Mastering Azure Synapse Analytics

Learn how to develop end-to-end analytics solutions with Azure Synapse Analytics

Debananda Ghosh

www.bpbonline.com

Every effort has been made in the preparation of this book to ensure the accuracy of the information presented. However, the information contained in this book is sold without warranty, either express or implied. Neither the author, nor BPB Online or its dealers and distributors, will be held liable for any damages caused or alleged to have been caused directly or indirectly by this book.

BPB Online has endeavored to provide trademark information about all of the companies and products mentioned in this book by the appropriate use of capitals. However, BPB Online cannot guarantee the accuracy of this information.

First published: 2023

Published by BPB Online
WeWork
119 Marylebone Road
London NW1 5PU

UK | UAE | INDIA | SINGAPORE

ISBN 978-93-55518-125

www.bpbonline.com

Dedicated to

My beloved wife:

Chandramita

&

My Son ***Devansh***

About the Author

Debananda Ghosh is a Data and AI specialist and has been working in the Data and AI field for the past 17 years. His expertise includes, Data Warehouse, Database admin, Data Engineering, Big Data & AI, Data architecture, and related cloud analytics practice. Prior to Microsoft he was in Rolls Royce Singapore Data lab developing Aviation Analytics products in Cloud Big Data and AI platforms. He completed his bachelor's in engineering from Jadavpur University, Kolkata, India, and a Post Graduate Program in Data Science and Business Analytics from McCombs School of Business at the University of Texas at Austin. He has worked with customers across multiple industries like finance, manufacturing, utilities, telecom, retail, e-commerce, and aviation. He is currently working with Microsoft cloud analytics products and helping industry partners achieve their digital transformation journey using advanced analytics and AI capability.

Debananda is also a distinguisher speaker, and blog writer, in cloud data and AI technology fields.

About the Reviewer

Himanshu Amodwala is a Cloud Solutions Architect focusing on Data & AI technologies, with over 11 years of experience in IT Services and Software Development. He has worked on several projects involving big data engineering, machine learning, and artificial intelligence on the cloud. He is passionate about helping businesses leverage the power and scalability of the cloud to solve complex problems and deliver innovative solutions. On a personal front, he is an uber class introvert, an avid reader and loves spending time with family.

Acknowledgement

I want to express my deepest gratitude to my family and friends for their unwavering support and encouragement throughout this book's writing, especially my wife Chandramita and my son Devansh.

I also want to thank my parents who have supported me through the thin and thick of my life

I am also grateful to BPB Publications for their guidance and expertise in bringing this book to fruition. It was a long journey of revising this book, with valuable participation and collaboration of reviewers, technical experts, and editors.

I would also like to acknowledge the valuable contributions of my colleagues and co-worker during many years working in the tech industry, who have taught me so much and provided valuable feedback on my work.

Finally, I would like to thank all the readers who have taken an interest in my book and for their support in making it a reality. Your encouragement has been invaluable.

Preface

Analytics developments in on-premise is a complex task and need extensive knowledge of different programming languages and open source and proprietary subject areas. Compared to on-premise, when it goes to cloud analytics platform development efforts and business/technical benefits increases significantly.

This book is designed to provide a detailed overview of one of the Microsoft platform cloud analytics core service, also known as Azure Synapse Analytics. This Book will cover the latest features and editions of the Azure Synapse Analytics platform, for example, Industry common data model-based Synapse database template, Synapse Spark-based Machine learning with Hyperspace Index, Synapse data explorer based observational analytics and IOT analytics.

Throughout this book, you will learn topics around advanced development using Synapse SQL Pool and Serverless SQL, and how to develop Synapse pipeline. Also, you will go through subject areas like Synapse Spark for Machine learning tasks. You will explore Synapse data explorer for Telemetry analysis, common data model-based database template of Synapse. You will be familiarised with how to query using T-SQL, KQL, and Spark SQL in Synapse. This book will also cover Azure cosmos database integration with Synapse, Microsoft Purview and Synapse integration, Power BI visualization within Synapse workspace, and the Industry architecture pattern of Azure cloud analytics.

Overall, this book is suitable for Cloud Data engineers who haVE some experience in Azure cloud computing. It is also suitable for Chief data officers (CDOs), and Data leadership to understand the technical benefits of Microsoft Azure cloud analytics platform.

By the end of this book, you will have comprehensive knowledge of Azure Synapse analytics service and related development practices. I hope you will find this book informative and helpful.

Chapter 1: Cloud Analytics Concept - In this chapter, we would discuss cloud computing fundamentals and why organizations are focussing on cloud investment. We will briefly discuss cloud computation benefits, and the data platform evolution story from the data warehouse and lake house. At the end of the chapter, we would introduce cloud analytics capability and how organizations can get benefit from it.

