A Practical Report Submitted in fulfillment of the Degree of

**MASTER OF SCIENCE**

In

**COMPUTER SCIENCE**

Year 2024-2025

By

**MR. Yogesh Vishwas Dhamke**

**(Application ID: 41752)**

Under guidance of

**DR. SUJATHA IYER**



Institute of Distance and Open Learning, Vidya Nagar, Kalina, Santacruz East- 400098.

University of Mumbai



**INSTITUTE OF DISTANCE AND OPEN LEARNING**

**Vidya Nagari, Kalina, Santacruz East – 400098 CERTIFICATE**

This is to certify that,

Mr. **Yogesh Vishwas Dhamke**, Application ID: **41752,**

Student of Master of Science in Computer Science has Satisfactorily Completed the Practical in

**Business Intelligence and Big Data Analytics**.

**Name Application ID**

Mr. Yogesh Vishwas Dhamke 41752

**Subject In-charge Examiner**

**INDEX**

|  |  |  |  |
| --- | --- | --- | --- |
| **Sr No.** | **Date** | **List of Practical** | **Signature** |
| 1. |  | Create tables using different applications. |  |
| 2. |  | Develop an application to design a warehouse by importing various tables from external  sources. |  |
| 3. |  | Develop an application to pre process data imported from external sources. |  |
| 4. |  | Develop an application to demonstrate operations: 1. Roll-up. 2. Cube |  |
| 5. |  | Create association rules by considering suitable parameters – Apriory Algorithm |  |
| 6. |  | Case study - the development of a data warehouse for an e-commerce company |  |
| 7. |  | Case study- Effectiveness of Hadoop as an open source companion to standard data warehouses |  |
| 8. |  | Case study- Emerging trends on cloud based business intelligence and analytics. |  |

# PRACTICAL 1:

**CREATE TABLES USING DIFFERENT APPLICATIONS.**

# Methods:

1. **Table Creation – using SQL**

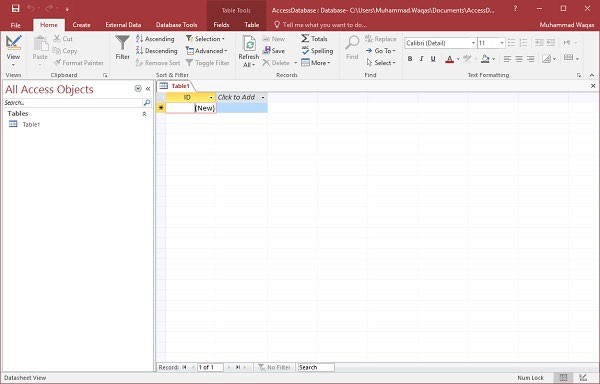
CREATE TABLE Persons ( PersonID int,

LastName varchar(255), FirstName varchar(255), Address varchar(255), City varchar(255)

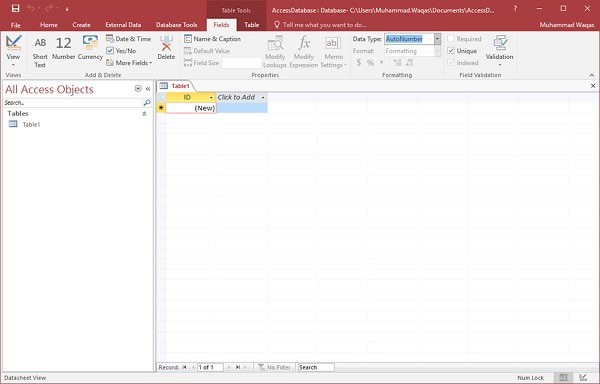
);

# Table Creation – using Ms Access

|  |  |
| --- | --- |
| Field Name | Data Type |
| EmployeelD | AutoNumber |
| FirstName | Short Text |
| LastName | Short Text |
| Address1 | Short Text |
| Address2 | Short Text |
| City | Short Text |
| State | Short Text |
| Zip | Short Text |
| Phone | Short Text |
| Phone Type | Short Text |

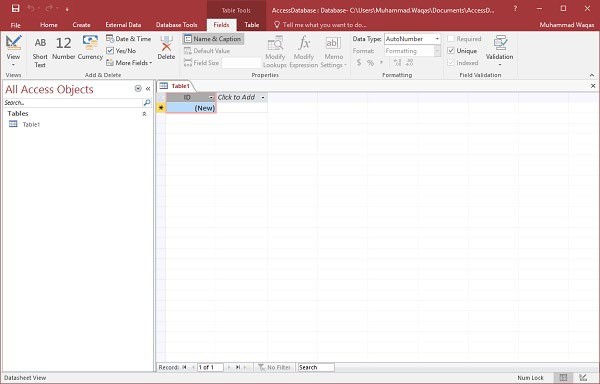


This is where we left things off. We created the database and then Access automatically opened up this table-one-datasheet view for a table.

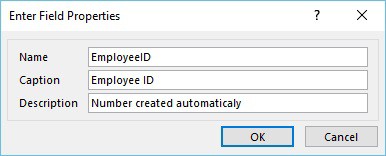


Let us now go to the Field tab and you will see that it is also automatically created. The ID which is an AutoNumber field acts as our unique identifier and is the primary key for this table.

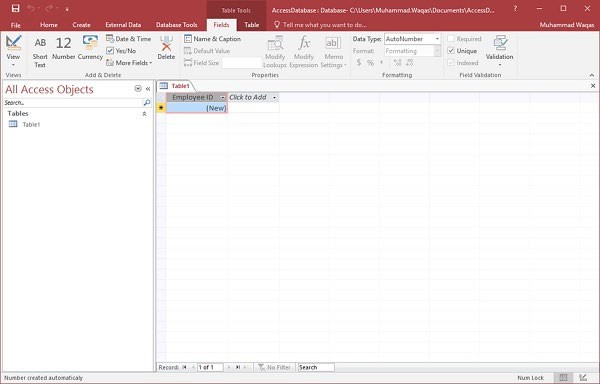
The ID field has already been created and we now want to rename it to suit our conditions. This is an employee table and this will be the unique identifier for our employees.



Click on the Name & Caption option in the Ribbon and you will see the following dialog box.

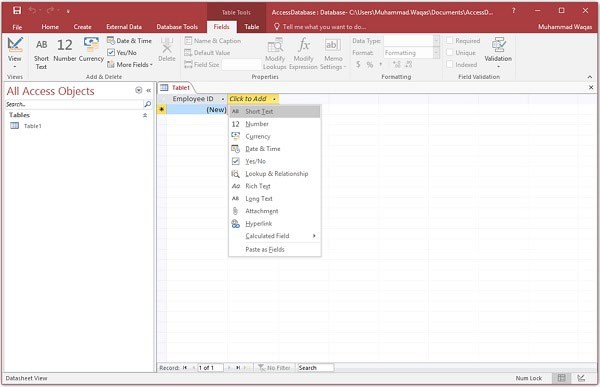


Change the name of this field to EmployeeID to make it more specific to this table. Enter the other optional information if you want and click Ok.

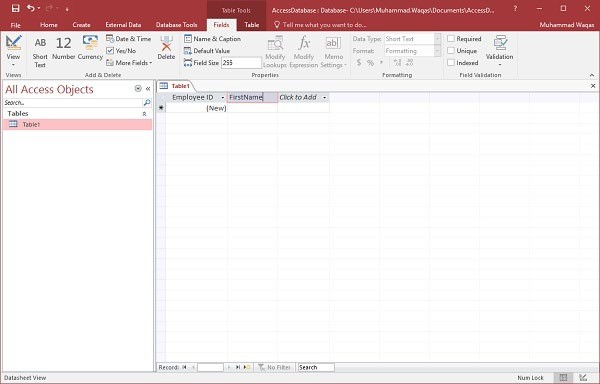


We now have our employee ID field with the caption Employee ID. This is automatically set to auto number so we don't really need to change the data type.

