

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645



Lt. Shree Chimanbhai Shukla

B.C.A./B.Sc.IT-SEMESTER-6

SUB :- DATA WAREHOUSE

Shree H.N.Shukla College
Street No. 2, Vaishali Nagar,
Nr. Amrapali Under Bridge,
Raiya Road, Rajkot.
Ph. (0281)2440478, 2472590



Shree H.N.Shukla College
Street No. 2, Vaishali Nagar,
Nr. Amrapali Under Bridge,
Raiya Road, Rajkot.
Ph. (0281)2440478, 2472590

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

CHAPTER-1:

Introduction to Data Warehouse

Sr No.	Topic
1.	What is Data Warehousing?
2.	How data warehouse works?
3.	Application of data warehousing
4.	OLTP
5.	Types of Data Warehouse
6.	Components of Data Warehouse
7.	Features of Data Warehouse
8.	Data Warehouse Architecture
9.	Advantages and disadvantages of Data Warehouse
10.	Data Flow Architecture
11.	Data Warehousing Today
12.	Future trends of Data Warehousing

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



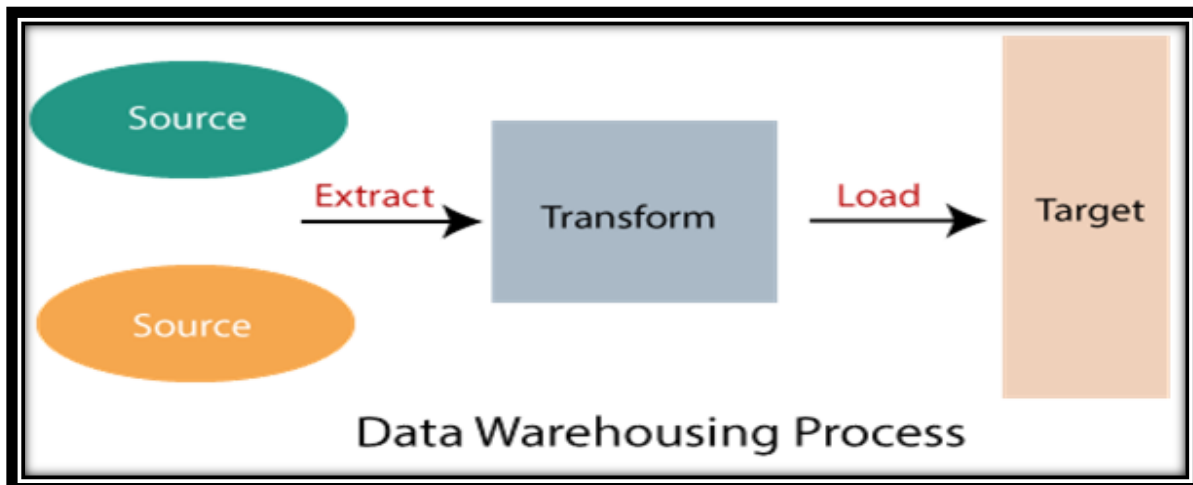
2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

Topic: What is Data Warehousing? (2019 Uni. Paper)

Ans:

- ❖ A **Data Warehousing** (DW) is process for collecting and managing data from varied sources to provide meaningful business insights.
- ❖ A Data warehouse is typically used to connect and analyze business data from different multiple sources.
- ❖ The data warehouse is the core of the BI system which is built for data analysis and reporting.
- ❖ It is a blend of technologies and components which aids the strategic use of data. It is electronic storage of a large amount of information by a business which is designed for query and analysis instead of transaction processing. It is a process of transforming data into information and making it available to users in a timely manner to make a difference.



Sr No.	M.C.Q	Answer
1.	DW/DWH stands for	Data Warehouse
2.	What is data warehouse?	A Data warehousing is the process for collecting and managing the data from varied sources to provide meaningful business insights

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

--	--	--

Topic: How data warehouse works?

Ans:

- A Data Warehouse works as a central repository where information arrives from one or more data sources. Data flows into a data warehouse from the transactional system and other relational databases.
- Data may be:
 1. Structured
 2. Semi-structured
 3. Unstructured data
- The data is processed, transformed, and ingested so that users can access the processed data in the Data Warehouse through Business Intelligence tools, SQL clients, and spreadsheets. A data warehouse merges information coming from different sources into one comprehensive database.
- By merging all of this information in one place, an organization can analyze its customers more holistically. This helps to ensure that it has considered all the information available. Data warehousing makes data mining possible.

Difference between Structured, Unstructured and Semi-Structured Data

Structured Data	Unstructured Data	Semi Structured Data
In this, data is well organized	In this, data is fully non-organized	In this, data is organized to some extent and rest data is non-organized
Structured data is organized by means of relational database	Unstructured data is organized on simple character and binary data	It is partially organized by XML
Example: Recipients, subject in Letter	Example: Text area in Email	Example: Email Message Box

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

Sr No.	M.C.Q	Answer
1.	text area of email is an example of	Unstructured data
2.	In, data is organised to some extent and rest data is non-organised	Semi-Structure Data

Topic: What are the applications of data warehousing? (2019 uni. paper)

Ans:

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

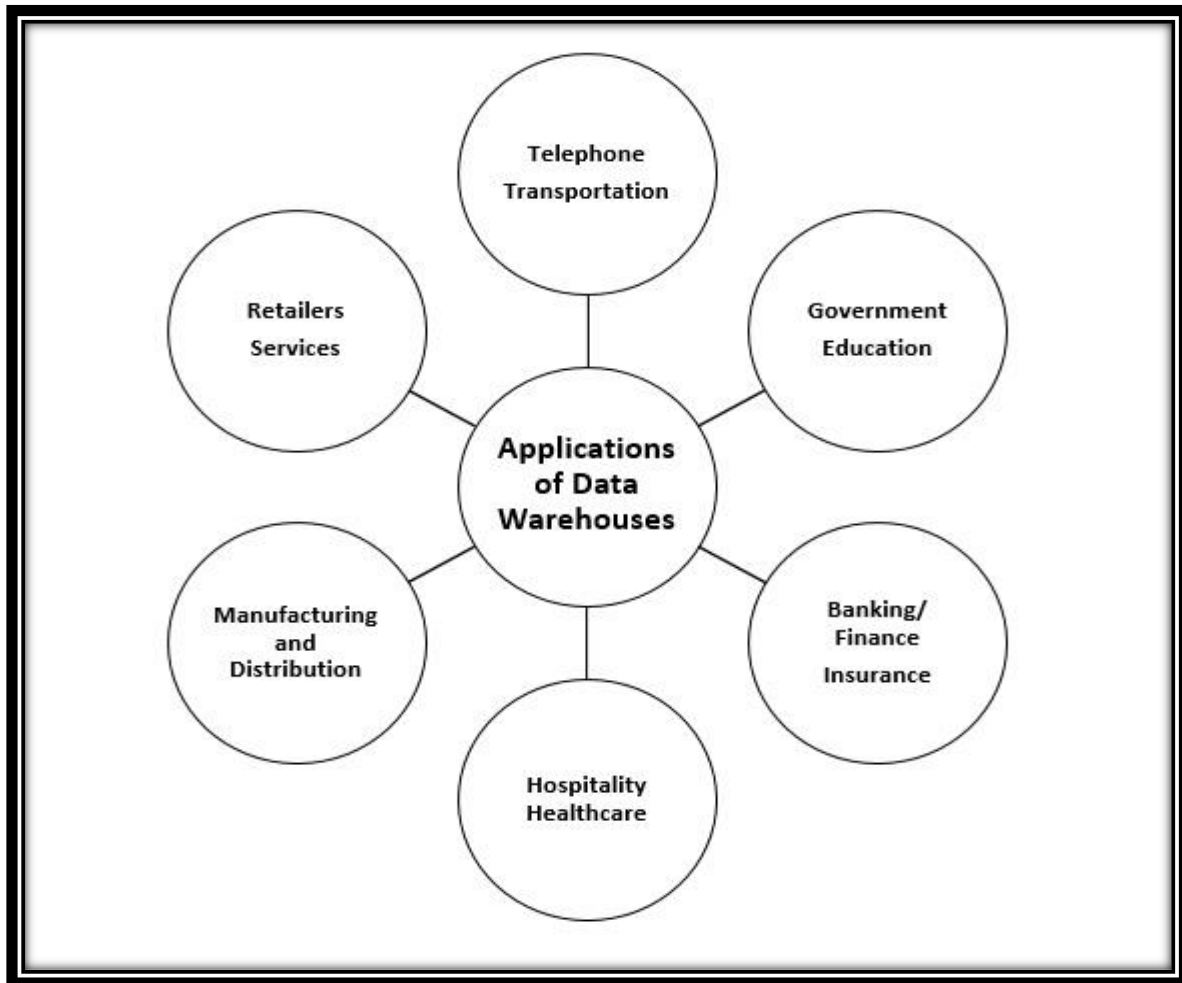
(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

