

Case Study on Atlas Lakehouse



TABLE OF CONTENTS

1 Atlas lake house	4
2 Why choose Atlas lake house	4
3 Features	5
4 Flexible for multiple sectors	7
5 Components	8
6 Use Case	10
7 Challenges	10
8 Proposed solution to customer	10



Atlas lake house:-

Atlas lake house delivers a user-friendly and reliable data integration platform for organizations with growing data needs. You can use Atlas Lake house to automate collecting data from multiple applications and databases, loading it to a data warehouse, and making it analytic-ready. This allows your analysts to set up faster analysis and reporting. The Atlas lake house product is delivered through an extremely intuitive user interface that eliminates the need for technical resources to set up and manage your data Pipelines. Atlas lake house's design approach goes beyond data Pipelines. Its analyst-friendly data transformation features are well integrated into the platform to further streamline analytic tasks.

Why choose Atlas lake house

Data Lakes allow you to import any amount of data that can come in real-time. Data is collected from multiple sources and moved into the data lake in its original format. This process allows you to scale to data of any size while saving time in defining data structures, schema, and transformations.







Features:

1. Data source registration module

- The first step in this module is to establish a connection with the source database. We will get server and port details from the user. If the connection is successful, we will proceed.
- We will use Rest API if the data source is a SAAS application, such as Google Analytics.
- In the case of database Sources, you can load data from one, multiple, or all databases. For each selected object or table, you can specify how the records to be ingested must be queried, through the Query Mode.



• Similarly, for SaaS Sources, you can select the objects or tables to be fetched or the report's data to be retrieved and specify the query mode.

2. Use case register module

 \cdot We will take below use case details from the user:

- ϕ The size of the data
- Ø Frequency of data change
- ϕ Frequency of data load- daily or real-time etc.

3. Key-Partition Module

- Our data will be placed on S3 once we receive it from different sources.
- The partitioning of the keys will need to be considered while storing them on S3.
- Partitioning allows the use of different file systems to be installed for different kinds of files. Separating user data from system data can prevent the system partition from becoming full and rendering the system unusable. Partitioning can also make backing up easier.

4. ETL Process Module

- Atlas lake house ETL Pipelines, fetch data from your data Source, perform in-flight Transformations based on the settings you configure, and load it to your Destination.
- In this module we will design an ETL process that will load data from S3 to the target.



- In this module we will perform the transformation.
- We will make the data compatible with target and use GLUE script and load them data to target.
- ETL is a process in Data Warehousing, and it stands for Extract, Transform, and Load. It is a process in which an ETL tool extracts the data from various data source systems, transforms it in the staging area, and then finally, loads it into the Data Warehouse system.

5. Visualization Module

Data visualization brings data to life, making you the master storyteller of the insights hidden within your numbers. Through live data dashboards, interactive reports, charts, graphs, and other visual representations, data visualization helps users develop powerful business insights quickly and effectively.

We will perform visualization using the following BI tools:

ØPower Bl ØTableau ØQuick Sight

Flexible for multiple sectors

ØBanking sector ØFinancial services ØRetail sector ØHealthcare



We are evolving and creating the Atlas lake house app with new and enhanced functionalities to better address the business needs and experiences of our customers.

Components:

A data lake holds a vast amount of raw data in its native format until it is needed.

1. Data ingestion:

A highly scalable ingestion-layer system that extracts data from various sources, such as websites, mobile apps, social media, IoT devices, and existing Data Management systems, is required. It should be flexible to run in batch, one-time, or real-time modes, and it should support all types of data along with new data sources.

2. Data Storage:

A highly scalable data storage system should be able to store and process raw data and support encryption and compression while remaining cost-effective.

3. Data Security:

Regardless of the type of data processed, data lakes should be highly secure from the use of multi-factor authentication, authorization, role-based access, data protection, etc.



4. Data Analytics:

After data is ingested, it should be quickly and efficiently analyzed using data analytics and machine learning tools to derive valuable insights and move vetted data into a data warehouse.

5. Data Governance:

The entire process of data ingestion, preparation, cataloging, integration, and query acceleration should be streamlined to produce enterprise-level data quality. It is also important to track the changes to key data elements for a data audit.





Use Case: -01

Financial Services Ltd has embarked on an initiative for Data Lake Solution on AWS

Challenges:

- ϕ Making product cloud agnostic
- ϕ Security of data in the cloud
- ϕ Add dynamic ETL script based on user option
- ϕ Create visualization on data as per user option

Proposed Solution to the customer:

- ϕ Quick Deployment for Data lake creation
- ϕ Easy to monitor and manage infrastructure
- ϕ User-friendly and customizable
- ϕ Simplistic hence any type of user can operate

 $\boldsymbol{\phi}$ Sits on top on customer cloud and not hosted outside customer

premises

- Ø Time Saving
- ϕ cost and manual work optimization



ABOUT ACC

ACC is an AWS Advance Partner with AWS Mobility Competency. Awarded The Best BFSI industry Consulting Partner for the year 2019, ACC has had several successful cloud migration and application development projects to its credit.

Our business offerings include Digitalisation, Cloud Services, Product Engineering, Big Data & Analytics and Cloud Security. ACC has developed several products to its credit. These include Ottohm – Enterprise Video and OTT Platform, Atlas API – API Management and Development Platform, Atlas CLM – Cloud Life Cycle Management, Atlas HCM – HR Digital Onboarding and Employee Management, Atlas ITSM – Vendor Onboarding and Service Management and Smart Contracts – Contract Automation and Management.



www.appliedcloudcomputing.com

Shubho Pramanik

(+91) 90297 20294

shubho@acc.ltd

Rogin Rappai

(+91) 8828478321

rogin.rappai@acc.ltd

also reach us at