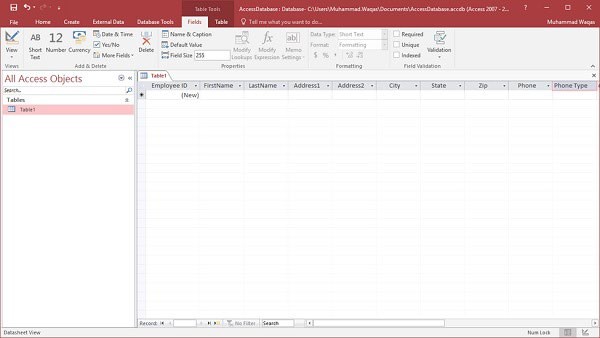
Let us now add some more fields by clicking on click to add.



Choose Short Text as the field. When you choose short text, Access will then highlight that field name automatically and all you have to do is type the field name.

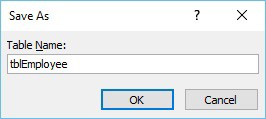


Type FirstName as the field name. Similarly, add all the required fields as shown in the following screenshot.

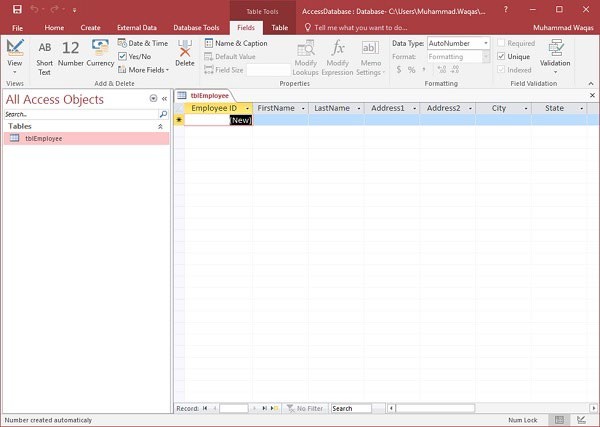


Once all the fields are added, click the Save icon.

You will now see the Save As dialog box, where you can enter a table name for the table.



Enter the name of your table in the Table Name field. Here the tbl prefix stands for table. Let us click Ok and you will see your table in the navigation pane.

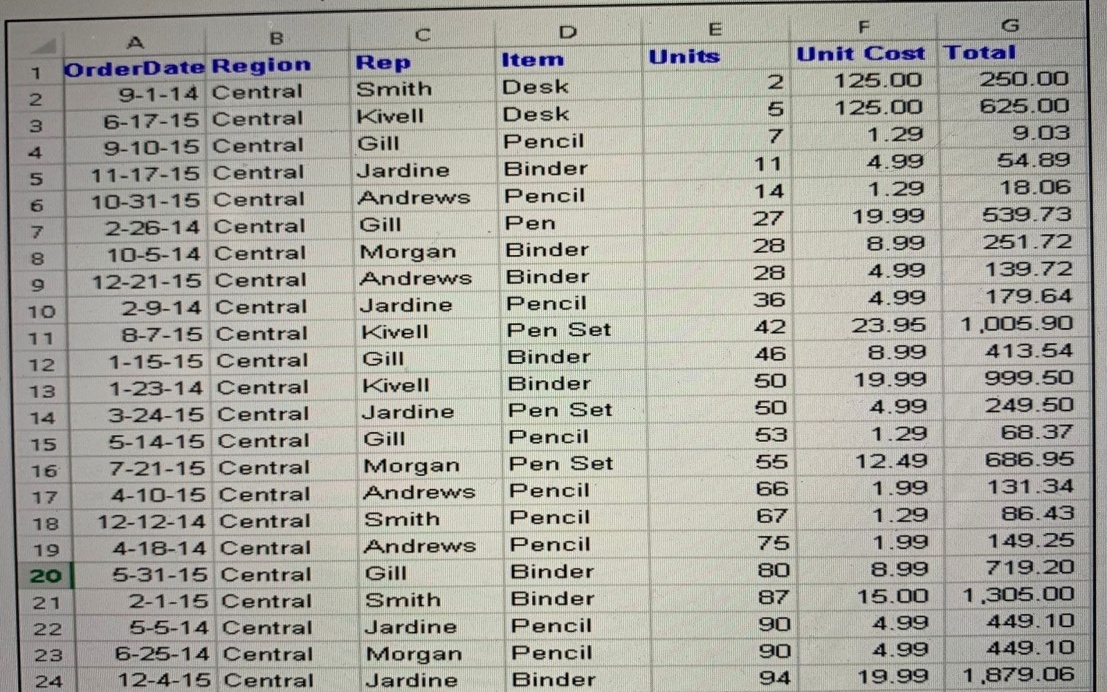


# PRACTICAL 2:

**DEVELOP AN APPLICATION TO DESIGN A WAREHOUSE BY IMPORTING VARIOUS TABLES FROM EXTERNAL SOURCES.**

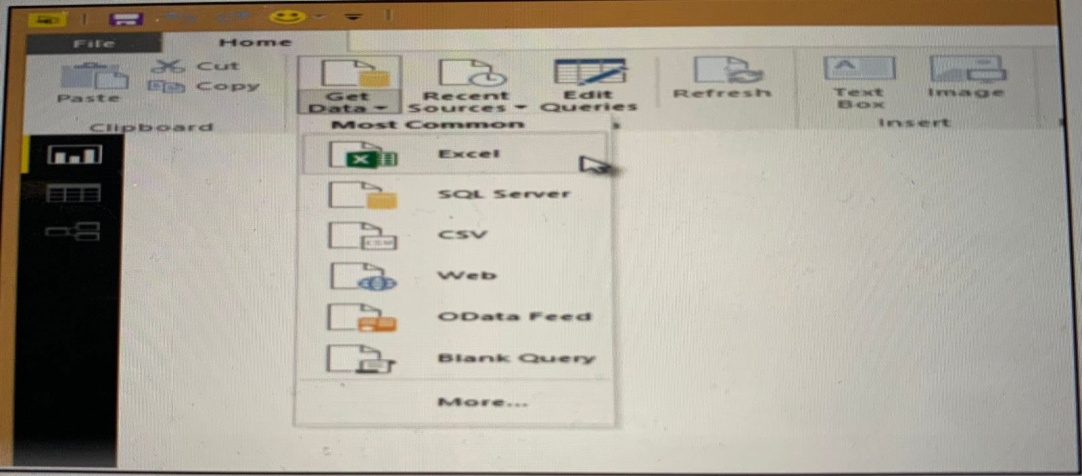
# Steps:

**Step 1:** Create an excel sheet with data

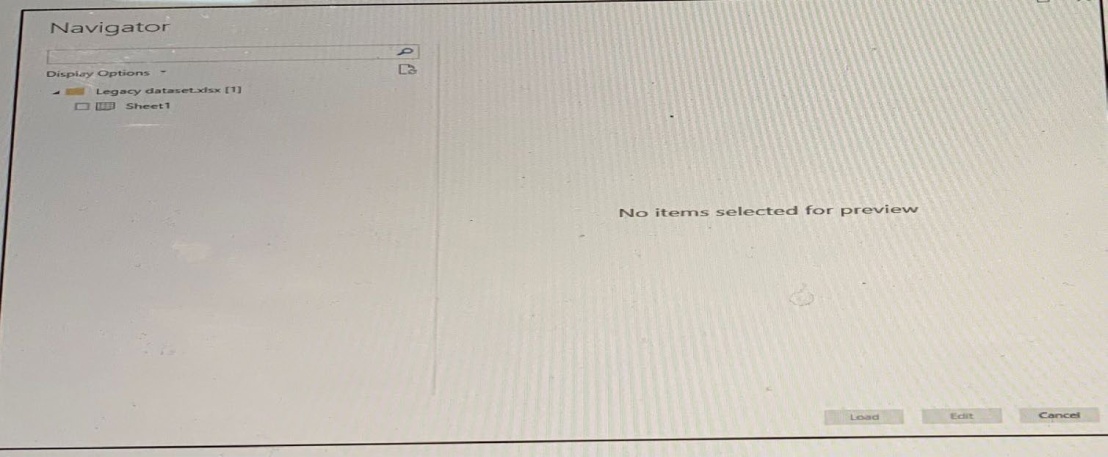


**Step 2**: Open Power BI desktop

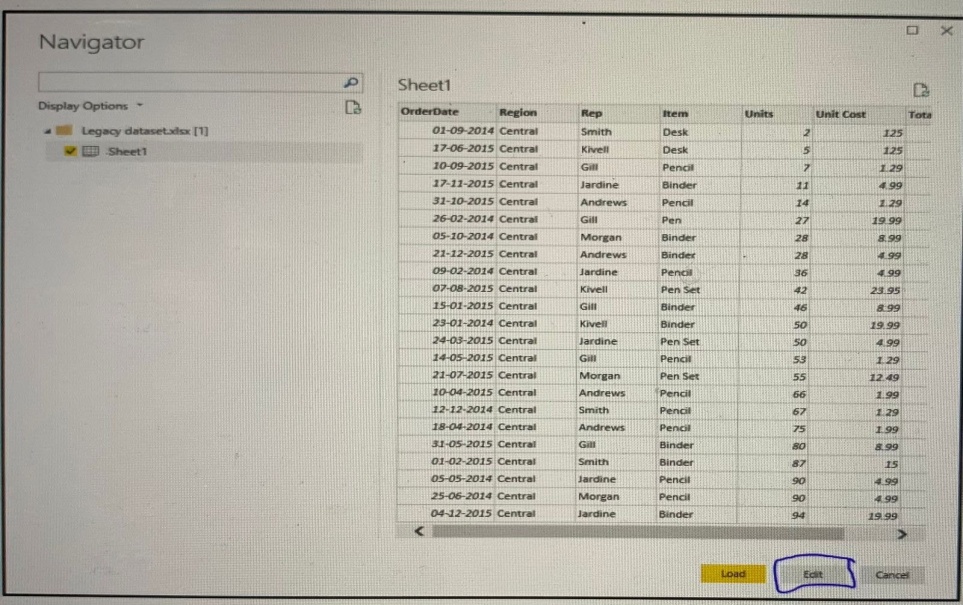
**Step 3:** Go to Home Ribbon-> Get Data-> Excel and browse your excel file



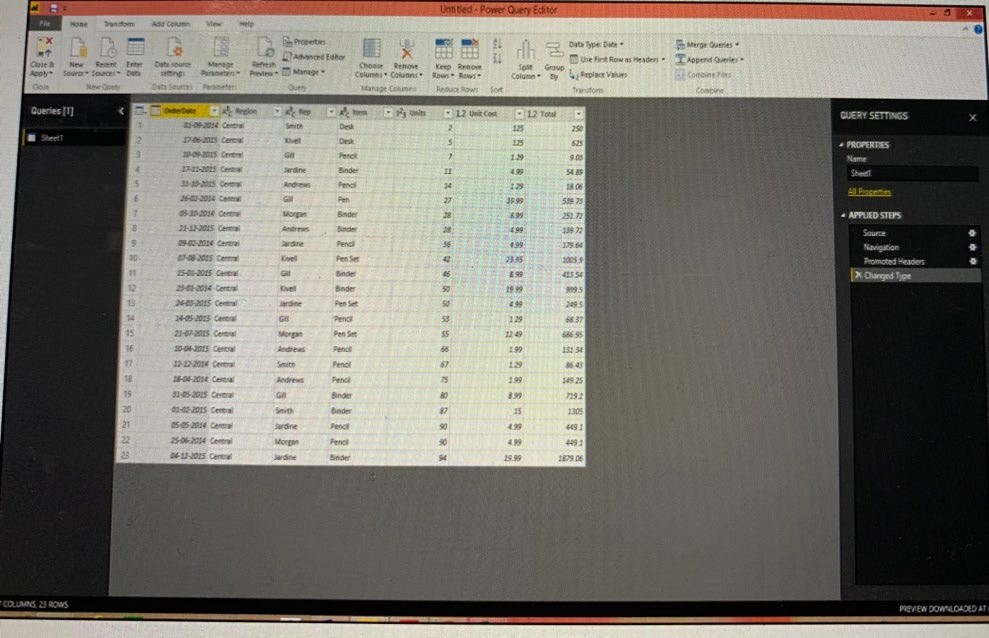
**Step 4**: In the Navigator tab, select your table (Sheet1) from your dataset (Legacy dataset.xlsx)



**Step 5**: Click Edit



**Step 6:** You will obtain this screen for queries



# DEVELOP AN APPLICATION TO PRE-PROCESS DATA IMPORTED FROM EXTERNAL SOURCES.

**Methods**:

# Method 1:

Importing Excel Data

1. Launch Power BI Desktop.
2. From the Home ribbon, select Get Data. Excel is one of the Most Common data connections, so you can select it directly from the Get Data menu.

In the Open File dialog box, select the Products.xlsx file.

5) In the Navigator pane, select the Products table and then select Edit.

# Method 2:

Importing Data from O Data Feed

In this task, you'll bring in order data. This step represents connecting to a sales system. You import data into Power BI Desktop from the sample Northwind OData feed at the following URL, which you can copy (and then paste) in the steps below: <http://services.odata.org/V3/Northwind/Northwind.svc/>

# Connect to an OData feed:

1. From the Home ribbon tab in Query Editor, select Get Data.
2. Browse to the OData Feed data source.
3. In the OData Feed dialog box, paste the URL for the Northwind OData feed.
4. Select OK.
5. In the Navigator pane, select the Orders table, and then select Edit.

