International Journal of Research Publications in Engineering, Technology and Management (IJRPETM)

|www.ijrpetm.com | ISSN: 2454-7875 | editor@ijrpetm.com |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 5, Issue 6, November - December 2022||

DOI:10.15662/IJRPETM.2022.0506022

Federated Learning BI across Multi-Cloud Data Silos

Aditi Namdeo

Independent Researcher, Northeastern University, Boston, USA

ABSTRACT: The enterprise data is sensitive, huge and spans across various cloud environments and quickly becoming a critical data source for Business Intelligence (BI) solutions. But, the need for a single data integration creates issues with data protection, data privacy, compliance, data ownership, latency and inter-cloud interoperability with traditional BI solutions. In the multi-cloud world with multiple and distributed data silos, a federated learning based BI framework is introduced to attain a collaborative analysis of these data silos without physically moving the raw data on the local data silos. The proposed system relies on a federation of local data repositories easily developed in the cloud (like Storage Service), a federation of models trained, a secure aggregation of the data involved and a layer of privacy-preserving computation, offering also a single BI visualization layer. This architecture may provide a way for models of analytics to be trained locally (using only data from each cloud) and for further shadowing (more specifically, transport) of the parameters of such models / information of analytics to the central co-ordination server but in an encrypted format. The global model is summarized and sent back to the Cloud nodes involved in the summarization, and can be utilized overtime for obtaining new BI insights without any data pollution. This framework enables customers' decision making at all clouds, customer analytics, financial forecasting and supply chain intelligence, healthcare reporting and enterprise risk monitoring. A variety of serious problems with centralized analytics, such as data duplication, compliance risks, bandwidth costs and vendor solutions can be overcome by combining federated learning and BI workflow. These include aspects of architecture, processes, security aspects, challenges and opportunities for the future intelligent multi-clouds BI ecosystem. The hinted solution path indicates how federated learning can transform the current landscape of BI – from producing and sending reports to a centralized data repository towards a distributed intelligence paradigm with privacy protection and scalability for today's information-driven enterprises.

KEYWORDS: Federated Learning, Business Intelligence, Multi-Cloud Computing, Data Silos, Privacy Preserving Analytics, Secure Aggregation and Distributed Machine Learning.

I. INTRODUCTION

Data-driven insights for planning, forecasting, performance monitoring, customer analysis, risk assessment and optimization of operations has transformed Business Intelligence (BI) to become a key component of an enterprise decision making process. The traditional BI (Business Intelligence) have been more or less handling the centralized data warehouse, moving data from various departments, applications and business unit's into the data warehouse for analysis and reporting in client's reporting dashboards through this ETL movement. But, today, with the advent of cloud computing, digital platforms, Internet of Things systems, enterprise applications and Artificial Intelligence, organizational data management is different from before. No longer in a single controlled environment do data reside. Rather, it becomes disaggregated and spread out among the myriad of cloud providers, regional computer data centres, software-as-a-service (SaaS) and ubiquitous edge and distributed systems and departmental repositories [1]. Moreover, the situation with multi-cloud data is not different – it remains stuck in other types of data silos, and the data in these silos is also inconsistent and not accurate – this has also created a problem for BI.

Cloud Computing is a paradigm that incorporates different kinds of resources, services or applications easy to expand with respect to requirements without big efforts in management. Cloud Computing is a concept built on the first principle, according to National Institute of Standards and Technology (NIST), as "Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (that provide networks, servers, storage, applications and services) that can be rapidly provisioned and released with minimal management effort or interaction with the provider. Many companies have begun to take the step of Multiple Cloud Providers (MCP) to limit their reliance on a single cloud provider, provide geographical redundancies, provide a level of resilience, provide legal compliance and save money. While multi-cloud instantaneity and scalable is great, it also

|www.ijrpetm.com | ISSN: 2454-7875 | editor@ijrpetm.com |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 5, Issue 6, November - December 2022||

DOI:10.15662/IJRPETM.2022.0506022

exacerbates fragmentation of data. It can reside in Amazon Web services, Microsoft Azure, Google Cloud, private cloud and sector-specific platforms, accommodate a variety of data types, follow different rules and regulations and be secure and operated under different security policies or access-control environments. As a consequence, the BI teams have a hard time decreasing it to a single vision of the organisation's performance [4] [5].

In particular, when discussing BI, Data Silos may be very relevant, because data is variable, and, if they don't always have the data, they cannot make appropriate decisions. The issue with data silos is they exist in multiple teams, in multiple systems and, in many cases, on a per-business-unit basis and don't offer enterprise visibility and, in turn, analytic benefits [6]. These 'silos' can also help to create duplication and report structure, as well as lack of or fragmented governance and a lack of trust in the decision making process. The problems are even more complex if you have a multi-cloud world with potentially costly slow technically and/or compliance driven cloud-to-cloud data transfers. For instance, a cloud with patient information, a separate cloud with diagnostic information and a regulated repository with insurance information – all in a single healthcare organization. Multinational banks could even use different clouds in different regions to store financial-transaction data to comply with laws of each region. Centralized BI pipelines with the need to move raw data can pose an increased risk for privacy, regulatory exposure, latency and operation costs in such scenarios [7] [8].

