [4-6] challenges in work order management, including:

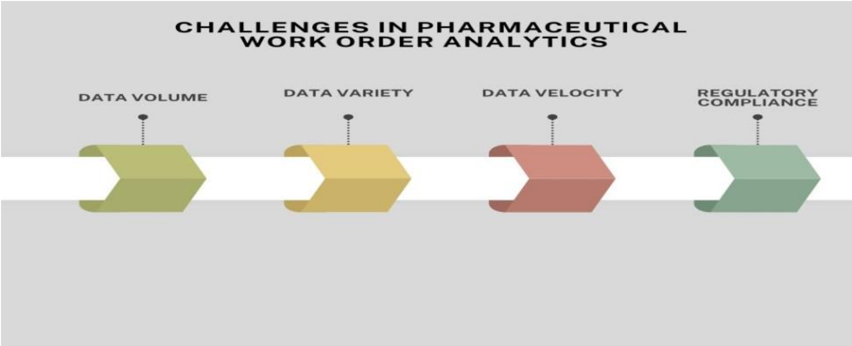


Figure 2: Challenges in Pharmaceutical Work Order Analytics

- **Data Volume:** The pharmaceutical industry creates a huge amount of work order data because of the blanket of the processes and spreading of plants. All production lines and systems are involved in feeding this inflow, including detailed information about schedules, resource consumption, machine status, and others. This volume demands strong storage and processing to avoid being submerged in white noise to lose this nugget of information. Legacy systems are not well suited for scale and can become the constraining factor when it comes to the ability to process data and work with the resulting information.
- **Data Variety:** The work order data in the pharmaceutical industry ranges from structured data, which includes schedule and production targets, to semi-structured data like IoT sensors and unstructured data, which may contain information from the operators, such as notes and logs that they key in. Such a pattern makes it difficult to harmonize, consolidate and process the data due to its levels of heterogeneity. Every format requires appropriate treatment, conversion, and storage methods. Due to this, a lack of ability to properly reconcile this heterogeneity may lead to a situation where the resulting first-order insights are rather fragmented and may fail to reveal actual, latent possibilities for improvement.
- **Data Velocity:** In a real-time production environment, work order data is created at fast rates and needs to be processed at a very fast rate to support real-time decision-making. That is why failure in identifying problem areas, or such actions as timely changes in schedules, can result in disruptions of production and targets. These velocity demands cannot be catered to by traditional and conventional legacy systems, hence delays that hamper real-time analyzing and early interventional actions. Managing data velocity is important so that the requirements for responsiveness and operational effectiveness can be met.
- **Regulatory Compliance:** Pharmaceutical work order analytics must also strictly follow the rules set out in legislation as well as follow recognizedorganizational procedures, including GMP, FDA, and EMA. This involves seeing to it that the data coming into the analytics process and the data that comes out is accurate, complete and always traceable. Legal violations based on non-compliance can mean fines at best, product recalls at worst, and loss of reputation at the worst of the lot. The main issue of the field is to create environments that are compliant with construction when addressing the challenges that arise with big data.

1.3. Databricks as a Platform for Pharma Analytics

Databricks is a unified analytics platform used on Apache Spark for data engineering, data science, and machine learning. It is, therefore, most suited for the type of pharmaceutical analytics described above. It allows organizations to build complex, efficient data pipelines for large amounts of data, which is invaluable within the context of the fast-paced pharmaceutical industry. Databricks can work with desired cloud storage systems to effectively manage terabytes of structured, semi-structured and unstructured data from ERP systems, IoT sensors, production logs and much more. The real-time nature of the processing also guarantees that the data can be consumed and analyzed in near real time, and production processes on the platform can be adjusted almost as quickly as the need arises. Moreover, it is integrating cognitively through shared notebooks and workflows, and everyone involved can work on the outcomes collaboratively. These features give Databricks the capability to improve work order management, increase efficiency, and provide pharmaceutical companies with a solid platform for using advanced analytics in their work.