Applications of Data warehouse:



1. Banking Industry:

- Data warehouse is used for analyzing consumer data, market trends, government regulations and reports, and more importantly financial decision making.
- Analysis of card holder's transactions

2. Finance Industry:

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

- Mainly revolve around evaluation and trends of customer expenses which aids in maximizing the profits earned by their clients.

3. Government

- The government uses data warehouse to maintain and analyze tax records, health policy records and their respective providers, and also their entire criminal law database is connected to the state's data warehouse. Criminal activity is predicted from the patterns and trends, results of the analysis of historical data associated with past criminals.

4. Education:

- Universities use warehouses for extracting of information used for the proposal of research grants, understanding their student demographics, and human resource management. The entire financial department of most universities depends on data warehouses, inclusive of the Financial Aid department.

5. Healthcare:

- All of their financial, clinical, and employee records are fed to warehouses as it helps them to strategize and predict outcomes, track and analyze their service feedback, generate patient reports, share data with tie-in insurance companies, medical aid services, etc.

6. Hospitality:

- A major proportion of this industry is dominated by hotel and restaurant services, car rental services, and holiday home services. They utilize warehouse services to design and evaluate their advertising and promotion campaigns where they target customers based on their feedback and travel patterns.

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

7. Insurance:

- As the saying goes in the insurance services sector, “Insurance can never be bought, it can be only be sold”, the warehouses are primarily used to analyze data patterns and customer trends, apart from maintaining records of already existing participants. The design of tailor-made customer offers and promotions is also possible through warehouses.

8. Manufacturing and Distribution Industry:

- A manufacturing organization has to take several make-or-buy decisions which can influence the future of the sector, which is why they utilize high-end OLAP tools as a part of data warehouses to predict market changes, analyze current business trends, detect warning conditions, view marketing developments, and ultimately take better decisions.
- They also use them for product shipment records, records of product portfolios, identify profitable product lines, analyze previous data and customer feedback to evaluate the weaker product lines and eliminate them.

9. Retailers:

- Retailers serve as middlemen between producers and consumers. It is important for them to maintain records of both the parties to ensure their existence in the market.
- They use warehouses to track items, their advertising promotions, and the consumers buying trends. They also analyze sales to determine fast selling and slow selling product lines and determine their shelf space through a process of elimination.

10. Services:

- Data warehouses find themselves to be of use in the service sector for maintenance of financial records, revenue patterns, customer profiling, resource management, and human resources.

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

11. Telephone Industry:

- ❖ The telephone industry operates over both offline and online data burdening them with a lot of historical data which has to be consolidated and integrated.
- ❖ Apart from those operations, analysis of fixed assets, analysis of customer's calling patterns for sales representatives to push advertising campaigns, and tracking of customer queries, all require the facilities of a data warehouse.

12. Transportation Industry:

- ❖ In the transportation industry, data warehouses record customer data enabling traders to experiment with target marketing where the marketing campaigns are designed by keeping customer requirements in mind.
- ❖ The internal environment of the industry uses them to analyze customer feedback, performance, manage crews on board as well as analyze customer financial reports for pricing strategies.

Topic: what is OLTP?

Ans:

- ❖ OLTP stands for online transactional Processing.
- ❖ It is a type of data processing that executes transaction focused task
- ❖ OLTP administers day to day transaction of an organization
- ❖ The primary objective of OLTP is data processing and not data analysis.
- ❖ OLTP uses traditional DBMS.
- ❖ An example of OLTP system is ATM centre.
- ❖ Other examples of OLTP applications are:
 - Online banking
 - Online airline ticket booking
 - Sending a text message
 - Order entry
 - Add a book to shopping cart

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

Benefits of OLTP method

- It administers daily transactions of an organization.
- OLTP widens the customer base of an organization by simplifying individual processes.

Drawbacks of OLTP method

- If OLTP system faces hardware failures, then online transactions get severely affected.
- OLTP systems allow multiple users to access and change the same data at the same time which many times created unprecedented situation.

KEY DIFFERENCE between OLTP and OLAP:

OLTP	Data Warehouse (OLAP)
1) OLTP stands for Online Transactional Processing	1) Data warehouse is not transaction oriented.
2) It supports transaction-oriented applications	2) It is a category of software tools that analyze the data stored in a database.
3) It uses traditional DBMS	3) Data warehouse (database) is created uniquely so that it can integrate different data sources for building a consolidated database
4) It holds current data	4) It holds historical data
5) Data is dynamic	5) Data is largely static
6) It is characterized by large numbers of short online transactions	6) It is characterized by large volume of data.

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

Sr No.	M.C.Q	Answer
1.	OLTP stands for...	Online Transaction Processing
2.	OLTP holds data	Current
3.	OLTP uses traditional DBMS (T/F)?	True
4.	OLTP focuses on and not on	Data Processing Data Analysis

Topic: Types of Data Warehouse:

Ans:

Three main types of Data Warehouses (DWH) are:

1) Enterprise Data Warehouse (EDW):

Enterprise Data Warehouse (EDW) is a centralized warehouse. It provides decision support service across the enterprise. It offers a unified approach for organizing and representing data. It also provides the ability to classify data according to the subject and give access according to those divisions.

2) Operational Data Store:

Operational Data Store, which is also called ODS, are nothing but data store required when neither Data warehouse nor OLTP systems support organizations reporting needs. In ODS, Data warehouse is refreshed in real time. Hence, it is widely preferred for routine activities like storing records of the Employees.

3) Data Mart:

A data mart is a subset of the data warehouse. It specially designed for a particular line of business, such as sales, finance, sales or finance. In an independent data mart, data can collect directly from sources.