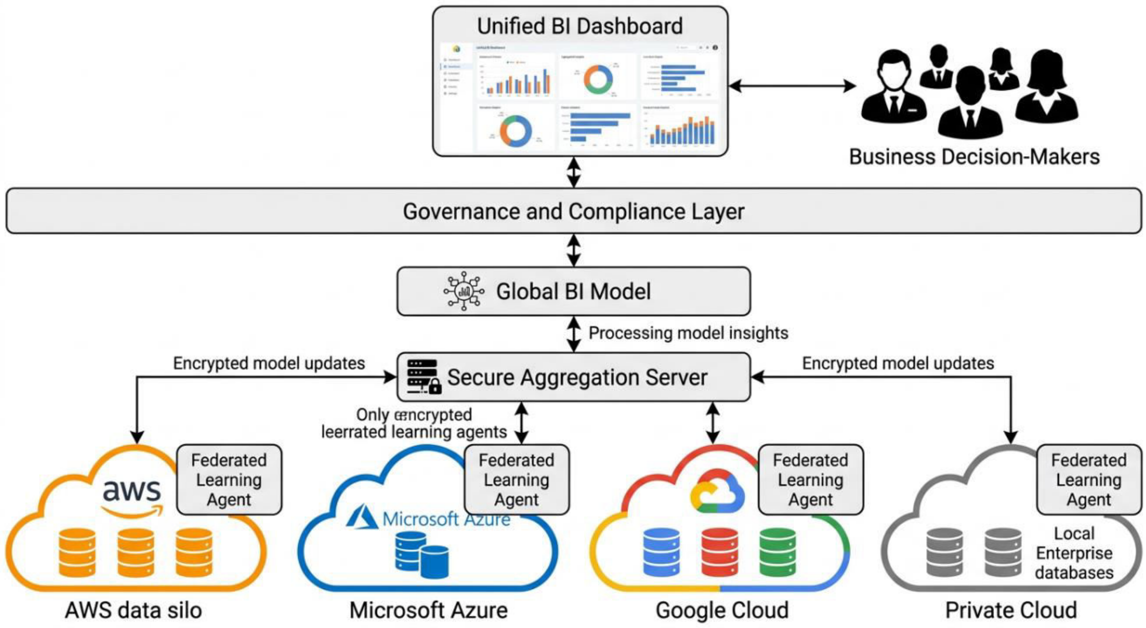

Fig. 1: Federated Learning BI Architecture Across Multi-Cloud Data Silos

By using some alternative approaches, this can be fixed, for example: Federated Learning (FL): This can be used for a model trained jointly without moving the raw data off-site. This involves an off-centralized machine learning model, which is achieved by storing the samples at the client device/institution and computing the updates by the participating client devices. While each participating cloud environment trains a local model on a specific subset of data, it only sends to the central mechanism parameters of the local model, gradients and/or encrypted updates. These are then aggregated and are passed forward to an 'aggregated global model' and pass back to the local node participants to be locally enriched. This makes FL very relevant when it comes to multi-cloud BI solutions as it allows an organization to have its knowledge in wooden piles of data and have a firmer hold on data ownership, privacy and regulation compliance [9].

Not just because of privacy issues, Federated Learning has its own promotion going, too. Plus it is efficient and scalable for communication and it's a large area for distributed analytics. Some recent studies have shown that federated training can be able to enable reduction in the number of communication rounds as it relates to standardized stochastic gradient descent (SGD) and thus can be significantly more viable in terms of learning over distributed environments at the early stages of FL. More recent research has also explored these notions of 'structured' updates, 'sketched' updates, compression and aggregation which might help in the federated world to lower the communication costs. Important for

|www.ijrpetm.com | ISSN: 2454-7875 | editor@ijrpetm.com |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 5, Issue 6, November - December 2022||

DOI:10.15662/IJRPETM.2022.0506022

BI in multi-cloud scenario, where a potential high amount of data may be moved around, adding data duplication, bandwidth as well as report turnaround time. Federated learning can help users render faster and more securely and scale analytics processing workflows without having to move the data to a centralized BI repository.

Why stop at the old reporting architecture though, when BI can enable a lot more than making an intelligent and distributed analytics architecture and revolutionising the traditional reporting architecture via federated learning? Historical trend analysis, dashboards and Key Performance Indicators (KPIs) that seek to report descriptive, or diagnostic, analysis are not paid attention to in the traditional BI. In the world of today's BI, however, a growing number of predictive/prescriptive types of analytics have been added, in conjunction with machine learning models. The Predictive BI models can contribute to this trend – they can learn from various kinds of data in a distributed set of clouds without spending any time centralising sensitive data, thanks to federated learning. A retail business, for instance, collecting sales data in different regions can utilize the sales data in various cloud regions as training sets to create demand forecasting models. Basically, without depending on patient level data within medical institutions in a hospital network, hospital-wide risk models can be developed for readmission. Both reduced fraud and compliance with data-localization laws can be achieved by a financial organisation, if it is able to learn pattern at multiple branches or subsidiaries.

The set of proposed research manuscript is aim at designing a framework, related to the Federated Learning- Business intelligence architecture and the target for this is the Multi-cloud Data Silos. It is made up of a set of key elements – distributed cloud data sources, local model training engines, privacy-preserving computation, secure aggregation, model coordination lying within the metropolitan area, metadata governance, and a single BI visualization layer. There are certain considerations to take note of when setting up the federated learning model using a collection of some federated learning silos which are detailed in Google Cloud's federated learning architecture guidance such as: Defining a federated learning consortium, Federated learning model and roles of the participants. This obviously has an impact on enterprise BI; getting these BI systems to interoperate is important, but so is trust, accountability and interpretability of the data within the cloud environments.

