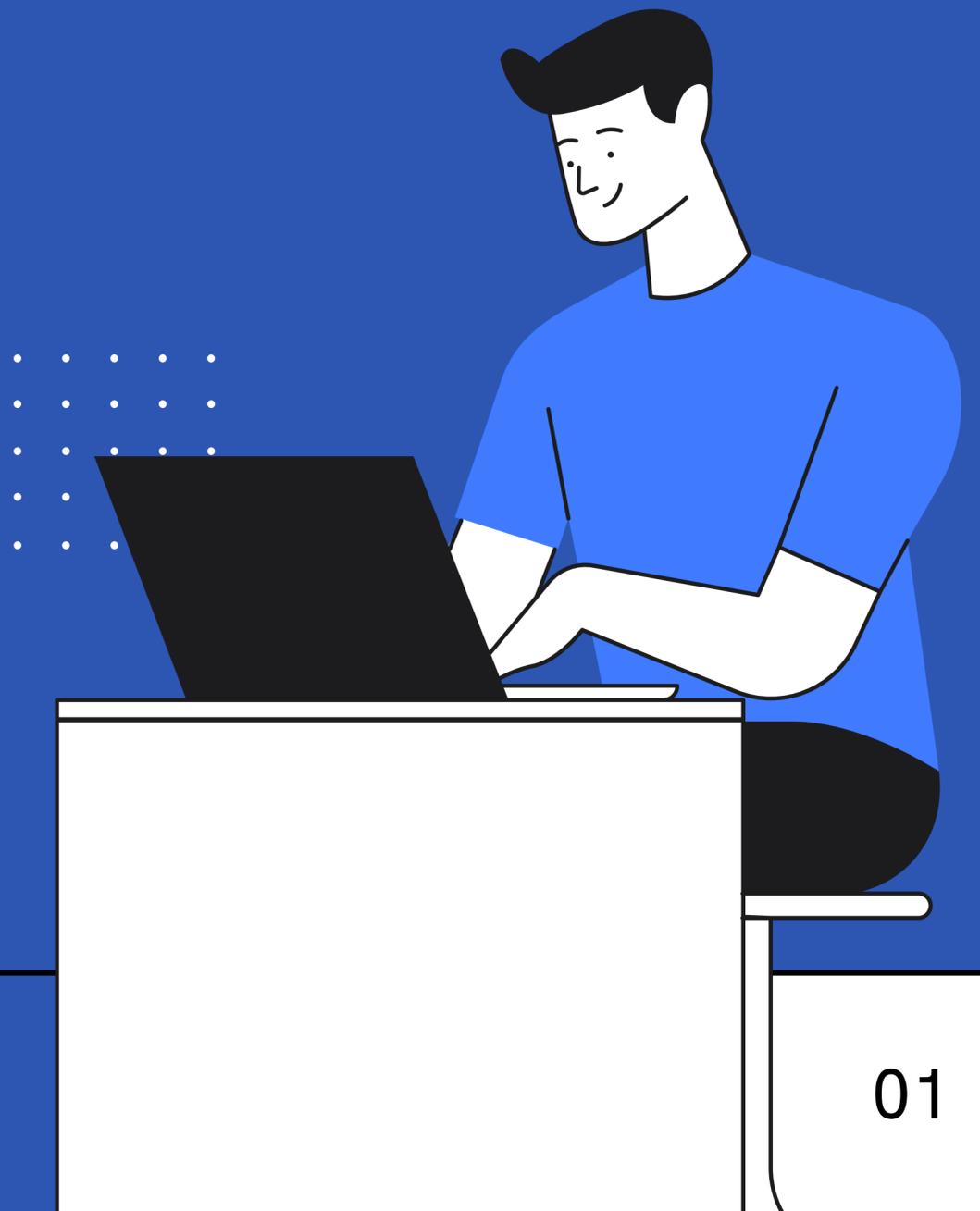
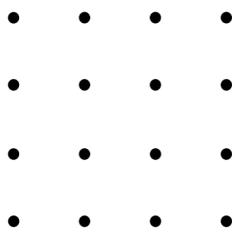


# Building your own modern Data Analytics Platform on AWS



# HARI PATEL

Managing Director - Softqube Technologies Pvt. Ltd.