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

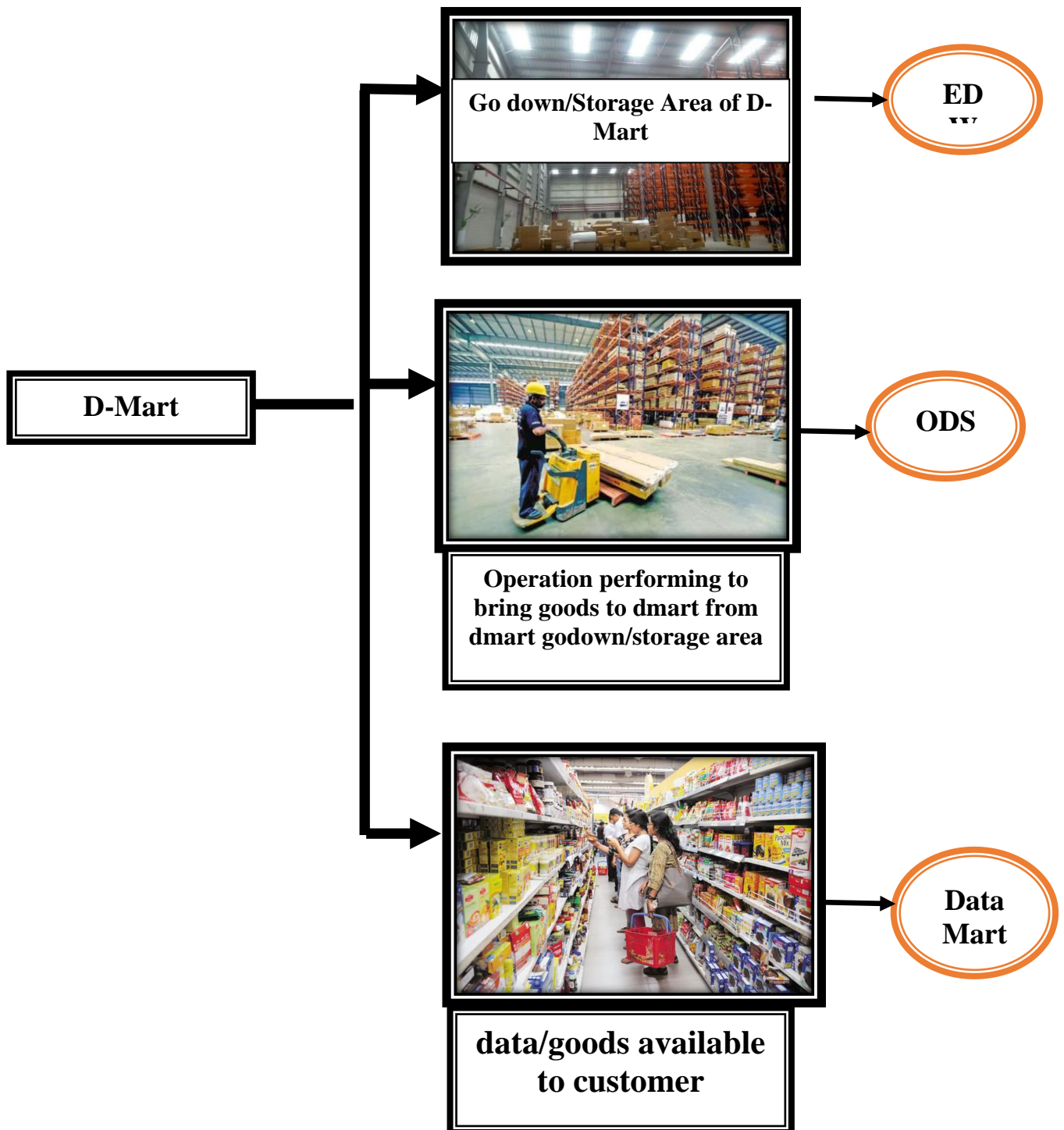
(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

Example:



SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

Sr No.	M.C.Q	Answer
1.	EDW stands for ...	Enterprise Data Warehouse
2.	ODS stands for....	Operational Data Store
3.	Datamart is the subset of Datawarehouse (T/F)?	True

Topic: Components of Data warehouse

Ans:

Four components of Data Warehouses are:

- 1) **Load manager:** Load manager is also called the front component. It performs with all the operations associated with the extraction and load of data into the warehouse. These operations include transformations to prepare the data for entering into the Data warehouse.
- 2) **Warehouse Manager:** Warehouse manager performs operations associated with the management of the data in the warehouse. It performs operations like analysis of data to ensure consistency, creation of indexes and views, generation of denormalization and aggregations, transformation and merging of source data and archiving and baking-up data.
- 3) **Query Manager:** Query manager is also known as backend component. It performs all the operation operations related to the management of user queries. The operations of these Data warehouse components are direct queries to the appropriate tables for scheduling the execution of queries.

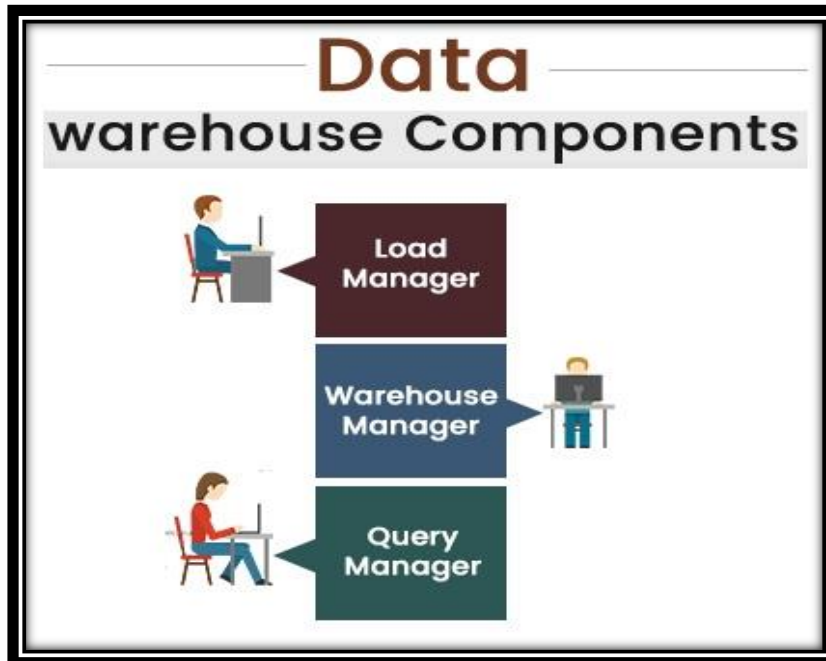
SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645



SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)

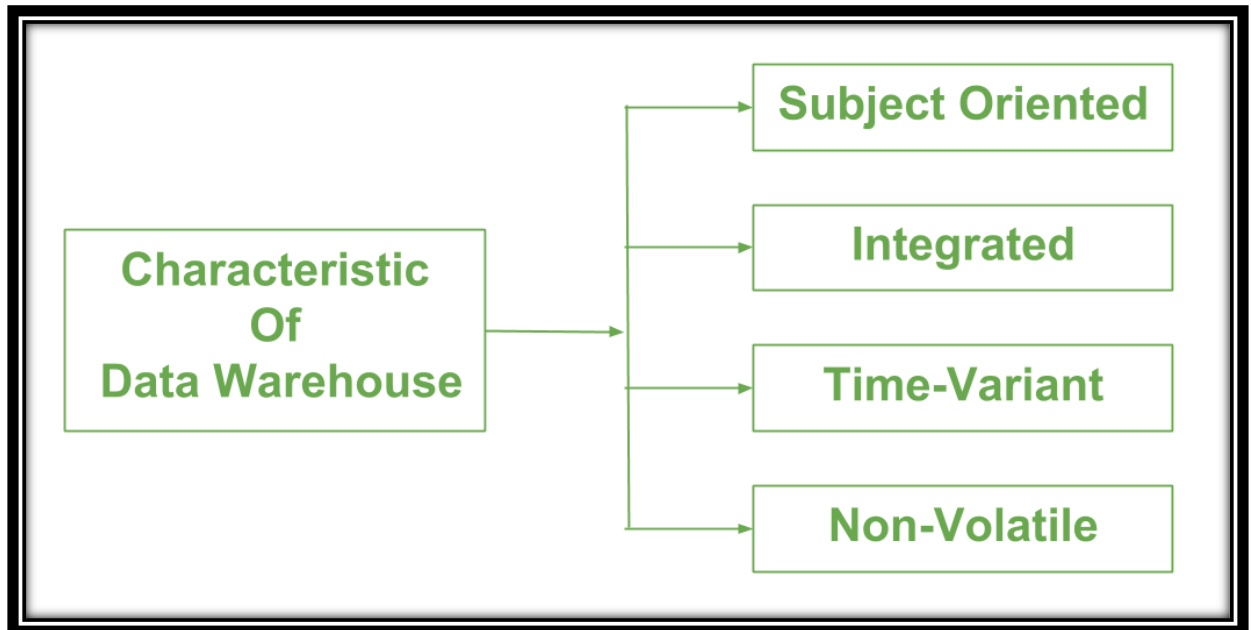


2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

Topic: Features of Data warehouse

Ans:



1) Subject-oriented:

- ❖ A data warehouse is always a subject oriented as it delivers information about a theme instead of organization's current operations.
- ❖ It can be achieved on specific theme. That means the data warehousing process is proposed to handle with a specific theme which is more defined.
- ❖ These themes can be sales, distributions, marketing etc.