Chapter 2: Introduction to Azure Synapse Analytics - This chapter talks more about Microsoft Azure Synapse Analytics service at a high level. This chapter also explains why we need to leverage limitless cloud analytics capabilities. It gives a complete end-to-end high-level overview of Synapse Analytics core pillars and how it addresses enterprise lake house requirements.

Chapter 3: Modern Data Warehouse with the Synapse SQL Pool - This chapter would zoom into specific capabilities inside the Synapse workspace and would focus on SQL-based MPP capability. We will attain a high-level understanding of Provisioned SQL engines.

Chapter 4: Query as a Service -Synapse Serverless SQL - This chapter will cover Serverless SQL computation capability within Synapse workspace. We will also get high-level understanding of Serverless SQL.

Chapter 5: Synapse Spark Pool Capability - Here we are covering Synapse spark latest developments and announcements. Also, we would cover in detail Synapse Spark-specific features that are complimentary to and sometimes beyond OSS Spark. It would cover machine learning development environment details inside the Synapse workspace. This chapter is predominantly about machine learning-related advanced analytics features.

Chapter 6: Synapse Spark and Data Science – In this chapter, we will be covering data science related to some basic tasks, so that we are familiar with Synapse Spark-related usage in the data science field.

Chapter 7: Learning Synapse Data Explorer - This chapter discusses data explorer capability in detail which is designed for Telemetry and IoT analytics purpose. We would briefly learn underlying Kusto query coding fundamentals which are essential to developing synapse-based Telemetry analytics.

Chapter 8: Synapse Data Integration - This chapter discusses Synapse's low code and no code-based pipeline development. Also by leveraging these features how an organization achieves developers’ productivity will be discussed in this chapter.

Chapter 9: Synapse Link for HTAP - Building a hybrid/Transaction analytics process is critical for an efficient analytics platform. In this chapter, we would focus on how to develop an HTAP system easily using the Synapse analytics environment.

Chapter 10: Azure Synapse-Unified Analytics Service -This chapter is mostly about the new synapse industry-based common data model-based database template which we can use in data lake as a one-click deployment. Also, this chapter would focus on synapse monitoring metrics and management pane for advanced users.

Chapter 11: Synapse Workspace Ecosystem Integration - This chapter is mostly about new synapse industry-based common data model-based database templates which can be used in data lake as a one-click deployment. Also, this chapter would focus on synapse monitoring metrics and management pane for advanced users.

Chapter 12: Azure Synapse Network Topology - Creating a secure data platform is crucial for any organization in order to align with industry compliance. In this chapter, we would discuss network security and isolation that we need to do as best practices for analytics service deployment.

Chapter 13: Industry Cloud Analytics - This chapter will briefly cover industry architecture related usecase and deployment patterns, and how synapse analytics is used across industries to deploy such architecture patterns.

Code Bundle and Coloured Images

Please follow the link to download the
Code Bundle and the ***Coloured Images*** of the book:

https://rebrand.ly/uyul2d4

The code bundle for the book is also hosted on GitHub at **https://github.com/bpbpublications/Mastering-Azure-Synapse-Analytics**. In case there's an update to the code, it will be updated on the existing GitHub repository.

We have code bundles from our rich catalogue of books and videos available at **https://github.com/bpbpublications**. Check them out!

Errata

We take immense pride in our work at BPB Publications and follow best practices to ensure the accuracy of our content to provide with an indulging reading experience to our subscribers. Our readers are our mirrors, and we use their inputs to reflect and improve upon human errors, if any, that may have occurred during the publishing processes involved. To let us maintain the quality and help us reach out to any readers who might be having difficulties due to any unforeseen errors, please write to us at :

errata@bpbonline.com

Your support, suggestions and feedbacks are highly appreciated by the BPB Publications' Family.

Piracy

If you come across any illegal copies of our works in any form on the internet, we would be grateful if you would provide us with the location address or website name. Please contact us at **business@bpbonline.com** with a link to the material.

If you are interested in becoming an author

If there is a topic that you have expertise in, and you are interested in either writing or contributing to a book, please visit **www.bpbonline.com**. We have worked with thousands of developers and tech professionals, just like you, to help them share their insights with the global tech community. You can make a general application, apply for a specific hot topic that we are recruiting an author for, or submit your own idea.

Reviews

Please leave a review. Once you have read and used this book, why not leave a review on the site that you purchased it from? Potential readers can then see and use your unbiased opinion to make purchase decisions. We at BPB can understand what you think about our products, and our authors can see your feedback on their book. Thank you!

For more information about BPB, please visit **www.bpbonline.com**.