# PRACTICAL 4:

**DEVELOP AN APPLICATION TO DEMONSTRATE OPERATIONS LIKE ROLL-UP.**

# Concept:

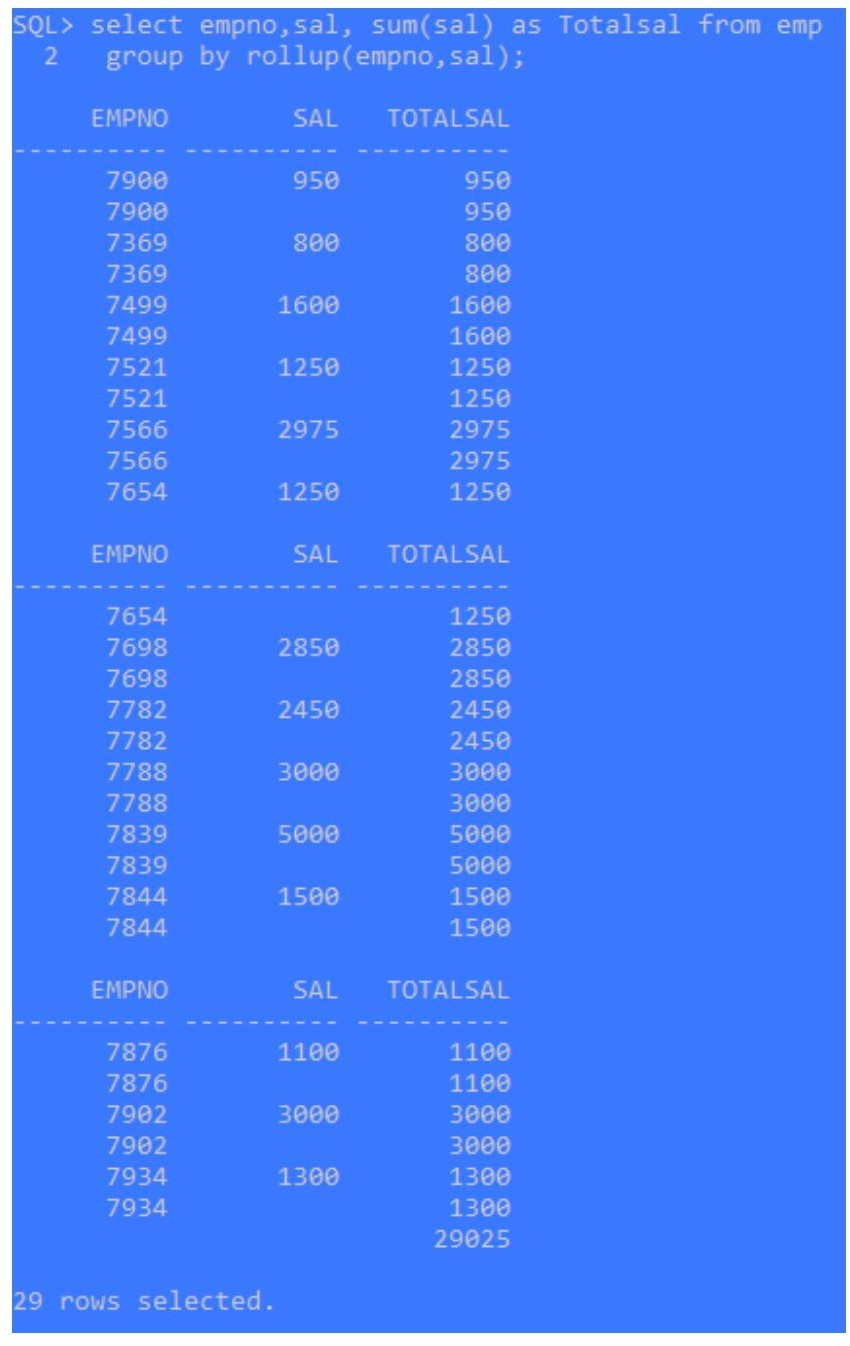
The last decade has seen a tremendous increase in the use of query, reporting, and on-line analytical processing (OLAP) tools, often in conjunction with data warehouses and data marts. Enterprises exploring new markets and facing greater competition expect these tools to provide the maximum possible decision-making value from their data resources. Oracle expands its long-standing support for analytical applications in Oracle8i release 8.1.5 with the CUBE and ROLLUP extensions to SQL.

Oracle also provides optimized performance and simplified syntax for Top-N queries. These enhancements make important calculations significantly easier and more efficient, enhancing database performance, scalability and simplicity. ROLLUP and CUBE are simple extensions to the SELECT statement's GROUP BY clause. ROLLUP creates subtotals at any level of aggregation needed, from the most detailed up to a grand total. CUBE is an extension similar to ROLLUP, enabling a single statement to calculate all possible combinations of subtotals.

# Syntax for rollup:

ROLLUP appears in the GROUP BY clause in a SELECT statement. Its form is:

SELECT ... GROUP BY ROLLUP (grouping\_column\_reference\_list) Example: select empno, sal, sum(sal) as Totalsal from emp group by rollup(sal);



# Concept of cube:

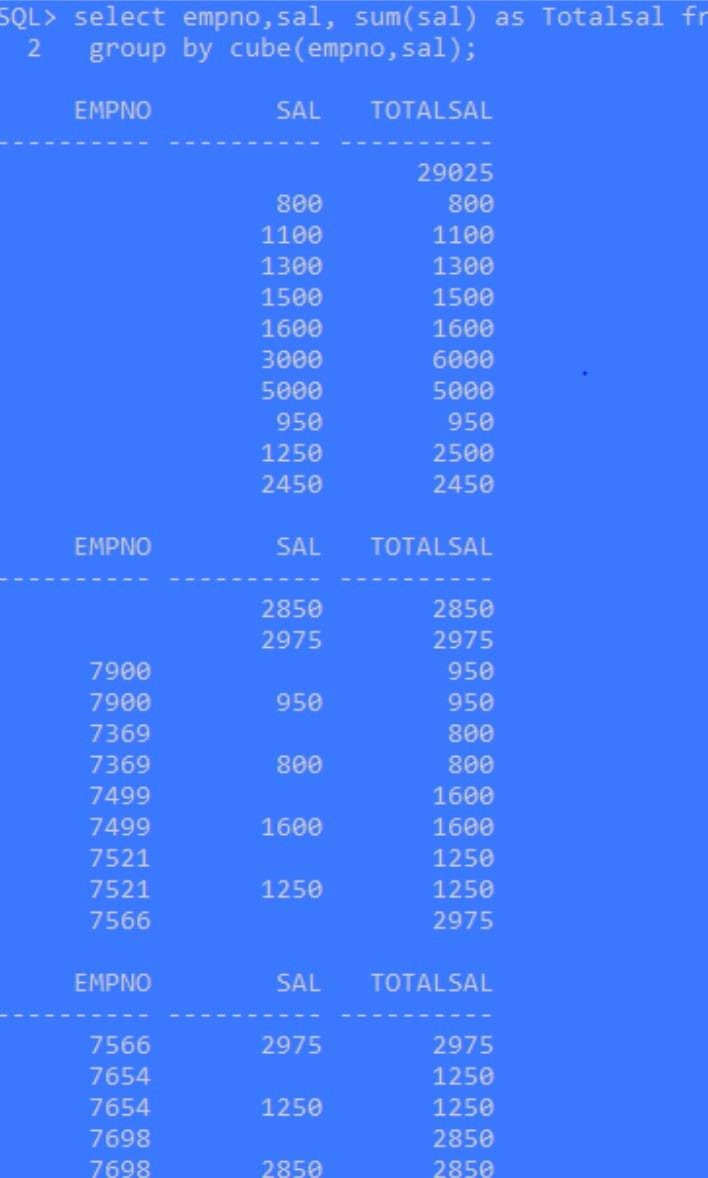
CUBE can generate the information needed in cross-tab reports with a single query. To enhance performance, both CUBE and ROLLUP are parallelized: multiple processes can simultaneously execute both types of statements.

CUBE appears in the GROUP BY clause in a SELECT statement.