2) Integrated:

- ❖ A data warehouse is built by integrating data from various sources of data such that a mainframe and a relational database.
- ❖ In addition, it must have reliable naming conventions, format and codes. Integration of data warehouse benefits in effective analysis of data.

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

3) Time-Variant –

- In this data is maintained via different intervals of time such as weekly, monthly, or annually etc.
- The data resided in data warehouse is predictable with a specific interval of time and delivers information from the historical perspective.
- It comprises elements of time explicitly or implicitly. Another feature of time-variance is that once data is stored in the data warehouse then it cannot be modified, alter, or updated

4) Non-Volatile –

- As the name defines the data resided in data warehouse is permanent. It also means that data is not erased or deleted when new data is inserted.
- In this, data is read-only and refreshed at particular intervals. This is beneficial in analysing historical data and in comprehension the functionality
- Functionalities such as delete, update, and insert that are done in an operational application are lost in data warehouse environment. Two types of data operations done in the data warehouse are:

Data Loading and Data Access

Topic: Write a short note on Data Warehouse Architecture

Ans:

- Data warehouse is collection of data from multiple data sources organised under single database.

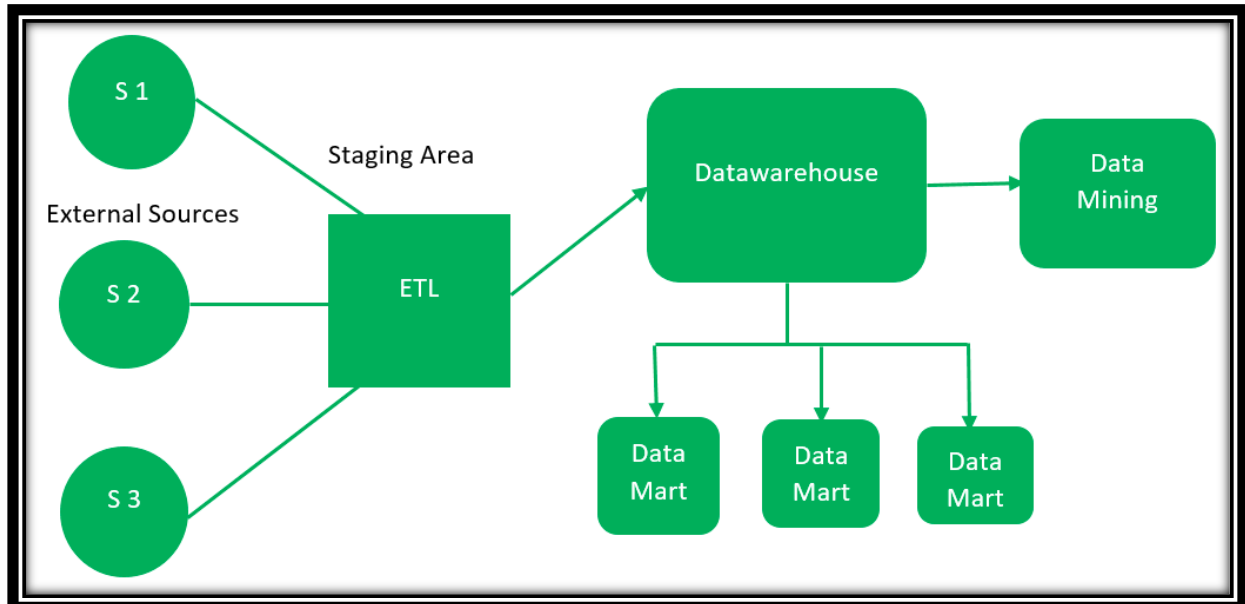
SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645



🌿 External Sources (Data Source Layer):

- 🌿 It is source from where data is collected irrespective of the type of data. Data can be structured, semi structured and unstructured as well.

🌿 Staging Area (Staging Layer):

- 🌿 Since the data, extracted from the external sources does not follow a particular format, so there is a need to validate this data to load into data warehouse.
- 🌿 For this purpose, it is recommended to use **ETL** tool.

E(Extracted): Data is extracted from External data source.

T(Transform): Data is transformed into the standard format.

L(Load): Data is loaded into data warehouse after transforming it into the standard format.

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.



(AFFILIATED TO SAURASHTRA UNIVERSITY)




2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645


Data Warehouse (Data Storage Layer):

-  After cleansing of data, it is stored in the data warehouse as central repository. It actually stores the Meta data and the actual data gets stored in the data marts.
-  **Note** that data warehouse stores the data in its purest form in this top-down approach.

Data Marts (Data Presentation Layer):

-  Data mart is also a part of storage component. It stores the information of a particular function of an organisation which is handled by single authority.

Data Mining:

-  In simple words, *data mining* is defined as a process used to extract usable data from a larger set of any raw data.

Sr No.	M.C.Q	Answer
1.	ETL stands for....	Extraction, Transform , Load
2.	What is Data Mining?	It is the process to get required or useful data from bulk of data
3.	To transform the data from in to particular format,is used	ETL

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

Topic: Write down advantages and disadvantages of Data Warehousing

Ans:

Advantages of Data Warehouse (DWH):

- ✿ Data warehouse allows business users to quickly access critical data from some sources all in one place.
- ✿ Data warehouse provides consistent information on various cross-functional activities. It is also supporting ad-hoc reporting and query.
- ✿ Data Warehouse helps to integrate many sources of data to reduce stress on the production system.
- ✿ Data warehouse helps to reduce total turnaround time for analysis and reporting.
- ✿ Restructuring and Integration make it easier for the user to use for reporting and analysis.
- ✿ Data warehouse allows users to access critical data from the number of sources in a single place. Therefore, it saves user's time of retrieving data from multiple sources.
- ✿ Data warehouse stores a large amount of historical data. This helps users to analyze different time periods and trends to make future predictions.

Disadvantages of Data Warehouse:

- ✿ Not an ideal option for unstructured data.
- ✿ Creation and Implementation of Data Warehouse is surely time confusing affair.
- ✿ Data Warehouse can be outdated relatively quickly
- ✿ Difficult to make changes in data types and ranges, data source schema, indexes, and queries.
- ✿ The data warehouse may seem easy, but actually, it is too complex for the average users.
- ✿ Despite best efforts at project management, data warehousing project scope will always increase.
- ✿ Sometime warehouse users will develop different business rules.
- ✿ Organisations need to spend lots of their resources for training and Implementation purpose.