Join our book's Discord space

Join the book's Discord Workspace for Latest updates, Offers, Tech happenings around the world, New Release and Sessions with the Authors:

https://discord.bpbonline.com

Table of Contents

CHAPTER 1
Cloud Analytics Concept

Introduction

The world is going through digital transformation and it's visible in our everyday activities. Today, we see tons of data being generated from multiple sources. Sensors, wearable devices, click stream, web applications, logging, monitoring, and intelligent application layer generates humongous volume of data every moment. To cater to such data explosion phenomenon, platform capability is continuously evolving over the last few decades. From a legacy transactional database to a data warehouse, data lake, and now lake house-like capability, it always helps achieve more towards business-related data growth and transformation. Every organization has embarked on a journey to adopt such data products so that they can achieve more towards their business goal. We see nowadays recent trends like cloud data analytics and AI adoption across the industry.

This chapter will introduce what is cloud analytics capability. It is essential to know the value of cloud analytics, before jumping into Azure Synapse Analytics products.

Structure

In this chapter, we will focus on the following topics:

- Data architecture evolution

- Data warehouse fundamentals and limitations
- Data Lake fundamentals and limitations
- Concept of Lakehouse, best of two worlds
- Introduction of cloud
- What is a cloud analytics platform?

Objectives

This chapter's objective is to take us to the data platform journey over the last few decades. In this chapter, we will do a high-level overview of all the phases of data platform evolution. At the end of this chapter, we will learn different phases of data management using cloud capability. Also, our goal is to learn what are the modern cloud analytics tech platform and its underlying building blocks.

Data architecture evolution

As we look back a few decades earlier, computers were mostly solving very simplistic data problems using programs. Storing the file data, and reading/writing the data sequentially, and hierarchically was the initial key problem statement which was addressed by legacy infrastructures. Key sequential data sets, and tapes used in tech stacks like IBM Mainframe, Cobol, and JCL programmatic approach was one way to deal with a large volume of data set processing effectively in a batch manner. The following figure shows a tape picture from the IBM archive:

***Figure 1.1**: Tape archival*

Gradually, programming languages and data computing capacity, both started evolving. Database management systems came up with solving problems of data storing structuring in the desired format, and data retrieval and manipulation. Database management system itself evolved from network database management system flavour (for example IBM IMS) to relational database management systems (DB2, Microsoft SQL Server, Oracle) to cater to high-performance retrieval for transactional operations. The relational database approach was a total shift of data management architecture, and such a database approach started providing ease of data retrieval using SQL (Structured query language) instead of legacy programmatic data retrieval approach, for example, COBOL-IDMS programs.

As we moved forward, data volume started growing exponentially across the organization. These were caused due to multiple business applications which started evolving in the organization to support various business process need. Accessing such large data in a single operation, and running complex queries within the database management system was neither cost-effective nor was it a healthy workload management operation. On the other hand, within the organization data started being scattered across different online transactional databases. Hence, it created data silo-related problems like data duplicity and many more challenges. In a data platform, the creation of a single version of the truth was important.

Hence, in the late 80's the concept of Datawarehouse evolved to address such challenges. Datawarehouse appeared as a unified data platform for all business application users to access structured data using SQL endpoints or via business intelligence tools mostly. DW appliances like Teradata, and Greenplum appeared in the market to provide such DW capability to organizations. As the Internet consumers started growing, application and device nature started evolving thus generating the PB (Petabyte) scale of data. As it happened, the traditional generic data warehouse framework started showing some limitations. This is discussed in the subsequent section in more detail within this chapter. Organizations needed to do data management in real time; hence, data velocity became important as well. Maintaining the veracity or accuracy of data became crucial. Hence, **5V** (**Volume, Variety, Velocity, Veracity, Value**) related challenges came in the industry which is also known as a big data problem. Bigdata platform evolved to address such problems. On-premises Hadoop ecosystem came up with a framework that supports such a data management process. Market players like Cloudera, Hortonworks, and MAPR created their distribution of Hadoop in late 2000/early 2010. The following figure depicts a high-level timeline of the data evolution architecture till the current cloud lake house trend:

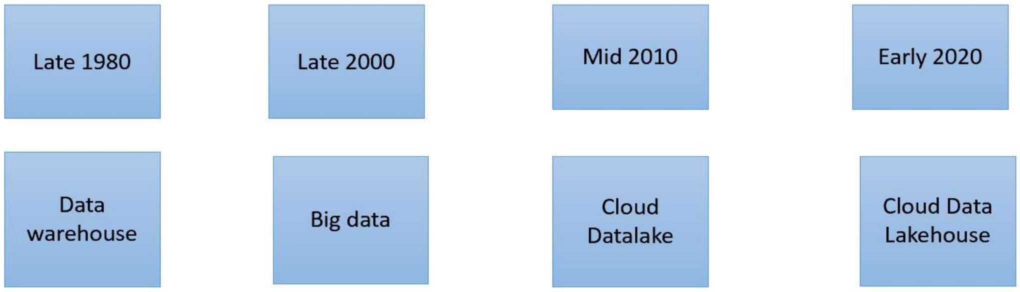

Figure 1.2*: Data architecture evolution*

Note that adopting Hadoop and a similar framework was also not a hassle-free journey since it had its limitations like security, and transactional consistency. The data lake platform started evolving and adopting the cloud framework in mid-2010. This addressed a few data lake limitations like scalability, ease of infra management, and cost effectiveness. Today's world is generating humongous data every moment; hence, the technology stack must evolve further. In the early 2020s, the cloud lake house capability was born to adopt all benefits of the data warehouse and cloud data lake. We will discuss each such framework in the subsequent sections. Note that the purpose of the subsequent section is not to provide too detailed an architectural explanation of each phase, but rather an understanding of the concept, and the reason behind such evolution phases.

Data warehouse fundamentals and limitations

In this section, we'll focus on why data warehouse platform key capabilities and why this platform evolved from the database. A database is usually designed for an **Online transactional processing system** (**OLTP**); hence, can accommodate a huge number of small transactions that do read update write. However, analytical processing that deals with a huge volume of data needs a different computation system. Usually, such processes may deal with the TB scale or more and the nature of the query is complex. Addressing silo data sources was another bigger concern for the industry. Hence, in the 90s the concept of a data warehouse evolved primarily to support the extensive scale of data analytics. Datawarehouse was designed to bring the following benefits:

- **Data mining**: Data mining on a large volume of data in the data warehouse is used to get useful patterns and was a strategic usage for business.

- **Cost-effective decision-making**: Data-driven decision-making should be cost-effective and provide business value.
- **Higher query performance**: Data mining in a larger data volume industry needed higher query performance and is dependent on fast retrieval of data.
- **Data security**: Secured platform is essential to segregate users and related authorization.

Usually, a data warehouse will have 3-tier architecture. The bottom tiers consist of data warehouse servers interacting with upstream sources. The middle tier usually hosts OLAP (Online Analytical Processing Server) Top tier is more of client-facing tools. *Figure 1.3* illustrates the concept of traditional data warehousing:

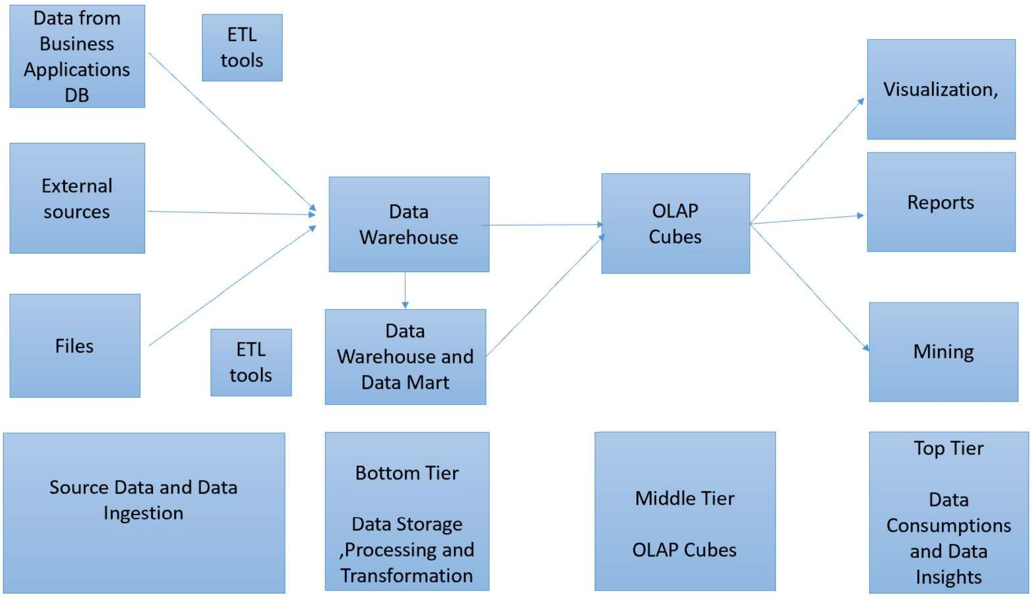
Data from Business Applications DB
ETL tools
External sources
Files
ETL tools
Data Warehouse
Data Warehouse and Data Mart
OLAP Cubes
Visualization,
Reports
Mining
Source Data and Data Ingestion
Bottom Tier
Data Storage ,Processing and Transformation
Middle Tier
OLAP Cubes
Top Tier
Data Consumptions and Data Insights

***Figure 1.3**: Data warehouse platform concept*

Let us now focus on why this framework must evolve further and the organization started adopting a data lake.