# Syntax for cube:

SELECT ... GROUP BY CUBE (grouping\_column\_reference\_list) Example:

select empno, sal, sum(sal) as Totalsal from emp group by cube(sal);



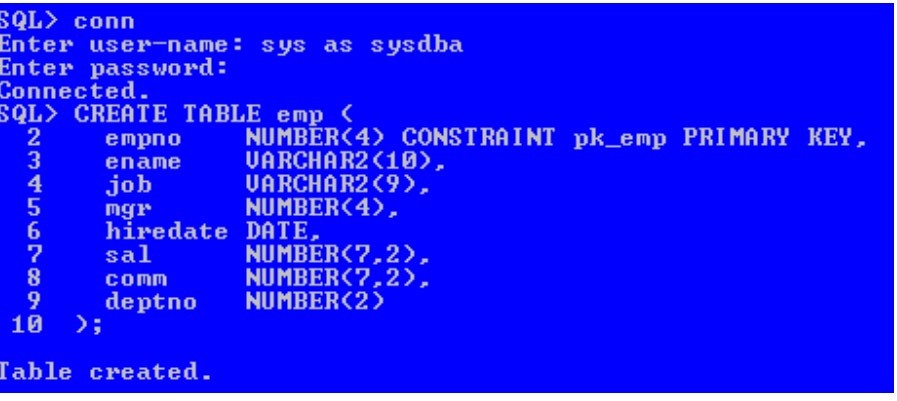
With analytic queries, we can combine data from multiple queries from the same or differing data sources into one result set. In some situations, we may need to draw data from several different sets of data, some of which might be stored in different data sources.

An analytic function computes values over a group of rows and returns a single result for each row.

**Example**: We will create a table named as EMP as follows.

# CREATE TABLE emp (empno NUMBER (4) CONSTRAINT pk\_emp PRIMARY

**KEY, ename VARCHAR2(10), job VARCHAR2(9), mgr NUMBER (4), hiredate DATE, sal NUMBER (7,2), comm NUMBER (7,2), deptno NUMBER (2));**



# PRACTICAL 5:

**CREATE ASSOCIATION RULES BY CONSIDERING SUITABLE PARAMETERS**

**Step 1:** Load required library

‘arules’ package provides the infrastructure for representing, manipulating, and analyzing transaction data and patterns.

library(arules)

’arulesviz’ package is used for visualizing Association Rules and Frequent Item sets. It extends the package ‘arules’ with various visualization techniques for association rules and item sets. The package also includes several interactive visualizations for rule exploration. library(arulesViz)

‘RColorBrewer‘ is a ColorBrewer Palette which provides color schemes for maps and other graphics.

library (RColorBrewer)

**Step 2:** Import the dataset

‘Groceries’ dataset is predefined in the R package. It is a set of 9835 records/ transactions, each having ‘n’ number of items, which were bought together from the grocery store. data("Groceries")

**Step 3:** Applying apriori () function

‘Apriori ()’ function is in-built in R to mine frequent item sets and association rules using the Apriori algorithm. Here, ‘Groceries’ is the transaction data. ‘parameter’ is a named list that specifies the minimum support and confidence for finding the association rules. The default behaviour is to mine the rules with minimum support of 0.1 and 0.8 as the minimum confidence. Here, we have specified the minimum support to be 0.01 and the minimum confidence to be 0.2.

rules <- apriori (Groceries, parameter = list (supp = 0.01, conf = 0.2))

**Step 4:** Applying inspect () function

inspect () function prints the internal representation of an R object or the result of an expression. Here, it displays the first 10 strong association rules.

inspect (rules [1:10])

**Step 5:** Applying itemFrequencyPlot () function

itemFrequencyPlot () creates a bar plot for item frequencies/ support. It creates an item frequency bar plot for inspecting the distribution of objects based on the transactions. The items are plotted ordered by descending support. Here, ‘topN=20’ means that 20 items with the highest item frequency/ lift will be plotted.

arules::itemFrequencyPlot(Groceries, topN = 20,

col = brewer.pal (8, 'Pastel2'), main = 'Relative Item Frequency Plot', type = "relative",

ylab = "Item Frequency (Relative)")

R code is given below:

# Loading Libraries library(arules) library(arulesViz) library(RColorBrewer)

# import dataset data("Groceries")

# using apriori() function rules <- apriori(Groceries,

parameter = list(supp = 0.01, conf = 0.2))

# using inspect() function inspect(rules[1:10])

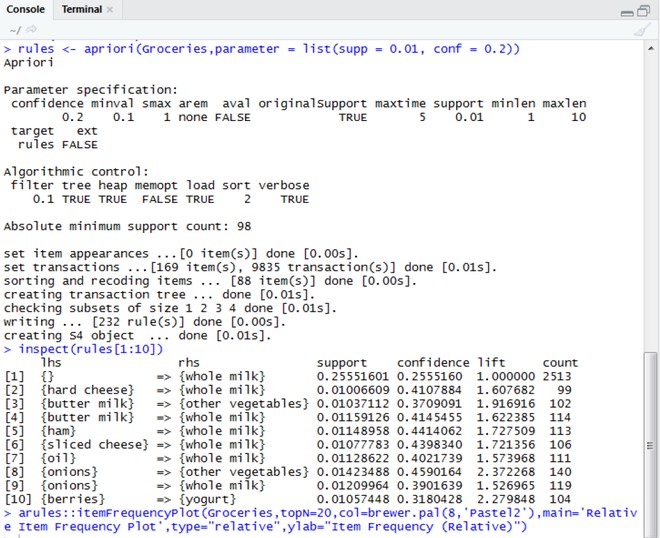
# using itemFrequencyPlot() function arules::itemFrequencyPlot(Groceries, topN = 20,

col = brewer.pal(8, 'Pastel2'),

main = 'Relative Item Frequency Plot', type = "relative",

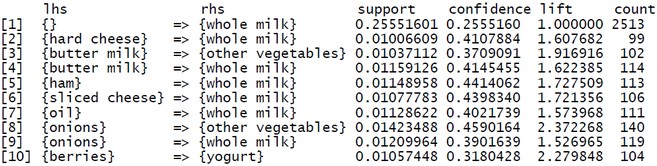
ylab = "Item Frequency (Relative)")

# Output:



**Strong Rules:**

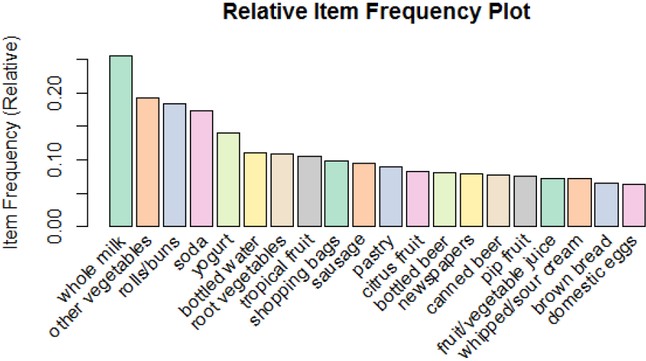
Strong Rules obtained after applying the Apriori Algorithm is as follows



After running the above code for the Apriori algorithm, we can see the following output, specifying the first 10 strongest Association rules, based on the support (minimum support of 0.01), confidence (minimum confidence of 0.2), and lift, along with mentioning the count of times the products occur together in the transactions.

Visualization:

Box Plot of the Top 20 Items having the Highest Item Frequency (Relative) using Lift as a Parameter



# PRACTICAL 6:

**CASE STUDY: THE DEVELOPMENT OF A DATA WAREHOUSE FOR AN E-COMMERCE COMPANY**

Assuming that you have been engaged to submit a blue print for the development of a data warehouse for an e-commerce company, make a project report detailing different steps to be taken for the purpose.

A data warehouse (DW) is a digital storage system that connects and harmonises large amounts of data from many different sources. Its purpose is to feed business intelligence (BI), reporting, and analytics, and support regulatory requirements – so companies can turn their data into insight and make smart, data-driven decisions. Data warehouses store current and historical data in one place

Data flows into a data warehouse from operational systems (like ERP and CRM), databases, and external sources such as partner systems, Internet of Things (IoT) devices, weather apps, and social media – usually on a regular cadence. The emergence of cloud computing has caused a shift in the landscape. In recent years, data storage locations have moved away from traditional on-premise infrastructure to multiple locations, including on premise, private cloud, and public cloud.

