

# Review of Unified DataLake or OneLake of Microsoft Fabrics- applied in a Data Analysis Platform

Ambika Ganapur\*, Dr. Sumati Ramakrishna Gowda

Department of Computer Science, Karnataka State Open University, Mysore, Karnataka, India

## ARTICLE INFO

### Article History:

Accepted : 25 Nov 2024

Published: 20 Dec 2024

### Publication Issue

Volume 10, Issue 6

November-December-2024

### Page Number

2106-2111

## ABSTRACT

Microsoft Fabric is an end-to-end data analytical platform designed for achieve a Unified solutions in a Business Intelligence and Enterprises. It incorporates data movements, processing, ingestion, transformation, real-time event routing, and report building. It provide services like Data Engineering, Data Factory, Data Science, Real-Time Analytics, Data Warehouse, and Databases. This paper examines Microsoft Fabric and its features especially about a One Lake. Microsoft Fabric applied in a real-time analytics, data storage, advanced machine learning capabilities, and robust security protocols. Collection of files, folders and tables which represent the database is a data house . A Data lake is a cloud-based big data analytics service. Which allows storing, processing, and analyzing large amounts of structured and unstructured data. It integrates with other Azure services to provide a full data analysis solution[1].

**Keywords :** Microsoft Fabrics, Data Engineering, Data Factory, Datascience, Data Warehouse, Onelake, Data House, Structured Database, Azure, Data Analytics, Cloud, Big Data.

## Introduction

Microsoft Fabrics is a User friendly platform that addresses the analytical requirements Within a single platform we can get multiple services rather than offering services from multiple vendors for Integration. It Operating on a Software as a Service (SaaS) model, Fabric brings simplicity and integration [2].Microsoft Fabric provides centralized data storage with One Lake that is the collection of the different databases.Microsoft Fabric provides centralized data storage with One Lake that is the collection of

different databases or data warehouses which eliminates the manual integration[2].

## Strengths's of the Microsoft Fabrics:

### Main key strength's of Microsoft fabrics are

- **Role-based workload:** It provides role based or designation based work load with in an organization and provides necessary required tools for individual role.
- **One Lake:** A unique data Lake which will provide easy access to the database [2].

- **Copilot support:** AI-driven features that assist users by providing intelligent suggestions and automating tasks.
- **Integration with Microsoft 365:** Seamless integration with Microsoft 365 tools, enhancing collaboration and productivity across the organization[2].
- **Azure AI Foundry:** Utilizes Azure AI Foundry for advanced AI and machine learning capabilities,enabling users to Synapse Analytics,AzureDatafactory, and Synapse Analytics .
- build and deploy AI models efficiently.
- Unified data management: Centralized data discovery that simplifies governance, sharing, and access[1].m,,

**Unification with SaaS foundation:**

Microsoft Fabric is the one of the cloud services that is Software As A Service (SaaS). It merges new and existing components by Power BI, Azure

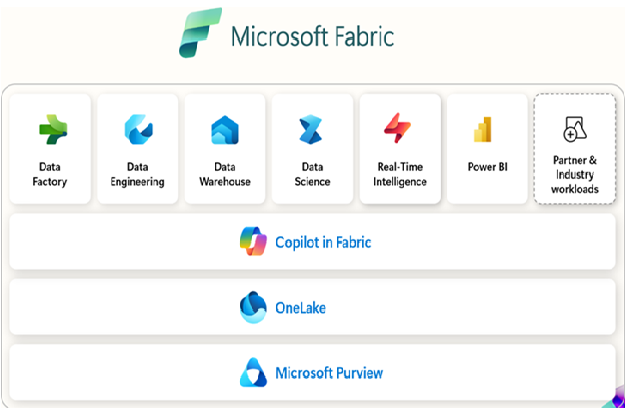


Fig 1 : Components of Microsoft Fabrics

**Power BI** – To connect database Previously we need to go with data adapters but With Power BI easily connects the datasources,visualize and discover and allow us to share with anyone . It allows us to integrate and access data fabric quickly and incorporate to take better decision[3].