I worked with many of companies from different industries (like Healthcare, Hospitality, Retail and Social Media) migrating their legacy data warehouses or analytical databases to the cloud. I saw the difficulty to let go of the monolithic thinking and design and to benefit from the modern cloud architecture fully. In this article, I'll share my pattern for a scalable, flexible, and cost-effective data analytics platform in the AWS cloud, which was successfully implemented in these companies.

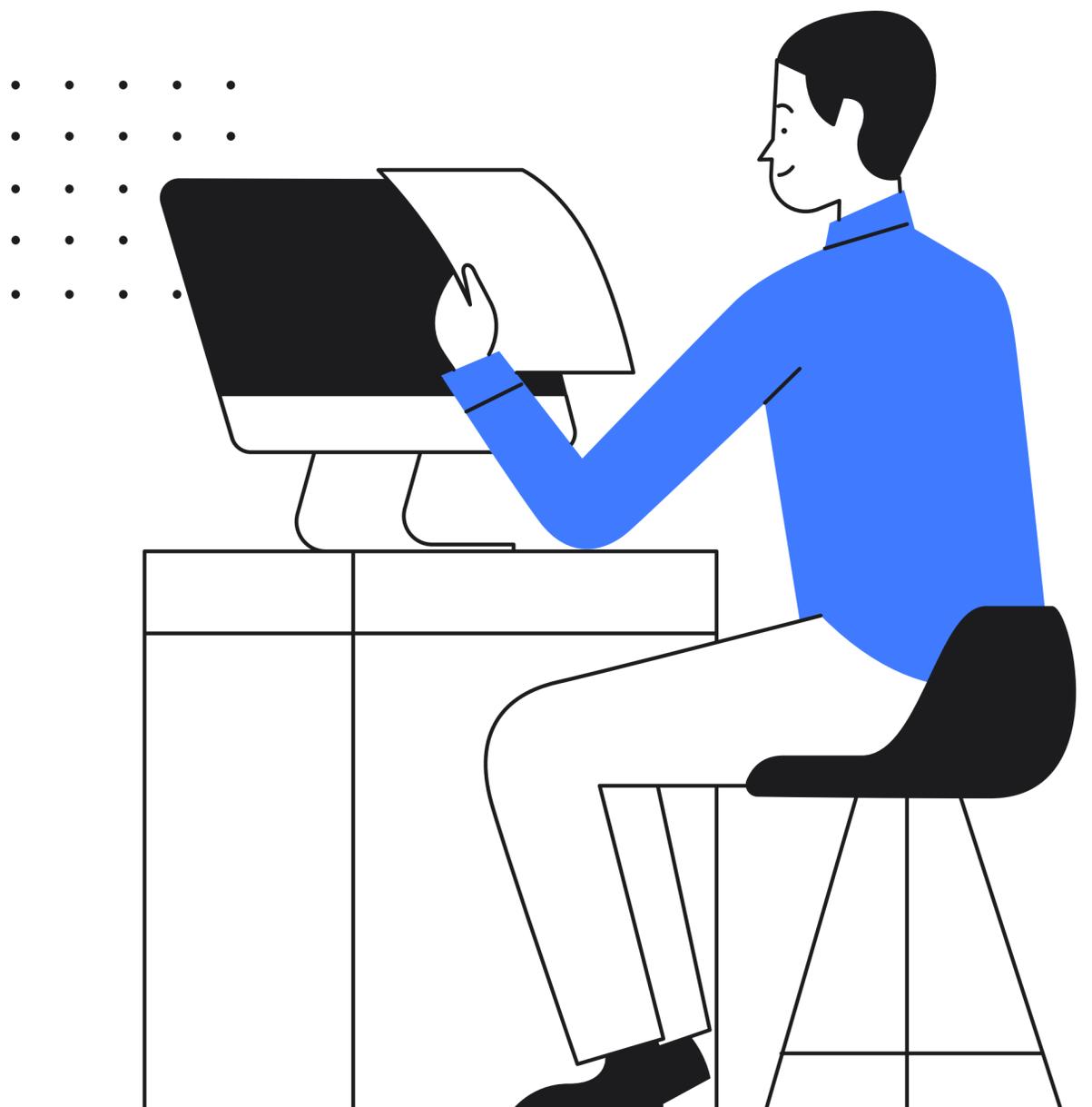


# CONTENT

Introduction .....	04
Process of Data Analysis .....	05
Output after Data Analysis .....	06
Building Data Analytics Platform .....	09
Data Lakes .....	10
AWS - Infrastructure Architecture .....	15
PMS moving to AWS .....	18
Conclusion .....	19

# Introduction

Data analytics is the science of analyzing raw data in order to make conclusions about that information. Many of the techniques and processes of data analytics have been automated into mechanical processes and algorithms that work over raw data for human consumption.