A federated BI framework should also provide a sound security measure. Model update(s) may be at a local level when the data sources are also at a local level, which could risk the security of the sensitive data. Implementing a federated system is necessary for having a trustworthy BI system and so it is required to have some features such as aggregation, encryption, differential privacy, access control along with audit logging and policy based governance. Intel says predictive model creation can span multiple data sources and don't even have to impact the underlying data sets, simply by “confidentializing” parameters of a predictive model. This is the main premise of the current article – the need for multi-cloud data silo BI shouldn't be solely reliant on data centralization – it is better to ensure privacy preserving distributed intelligence.

But, hurdles need to be overcome in federated learning with BI. But, with the multi-cloud models there can be differences regarding data scheme level, cloud-based services, IDM, network practicality, and storage format / compliance requirement: In federated models, there can be non-independent and non-identically distributed data – different types of users or business process per cloud or thought regional variations. This may impact the accuracy and fairness of models. Federated learning-models may also be complex and explainable, BI end users will require outputs to be explained, reliable dashboards, and all metrics will need to be the same throughout the BI. But for it to be effective there must be a solution assuring machine learning performance and enabling BI usability, governance, monitoring and explainability.

This research is relevant because the lack of enterprise BI needs and expectations in the multi-cloud world of data. Centralized BI architecture is still valid, but may not support enterprise-level business environments that are large, distributed and need to be privacy first and follow regulatory requirements. Collaborative analytics with federated learning will be a first step to allow organisations to profit from the collective wisdom without giving up control of the institutions raw data. The purpose of this article then, is to present this federated BI template, provide some examples of building blocks that federated BI might have, discuss a federated BI workflow, look at some security and governance considerations, and provide some perspective on the difficulties of realizing federated BI between multi-cloud data-silos. The proposed federated learning in BI can help pave the way towards a future where the enterprise intelligence will be more privacy aware, scalable, interoperable and will be more adaptively behaving in distributed cloud.

International Journal of Research Publications in Engineering, Technology and Management (IJRPETM)

|www.ijrpetm.com | ISSN: 2454-7875 | editor@ijrpetm.com |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 5, Issue 6, November - December 2022||

DOI:10.15662/IJRPETM.2022.0506022

II. FRAMEWORK FOR FEDERATED LEARNING BI ACROSS MULTI-CLOUD DATA SILOS

Proposed Federated Learning Business Intelligence (FL-BI) in data silos of multiple cloud infrastructure will allow businesses to implement collaborative analytics without their enterprise data being centralised. In traditional BI, data from a multitude of sources need to be extracted, altered and loaded (ETL) into a central data store like a data warehouse or data lake for BI reporting and analysis. Multi-cloud isn't a good model, however, because enterprise data might be broken up and hosted on different public clouds, a private cloud or hybrid cloud, regional data centers, SaaS or even at the departmental database. Although at such an early stage, clouds are flexible and provisionable by the sharing of provisionable resources (networks, server, storage, applications & services), but if a large number of clouds need to be interconnected, it's a little more complicated [11] [12].

To address this challenge, a proposed FL-BI framework to enable the advent of distributed model learning. All of the data are placed locally, each cloud deploy a model and train the deployed model in their own silo and only send the model parameters and/or gradients/encrypted updates to a central/federated coordination-layer. All participating cloud nodes are then re-distributed the aggregated model and the aggregated model is utilized for BI reporting, prediction, forecasting, anomaly detection and decision support. This is consistent with the original concept of federated learning that models are averaged (or federated) and not the data.

2.1 Framework Overview

It consists of seven main layers: multi-cloud data source layer, local preprocessing layer, federated training layer, secure aggregation layer, global intelligence layer, BI visualization layer and governance/monitoring layer. These layers can interact and communicate seamlessly, as they can bring together fragmented enterprise data from the cloud to help collectives enhance their collective intelligence and can provide privacy, compliance and data ownership.

Basically, anything cloud – from foundation up – can be an independent data silo. Typical examples of this are databases, sale data may be cloud anywhere, customer relations management data cloud somewhere else, financial data may reside in a regional cloud or private cloud, or supply-chain data in a regional cloud and operational logs in an edge connected environment. Each cloud node is considered to be an independent participant. An important point about these nodes is that they don't necessarily need to be available in order for these records to a central repository. The reverse is each participant gives the global system a model of which he is the player, and controlled system.

An excellent option for large organisations with several business units, regulated industry/jurisdictions. Let's say, in the retail sales business a company may have this data localisation requirement in each of the countries in which it has a presence and sells its products, hence the data (in this particular case Customer data) may change, depending on the country. A bank could have separate data records for the transactions at one branch or subsidiary and for the transactions at the other branch or subsidiary. This means that hospitals may have their own cloud and patient information may reside in their hospitals' cloud. In all these cases, there is a logical learning that is reduced to distributed silos possible in FL-Bi without compromising local control of the sensitive data.