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)

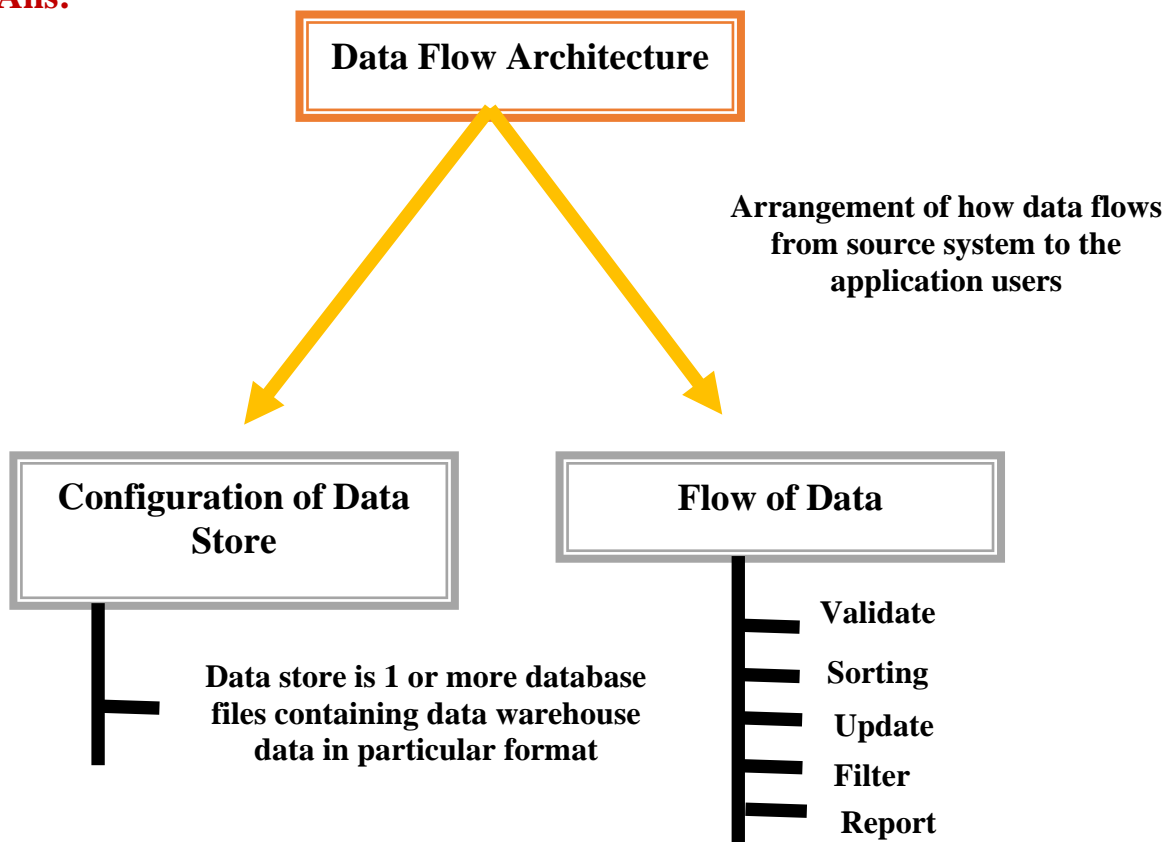


2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

Topic: Write a short note on Data Flow Architecture

Ans:



SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

 Following are the types of Data Store :

user-facing data store

A **user-facing data store** is a data store that is available to end users and is queried by the end users and end-user applications

An internal data store

An **internal data store** is a data store that is used internally by data warehouse components for the purpose of integrating, cleansing, logging, and preparing data, and it is not open for query by the end users and end-user applications.

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

hybrid data store

A **hybrid data store** is used for both internal data warehouse mechanisms and for query by the end users and end-user applications.

Components of Data Store :

Before the data is loaded to other data stores in a data warehouse.

✓ A normalized data store (NDS) is an internal master data store in the form of one or more normalized relational databases for the purpose of integrating data from various source systems captured in a stage, before the data is loaded to a user-facing data store.

✓ An operational data store (ODS) is a hybrid data store in the form of one or more normalized relational databases, containing the transaction data and the most recent version of master data, for the purpose of supporting operational applications.

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

✓ A dimensional data store (DDS) is a user-facing data store, in the form of one or more relational databases, where the data is arranged in dimensional format for the purpose of supporting analytical queries.



MDB (Multi dimensional Database)
Data is stored in cells
It is populated from
DDS

Sr No.	M.C.Q	Answer
1.	NDS stands for....	Normalised Data Store
2.	ODS stands for...	Operational Data Store
3.	DDS stands for	Dimensional Data Store
4.	NDS supportstype of .data store.	Internal
5.	DDS supportstype of data store	User Facing

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

Topic: Explain data warehousing today

Ans:

- Data warehouse is collection of data from multiple data sources organised under single database.
- In today's era, Data ware housing is very much useful.
- It is useful for decision making.
- It is very much useful to support analytical reporting and perform the query
- In any organization, it is helpful to perform production strategies, by comparing new or today's data with old stored data.
- There are many data warehouse applications which are used in different fields like: Financial Service, Banking Service, Manufacturing and distribution etc.
- The following are the functions of data warehouse tools and utilities –
 - 1) **Data Extraction** – Involves gathering data from multiple heterogeneous sources.
 - 2) **Data Cleaning** – Involves finding and correcting the errors in data.
 - 3) **Data Transformation** – Involves converting the data from legacy format to warehouse format.
 - 4) **Data Loading** – Involves sorting, summarizing, consolidating, checking integrity, and building indices and partitions.
 - 5) **Refreshing** – Involves updating from data sources to warehouse.

Topic: Explain Future Trends of Data Warehousing/ Explain IOT and Cloud in terms of data warehousing

Ans:

1. BI (Business Intelligence)

Example:
Any ERP (Enterprise Resource Planning) Software

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

2. IOT (Internet Of Things)

Example:
Smart Watch, Air-Conditioner

6. AI (Artificial Intelligence)

Example:
Robot in Covid-19 Situation

5. Cloud

Example:
Photos, Gmail Drive

4. Big Data

Focus on 3V's
Example: Hadoop, R Lang.

3. SOA
(Service Oriented Architecture)

Example: Any College website

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

Topic: What is Third-Party Data?

Ans:

- Third-party data is collected by entities that often do not have a direct relationship with end users.
- To obtain third party data, you need to purchase it from data providers.
- You can find these providers through DMP (Data Management Platforms)

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

CHAPTER-2:

Designing and Implementation of Data Warehousing

Sr No.	Topic
1.	Logical Design for Data Warehouse
2.	Physical Design for Data Warehouse
3.	Design Table, Fact Table for Data Warehouse
4.	Design and Implement effective physical data structure of data warehouse

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

Topic: Write a short note on logical design and physical design of data warehouse

Ans:

Logical Design:

- The client will define their business and functionality of their business for a data warehouse project.
- Logical design deals with the logical relationships between objects.
- ER (Entity Relationship) modelling technique can be used for logical design of data warehouse.
- ER modelling involves entities, attributes and relationship among them.
- **Entity:** It represents a chunk of data, which maps to a table in database.
- **Attribute:** It is part of an entity, which maps to a column in database.
- **Unique Identifier:** To define records uniquely.
- During the logical design phase, following are the different steps:
 - Define a model for your data warehouse
 - Select entities, attributes and relationships in a model
 - Select different types of information that you need in future
 - Check the logical relationships of different objects
 - Arrange the data into a series
 - See the attributes those are used to describe the entities
 - Set the unique identifier (UID) distinguishes between one instance of an entity and another

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

Physical Design:

- Physical design deals with the effective way of storing and retrieving the data.
- During the physical design process, we convert the logical design into physical by using database structures.
- During the physical design phase, following are the different steps:
 - Entities to tables
 - Relationships to foreign key constraints
 - Attributes to columns
 - Primary Unique identifier to primary key constraints
 - Unique identifiers to unique key constraints.