# What is the Process of Data Analysis?

Data Analysis is a process of

- **Inspecting**
- **Cleansing**
- **Transforming**
- **Modelling**

data with the goal of discovering useful information, suggesting conclusions, and supporting decision-making.

Data Analyst needs some skills must have high technical ability as much of the role focuses on complex databases, statistics, and formulas that need skills to analyze data like SQL, DATA mining, OLAP, Reports, statistics etc.

## What will be

# Output After Data Analysis?

- **Predictive Analytics** : This method basically looks at future outcomes using historical **data**. The goal is to determine what might happen in the future so that companies can make better decisions. This term specially used in **Aviation** and **Hospitality** to predict day to day pricing based on occupancy, demand and inventory.
- **Business Analytics** : Business analytics (BA) refers to all the methods and techniques that are used by an organization to measure performance. Business analytics are made up of statistical methods that can be applied to a specific project, process or product. Business analytics can also be used to evaluate an entire company. Business analytics are performed in

order to identify weaknesses in existing processes and highlight meaningful data that will help an organization prepare for future growth and challenges. This term specially used in **Healthcare** for analyzing insurance claim.

- **Customer Analytics** : Customer analytics uses data collection and subsequent software analysis to zero-in on customers' online order transactions for the purpose of sorting out specific customer demographics, shopping patterns, internet usage and applying predictive analyses to allow marketers to take measures to increase online business profit margins. This term specially used in **CRM** kind of System and **Retail industry**.
- **Business Impact Analysis** : A business impact analysis can work on several levels. An organization-wide BIA will identify large risks and their fallout, such as losing

all the data on the company servers. A branch or department BIA will highlight the particular systems that are critical to those users and what would happen if they were unable to access them. The overall business continuity plan will contain steps for addressing any problems, but a BIA can be very helpful in spotting obvious weaknesses before they become serious issues. This term specially used in **Hotel Industry.**



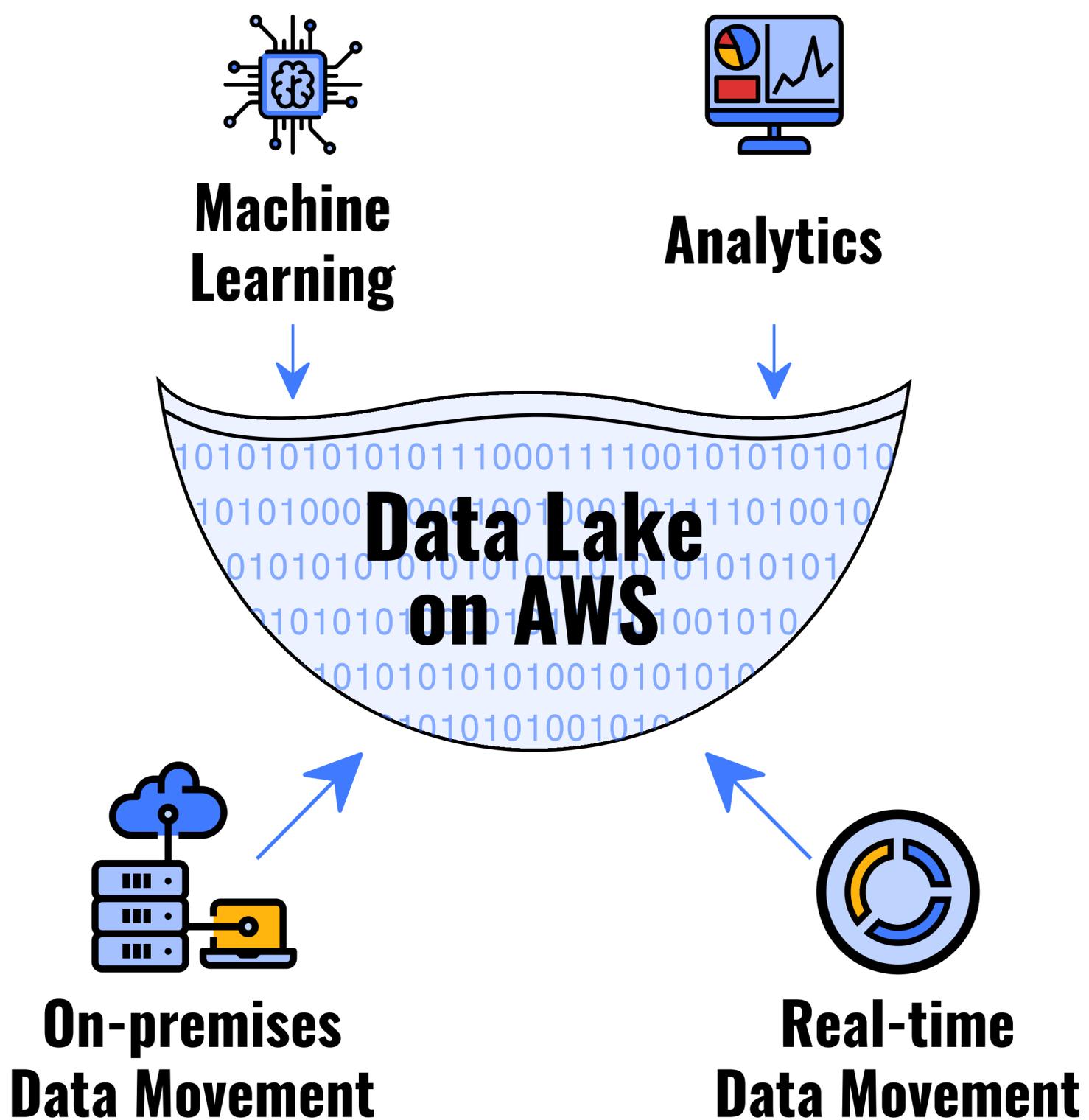
# How to use AWS Cloud Computing for Building Data Analytics Platform?