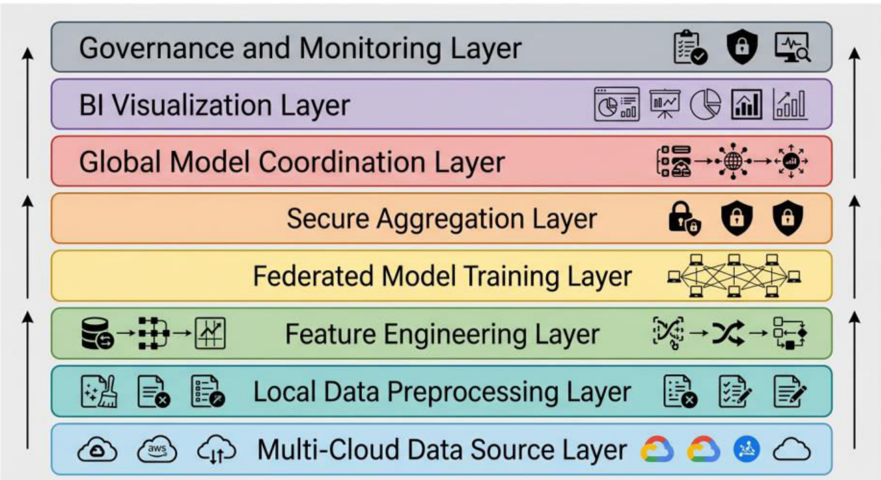

Fig. 2: Layered Framework of Federated Learning BI

International Journal of Research Publications in Engineering, Technology and Management (IJRPETM)

|www.ijrpetm.com | ISSN: 2454-7875 | editor@ijrpetm.com |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 5, Issue 6, November - December 2022||

DOI:10.15662/IJRPETM.2022.0506022

2.2 Multi-Cloud Data Source Layer

The bottom-most layer of the framework is the multi-cloud data source layer. In this layer all types of distributed data repositories come into play that are utilized during the federated BI process. These can be organized, semi-organized or not organized. These data could be made up of customer transactions, sales data, inventory data, financial information or any other subset of operational KPI(s) that the enterprise is structured with. Examples of semi structured data include IoT output, API response, clickstream data and JSON logs. Service tickets, emails, reports and customer feedback and documents are all examples of what can be classified as unstructured data.

The type and quality of the data upon which the databases are operating and used to store data – access-control policy and other security tools in the cloud as well as for the data – and standards for data metadata may be different for various cloud providers. Thus dealing with a mechanism that can establish connectivity with the clouds irrespective of which cloud is being customized is required in the framework. Each local cloud node can communicate to the federated learning engine through these connectors, without having to impose on a common physical database architecture. Facilitate local model training in connector layer by using APIs, data virtualization or metadata mapping or cloud native integration services.

What's important is that data remains in the cloud when it's created. Only essential feature/labels/inputs to train models are created locally. This not only cuts down on the cost of data transit, but also offers decreased danger of over-provisioning and cuts down on inner governance policy compliance problems.

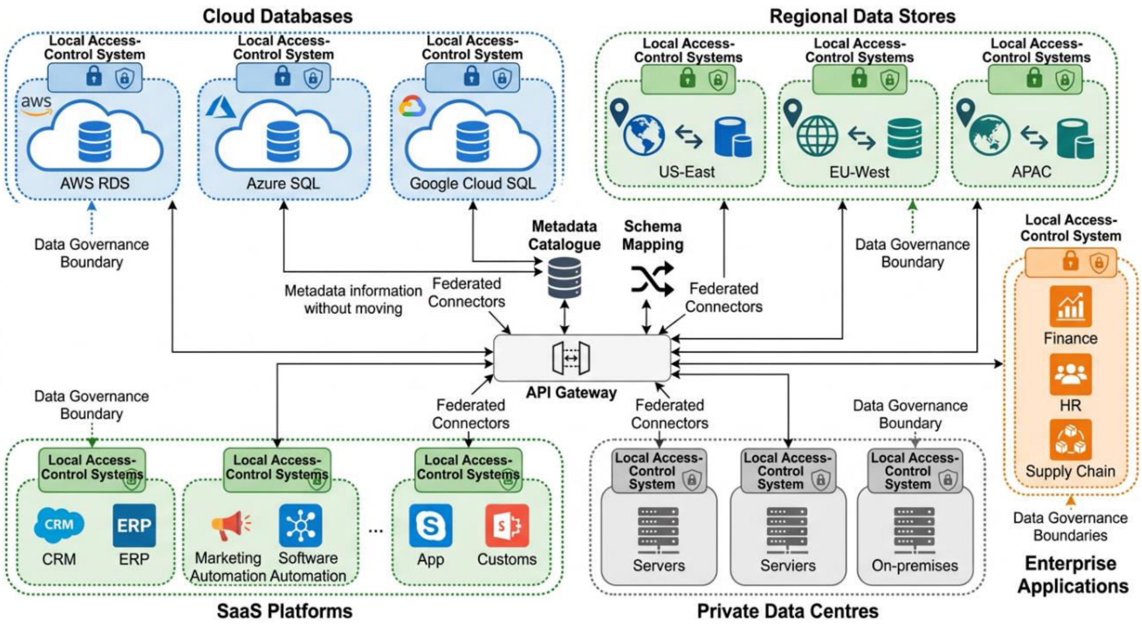

Fig. 3: Multi-Cloud Data Silo Integration Model

2.3 Local Data Processing and Feature Engineering Layer

The second layer is local pre-processing \& local feature engineering layer. The data could be different for each local participant as each local participant will have their own cloud and have to be cleaned, normalized, have features extracted and transformed before the model can be trained. These include outlier data and data aligned to the federated learning discussed objective, data cleaning and data categorifying/coding (categorify) data, and data scaling (numerical) data.