Modern data warehouses are designed to handle both structured and unstructured data, like videos, image files, and sensor data. Some leverage integrated analytics and in-memory database technology (which holds the data set in computer memory rather than in disk storage) to provide real-time access to trusted data and drive confident decision-making. Without data warehousing, it’s very difficult to combine data from heterogeneous sources, ensure it’s in the right format for analytics, and get both a current and long-range view of data over time.

A data warehouse will help you to build accurate forecasting models and identify impactful trends. When building a data warehouse, it’s important to recognise the following steps and thoroughly address each one.

# Defining Business Requirements

Since a data warehouse encompasses all areas of your business it’s vital that every department is involved with the design process. The process of requirements gathering involves all stakeholders and helps every department understand the purpose of the data warehouse, how they will benefit, and what results they can expect.

Requirement gathering can happen as one-to-one or collective meetings. This phase often turns out to be one of the hardest parts of data warehousing implementation. Because a data warehouse includes data from so many sources, spanning multiple departments, there can often be negotiations over information sharing and prioritisation. A skilled business analyst can act as an external mediator to ensure all stakeholders are happy with the defined project goals.

The requirements gathering phase is so important in ensuring that department goals are aligned with the overall project and that buy-in from all the relevant stakeholders is achieved. It can also help to highlight current and future needs from taking a deep dive into the data which will be used for analysis, which will likely uncover where your data is and isn’t being used effectively.

# Setting Up Physical Environments

Your data warehouse will typically have three environments which mimics software development best practice. The three environments are development, testing, and production and these are used in tandem to ensure changes are tested for integrity and security before they are pushed to live in the production environment. They allow for development and Quality Assurance to occur without affecting the productive environment.

Three separate environments are also needed to run test data, identify breakpoints that need to be rectified, and to reduce stress on server workloads. It is not enough to simply have different physical environments set up. The different processes (such as ETL, OLAP cube and reporting) also need to be set up properly for each environment.

# Introducing Data Modelling

Data modelling is the blueprint from which the data warehouse is built. It can help you visualise data relationships, standardise naming conventions and establishing security process compliance.

This is known as the most complex phase of data warehouse design. A good data model will allow the data warehousing system to grow easily and perform well!

Data modelling normally takes place at the data mart level and branches out into the data warehouse. The three most popular data models for warehouses are the Snowflake, Star, and Galaxy Schemas. The chosen model will impact the structure of your data warehouse and data marts, and it will help to guide the overall architecture within the warehouse.

# Choosing Your ETL Solution

ETL stands for Extract, Transform and Load and represents the collection and processing of data from various sources into one central data store where it can be later analysed. Your business has access to many data sources but often it’s presented in a way that is hard or impossible to consume.

A good ETL process can be the difference between a slow and hard-to-use data warehouse and a sleek warehouse that adds value to every part of your organisation. For this reason, it’s vital that the right ETL solution is selected.

# Online Analytics Processing (OLAP) Cube

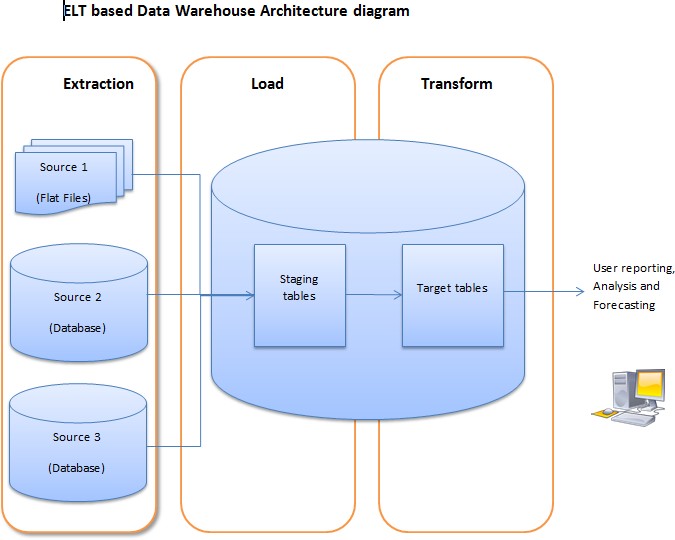
An OLAP cube helps you to analyse the data in your data warehouse or data mart. Since your warehouse will be sorting data from multiple sources, the OLAP cube helps you to organise all of that data in a multi-dimensional format that makes it easier to analyse.

# Creating A Front End

This stage refers to the front-end visualisation, where users can understand and apply the results of data queries. If users cannot visualise the reports, the data warehouse is likely to provide little value to them, making front end development an important part of a data warehouse initiative. Another area of importance is the complexity of the reporting tool. How often do reports need to be published?

Do they require specific formatting?

And does the user require an interface that allows for customisable reports



# Optimising Queries

The more data returned from a query, the more resources the database needs to expand to process and store this data. This is why it’s important to only minimise data retrieval, especially if you’re paying for query power separately.

This stage is very specific to each organisation’s needs.

# Rolling Out the End Product

The hard work has now been done and you’re close to getting value from your shiny, new data warehouse. It’s at this point that team members will need to be trained in using it.

Throughout the process Quality Assurance and Testing have been ensuring there are no bugs or usability issues.

Although these are the standard steps in creating a data warehouse, it’s important to remember that every scenario is different. There may be additional steps that your business needs to take based on the requirements or complexity of your organisation’s needs.

Ultimately, a successfully implemented data warehouse will deliver value at every level of your organisation.

# PRACTICAL 7

**CASE STUDY: EFFECTIVENESS OF HADOOP AS AN OPEN-SOURCE COMPANION TO STANDARD DATA WAREHOUSES.**

# What is Hadoop?

Apache Hadoop is an open source, Java-based software platform that manages data processing and storage for big data applications. The platform works by distributing Hadoop big data and analytics jobs across nodes in a computing cluster, breaking them down into smaller workloads that can be run in parallel. Some key benefits of Hadoop are scalability, resilience and flexibility. The Hadoop Distributed File System (HDFS) provides reliability and resiliency by replicating any node of the cluster to the other nodes of the cluster to protect against hardware or software failures. Hadoop flexibility allows the storage of any data format including structured and unstructured data.

# Hadoop is used for:

When it comes to Hadoop, the possible use cases are almost endless. Retail

Large organizations have more customer data available on hand than ever before. But often, it’s difficult to make connections between large amounts of seemingly unrelated data. When British retailer M&S deployed the Hadoop-powered Cloudera Enterprise, they were more than impressed with the results.

Cloudera uses Hadoop-based support and services for the managing and processing of data. Shortly after implementing the cloud-based platform, M&S found they were able to successfully leverage their data for much improved predictive analytics.

This led them to more efficient warehouse use and prevented stock-outs during “unexpected” peaks in demand and gaining a huge advantage over the competition.

Finance

Hadoop is perhaps more suited to the finance sector than any other. Early on, the software framework was quickly pegged for primary use in handling the advanced algorithms involved with risk modelling. It’s exactly the type of risk management that could help avoid the credit swap disaster that led to the 2008 recession.

Banks have also realized this same logic also applies to managing risk for customer portfolios. Today, it’s common for financial institutions to implement Hadoop to better manage the financial security and performance of their client’s assets. JPMorgan Chase is just one of many industry giants that use Hadoop to manage exponentially increasing amounts of customer data from across the globe.