AWS delivers an integrated suite of services that provide everything needed to quickly and easily build and manage a data lake for analytics. AWS-powered data lakes can handle the scale, agility, and flexibility required to combine different types of data and analytics approaches to gain deeper insights, in ways that traditional data silos and data warehouses cannot.

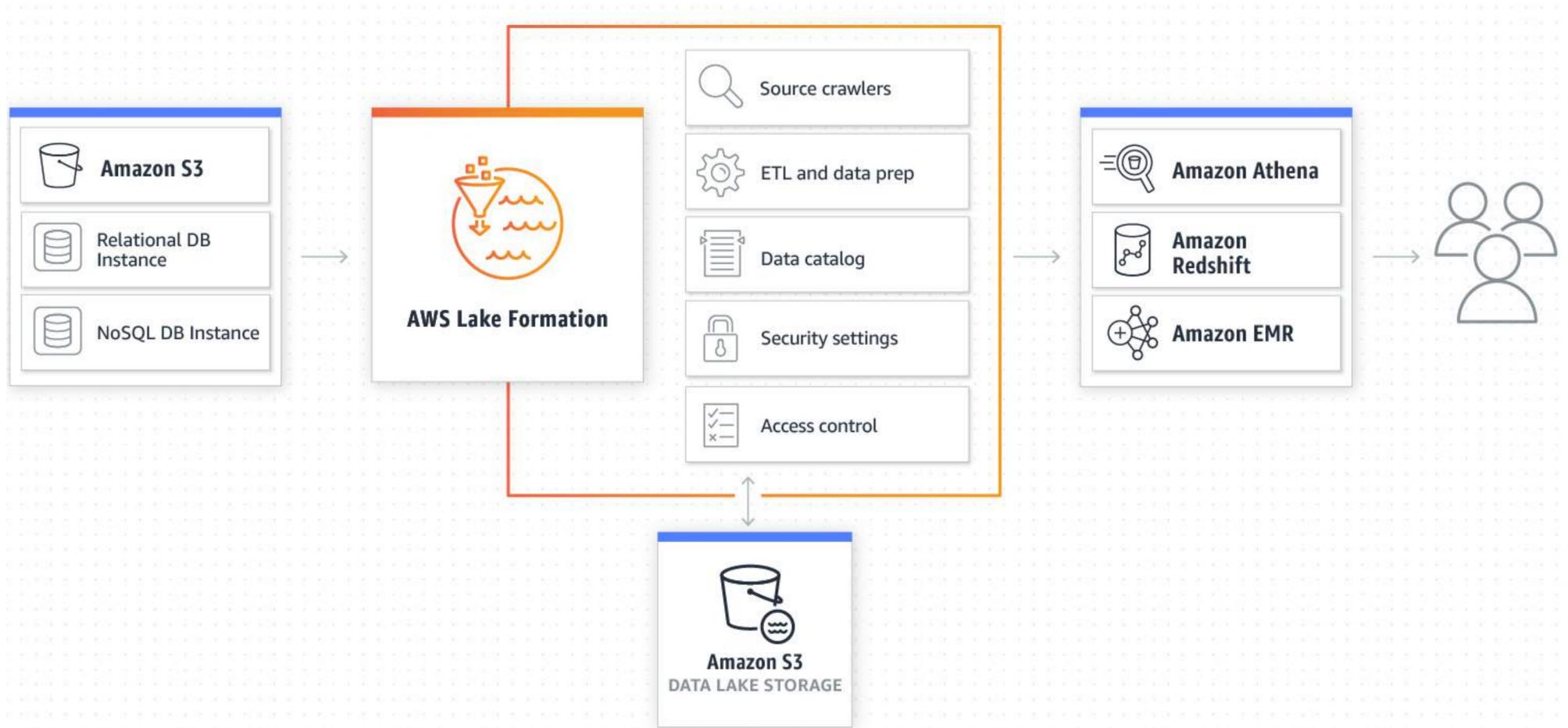


# What is Data Lakes?

To build your data lakes and analytics solution, AWS provides the most comprehensive set of services to move, store, and analyze your data.



**Data Movement : Extracting data from different sources like (SFTP, FTP, AWS S3 Bucket, Dropbox, Google Drive or On Premise HDD) and different types of data structures like (XML, CSV, PDF, DOC, EXCEL, JSON or TEXT).**



The first step to building data lakes on AWS is to move data to the cloud. The physical limitations of bandwidth and transfer speeds restrict the ability to move data without major disruption, high costs, and time. To make data transfer easy and flexible, AWS provides the widest range of options to transfer data to the cloud. To build ETL jobs and ML Transforms for your data lake via **SSIS** or **AWS Glue Services**.

The AWS Services we can use for Data Movement are as below:

- [Direct Connect For On-Premise Data Movement](#)
- [IoT For Real-Time Data Connect](#)