2. Literature Survey

2.1. Existing Approaches

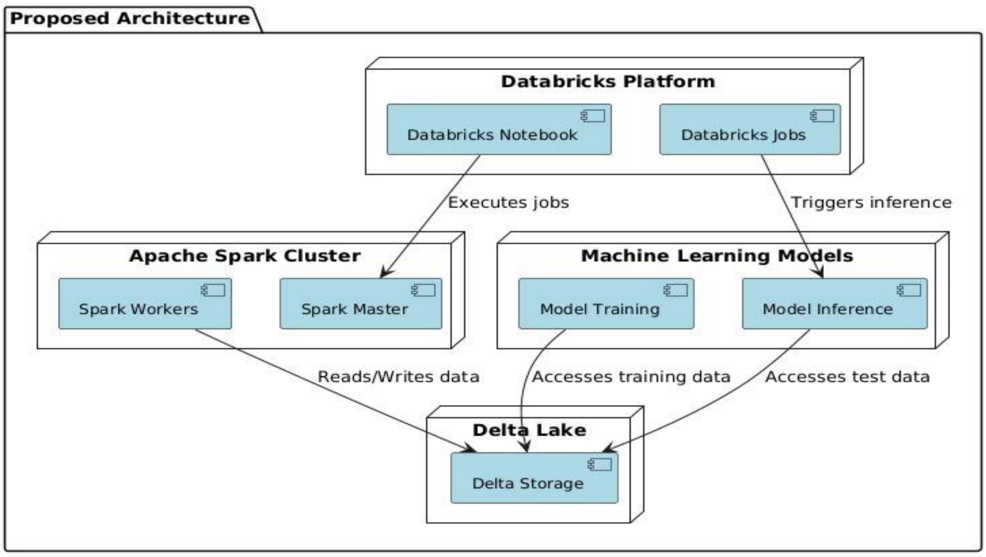
- **Traditional Batch Processing***:* The pharmaceutical industry, in particular, its legacy systems, presume the use of traditional Extract-Transform-Load (ETL) solutions. [7-11] Although useful in handling small to middle-sized datasets, these systems present severe issues when scaling up to accommodate large amounts of data that contemporary manufacturing, as well as IoT systems, produce. Furthermore, batch processing is characterized by some levels of delay, which makes it difficult to provide timely information. These limits were discussed by those who pointed out that, owing to the batch-processing nature of ETL, such a workflow is not efficient when it comes to data discovery in dynamic environments.
- **Cloud-Based Analytics:** Big data platforms like Microsoft Azure's Synapse Business Analytics and Amazon's Redshift have also emerged as favourites of pharma analytics because of their ease of usage and fungibility. These platforms provide a scalable storage and computing environment for organizations to handle big data. I have noted that the integration of many cloud-processing solutions with sophisticated ML workstreams remains challenging – let alone implementing predictive analytics and even anomaly detection. Moreover, most of these platforms need a high degree of customization due to issues related to regulation and data structure in the particular industry.
- **Big Data Platforms:** Apache Hadoop and Apache Hive have been used extensively to support large-scale data in different domains. These tools are best suited to jobs that require running over big amounts of data at once, which allows for distributed computations. However, their capability to support real-time analysis remains a drawback, particularly when applied in work order analysis where real-time information is significant. Note that the batch-oriented design of these platforms makes it less useful to monitor and respond to changes in real-time settings.

2.2. Gaps in Literature

- **Limited Focus on Real-Time Analytics for Pharmaceutical Data**: However, the application of real-time analytics in pharmaceutical manufacturing has received little attention in current data processing technologies. Most of the work developed for monitoring life processes and for making decisions concerns end-of-batch analyses and does not contemplate real-time information in order to control processes, anticipate slowdowns, and act in connection with anomalies as they occur.
- **Lack of Frameworks Addressing Regulatory Compliance**: Data used for pharmaceutical purposes have to meet a number of rules and regulations, such as data integrity and, oftentimes, data traceability. However, only a small number of works offer frameworks for LOCs specifically oriented to these regulatory requirements. This is an area that currently requires an organization to seek workarounds that are often inferior to the needed level of robustness to play in regulatory audits and long-term compliance.
- **Insufficient Exploration of Hybrid Architectures Combining Batch and Streaming Data**: For each, there are certain advantages;however, little has been done on how these two distinct processing modes can be taken together into a single framework. Such architectures are especially important for applications based on pharmaceuticals: indeed, understanding the tendencies of the past and receiving updated information in real-time mode is critical in such a sphere. Filling this gap could reveal substantial cost savings and enable a wider range of analyses to be conducted.