**Databases** – Which allow us to create developer friendly transactional Databases in Microsoft Fabric as Azure SQL Database, Using the mirroring capability, we can integrate data into OneLake platform extracted from various systems[4] .

**Data Factory** –It provides data collaboration, prepare and transfer data from rich set of data sources [4].

**Data Activator:** Email notification can be generated[5].

**Industry Solutions** – Industry specific unified data solutions can be provided by Fabric and it addresses for industry need and challenges and also incorporates with data management, analytics, and decision-making[5].

- **Real-Time Intelligence** –Robust Real-Time Data Processing can be done by the fabrics.Microsoft Fabric is adept at processing and analyzing streaming data. It applies in Fraud detection. IoT analytics for operational insights, and real-time monitoring of business processes[5].
- **Data Engineering** –It provides systematic engineering to create manage and Fabric Data Engineering provides a Spark platform with great authoring experiences. It enables you to create, manage, and optimize infrastructures for collecting, storing, processing, and analyzing vast data volumes[5].
- **Data Science** - Fabric Data Science enables you to build, deploy and operationalize machine learning models from Fabric. Data analyst integrates predictions in to the BI which has been generated by Data science with the help of Fabrics.
- **Data Warehouse** - Fabric Data Warehouse provides industry leading SQL performance and scale. It separates compute from storage, enabling Independent scaling of both components[5].

**Data Lake Evalution:**

The evolution of data Lake has done very rapidly. Previously for distributed storage from varieties of data sources and to process the Big data many Organizations adopted Hadoop which is an open-source frame work[9]. Hadoop stores massive amount of data at the small cost of traditional data ware house.

But still Hadoop had limitations and it is quite complex to configure and setup .It required skilled

employees to deal effectively therefore many Organizations faced many problems to ensure with the Hadoop. In this moment flexible and scalable Lake house or data lake house has introduced .

Where Lake house is built on modern cloud based platform which supports Azure, WS or GCP.

It is very easy and flexible to use and manage even for no-technical users also Irrespective of format and location it specifies the Unified view to the data.

#### **Description about OneLake:**

- Unified, secured data storage for the entire enterprise on Microsoft Fabric based on Azure data Lake storage Gen2 stands as a One Lake.
- Professionals can be used as a SaaS and Developers can be used as a data storage( pay as per use).
- It is used as a centralized data storage.
- It makes easy to access ,share and analyse the information through out the organization[6].

#### **Advantages of OneLake:**

- One Lake is an excellent tool that incorporates Organizations flexible data management and data analysis.
- It is used in Data science and real time analysis.
- It helps in BI .
- It is the core of the Microsoft Fabric platform.
- It allows single namespace for all data irrespective of its location.
- Multi Engine supportive Platform.
- Provides strong security and management features to protect our data.
- One Lake is designed to fulfill the needs of Organizations with high performance[6].
- Data Mesh as a service.

#### **Disadvantages :**

- Present OneLake has limited T-SQL support. It can't support all T-SQL features .
- There are some limitations in the features of Data Lake those are need to be enhanced.
- One Lake is bit expensive for the Organizations which have large database.
- Setup and managing of One lake is complicated at initial stage. Employees need to be trained.

#### **Security Concerns of One Lake:**

- One lake works with Microsoft Azure Web Service. And Azure provides credentials that is Authentication and Authorization for the User By that ,only authorized users can access the One Lake .
- Unauthorized user can be restricted to use.
- Role Specific Access Control (RSAC), by this data can't be misused and data maintained confidentially.
- OneLake provides the secured environment for accessing the tables, files, and workspaces.
- One Lake provides Data Encryption .
- One Lake supports the Transactional logs by this transparency increased and tracked easily.
- One Lake held up the Suspicious threats and unauthorized access like a firewall.
- It allows the multi factor Authentication for the user.

#### **Data Lake or One lake as A service:**

- One Lake is Served as a One drive.
- It is a Mono unified logical SaaS data lake for the Whole Organization[7].
- One of the most exciting things in Fabric I found is OneLake.
- I was amazed to discover how OneLake is simplified just like OneDrive.