**Data Lake : Store any type of data securely, on different Database Systems like (MongoDB, MySQL, MS SQL, ORACLE, DynamoDB) from gigabytes to exabytes.**

Once data is ready for the cloud, AWS makes it easy to store data in any format, securely, and at massive scale with Amazon S3, AWS Redshift or Amazon Glacier. To make it easy for end users to discover the relevant data to use in their analysis, AWS Glue automatically creates a single catalog that is searchable, and queryable by users.

The AWS Services we can use for Data Lake are as below :

- [S3 For Cloud Storage](#)
- [Glaciar For BackUp And Archive](#)
- [Glue For Data Catalogue](#)

## **Analytics : Analyze your data with the broadest selection of analytics services or algorithm.**

AWS provides the broadest, and most cost-effective set of analytic services that run on the data lake. Each analytic service is purpose-built for a wide range of analytics use cases such as interactive analysis, big data processing using Apache Spark and Hadoop, data warehousing, real-time analytics, operational analytics, dashboards, and visualizations.

The AWS Services we can use for Analytics are as below :

- **Athena For Interactive Analysis**
- **EMR For Big Data Processing**
- **Redshift For Data Warehousing**
- **Kinesis For Realtime Analytics**
- **Elasticsearch For Operational Analytics**
- **QuickSight For Dashboard And Data Visualization**

## **Machine Learning: Predict future outcomes, and prescribe actions for rapid response.**

For predictive analytics use cases, AWS provides a broad set of machine learning services, and tools that run on your data lake on AWS. Our services come from the knowledge and capability we've built up at Amazon, where ML has powered Amazon.com's recommendation engines, supply chain, forecasting, fulfillment centers, and capacity planning.