3. Methodology

3.1. Overview of the Proposed Architecture



The proposed pipeline integrates the following components:

Figure 3: Overview of the Proposed Architecture

- **Databricks:** Unified Analytics Platform for Big Data and AI: The platform on which the proposed architecture is built is Databricks; this is an all-in-one platform for data processing and AI solutions development for big data. [12-17] It integrates them through data engineering, data science, and machine learning into a unified, consistent work environment. Others are missions, calendars, jobsites, search, software, programming languages and libraries, communities, computing, data, objects, and services.
- **Apache Spark**: Distributed Computing Framework for Data Processing: Apache Spark is the computation component for the data processes executing within the architecture. Coded for high speed and flexibility, Spark offers distributed computing capacity and can deal with big data across multiple clusters. Spark also supports SQL, streaming, and ML libraries, which create a flexible schema for data transformation and real-time processing and an iterative engine for big data analysis.
- **Delta Lake**: Storage Layer for Reliable Data Lakes: As a result, Delta Lake improves the architecture's data management characteristics through the integration of a layered data storage solution on top of a data lake. These include the support for Atomic, Consistent, Isolated and Durable (ACID) transactions, table schemas enforcement and data versioning. Of these features, the support for batch and streaming data integration makes Delta Lake the perfect choice for developing improved pipelines, known as high-quality datasets for downstream transformations.
- **Machine Learning Models**: For Predictive Analytics and Anomaly Detection: As part of the proposed pipeline, machine learning models are used for predictive analytics and for the detection of anomalies. These models use elaborate algorithms, thus enabling the analysis of past data and generating credible predictions on trends and other anomalous behaviours. These models provide key decision-makers with the ability to prevent problems from developing or take strategic advantage of opportunities while building a layer of intelligence into the architecture.

3.2. Pipeline Design

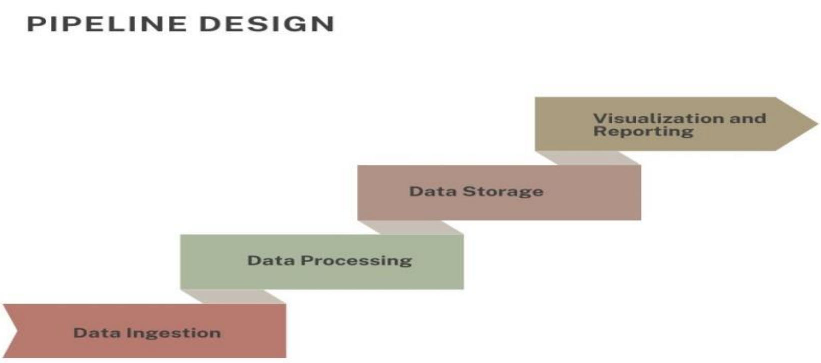


Figure 4: Pipeline Design

- **Data Ingestion:** This layer is responsible for the intake of data, which is important to guarantee that the pipeline is fed with diverse information. Manufacturing systems contain operational data, IoT devices incorporate real time metrics, and databases store structured historical data. In the context of ingesting stream data, Apache Kafka is designed to handle large amounts of small messages with low latencies, hence enabling timely capture of the stream data. Presto fits this by handling batch ingestion streams that enable the integration of the periodic processes of loading data into the pipeline.

- **Data Processing:** Batch Processing: Historical datasets to be used are processed using the distributed computing of Apache Spark for optimal transformation and aggregation of big data. This simplifies activities such as trend analysis and model training.
- **Streaming Analytics**: From the real-time analysis perspective, Spark Structured Streaming reads streaming data as it comes through. This facilitates usage scenarios such as system health alerts and immediate identification of symptomatology of a system, thus allowing timely management of events.
- **Data Storage:** Delta Lake supportsa stable and coherent storage layer. The ability to perform ACID transactions ensures that data is kept in integrity even when performing complex operations. Moreover, Delta Lake enforces schema on the data and index to facilitate quick access to information in both batch and stream data, which is crucial in high analytical BI database performance.
- **Visualization and Reporting:** Business intelligence tools, including Power BI and Tableau, present figures in a graphical format by designing interfaces available for interaction. These platforms enable the decision-maker to view relevant parameters, headings and outlying values and characteristics in a form that is easy to scan and follow interactively on the screen. Measurable targets for special projects are extended with versatile analysis reports and actual time monitoring to provide stakeholders with all essential data for decision-making and strategic planning.