- It's a single unified logical SaaS data lake for the whole organization (no data silos).
- Over the past couple of months, I've had the incredible opportunity to engage with the product team and dive into the private preview of Microsoft Fabric

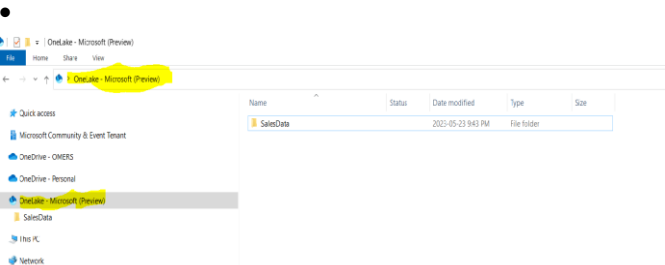


Fig 2: One Lake as a Onedrive.

OneLake Served as a Unified Solution and Platform for Entire Organization.

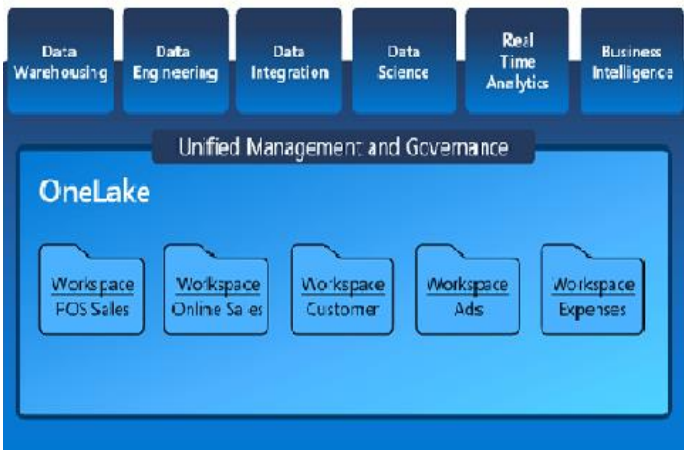


Fig 3: One Lake as a Unified Management and Governance[7].

- OneLake is a default service with Microsoft Fabrics.
- No infrastructure need to be managed.
- Any data in OneLake works with out-of-the-box governance such as data lineage, data protection, certification, catalog integration[7].
- It is a Role specific platform.
- It allows centralized database with distributed ownership.
- Individual workspace can have its respective administrator ,access control, region and capacity for billing[7].

- One Lake used in BI as a Business Analytical Software:
- Self-Service BI: It provides the data analysis which will help in taking decisions and Provides the report.
  - Advanced Analytics & Machine Learning: Leverage OneLake's integration with various analytical engines like Azure Synapse Analytics and Databricks for advanced data processing, machine learning, and AI development[6].
  - Unified Data Source for Reporting: Eliminate data redundancy and provide a consistent, trusted data source for various reporting tools like Power BI and others[6].

One Lake would become the core of Datawarehouse in future:

According to analysis each and every organization is facing the challenges of data silos problem( each organization have split database like data stored in separate systems and separate departments ,making integrated and updating invoice is quite complicated in distributed centralized database for those organizations looking for such a tool which works as data mesh

One lake perfectly suitable and addresses the problem of distributed database .

It incorporates the single unified solution for individual data with same repository for namespaces. In fact, statistics reveal that an average company has over 2,000 silos of information. And as per the Customer Journey report, this data silos is the greatest problem when obtaining insights from data[8]. Data Silos elimination is the only solution that can be achieved by the Microsoft’s Data Lake by integration of data from multiple data sources in to a unified repository which allows data integration and collaboration.

Applications of Data Lake:

- 1). Data Lake applied in the Power BI best analytical tool which ensures to combine data from various

- datasources in to single repository- Organizations current demand.
- 2). Applied as a Centralized logical data Lake for the entire Organization to Fulfill its needs.
- 3). Served as a Powerful analytic service with the same platform which able to simplifies the processing of data and enhancement of data insights as a Azure Databricks, Power BI[8].
- 4). It applied just like “ Multi tier as a One tier Architecture” that mean the distributed workspace with mono control ownership model.