# Healthcare

Whether nationalized or privatized, healthcare providers of any size deal with huge volumes of data and customer information. Hadoop frameworks allow for doctors, nurses and carers to have easy access to the information they need when they need it and it also makes it easy to aggregate data that provides actionable insights. This can apply to matters of public health, better diagnostics, improved treatments and more.

Academic and research institutions can also leverage a Hadoop framework to boost their efforts. Take for instance, the field of genetic disease which includes cancer. We have the human genome mapped out and there are nearly three billion base pairs in total. In theory, everything to cure an army of diseases is now right in front of our faces. But to identify

complex relationships, systems like Hadoop will be necessary to process such a large amount of information.

# Security and law enforcement

Hadoop can help improve the effectiveness of national and local security, too. When it comes to solving related crimes spread across multiple regions, a Hadoop framework can streamline the process for law enforcement by connecting two seemingly isolated events. By cutting down on the time to make case connections, agencies will be able to put out alerts to other agencies and the public as quickly as possible.

In 2013, The National Security Agency (NSA) concluded that the open-source Hadoop software was superior to the expensive alternatives they’d been implementing. They now use the framework to aid in the detection of terrorism, cybercrime and other threats.

# How does Hadoop work?

Hadoop is a framework that allows for the distribution of giant data sets across a cluster of commodity hardware. Hadoop processing is performed in parallel on multiple servers simultaneously.

Clients submit data and programs to Hadoop. In simple terms, HDFS (a core component of Hadoop) handles the Metadata and distributed file system. Next, Hadoop MapReduce processes and converts the input/output data. Lastly, YARN divides the tasks across the cluster.

With Hadoop, clients can expect much more efficient use of commodity resources with high availability and a built-in point of failure detection. Additionally, clients can expect quick response times when performing queries with connected business systems.

In all, Hadoop provides a relatively easy solution for organizations looking to make the most out of big data.

# How to use Hadoop for analytics

Depending on data sources and organizational needs, there are three main ways to use the Hadoop framework for analytics.

# Deploy in your corporate data centre(s)

This is often a time-effective and financially sound option for those businesses with the necessary existing resources. Otherwise, setting up the technical equipment and IT staff required may overextend monetary and team resources. This option does give businesses greater control over the security and privacy of data.

# Go with the cloud

Businesses that desire a much more rapid implementation, lower upfront costs and lower maintenance requirements will want to leverage a cloud-based service. With cloud provider, data and analytics are run on commodity hardware that exists in the cloud. These services streamline the processing of big data at an affordable price but come with certain drawbacks. Firstly, anything that’s on the public internet is fair game for hackers and the like. Secondly, service outages to the internet and network providers can grind your business systems to a halt. For existing framework users, they may involve something like needing to migrate from Hadoop to the Lake how Architecture.

# On-premise providers

Those opting for better uptime, privacy and security will find all three things with an on- premise Hadoop provider. These vendors offer the best of both worlds. They can streamline the process by providing all equipment, software and service. But since the infrastructure is on-premises, you gain all the benefits that large corporations get from having data centres.

# What are the benefits of Hadoop?

**Scalability** - Unlike traditional systems that limit data storage, Hadoop is scalable as it operates in a distributed environment. This allowed data architects to build early data lakes on Hadoop. Learn more about the history and evolution of data lakes.

**Resilience** - The Hadoop Distributed File System (HDFS) is fundamentally resilient. Data stored on any node of a Hadoop cluster is also replicated on other nodes of the cluster to prepare for the possibility of hardware or software failures. This intentionally redundant design ensures fault tolerance. If one node goes down, there is always a backup of the data available in the cluster.

**Flexibility** - Differing from relational database management systems, when working with Hadoop, you can store data in any format, including semi-structured or unstructured formats. Hadoop enables businesses to easily access new data sources and tap into different types of data.

# What are the challenges with Hadoop architectures?

**Complexity** - Hadoop is a low-level, Java-based framework that can be overly complex and difficult for end-users to work with. Hadoop architectures can also require significant expertise and resources to set up, maintain, and upgrade.

**Performance** - Hadoop uses frequent reads and writes to disk to perform computations, which is time-consuming and inefficient compared to frameworks that aim to store and process data in memory as much as possible, like Apache Spark.

**Long-term viability** - In 2019, the world saw a massive unravelling within the Hadoop sphere. Google, whose seminal 2004 paper on MapReduce underpinned the creation of Apache Hadoop, stopped using MapReduce altogether, as tweeted by Google SVP of Technical Infrastructure, Urs Hölzle. There were also some very high-

profile mergers and acquisitions in the world of Hadoop.

Furthermore, in 2020, a leading Hadoop provider shifted its product set away from being Hadoop-centric, as Hadoop is now thought of as “more of a philosophy than a technology.” Lastly, 2021 has been a year of interesting changes. In April 2021, the Apache Software Foundation announced the retirement of ten projects from the Hadoop ecosystem. Then in June 2021, Cloudera agrees to private. The impact of this decision on Hadoop users is still to be seen. This growing collection of concerns paired with the accelerated need to digitize has encouraged many companies to re-evaluate their relationship with Hadoop.

# Which companies use Hadoop?

Hadoop adoption is becoming the standard for successful multinational companies and enterprises. The following is a list of companies that utilize Hadoop today:

**Adobe** - the software and service providers use Apache Hadoop and HBase for data storage and other services.

**eBay** - uses the framework for search engine optimization and research.

**A9** - a subsidiary of Amazon that is responsible for technologies related to search engines and search-related advertising.

**LinkedIn** - as one of the most popular social and professional networking sites, the company uses many Apache modules including Hadoop, Hive, Kafka, Avro, and DataFu.

**Spotify** - the Swedish music streaming giant used the Hadoop framework for analytics and reporting as well content generation and listening recommendations.

**Facebook** - the social media giant maintains the largest Hadoop cluster in the world, with a dataset that grows a reported half of a PB per day.

**InMobi** - the mobile marketing platform utilizes HDFS and Apache Pig/MRUnit tasks involving analytics, data science and machine learning.

# How much does Hadoop cost?

The Hadoop framework itself is an open-source Java-based application. This means, unlike other big data alternatives, it’s free of charge. Of course, the cost of the required commodity software depends on what scale.

When it comes to services that implement Hadoop frameworks you will have several pricing options:

Per Node- most common Per TB

Freemium product with or without subscription-only tech support All-in-one package deal including all hardware and software

Cloud-based service with its own broken down pricing options- can essentially pay for what you need or pay as you go

# PRACTICAL 8

**CASE STUDY: EMERGING TRENDS ON CLOUD-BASED BUSINESS INTELLIGENCE AND ANALYTICS.**

The business intelligence landscape has been completely revolutionized over the past ten years, and access to the cloud has become a common phenomenon. Interactive business dashboards and insightful data visualizations replaced plain spreadsheets, and the use of data exploded. The past few years saw a massive boom in the business intelligence industry. The data product chain became democratized due to the rise of self-service analytics. Data began to be handled not just by analysts.

Some of the trends continue to be a part of the business intelligence industry. But in the current year, the strategies and tools have become even more personalized, and some new trends have emerged too. Businesses- small or big, are no longer contemplating the need for business intelligence analytics but are rather concerned about which BI solution fits the best within their business needs.

Every business is now aware that they need data visualizations for better analysis. So, they are busy figuring out how to leverage modern BI dashboard software to present data stories in the best possible manner. The theme of business intelligence this year will revolve around data discovery and data security: clean and secure data presented but powerfully. There will be collaborations between artificial intelligence and business intelligence.