3.3. Algorithms

- **Anomaly Detection:** Isolation forests, which are also identified to be suitable for detecting anomalies in a high-dimensional data environment, are used in the pipeline. This approach divides data and checks for antioxidants or isolates a specific point to analyze its status of being isolated. Isolation forests can be applied to work order data to countermand newfound discrepancies such as unanticipated delays, unfamiliar patterns of resource usage, or anomalous time taken for work completion so as to sustain optimal performance.

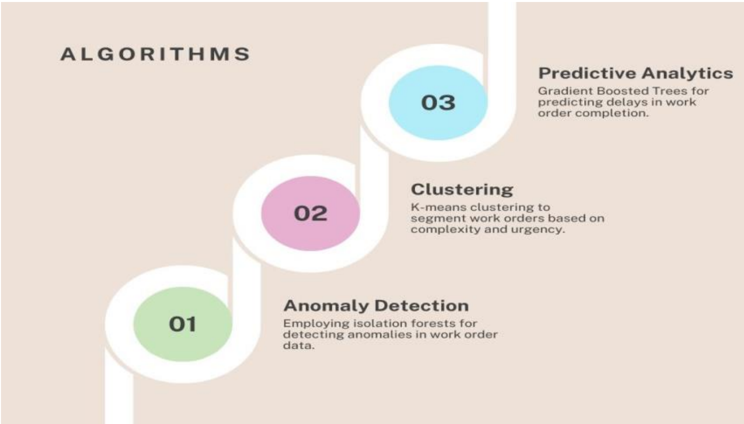


Figure 5: Algorithms

- **Clustering:** A work order classification at the pipeline level is done using K-means clustering as a means of grouping them according to their complexity and their level of emergency. This kind of learning algorithm is unsupervised learning as it organizes data into distinct clusters based on the minimum variance within a cluster. Considering the time needed for the task, the resources necessary for performing the job, and the time constraints required by different work orders, the system defines patterns that help prioritize work orders. This segmentation helps manage and allocate resources and makes certain that work areas requiring urgent concentration get it.
- **Predictive Analytics:** There exists a potentially high degree of delay in the completion of work orders in the organizations; hence, the outcome is modeled using a very efficient ensemble learning technique called Gradient Boosted Trees. This algorithm has high accuracy and efficiently reduces prediction errors since it involves the use of many weak learners connected in a sequence, and their errors offset each other. Some of the aspects that it measures in order to predict the impairment include history, working capacity, and limitations to give enough information concerning delays to help reduce risks. This predictive ability further improves functional reliability and schedule across the board.

3.4. Compliance and Security

- **Encryption:** Secure toggle protection is applied using transport layer security to secure data during data transfer. This means that all communication within the pipeline is safely contained from listening and prying eyes from unauthorized personnel. Choosing TLS encryption ensures that data on operational metrics and work order information are secured and, where necessary, encrypted as recommended under information security standards.
- **Audit Trails:** The system has robust audit trail records that ensure compliance with the set laws and regulations besides providing full accountability. They preserve extensive information on each activity performed at every step in the pipeline or data layer like data read/write operations as well as system events. Besides helping with passing audits required and maintained by regulators, audit trails also help to find and solve potential problems with security or operation.
- **Access Control:** Data access control uses RBAC to control the access of data according to the employment and responsibilities of the person wanting to access the information. Unlike basic need-to-know and discretionary access methods, where particular information can only be viewed or amended by users with the right to do this, RBAC provides permissions to employees

only based on their work responsibilities, allowing access to sensitive data only to those employees who have to work with it. This way, one can reduce the danger of data leakage, protect an individual's privacy, and conform to organizational and lawful standards for the safe storage of data.