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

Logical Design	Physical Design
1) Deals with Client business and their functionality	1) Physical Design deals with storing and retrieving the data
2) ER Modelling can be used for logical design of data warehouse	2) Database and Tables are used for physical design of data warehouse
3) Different entities in the logical design are specified	3) Entities of logical design are converted into tables in physical design
4) Different attributes of respective entities are specified.	4) Attributes of logical design are converted into columns of the tables in physical design
5) UID is used to uniquely identify the records	5) UI of logical design is converted into primary key and foreign key constraints
6) Simpler than physical design	6) Complex than logical design

Feature	Logical	Physical
Entity Names	✓	
Entity Relationships	✓	
Attributes	✓	
Primary Keys	✓	✓
Foreign Keys	✓	✓
Table Names		✓
Column Names		✓
Column Data Types		✓

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

Sr No.	MCQ	Answer
1	Entities of logical design are converted into in physical design	tables
2	Attributes of logical design are converted into in physical design	Columns
3	UID stands for.....	Unique Identifier

Topic: Write a short note on Design Dimension table, fact table for data warehouse (M.I.M.P)

Ans:

Fact Table:

- ❑ A fact table is a primary table in a dimensional model.
- ❑ A Fact Table contains:
 - 1) Measurements/facts
 - 2) Foreign key to dimension table
- ❑ Facts are the continuous values
- ❑ Fact table contains two or more foreign keys that connects to the dimension table's primary keys

Dimension Table:

- ❑ Describe “who, what, when, how, and why” associated with the event
- ❑ A dimension table contains dimensions of a fact.
- ❑ They are joined to fact table via a foreign key.
- ❑ Dimension tables are de-normalized tables.
- ❑ The Dimension Attributes are the various columns in a dimension table
- ❑ Dimensions offers descriptive characteristics of the facts with the help of their attributes
- ❑ No set limit set for given for number of dimensions
- ❑ The dimension can also contain one or more hierarchical relationships

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

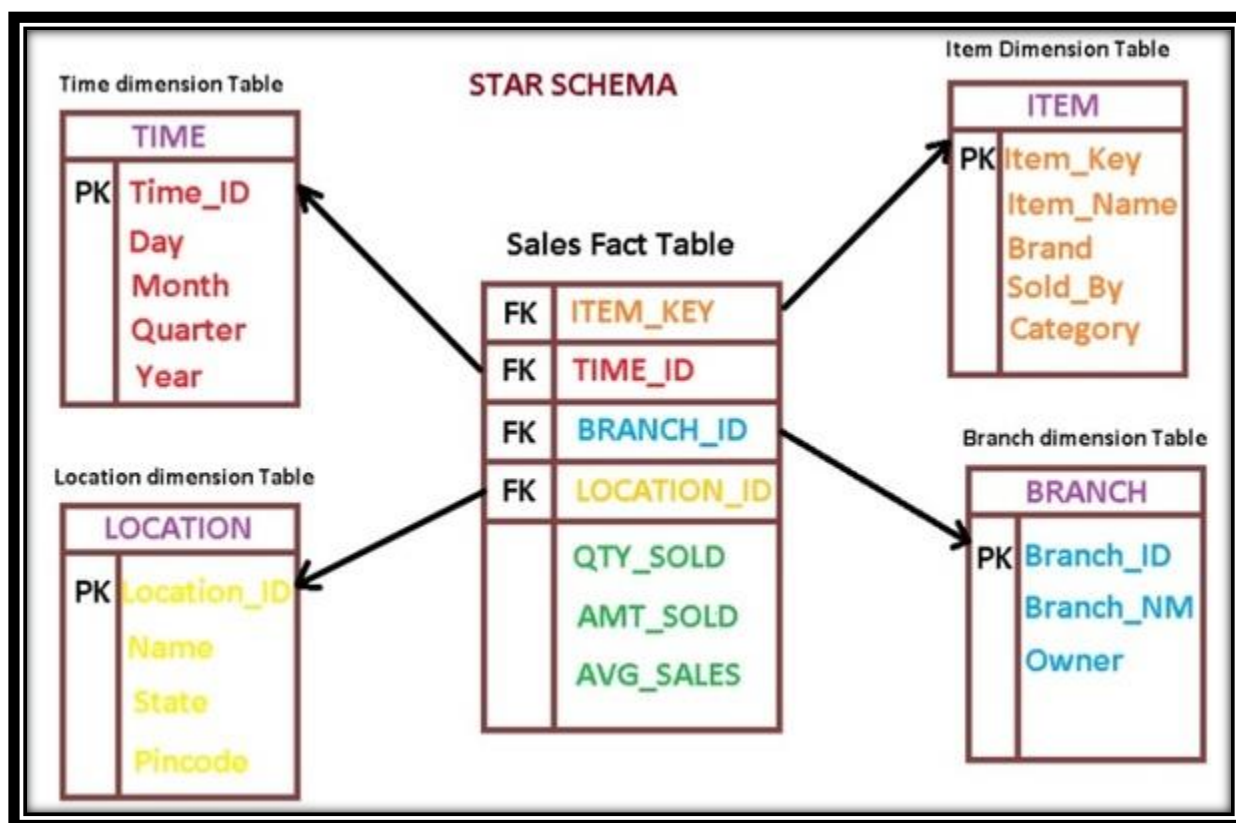
(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

Example:



- ❖ In the above example, Company sells items or product to the customers. A fact table holds the data to be analyzed, and a dimension table stores data about the ways in which the data in the fact table can be analysed. If we want to fetch the data that which product has been sold in particular place can be achieved from the above mentioned table.
- ❖ **Star Schema** in data warehouse, in which the center of the star can have one fact table and a number of associated dimension tables. It is known as star schema as its structure resembles a star.

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

Difference between Fact Table and Dimension Table:

Parameters	Fact Table	Dimension Table
Definition	Contains facts or measurement values	Contain description/text
Characteristics	Located at the centre of star topology or schema	Connected to fact table and located at the edges of the star schema or topology
Design	Defined by their grain (Level of detail stored in data)	Should be descriptive
Type of Data	Contains information like sales	Contains information like product name, category etc.
Key	Act as foreign key in fact table	Primary key of dimension table is mapped as foreign key to the fact table.
Hierarchy	Does not contain hierarchy	Contains Hierarchy. Example: Location: 1) Country 2) City 3) Pin code

Sr No.	MCQ	Answer
1	Fact table contains	Measurements or values
2	Dimension Table contains	Descriptive or text
3	Fact table is located atof star schema	Centre
4	Fact Table does not contain hierarchy (T/F)?	True

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

Topic: Write a short note on types of facts.

Types of Facts:

- A fact table stores Quantative information for analysis.
- Following are the types of facts in fact table

1) Additive Fact:

- Measures should be added to all dimensions
- Consider the following retail fact table:

STORE-KEY	PRODUCT-KEY	TRANSACTION - DATE	REVENUE
S1	P1	27 AUG 2010	100
S2	P1	27 AUG 2010	150
S3	P1	27 AUG 2010	300
S1	P2	28 AUG 2010	600
S2	P2	28 AUG 2010	300
S1	P1	29 AUG 2010	200
S2	P2	29 AUG 2010	100
S3	P3	29 AUG 2010	400

Store wise sales

S1	900
S2	550
S3	700

Product wise sales

P1	750
P2	1000
P3	400

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

Daily Sales

27 Aug 2010	550
28 Aug 2010	900
29 Aug 2010	700

2) Semi-Additive Fact:

- In this type of fact, measures may be added to some dimensions and not with others
- Consider the following retail fact table:

Acct_Id	Trans_Dt	Curr_Bal	Profit margin
21653	27Aug09	80000	0.06
21654	27Aug09	120000	0.08
21653	28Aug09	22000	0.08
21654	28Aug09	48000	0.12

- In the above table, we cannot sum up current balance across Acct Id
- If we ask balance for Id 21653 we will say that 22000, not 22000+80000

3) Non-Additive Fact:

- In this type of fact, facts cannot be summed up across any dimension key
- Example: % and ratio columns are non-additive facts.