Local feature engineering is important in BI applications as the business information in various departments and regions may vary. For example, for geographic area, customers can be subdivided into income groups, for other geographic area, customers can be subdivided by levels of frequency of purchasing done by the customers, and other geographic area by loyalty level of the customers. Likewise, the digits of a monetary system may change or perhaps the sort of information connected with it might change. A semantically mapped correlation mechanism independent of the integration of data sets, to correlate equivalent business variables has been proposed in the framework.

|www.ijrpetm.com | ISSN: 2454-7875 | editor@ijrpetm.com |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 5, Issue 6, November - December 2022||

DOI:10.15662/IJRPETM.2022.0506022

Further veridiction for local veridicity: This layer can be used for further checks of veridicity, local veridicity. At every node, you determine if your data that you would like to share in the federated training has a minimum quality. At each cloud node, do Proficient Data Profiling/Data Schema Validation/Data Quality Scoring of the data, in case of having poor quality data, it could impact the global model. Local level scores of quality can be reported to the governance level with the raw data to be stored at the local level.

2.4 Federated Model Training Layer

The third layer that is main component of the proposed model federated training model layer. From here, this layer stores models training that are going to localise to each Cloud Silo, using each data set. Depending on the BI objective, the local model may be utilized for classification, regression, clustering, forecasting, anomaly detection or to recommend. They can include a prediction by the customer that the customer is likely to drop their service, sales predictions, fraud analysis, distribution resource utilisation predictions, supplier performance analysis, etc.

To make the participating cloud nodes obtain the global model constructed in the central coordinator, the training process starts with distributing the global model to the participating cloud nodes. Locally models are trained with the information of each node for a certain number of epochs at each node. The node emits changes in the model, after the training process (a gradient, the weight or a change in model parameters). These are the changes that are up-hiked up to the layer that aggregates enclosed. No information flows on the first is sent to the initial business records, customer identities, the transaction value or any other attribute which could be considered confidential.

This can be achieved by developing a better global model by FedAvg of locally created updates to the model. Federated Averaging is widely practised as a feasible solution to decentralised learning from data, and is case agnostic for learning from NIID (non independent and non identically distributed) data, which is an often-occurring situation in real life distributed environment.

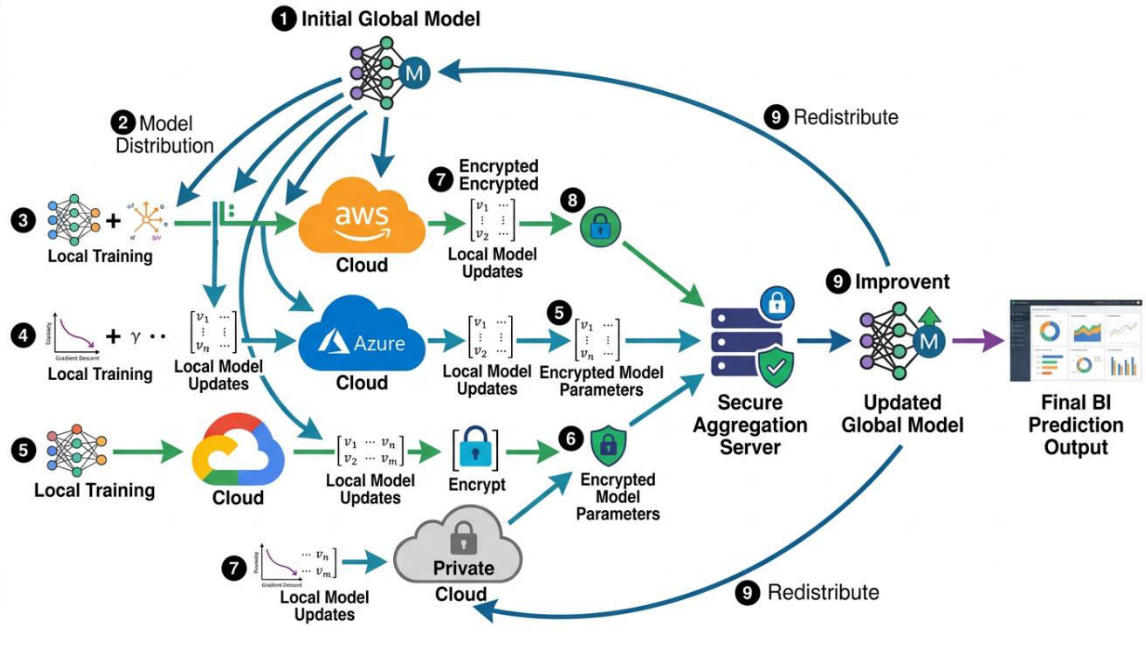

Fig. 4: Federated Learning Workflow for BI Model Training

2.5 Secure Aggregation and Privacy Layer

They have a safe aggregation and privacy layer as 4th layer. While federated learning decreases the volume of raw data that's swarming around, sensitive information could be compromised in the updates to the model if they were not properly protected. As a result, capabilities relating to privacy such as secure aggregation, encryption, access control, differential privacy, authentication, audit logging are also integrated to this framework.

Secure aggregation allows computing the aggregation, without revealing the individual change of participants. Important in the scenario where you might have several different departments, companies, hospitals, banks or cloud

|www.ijrpetm.com | ISSN: 2454-7875 | editor@ijrpetm.com |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 5, Issue 6, November - December 2022||

DOI:10.15662/IJRPETM.2022.0506022

regions with participants. Secure aggregation is one of the approaches used in privacy preserving federated learning (P2FL), which have been demonstrated to perform secure aggregation among a number of parties.