The AWS Services we can use for Machine Learning (ML) are as below :

- **Deep Learning AMIs For Frameworks And Interfaces**
- **Sagemaker For Platform Services**

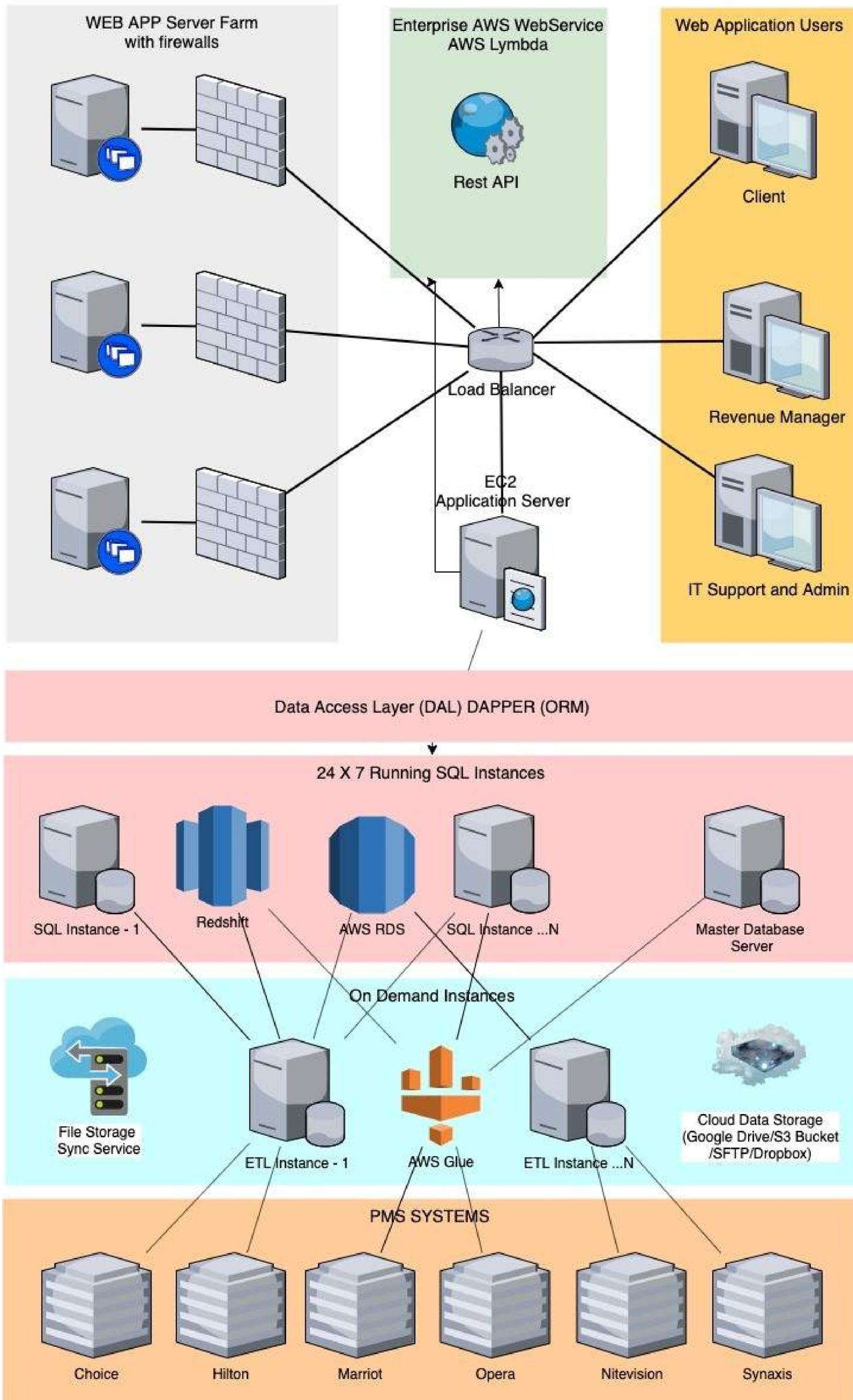
# Extract, Store and Analyze Your Hotel Data with AWS

Hotels' data is generated at an ever-increasing rate and is predicted to reach 35 exabytes by 2020. Being able to cost-effectively and securely manage this data whether for reservation data, guest data or rate shop data is increasingly important for different PMS vendors.

AWS is well-suited to this data deluge with a wide variety of ingestion, storage and security services (e.g. AWS Direct Connect, Amazon Kinesis Streams, Amazon S3, Amazon Macie) for customers to handle their Hotel Reservation data.