Data Lake fundamentals and limitations

In the past two decades, the amount of data that is generated is more than what mankind generated in history. In 2006, a British mathematician coined the phrase, "Data is the new oil." We observed the data storm when smart devices and smart applications started evolving like iPhone, Uber/Grab, YouTube, Netflix, Facebook, and WhatsApp. The latest smartphones generate tons of data, including photos, videos, global positioning data, application Telemetry and many more. Devices like television, watches, fridges, and wearable devices like billions of consumer devices start connecting with the Internet, hence data platforms ended up with a variety of source data. The nature of this data was quite different from the structured type. Soon organizations felt a need to analyse such high-volume data, image files, video files, and Telemetry-related semi-structured files to gain more insights. *Figure 1.4* shows 188 Zeta bytes as 2025 world data volume prediction as per Statistica 2022 resources:

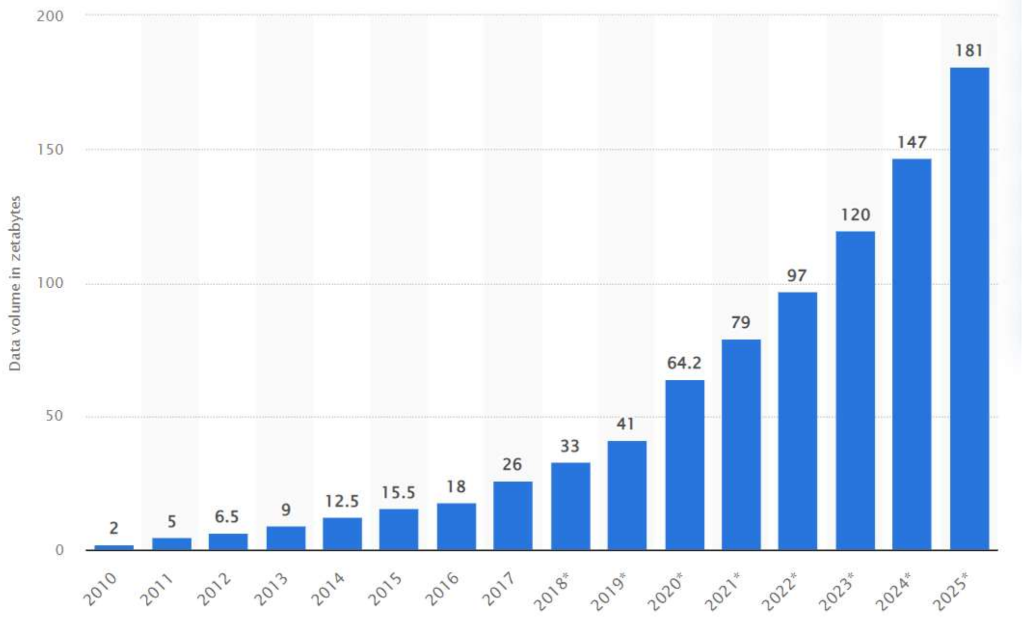

***Figure 1.4**: Worldwide data volume as per Statistica 2022*

Here are some fun facts on modern data trends from the findstack website, refer to the *Further read* section for more similar facts.

- Every human created about 1.7 mb of data per second in 2020.
- Companies generate around 2,000,000,000,000,000,000 bytes of data a day.
- It would take 181 million years to download all the data of the internet that exist today.
- As per IDC (International Data Corporation) there will be 41.6 billion IOT (internet of things) devices connected to the Internet by 2025.

Traditional data warehouse technology started showing cost versus performance challenges for such volumes of data. Also, data consumers needed a platform that can access raw data quickly, and apply complex logic, and algorithms as required to get desired output in real-time in a cost-effective manner. These features were not predominantly present in traditional data warehouse appliances. While some people also consider it as only the pre-staging area of a data warehouse, the data lake platform provides the following capabilities:

- **Raw Data flexibility**: The ability to access and apply computation on raw data files. Also, provides ease of use and access for all types of data (structured, semi-structured like JSON, XML, free text, and unstructured data like image files, and video files) and not just accessing processed structured data in tabular format.
- **Data Fidelity**: Since it keeps the data in the AS-IS format of business, it provides data fidelity to consumers.
- **Processing Capability**: This type of platform helps advanced data engineers apply related big data frameworks on raw data and thus process the PB scale of data.
- **Meant for all Data consumers**: Helps data scientists process algorithms on raw data based on artificial intelligence needs.
- **Support all file types**: Traditional Datawarehouse was lagging with processing capabilities for Video files, Image files which were solved by the Datalake problem.