Hence in the proposed method each node in the cloud updates the local model by masking and/or encrypting data to be sent. Protected updates will be directed to an aggregation server, which will generate the global update, without ever being aware of updates generated by each individual cloud silo. The noise can also be applied on the noise, i.e., model updates can also be corrupted with the noise. This will minimise the chances of "diffing out" some sensitive gradient(s) and/or parameter(s). The privacy layer in addition should include a role based access control mechanism, an identity verification on reception of models and a key management layer and logging of all events related to the exchange of models.

2.6 Global Model Coordination Layer

The fifth layer is the 'Global model coordinating layer. This layer is to handle the federated training process. It can be used as node selection for model training in the cloud, node transmission for model training, acquisition of node training results, aggregation, model evaluation and node retraining with the upgraded model in the world.

The coordinator could be centralized, decentralized or the consortia. Centralized aggregation server is a single, trusted enterprise (or cloud). For decentralised model, the aggregation responsibility can be divided into various entities, or through peer coordinating. Communication and negotiation take place with all partners regarding the norms of collaboration, the portion of the partners, the management of the process between the consortium and the rules of governance (consortium variant). Rounding the list off is Google Cloud's federated learning tips that includes the advice to build a team of federated learning members, define roles for the members, and establish a protocol.

This is also a process that can be achieved at the coordination layer with a coordinated versioning of the models. The version number along with the list of nodes participating in the governance, the version's performance and train history should be included in all globalmodel versions. This is an important feature for BI since a business user (and auditor too!) should be able to know what model underlying the given set of results displayed on a chart on a dashboard is the result from, or what model would a given prediction be from.

2.7 BI Analytics and Visualization Layer

BI Analytics and Visualization layer is the 6th Layer. This layer transforms output of the federated model in to business insights. It 'brings to surface' the global model, on an easy to use dashboard, report, score card and also as an alert and decision support tool. The FL-BI framework provides predictive Intelligence BI which can't be provided with the traditional BI based on aggregating historical data.

For example, on the BI dashboard, they can be presented with the probability of a sale for the next time periods, the probability that the customer will be a churn customer, the risk associated with frauds, the risk associated for delay and/or forecast results on supplier by regions as well as the risk of stock not available and the probability of sale. Combined multiple sources of data from various types of cloud silos to train this model, allowing her to make predictions that you wouldn't have been able to do otherwise with one source of data alone. These raw data, meanwhile, are safe in each of the cloud environments.

The BI layers should have "explainer" features. A forerunner of a model should state why a specific prediction/recommendation is made and caveat should be given to the recipient/receiving system. So feature significance scores, description statement about the trend, confidence and why approximately the anomaly occurred are collected and posted on a dashboard/monitor as are the indicators of the quality of the model. This aids customers to trust and make decisions by the managers.

2.8 Governance, Compliance, and Monitoring Layer

This layer guarantees the functioning of the FL-BI framework both from an organizational, legal, ethical and technical perspective. Comprising of data governance policies, Model governance, compliance monitoring, participant authorization, audit trail and risk assessment and performance monitoring.

In the case of multi-cloud BI, there isn't any negotiating about the setting of governance – it will have its own associated legal and security requirement, simply because of the fact that it's a cloud silo. Locally, deciding who may want to be included under the umbrella of participatory data collectors/supporters, what data will be included in

|www.ijrpetm.com | ISSN: 2454-7875 | editor@ijrpetm.com |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 5, Issue 6, November - December 2022||

DOI:10.15662/IJRPETM.2022.0506022

training the model and sharing results of the model training, how often the model is trained, and verifying that the model results are accurate and correcting any errors is important. Further, it should outline guidelines and procedures for the case a participant is removed, rolled back, incident is encountered and/or the violation of any policy is observed.

It is as important to monitor! Monitoring of nodes availability, communication cost, training failure, and accuracy, drift, bias and latency of the model should be done. Maybe there is a performance difference between data models, since participants have different Data Models or they differ in the kinds of cloud silos in which they're deployed. This process of evaluation of the performance at global and local level, should therefore be integrated into the framework. When the overall performance is good and the region/business unit performance is poor - it may need to be adapted locally or a customization to the federated learning architecture.

III. FRAMEWORK EVALUATION

Assessing the proposed Federated Learning Business Intelligence (FL-BI) framework on multi-cloud data silos is vital because the ability of the framework to provide reliable business oriented analytics in a distributed cloud environment in a secure and scalable manner needs to be verified. Unfortunately there is no centralization of all the data, so it would be unfair to check the model's accuracy as the test of the model. It should also take into account privacy protection, communications efficiency, interoperability, governance, ease of use with the dashboard and the benefits of decisions that can be made using the dashboard. Hence, the system is subjected to various dimensions, based on the technical performance and effectiveness of Business Intelligence.

3.1 Evaluation Objectives

The goal of an evaluation of the framework is to ensure that federated learning can benefit BI results across multi-cloud BI data silos without impacting and violating the privacy and ownership of data. Overall assessment of this evaluation is then taken across learning at local cloud nodes in the model and efficiency in transferring model to finally create a model that can be used for forecasting, classification, anomaly detection etc. in BI applications. A second aim is to determine whether that scheme might be developed for a mixed up cloud set-up where every cloud might be entirely unique, various topology, differing technology used to store information, varying band-width of connections, different type of governances required etc.

3.2 Model Performance Evaluation

If one criterion relates to the model performance, it is most relevant criterion, if not, the criterion of FL-BI model performance is most relevant criterion. There are different performance metrics to use, depending on the BI use case. Accuracies, precision, recall, F1-score and the area under the receiver operating characteristic (ROC) curve can all be used for classification related BI tasks, such as fraud detection or customer churn prediction. Mean absolute error and root mean square error and mean absolute percentage error can be used as forecasting (sales, demand forecasting etc.) measurement. These measures can aid in anomaly detection such as detection rate, FPR and precision of alert.