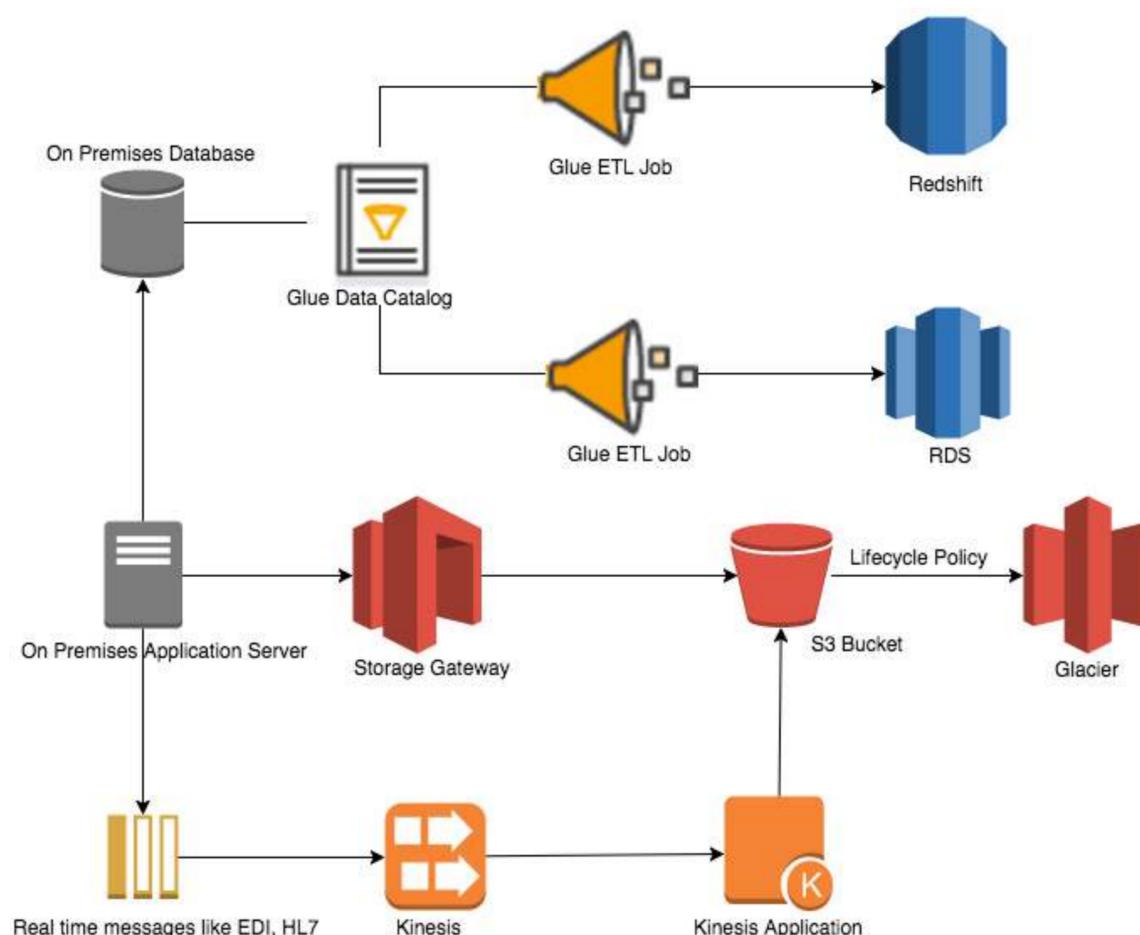
I predict that five years from now none of us will have datacenters. We're going to go out to the cloud to find forecast, revenue and predictive analytics.

# AWS - Infrastructure Architecture



## Ingestion :

A typical provider data center will consist of many systems with varied datasets. AWS provides multiple tools and services to effectively and securely connect to these data sources and ingest data in various formats. The customers can choose from a range of services and use them in accordance with the use case.



For use cases involving one-time (or periodic), very large data migrations into AWS, customers can take advantage of AWS Snowball devices. These devices come in two sizes, 50 TB and 80 TB and can be combined together to create a petabyte scale data transfer solution.

# PMS moving to AWS

- [Wyndham Hotels & Resorts Completely Moving To AWS Cloud](#)
- [Choice Hotels Goes All In On AWS Cloud](#)
- [Best Western Hotels And Resorts Go All-In On AWS](#)

## Key Terms

**Algorithm** : An algorithm is set of instructions for solving a problem or accomplishing a task. Every computerized device uses algorithms to perform its functions.

# Conclusion

As long as you are curious and able to learn new and better technologies, in the rapid pace of technological advancements we live in, you can build and operate a powerful and modern data platform. This data platform is an essential part of the **digital transformation** and **AI transformation** of every company that wants to stay relevant and competitive today.