Noble features of One Lake:

Multiple Lake Houses: Within the one Lake user allowed to create multiple “ Lake Houses” as one to many by this feature Organizations can maintain private and confidential data separately and can controlled under the single access policy.

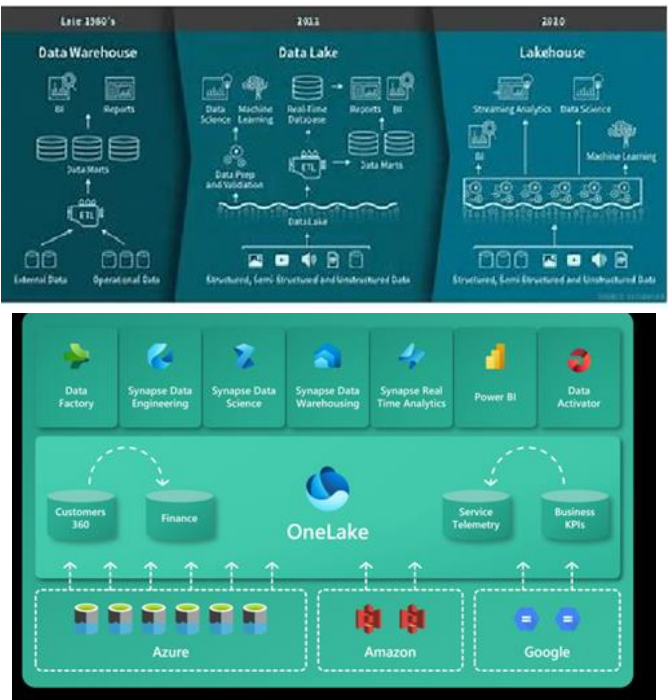


Fig 4: Lake House

- Shortcuts: Organizations can able to share the data between their users and applications easily and quickly without replicating it.

- One copy for multiple analytical engines: OneLake allows for reusability of data for multiple applications ( need not to generate data again and again).

Challenges of OneLake :

- One of the highly challenge associated with the Data Lake is less Configured than Azure data storage.
- Security matters when it allows the Multi tenant Fabrication feature.
- When an instances of one lake separates from it, then each instance would need its own Azure Active Directory(AAD) tenants. This is bit Costly process.

Conclusion

- OneLake is developing as a best solution for data Engineering.
- It is default platform for every Microsoft Fabric tenant with centralized data storage
- Across Amazon, Azure, and more platforms “OneLake, the OneDrive for data.
- Data Lake is a wonderful solution for a many data needs.
- It provides unified data integration and ETL (Extract, Transform, Load) processes which allows to access the data , prepares the data and transforms the data.
- Data Lake become the ideal choice for data management.
- Data lake ensures with the supporting, managed data and unmanaged data ,this feature became as a innovative feature.

Future Scope:

- Need to enhance its interoperability with the Azure.
- Data Lake need to enhance its power to fully support the T-SQL.
- DataLake still need to enhance its performance in Multi

## References

- [1]. “Microsoft Fabric Review: Exploring Microsoft's New Data Analytics Platform” - Praveen Borra
- [2]. <https://www.geeksforgeeks.org/one-lake-in-microsoft-fabric-> “One Lake In Microsoft Fabric - GeeksforGeeks”.
- [3]. <https://www.sqlservercentral.com> – “ what-is-onelake-in-microsoft-fabric”
- [4]. “<https://algoscale.com/blog>”-Future of Data Warehousing
- [5]. <https://lakehouse.app/article>“ The\_Future\_of\_Data\_Lake\_Trends\_and\_Predictions”
- [6]. [https://learn.microsoft.com/en-us/fabric/get-started/microsoft-fabric-overview-](https://learn.microsoft.com/en-us/fabric/get-started/microsoft-fabric-overview)“
- [7]. <https://learn.microsoft.com/en-us/fabric/get-started/>”microsoft-fabric-overview”