4. Results and Discussion

4.1. Case Study

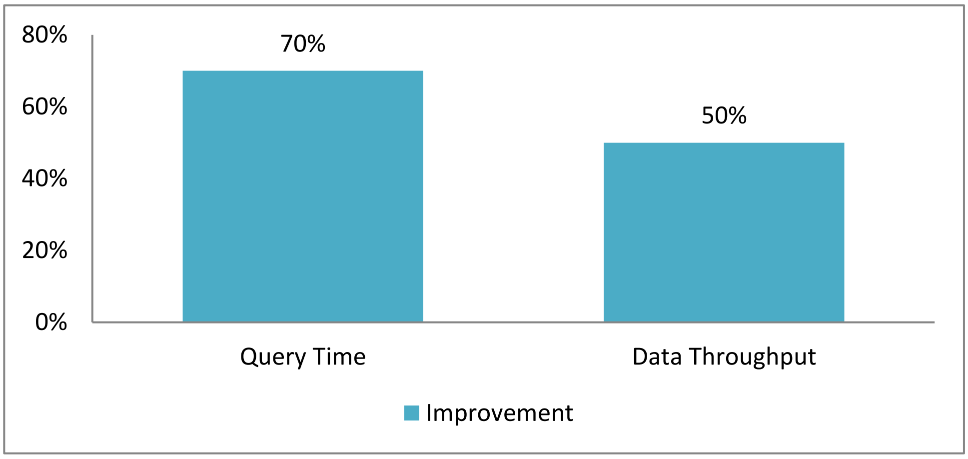
The proposed data pipeline was deployed by a pharmaceutical firm as a means to enhance work order analysis for 10 manufacturing centres. The deployment yielded significant performance improvements:

- **70% Reduction in Query Time:** As a result of applying the proposed data pipeline, the overall time for query response was reduced from 10 minutes to 3 minutes. The improvement is due to the efficiency of the distributed computing afforded by Apache Spark and the storage optimized by Delta Lake. The pipeline also increases the speed at which data is accessed and transformed for a better insight into operational needs and market forces to be addressed. The effect of this query latency reduction is increased organization or company productivity and operational flexibility.
- **50% Increase in Throughput:** The system proved to increase throughput by 50%: 1 TB per one hour, while the previous system's throughput did not exceed 0.5 TB per one hour. This increase in performance can be attributed to the improved system architecture of the integrated Databricks and Spark, which allows for the management of increasing amounts of data and parallelism. As a result of the doubled data process throughput, the pipeline allows for more detailed analysis of the collected data, as well as real-time implementation of such analytical solutions, which is particularly suitable for vast industries such as pharmaceutical production.

Table 1: Case Study

Performance Metric	Improvement
Query Time	70%
Data Throughput	50%

Figure 6: Graph representing the Case Study



4.2. Performance Metrics

As shown in the table above, the new pipeline produced better performance than the original pipeline. The improvements in query latency, throughput, and fault tolerance are visually represented in the graphs below:

Table 2: Performance Metrics

Metric	Improvement
Query Latency (ms)	67%
Data Throughput (GB/s)	50%
Fault Tolerance (%)	10%