The following figure illustrates the Data Lake platform concept:

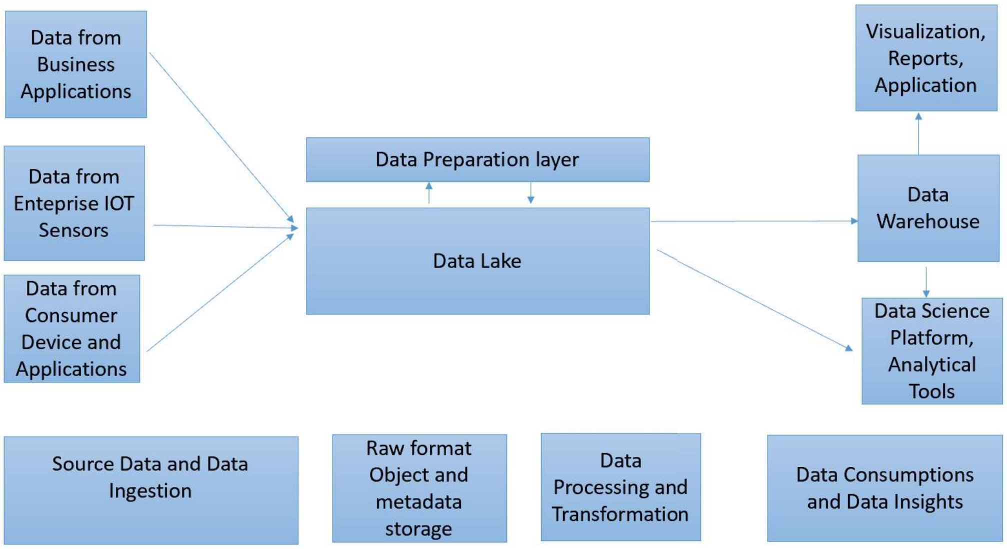
Data from Business Applications
Data from Enteprise IOT Sensors
Data from Consumer Device and Applications
Data Preparation layer
Data Lake
Visualization, Reports, Application
Data Warehouse
Data Science Platform, Analytical Tools
Source Data and Data Ingestion
Raw format Object and metadata storage
Data Processing and Transformation
Data Consumptions and Data Insights

Figure 1.5: *Data Lake concept*

Worldwide industry started showing high adoption. Here are very few high-level business scenarios in data lake practice across the industry.

- **Health Industry**: Analyzing clinical notes is important, however, it comes in different formats since it originates and stays in a different system. Analysing such data to get contextual information is quite helpful for medical practitioners. They can understand the profile of the patient easily and understand more what the diseases patient had, the severity of the illness and past medical history. This industry is transforming with super app-based telemedicine, teleconsulting, tele medicine-based delivery capability. Such intelligent app platforms are using cloud data lakes as a foundation to support this digital transformation.

- **Manufacturing Industry**: Industry 4.0 is a digital revolution for which the fundamental pillar is Industrial IOT (Internet of Things) supported by Analytics, Artificial intelligence, cloud, and other tech platforms. Smart and connected factories and intelligent and real-time supply chain visibility are a few capabilities which use data lake for analytics and AI computation purposes.

- **Automotive Industry**: Today's automotive industry brings a different experience to consumers. Connected vehicles provide real-time Telemetry information to all vehicle stakeholders starting from owner to care manufacturer for a better experience. The core of this industry digital data uses data lake for its data storage and computing need. Learn details on connected vehicle geospatial analytics use cases in the *Further read* section.

- **Aviation Industry**: Airlines generate huge volumes of data. Especially, when a flight moves from one location to another location it generates a TB scale of data. Using flight black box data building engine health, fuel efficiency, aircraft safety, risk predictive analytics, and prescriptive pilot training are some key use cases in the aviation analytics field and Datalake is always an integral part of such use cases to support these use cases.

Likewise, financial services, retail and all other industries use data lakes as their core pillar of digital transformation today. While data lake can deal with PB scale data-related problems, this framework also has its limitation. Data Lake started lagging in the following technical areas.

- **Transactional support**: Maintaining transactional logs, and concurrency and thus providing a certain level of consistency to data consumers.

- **Data Quality**: Since data continued to push in raw format, it is not any more curated data hence data quality started becoming concerned.

- **Data governance**: Industry-standard data governance was not always quite supported by the data lake platform.

- **Data swamp**: Since there was ingest first policy (and figuring out later what to do with that data) soon it started becoming Data Swamp due to a lack of governance, and quality.

We will learn about the latest phase of data evolution in the next section which will address a few such data lake limitations.