It should be noted: The global federated to local models requires local models to be trained separately (in each cloud silo) from each other. The overall world model will be better off if we can believe that collaborative learning has a value, than the stand-alone local models. In addition, the man-made/anonymous data could be used to compare the model, against central benchmark model (if applicable). In a lot of real-life use cases, however, it may not be feasible to do a centralized comparison – privacy is one of the primary concerns.

3.3 Assessment of Privacy and security issues

As it is designed for privacy preserving BI privacy & security evaluation is an essential part of the framework. This includes analysing for training raw data, localised into the cloud. Topics that should be considered for the score include consideration of the strength of the encryption technique(s) as well as protection against secure aggregation, access control, authentication, audit logging, gradient leakage attacks and model inversion. A design that no cloud node (or even the central coordinator) should be able to recreate valuable records about the business should be the one based on the updates disributed across cloud nodes.

One main way of assessing the differential privacy is to compare with the model's performance level. Adding too much noise, may lead to a less accurate model. Hence, the framework needs to be both useful for analysis as well as accommodating to privacy.

|www.ijrpetm.com | ISSN: 2454-7875 | editor@ijrpetm.com |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 5, Issue 6, November - December 2022||

DOI:10.15662/IJRPETM.2022.0506022

3.4 Communication and Computational Efficiency

The effectiveness of the framework affects the amount of communications that will be required between nodes in the cloud and the aggregation server. Costs for communication aspects may be applicable as well, since the data centres might be in different geographical locations, and might have different network conditions. Hence the assessment should be done according to the number of communications rounds, size of the updates, bandwidth used and training time and delay of synchronization.

An additional measure of the effectiveness of such a computational operation should be whether it is efficient in an individual cloud node on a local basis. Cloud resources can range from large compute resources to running large GPUs or TPUs. The framework needs to accommodate differences in the computational ability of the resources in the framework without introducing any pathawks into the BI process.

3.5 Multi-Cloud Interoperability Evaluation

Interoperability evaluation is to study interoperability of the framework cloud provider and platforms. Objectively, a multi-cloud BI environment could be AWS and Azure, Google and private cloud, SaaS and more, requiring different storage formats and APIs, security policies and identity systems to operate. These factors involve ease of node enrollment, affinity to the schema, ability to update the model and integration with other BI applications.

Be careful not to jump on the single cloud provider, single solution bandwagon or make an assumption that a single cloud vendor will provide a single solution for every department or organization that has thrived with a successful FL-BI solution. Rather it must enable each participant to contribute, with his own intelligence, infrastructures, to the collective intelligence.

3.6 Assessment of usability & decision-support

Of course, it is worth pointing out, that if the framework could not be used to support the business in its decision, then it will have no use for the business. The following is a list of outcomes that can be assessed only through: Dashboard clarity, Dashboard interpretability, Response time, Report consistency and managerial relevance. In the Business User view, users will be able to see predictions, top scores, trends and alerts, related to the Federated Model.

An explainable output, also desirable for the decision maker to understand how factors can influence the prediction, should be also given as output of the system. A predictive dashboard for forecasting sales should, for example, show if the following factors are affecting the demand for the product: region, season and the level of stock and customer segment. This increases the chances of building trust and actuality adoption.

3.7 Overall Evaluation Outcome

In summary, this configuration will work as long as it can provide the right BI information, protect the privacy of the data, reduce the amount of undesirable data migration, be platform independent, and allow for beneficial decision sheds to be made. Evaluations should communicate how federated learning goes beyond a machine learning approach and useful architectural pattern for today's BI systems in multi-cloud data silos

IV. INDUSTRY APPLICATIONS OF THE FRAMEWORK

In many cases of distributed, sensitive and challenging data centralisation in the different industry sectors the Federated Learning Business Intelligence (FL-BI) framework over multi-cloud data silos is highly applicable. In the healthcare industry, patient information is typically held on cloud resources at various locations such as hospitals, diagnostic imaging centers, insurance companies and research institutions. The model can include disease transmission dynamics, patient re-admissions and effectively manage treatment and hospital resource management whilst maintaining patient privacy.

It can be used in the banking and financial services industry for fraud detection, credit-risk assessment, anti-money laundering surveillance, customer segmentation, and investment-risk analysis. For financial institutions it's about federated intelligence: every institution has to report based on constantly evolving regulations and data is distributed and stored in regional/departmental systems.

By serving retailers in regions and/or cloud, retail/ecommerce organizations can also use the framework for many other scenarios such as demand forecasting, customer behaviour analysis, providing customers with personalized

|www.ijrpetm.com | ISSN: 2454-7875 | editor@ijrpetm.com |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 5, Issue 6, November - December 2022||

DOI:10.15662/IJRPETM.2022.0506022

recommendations, inventory optimization and dynamic pricing etc. without having to move any raw customer or sales data into a centralized location.

Within manufacturing and supply chain management federated BI can be used for energy forecasting, logistics optimization, supplier-risk analysis, monitoring production quality and for predictive maintenance. It's possible to train analytical models in various factories, warehouses and vendor systems while maintaining the operational data but with each system.

In addition, the framework can be used in the telecom industry, where a proliferation of cloud and edge deployments underpin the traffic records system and a lot of customer usage data is stored on individual clusters. It can be utilized for churn detection, and for discovering anomalies in the network, bandwidth planning and even service-personalization.