# Top Business Intelligence Trends

Some of the biggest business intelligence trends are a no-brainer. Artificial intelligence leads the race with workforce automation and digital transformation sectors seeing massive improvements. Read on to know why.

# Artificial Intelligence

Artificial intelligence aims to make machines perform those complex tasks on their own that can only be executed through human intelligence. Our interactions with analytics and data management are getting revolutionized through artificial intelligence. According to the Strategic Technology Trends report, the trend will combine engineering and hyper- automation with AI with a high focus on possible security risks and vulnerable attack points.

Businesses can enjoy real-time alerts about what is happening every second and get immediately notified about unexpected events. Integrating AI in BI solutions will assist in automatic and comprehensive analysis of the full dataset from any data source without human effort. You can instantly access business reports on growth or trends or forecast, anomalies, what-if analysis, value drivers, key segments correlations, etc. AI can also be utilized for online verification processes, like CAPTCHA technology, with the help of generative adversarial networks (GANs).

# Data Discovery

Data discovery means discovering patterns and discrepancies in data. It is the process of using advanced analytics and visualizations to present all the data collected from different internal and external sources. It has great benefits in keeping relevant stakeholders involved with the data since it allows them to extract actionable insights and intuitively manipulate and analyse the data. The demand for data discovery tools across businesses of all sizes has boomed due to the increasing need for data usage and insights.

Generating insights that add value to the business requires a deep understanding of the relationship between data in the form of guided, advanced analytics, visual analysis, and data preparation.

Online data visualization and discovery tools are helping businesses create a sustainable decision-making process. The detailed and interactive reports or sales charts presented with several graphs will help teams spot crucial outliers and trends within minutes. Since it is a fact that humans process visual data better, in 2022, the usage of the dashboard as a visual communication and collaboration tool will increase.

In-depth data analysis through interactivity and augmented analytics will replace simple KPI monitoring. KPI dashboards will have other interactive features, too, such as real-time data and AI-based alarms.

# Business users need software for this purpose that is:

* Flexible and agile
* User friendly
* Helpful in reducing time to insight
* Beneficial in handling a variety of data at high volumes

# Data Literacy

The ability of businesses to use data analytics and insights in their decision-making has become a core factor in determining the business’s success. From goal setting to strategizing to taking action, businesses require data at every step. No wonder data literacy is of utmost significance for every business. It is the reading, writing, analyzing, and communicating data in a particular context. Data literacy requires a deep understanding of all the tools and technologies adopted and techniques and methods implemented for data analysis.

Business leaders must equip all the organization members with the training and tools required for working with data and analytics. Managers need to assess the skill sets of employees, and managers need to identify gaps and weak spots. Team members fluent in data can be appointed as mediators for non-skilled groups. With the right tools and quality training, all the members will acquire enough data literacy to use data as the key language and perform advanced analysis. By 2025, prediction says that data literacy will be so widespread that businesses will no longer require data scientists to progress technologically.

# Data Quality Management

Currently, there is abundant data in every business flowing in from literally everywhere, and it has become crucial to assess data quality before using it. Poor quality data can cost businesses around $9.7 and $14.2 million per year. No wonder data quality management is an increasingly significant trend. Poor data quality can lead to a poor understanding of consumer behaviour, wrong estimation of conversion rates, poorly generated marketing budgets, incorrect resource allocation, bad investments, and other errors that can harm businesses significantly.

Data quality management is the solution to all these problems. It ensures that businesses only use the correct data for analytical purposes to arrive at the right data-driven decisions. Data quality depends upon how complete, timely, accurate, consistent, and compliant it is. There can be no outdated data that does not fit within the timeline or duplicate or missing values. Companies are collecting complex data from several sources regularly, and managing these data using the right tools and processes has become critical.

# Predictive & Prescriptive Analytics Tools

Predictive analysis means forecasting future possibilities by extracting information from existing data sets. It is data mining of past data. Companies get an insight into their future along with alternative scenarios and risk assessments that are reliable enough. It helps companies better understand their customers, products, and partners and identify potential risks and opportunities. For instance, the airline industry can use it to determine how many tickets to sell at a particular price. The hotel industry can gain insight into how many guests can be expected on a day so that hotels can adjust their availability accordingly.

Marketers can use this trend to predict customer purchases or responses to locate cross- selling opportunities, and bankers can generate credit scores. The prescriptive analysis goes even a step further. It uses techniques like graph analysis, complex event processing, simulations, neural networks, recommendation engines, machine learning, and heuristics to determine the appropriate business decisions and steps for achieving a particular goal. It incorporates future outcomes in decision-making that improve the decision-making quality related to optimizing scheduling, inventory, production, and supply chain design to enhance customer experience.

# Real-time Data & Analytics

Since the pandemic arrived, the need for accurate updates and real-time data has become crucial in strategizing and responding to crises. It has played a crucial role in the best possible decision-making for risk aversion and survival during such risky times. Even in the future, developing proper business responses and strategies will require forecasting and alarms.

Live dashboards implemented across companies will provide immediate access to relevant information regarding their business and solve any potential issues. Businesses are staying on top of changes and adapting to immense challenges with the creation of ad hoc analyses. No wonder companies need to gear up rapidly to the increasing use of up-to-date data.

# Collaborative Business Intelligence

Businesses are getting more competitive, and thus the need for collaborative business intelligence has enhanced. It combines collaboration tools like online BI tools, including social media and other 2.0 technologies. Fast-track businesses where analysis is done and

reports edited are massive pose unique challenges that only enhanced collaboration can solve. These online BI tools generate automated reports that can be scheduled at specific times for specific people.

For instance, setting up business intelligence alerts helps to share embedded or public dashboards that are highly interactive and flexible. Such a collaborative environment is especially useful in the current work-from-home setup of organizations. The business world now is more dynamic than ever, and such high levels of collaboration are necessary for problem-solving. It is not limited to document updates or exchanges but extends to the progress of meetings, e-mail exchanges, calls, and ideas collection. Studies predict that in the future collaborative business intelligence will become accessible by larger sets of users and more connected to bigger systems.

# Data Automation

During the past decade, so much data has been produced, stored, and analysed that companies felt the need for data automation solutions to handle the massive volumes of data collected. Business intelligence allows users to consolidate all the data managed by a company. It provides techniques to discover, measure, analyse, monitor, and evaluate large- scale data. The new trend points toward businesses automating the maximum possible processes using multiple technologies and tools such as AI, no-code tools, machine learning, low-code, etc.

The barriers between data scientists and business users are slowly demolishing. Right now, businesses have a one-stop destination for any data requirement, like analyzing, collecting, monitoring, analyzing, sharing findings, and reporting. Predictive analytics and automated reports have enhanced the capability of businesses to automate data without depending on IT departments. Data scientists predict that over the next decade, one of the significant trends in business intelligence will be the automation of data science tasks.

# Embedded Analytics

Businesses have become much more productive and capable of improved decision-making by embedding various BI components such as reports or dashboards into their applications. These embedded dashboards add much more value to businesses than conventional

spreadsheets. Studies by Allied Market research indicate that the embedded analytics market will reach $77.52 BN by 2026, with a CAGR of 13.6%.

Organizations can offer more polished presentations and report to customers by white labelling the selected applications. Embedding analytics to an application gives scope for enhanced collaboration and increases the involvement of every stakeholder rather than just embedding a dashboard or BI features. It equips employees and clients to manipulate the data in a well-monitored environment that facilitates better extraction of insights from every area of your business.