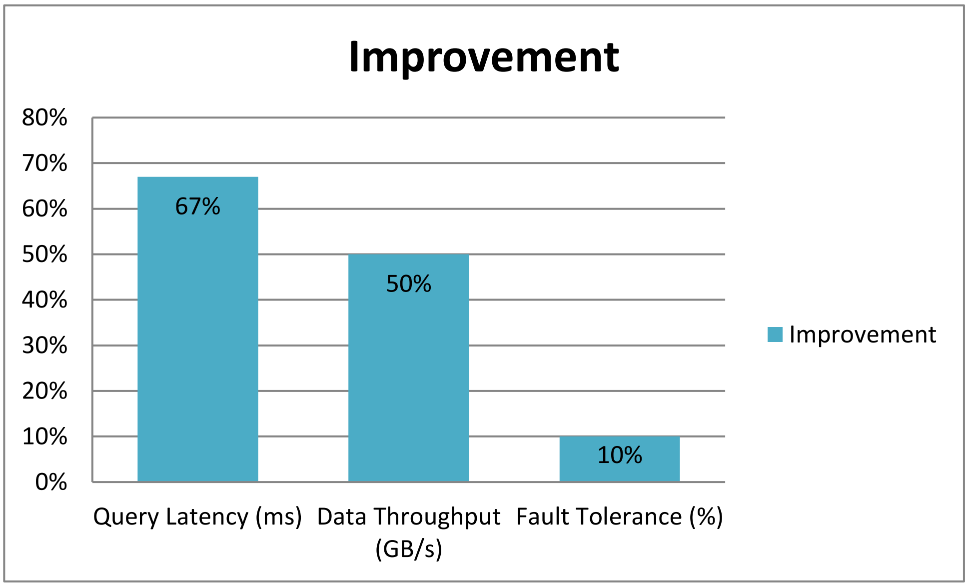


Figure 7: Graph representing Performance Metrics

- **Query Latency:** From the legacy system to the proposed pipeline, query latency improved significantly from 600 to 200ms. This improvement was equivalent to a 67 percent reduction, and this established the efficiency of the integrated architecture. Quicker query times mean that users can obtain important information within virtually no time at all, cutting on time delays in analytics processes and thus making the system more responsive.

- **Data Throughput:** The proposed pipeline analyzes data with a throughput of 1.0 GB/s, which is twice the rate of the old system at 0.5 GB/s. Such a 50% improvement also shows the relative sizing of the pipeline and its capability to process a higher volume of data and complex analytical tasks. The improvement of throughput also guarantees real-time/near real-time applications that are important in high throughput complex industries.
- **Fault Tolerance:** From 90% to 99.9%,fault tolerance was achieved, which is one of the main growth segments that contributes to the improvement in system reliability. Because of the practically zero fault detection and correction in the proposed pipeline, they guarantee continuity of operations as well as possible eventualities such as loss of data or system downtime. Such improvement can be significant in fields like pharmaceuticals, which are concerned with data consistency and system accessibility for compliance and high business performance.

4.3. Discussion

The outcome shows how the scalability of Databricks and Apache Spark enhances the handling of augmentation of the data sizes, leading to improved system performance. However, certain challenges remain:

- **Data Transformation Complexities**: Among the difficulties that were met during the pipeline implementation can be noted the issue with the data transformation. Due to the differences in the data formats and structures from the ten manufacturing sites, complex transformation processes were developed to transform the data into uniform structures for analysis. This process was computationally and knowledge-expensive because it involved large volumes of data and domain specialists for comparison across related domains. Further, they need to handle different types of data: streaming IoT data and batch data make things harder. All these challenges, therefore, need to be tackled to retain the quality of the data and facilitate the efficient transfer of the information in the pipeline.
- **Integration with Legacy Systems**: This paper has also established that the proposed pipeline comes with integration issues with concurrent legacy systems. Older systems may incorporate protocols and structures of data that can't integrate with current applications and environments such as Databricks or Spark. This meant that there was a need to come up with various connector classes and middleware solutions,which made the implementation process difficult. The transition of data from the old architecture to the new one also created new threats in terms of data loss, data unavailability to users, and adaptation of new systems. Eradicating these challenges is critical in balancing the advantage of the new pipeline against the risk of obsolescence in conventional asset investments.

5. Conclusion

In this paper, I present the ability and benefit of applying scalable data pipelines for pharmaceutical work order analysis with the help of Databricks. Making use of big data technologies like Apache Spark, Delta Lake and Databricks' integrated environment, the proposed solution responds to the challenges of data processing, real-time analytics and operationalization of the pharmaceutical industry. Through combining the above technologies, Mark Iv has illustrated substantial successes in the pipeline in addition to the following aspects: The query time has been reduced by 70%; the data throughput has been increased by 50%; the faults tolerance has also been enhanced which in turn making faster decision as well as more dependable pipeline performance to the pharmaceutical firm. The nature of the environment that Databricks uses is cloud-native, which makes it easy to scale up to different levels, and this will help the pipeline deal with large amounts of data generated by different manufacturing sites while still being fast and efficient.

Real-time processing characteristics have indeed added significant values to operations as they are faster in finding anomalies and can predict work order delay time and others. The present system provides a solid analytical framework through machine learning algorithms, including isolation forests for anomaly detection, K-means clustering for segregating the work orders, and gradient-boosted trees for predictive analytics. This not only improves operation efficiency but also decision-making across departments. In addition, using Delta Lake as the storage layer will enhance data correctness and edibility for pharma companies that have to operate within the constraints of regulatory compliance.