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



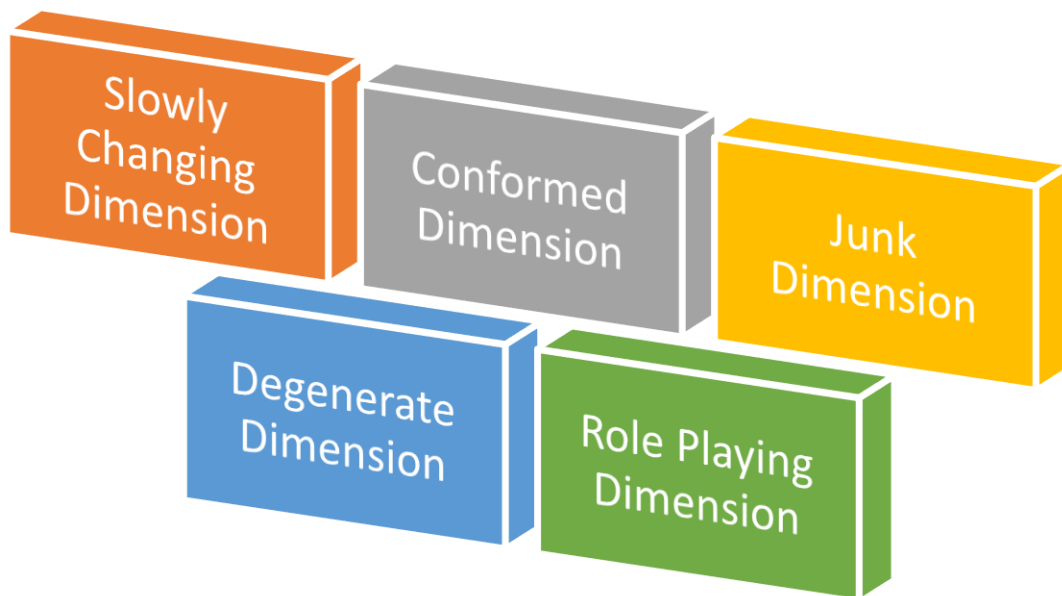
2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

Sr No.	MCQ	Answer
1	Measures added to some dimension and not with others is known as	Semi-Additive
2	% and ratio columns are example of	Non-Additive facts

Topic: Write a short note on types of Dimension

Ans:



Slowly Changing Dimension

- Dimension attributes that change slowly over a period of time rather than changing regularly is known as SCD
- Example: Name, Address

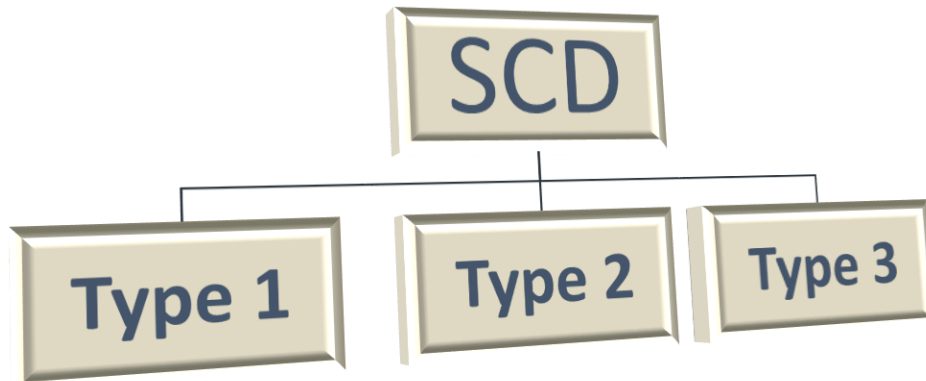
SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645



Type 1: Overwriting the Old Value:

	ID	Name	City
-1 st Record	1	Karishma	Rajkot
After Overwriting the value			
	ID	Name	City
1 st Record	1	Karishma	Bangalore

Advantage:

- 1) Less space is required as the data is overwritten

Disadvantage:

- 1) History of data is not maintained in this type of SCD

Type 2: Adding new row

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

	ID	Name	City
1 st Record	1	Karishma	Rajkot
After adding the new row			
	ID	Name	City
1 st Record	1	Karishma	Rajkot
New Row	1	Karishma	Bangalore

Advantage:

- 1) History of data is maintained in this type of SCD

Disadvantage:

- 2) This type of SCD requires more space

Type 3: Adding new Column

	ID	Name	City	
1 st Record	1	Karishma	Rajkot	
After adding the new Column				
	ID	Name	Old City	New City
After adding Column	1	Karishma	Rajkot	Bangalore

Advantage:

- 1) History of data is maintained in this type of SCD

Disadvantage:

- 3) This type of SCD requires more space

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



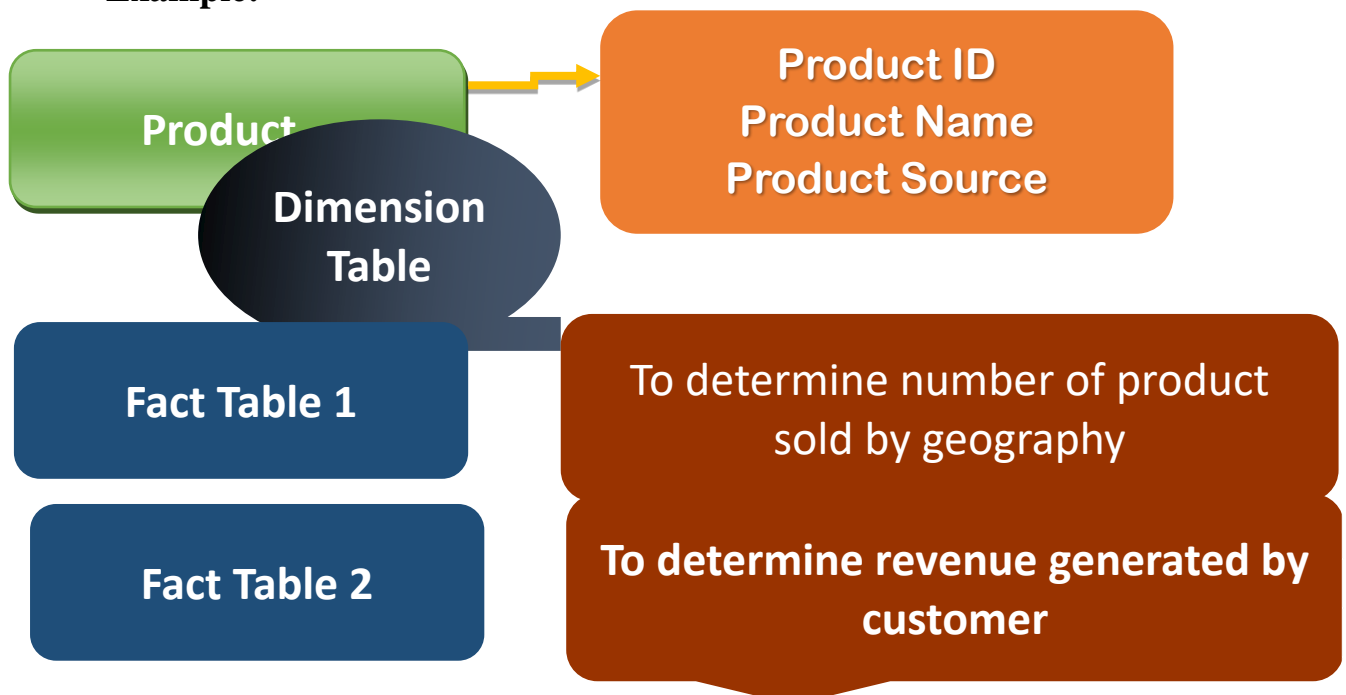
2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

Conformed Dimension

- It is used in multiple locations
- It maintains consistency of data that is data is not changed during processing or transfer
- Same dimension can be used across fact tables and can be used to create different fact tables

Example:



In the above example, same dimension can be used across fact tables that is revenue generated by customer and product sold are using data from product table.