Concept of Lake house, best of two worlds

In this section of the chapter, we will focus on Data Lakehouse = Datawarehouse + Data Lake. This capability brings benefits to both the data warehouse and data lake world. It also aims to solve some of the problems which were introduced by either the Data warehouse or Data Lake. The following figure illustrates the Data Lake platform concept:

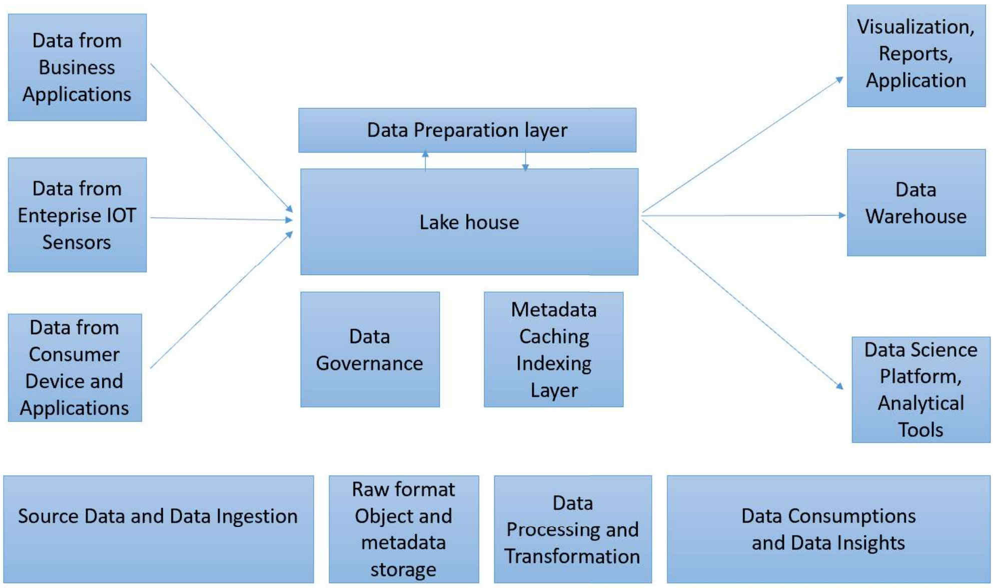
Data from Business Applications
Data from Enteprise IOT Sensors
Data from Consumer Device and Applications
Data Preparation layer
Lake house
Data Governance
Metadata Caching Indexing Layer
Visualization, Reports, Application
Data Warehouse
Data Science Platform, Analytical Tools
Source Data and Data Ingestion
Raw format Object and metadata storage
Data Processing and Transformation
Data Consumptions and Data Insights

***Figure 1.6**: Lake House concept*

Let's walk through some of the problems solved in Lakehouse:

- **Transaction-related consistency**: Atomicity, consistency, isolation, and durability capability is essential in any enterprise data platform. As highlighted in the previous section, this was present in the data warehouse but was missing in the Data Lake architecture. Lakehouse solution inherited this capability from the data warehouse and help achieve concurrent read and write operations for high-volume data scenarios. Also, it supports schema enforcement, and updates insert (upsert) capability to the data platform.

- **Data Governance**: Every enterprise data platform needs industry-standard security and data governance. Data Lake solution had a lack of data governance and data security. For a regulated industry implementing a data lake would need a complex ecosystem to support such robust data governance. Lakehouse brought better data versioning and security and enforce industry-standard data governance.
- **Cost versus performance**: Data warehouse computation's cost was higher for larger data volume operations. Data Lake implemented cost-optimized storage capability and hence made such large data volume-based computation operations cost-effective and subsequently, this capability is introduced in Lake house.
- **Openness**: It leverages industry standard open formats like parquet so that all data frameworks and tools can access the data seamlessly.
- **Artificial Intelligence support**: Lakehouse provides native support to analytics, machine learning and artificial intelligence (AI) tools by directly allowing them to access source data using its native open data formats and APIs.
- **Support all types of data**: Ability to store raw data in structured, semi-structured, and unstructured formats like datalake capability.
- **Real-time analytics**: Streaming support is essential for a well-architected and simplified data platform. However, often it was not present in data warehouse products. On the contrary, Data Lake architecture provided support for such streaming capabilities and generate insights from the high volume of streaming data. This capability is inherited from the Data Lakehouse to support high velocity streaming data.

Note that lake houses are born in the clouds only. In our next section, we will focus on the basics of the cloud. Understanding cloud capability is essential before jumping into cloud analytics services.

Introduction of Cloud

Before we learn the cloud analytics capability provided by the cloud, let us take one step back, what is cloud computing? As per Microsoft documentation '*cloud computing is the delivery of computing services—including servers, storage, databases, networking, software, analytics, and intelligence—over the Internet ("the cloud") to offer faster innovation, flexible resources, and economies of scale*'. Simply put, instead of using our /organization's infrastructure we lease infrastructure as per our need from other service providers There are few major cloud service providers in the market and the following list depicts major public cloud service providers and corresponding cloud URLs.