Yet in the public sector, in smart city space, various departments and disciplines such as transport or health, utilities or public safety can all utilise analytics without having to combine sensitive citizen level data. The overall architecture allows industries to accelerate transitioning from one-off reporting in BI to a safer, shared & private intelligence in multi-cloud environments.

V. CONCLUSION AND FUTURE WORK

This is a research paper which proposes Federated Learning Business Intelligence (FL-BI) framework for analytics over multi-cloud data silos. Thus this study was suited to the traditional BI systems, as one of the greatest issues facing such systems was the need to be able to collect data from all systems in a central location. Now organisations can see their data being kept in a web of cloud services, on their own data centres, at regional data centres, or on SaaS (software-as-a-service) applications or other various repositories of data in other departments. All of these can be challenging issues with privacy, compliance, latency, cost, ownership and interoperability when consolidated at a single warehouse. Hence, a new approach in BI needs to be developed which can set the way for Collective Intelligence but does not dictate any data centralisation.

For the needs of federated learning and BI framework, the proposed architecture could be a potential "best practice" to provide distributed, secure and privacy-preserving analytics or analytics piece. In this context, only the cloud silo that sends the data will use their data to train the models, and will only send the updates for the trained models to the aggregation layer and/or only the encrypted parameters of the models for training. Then, the aggregation is securely performed to enhance globally constructed model, and sent back to the cloud nodes participating in the aggregation. It allows organisations to have predictive and intelligent BI capabilities whilst securely handling sensitive data.

It also adds some other key functionality such as federated training of models over the cloud, secure model aggregation, model orchestration across the world, BI visualization for models, and multiple cloud data sources and data processing 'on-the-go'. Combined, they make applications; such as predicting customer churn, detecting fraud, forecasting demand, other predictive maintenance applications, healthcare analytics and, in the smart city world, decision-making – just to name a few. The performance evaluation section of the report also demonstrates that besides model accuracy, this types of framework must be evaluated by using other metrics like privacy protection, communication efficiency, interoperability, scalability, explainability and business value.

FL, in a nutshell, could drive BI forward on a path towards an ecosystem of distributed intelligence. It allows organisations to work across the cloud boundaries and minimises not only the amount of unnecessary data transfer but also enhances trust in analytics.

5.2 Future Work

Regarding future studies, the proposed FL-BI framework misses a way of actual deployment and further improvements can be made to improve its technical maturity and scalability. Firstly, the applicability of these new privacy preserving techniques such as differential privacy, homomorphic encryption and trusted processing environment, need to be further enhanced and strengthened to enhance security concerns related to leakage of models and attacks due to inference reflections. Second, future research activities may develop an adaptive aggregation technique to react to varying data quality, unbalanced cloud resources and varying data distribution of participating cloud silos.

|www.ijrpetm.com | ISSN: 2454-7875 | editor@ijrpetm.com |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 5, Issue 6, November - December 2022||

DOI:10.15662/IJRPETM.2022.0506022

Another key one is to create federated (explainable) BI dashboards. The users of business realize the need for an explanation of the prediction instead of just the prediction. In the future system, the ability to explain the decisions made, level of confidence in the decisions, discover bias and to increase the model interpretability should be included.

Additional investigations for future work could consider real time federated BI for the more lively BIs of economical areas, such as finance, healthcare, logistics or telecommunications applications. Lastly, experiments using real enterprise multi-cloud data datasets would be helpful to quantify the actual performances of the proposed framework, its deployment cost, security level and managerial value

REFERENCES

[1] H. B. McMahan *et al*., "Communication-efficient learning of deep networks from decentralized data," in *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2017.
[2] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," *Proceedings of Machine Learning and Systems (PMLR)*, vol. 2, pp. 429–450, 2020.
[3] S. P. Karimireddy *et al*., "Scaffold: Stochastic controlled averaging for on-device federated learning," in *International Conference on Machine Learning (ICML)*, 2020.
[4] S. U. Stich, "Local SGD converges fast and communicates little," in *International Conference on Learning Representations (ICLR)*, 2019.
[5] J. Konečný, H. B. McMahan, D. Ramage, and P. Richtárik, "Federated optimization: Distributed machine learning for on-device intelligence," *arXiv preprint arXiv:1610.02527*, 2016.
[6] P. Kairouz *et al*., "Advances and open problems in federated learning," *arXiv preprint arXiv:1912.04977*, 2019.
[7] C. He *et al*., "Group knowledge transfer: Federated learning of large CNNs at the edge," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.
[8] Q. Li, Y. Diao, Q. Chen, and B. He, "Federated learning on non-IID data silos: An experimental study," *arXiv preprint arXiv:2102.02079*, 2021.
[9] J. Wang and G. Joshi, "Cooperative SGD: A unified framework for the design and analysis of local-update SGD algorithms," *Journal of Machine Learning Research*, vol. 22, no. 213, pp. 1–50, 2021.
[10] L. Liu, J. Zhang, S. H. Song, and K. B. Letaief, "Client-edge-cloud hierarchical federated learning," in *IEEE International Conference on Communications (ICC)*, 2020.
[11] Y. Wang *et al*., "Federated latent Dirichlet allocation: A local differential privacy based framework," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, pp. 6283–6290, 2020.
[12] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning: Concept and applications," *ACM Transactions on Intelligent Systems and Technology*, vol. 10, no. 2, pp. 1–19, 2019.