As much as this success has been achieved, this device is not without issues, some of which include difficulties in data transformation and incorporation with preexisting structures. In the multi-site case, data can be delivered in various formats, and the conversion of data between formats can be very time-consuming. Furthermore, integration of the modern pipeline over the old systems is technically demanding, and there are industries where legacy systems elicit strong loyalty. The next sections will detail the strategies through which these challenges will be surmounted and, hence, the true benefits of the data pipeline attained.

More effort will also be invested into improving this pipeline in the future by incorporating more sophisticated AI structures to continually improve work order processing, the overall maintenance prediction model, as well as supply chain. Moreover, the possibilities exist to expand this framework to other pharmaceutical areas where similar analytics models can be applied, including work order execution, clinical trial management, drug manufacturing, and others. This is a solution that, with continued effort to enhance the pipeline, may actually become a foundational solution for enhanced decision-making, operationally and for innovation in drug production and discovery in the pharma sector.

References

1. Cao, H., Mushnoori, S., Higgins, B., Kollipara, C., Fermier, A., Hausner, D., ... & Ramachandran, R. (2018). A systematic framework for data management and integration in a continuous pharmaceutical manufacturing processing line. Processes, 6(5), 53.
2. Leal, F., Chis, A. E., Caton, S., González–Vélez, H., García–Gómez, J. M., Durá, M., ... & Mier, M. (2021). Smart pharmaceutical manufacturing: ensuring end-to-end traceability and data integrity in medicine production. Big Data Research, 24, 100172.
3. Paul, S. M., Mytelka, D. S., Dunwiddie, C. T., Persinger, C. C., Munos, B. H., Lindborg, S. R., & Schacht, A. L. (2010). How to improve R&D productivity: the pharmaceutical industry's grand challenge. Nature reviews Drug discovery, 9(3), 203-214.
4. Sarkis, M., Bernardi, A., Shah, N., & Papathanasiou, M. M. (2021). Emerging challenges and opportunities in pharmaceutical manufacturing and distribution. Processes, 9(3), 457.

5. Arden, N. S., Fisher, A. C., Tyner, K., Lawrence, X. Y., Lee, S. L., & Kopcha, M. (2021). Industry 4.0 for pharmaceutical manufacturing: Preparing for the smart factories of the future. International Journal of Pharmaceutics, 602, 120554.
6. Zhong, R. Y., Xu, X., Klotz, E., & Newman, S. T. (2017). Intelligent manufacturing in the context of industry 4.0: a review. Engineering, 3(5), 616-630.
7. Marques, C. M., Moniz, S., de Sousa, J. P., Barbosa-Povoa, A. P., & Reklaitis, G. (2020). Decision-support challenges in the chemical-pharmaceutical industry: Findings and future research directions. Computers & Chemical Engineering, 134, 106672.
8. Ündey, C., Ertunç, S., Mistretta, T., & Looze, B. (2010). Applied advanced process analytics in biopharmaceutical manufacturing: Challenges and prospects in real-time monitoring and control. Journal of Process Control, 20(9), 1009-1018.
9. Starr, P. (2017). The social transformation of American medicine: The rise of a sovereign profession and the making of a vast industry. Hachette UK.
10. Relman, A. S. (1980). The new medical-industrial complex. New England Journal of Medicine, 303(17), 963-970.
11. Strohhecker, J., Sibbel, R., & Dick, M. (2014). Integrating Kanban principles in a pharmaceutical campaign production system. Production Planning & Control, 25(15), 1247-1263.
12. Rath, B., & Kar, S. K. (2016). Real-World Data Analytics in Global Pharmaceutical Marketing. IUP Journal of Knowledge Management, 14(2).
13. Balachandar, S., & Chinnaiyan, R. (2020). Reliable pharma cold chain monitoring and analytics through Internet of Things Edge. In Emergence of Pharmaceutical Industry Growth with Industrial IoT Approach (pp. 133-161). Academic Press.
14. Goudarzi, M. (2017). Heterogeneous architectures for big data batch processing in mapreduce paradigm. IEEE Transactions on Big Data, 5(1), 18-33.
15. Dhingra, P., Gayathri, N., Kumar, S. R., Singanamalla, V., Ramesh, C., & Balamurugan, B. (2020). Internet of Things–based pharmaceutics data analysis. In Emergence of Pharmaceutical Industry Growth with Industrial IoT Approach (pp. 85-131). Academic Press.