- **Microsoft Azure (Azure): https://portal.azure.com**
- **Amazon web services (AWS): https://aws.amazon.com/**
- **Google Cloud platform (GCP): https://console.cloud.google.com/**

The following figure contains a logo to depict a few more cloud service providers:

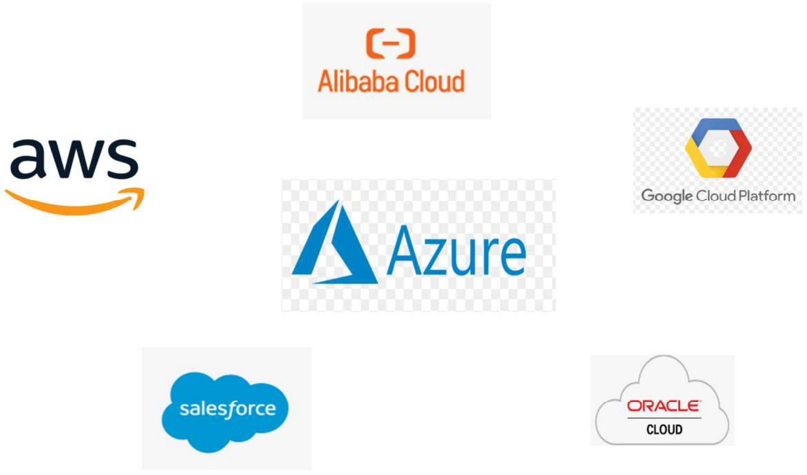

***Figure 1.7**: Major Public cloud provider*

Cloud tech stack provides a wide variety of **Infrastructure as service** (**IAAS**), **Platform as service** (**PAAS**) services in the network, infra, application, database, Analytics, AI, and so on the field. Overall high-level cloud data platform provides huge benefits when compared to on-premise data platform products, few such benefits are depicted as follows:

- **Elastic Scalability**: Most of the cloud-native services can be scaled up or scaled out, hence providing platform elasticity. These scalings are manual or automated based on service by service. Usually, on-premises servers and appliances lack such scaling.
- **Storage and computation decoupled**: Cloud services are usually decoupled in nature, and storage and computation are decoupled. Hence multiple flavours of computation like spark-based computation, sql based computation can work on the same storage layer to achieve the right outcome. Thus, the cloud platform provides more flexibility around computation service and also avoids any specific vendor lock-in.
- **Pay-per-use model**: Cloud-native services provide flexibility and charge mostly based on pay per usage model, hence data platform cost optimization can be achieved by wisely using computation as and when required.

- **Industry compliance:** Industry standard security and compliance maintained by cloud-native products. Learn more about Microsoft industry compliance in the *further read* section.
- **Data centre resiliency**: Cloud data centre infrastructures provide resiliency as per the facility architecture plan.
- **Ease of Infrastructure management**: It's easy to create an infrastructure in the cloud without having much infra management.

In the following section, we will be focusing on the cloud analytics field.

What is a Cloud Analytics platform?

Cloud analytics platform leverages cloud data platforms to perform analytics-related data storing and analyse operations, subsequently gets business actionable insights. Since cloud analytics inherit all such previous section mentioned cloud computing benefits like scalability, detached computation, pay-per-usage model, enterprise standard security, and industry compliance, it provides a resilient, cost-effective analytics platform and usually outperforms while compared with on-premises analytics tools. Usually, lake house provides SQL endpoint like traditional data warehouse hence the usage of the lake house is simple for Data analysts like Datawarehouse usage however in a cost-effective manner. The following figure illustrates the concept of cloud analytics:

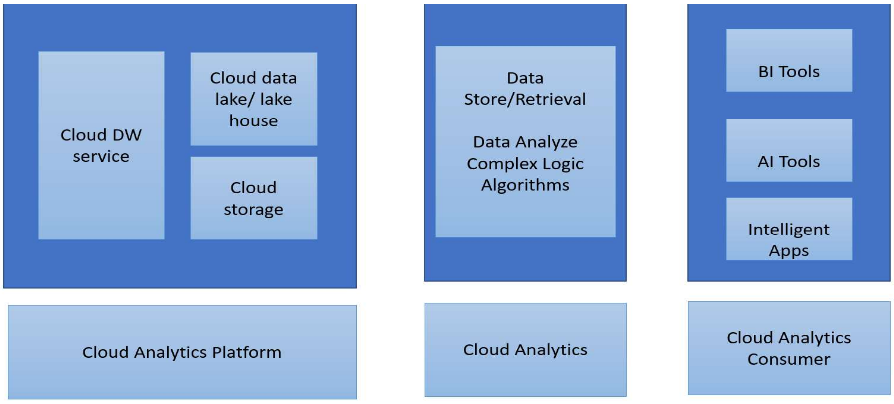

***Figure 1.8:** Cloud analytics concept*

Today the market is predominantly using various offerings of cloud data warehouse, cloud data lake and cloud lake house to fulfil their platform need. Microsoft offers Azure synapse analytics, Azure HDinsight, AWS offerings Redshift, EMR, Google offerings big query, big lake, Databricks, Snowflake is a few such key players. In our book, we will focus on Azure synapse analytics key capability from the next chapter onwards.