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)

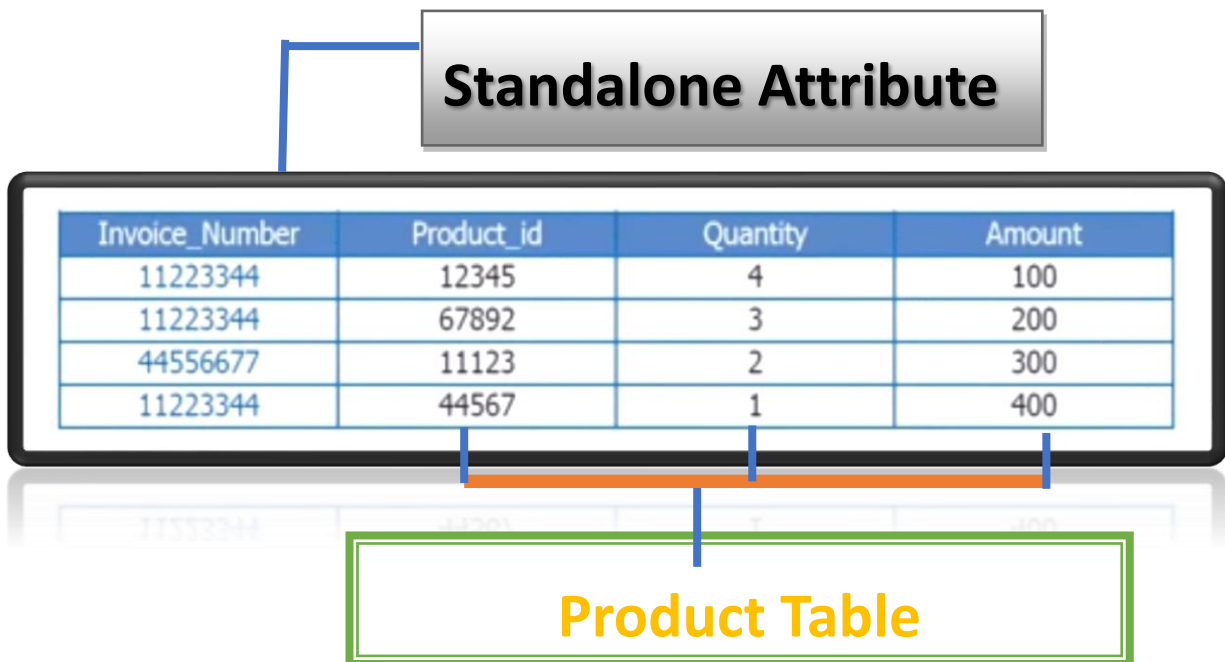


2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

Degenerate Dimension

- Dimension which is stored in fact table and has no associated dimension table is known as Degenerate Dimension



- In the above example, invoice number field is known as degenerate dimension because invoice number does not have associated dimension table.

Junk Dimension

- It is a single table with a combination of different and unrelated attributes

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

Example:

ATM Transaction:

Txn Successful	Txt Failed	No Cash	Invalid Pin	No Balance
1	0	0	0	0
0	1	1	0	0
0	1	0	1	0
0	1	0	0	1

- There are multiple flags that are set when a transaction takes place. If fields or flags do not integrate with the common dimensions like customer, product etc. (Not frequently required) as these are related to the system and not to the customer or product.
- Accommodating them in a table or binding them together in a new dimension table is known as Junk dimension because accommodating them in a fact table will make number of field in fact table huge and result in performance degradation.

Role Playing Dimension

- A dimensional table that has multiple valid relationship with fact table is known as Role playing Dimension
- For Example: Date Dimension can be used to fetch order placed date, order shipment date, order closure date

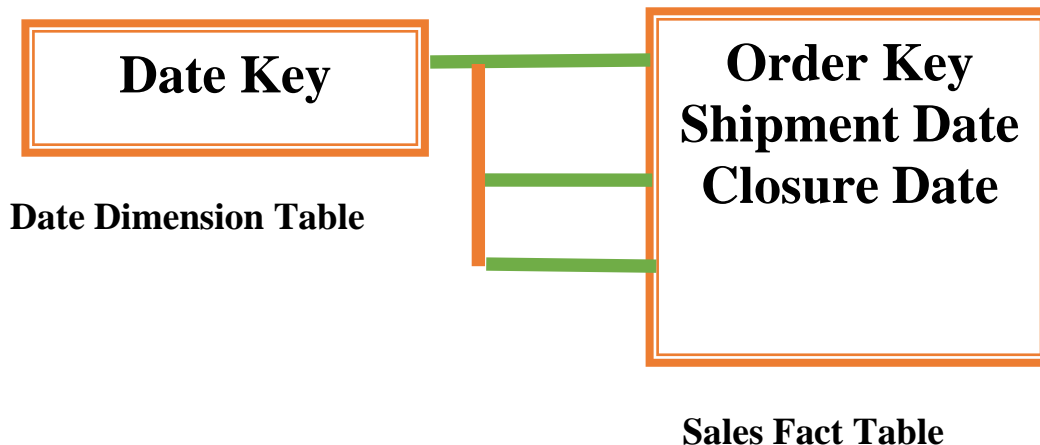
SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645



- 🌸 Hence, we need not create 3 dimension tables to link 3 columns in fact table
- 🌸 Dimension is processed only once and also stored only once. It saves time and memory space

Sr No.	MCQ	Answer
1	SCD stands for....	Slowly Changing Dimension
2	In SCD, inhistory of data is not maintained	Type 1
3	The dimension that do not have associated dimensional table is known as	Degenerate Dimension
4	The dimensions that are rarely used by customer or product are accommodated in one table is known as.....	Junk Dimension
5	A dimensional table that has multiple valid relationship with fact table is known as	Role Playing Dimension

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



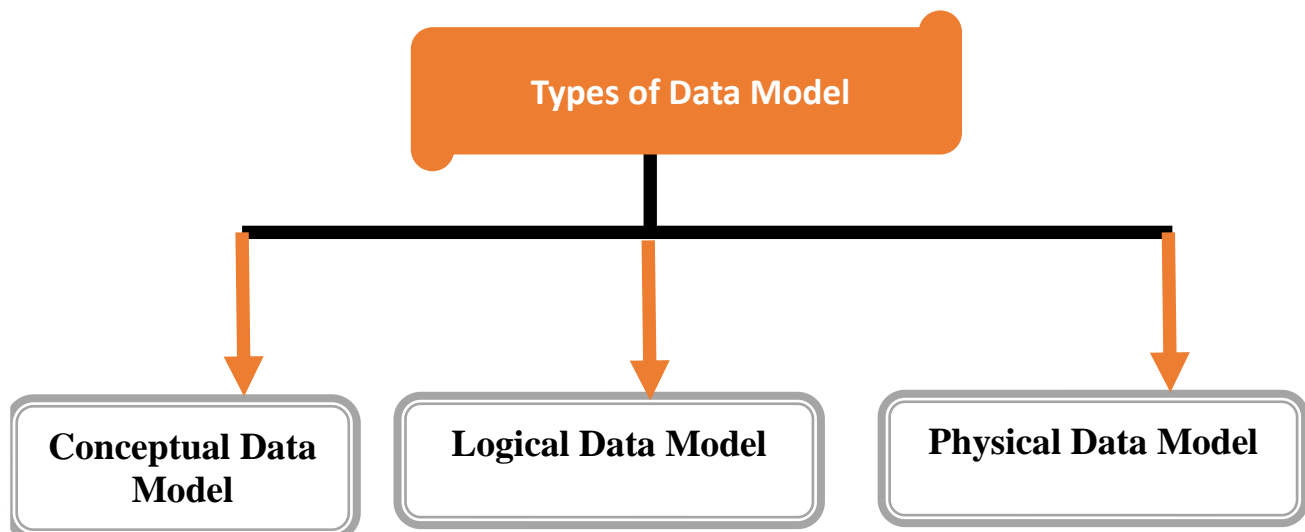
2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

Topic: Write a short note on physical data structure of data warehouse

Ans:

- Physical data structure means designing data and database.
- To design data and database the data-modelling concept is used.
- The Data Model is defined as an abstract model that organizes data description, data semantics, and consistency constraints of data. The data model emphasizes on what data is needed and how it should be organized instead of what operations will be performed on data.
- It is also helpful to identify missing and redundant data.
- Following are the types of data models:



1) Conceptual Data Model:

- The purpose of creating a conceptual data model is to establish entities, their attributes, and relationships. In this data modelling level, there is hardly any detail available on the actual database structure.
- Business stakeholders and data architects typically create a conceptual data model.
- 3 Main components of conceptual data model are:
 - 1) Entity: A real world thing. Example: Person, place, object etc.

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)

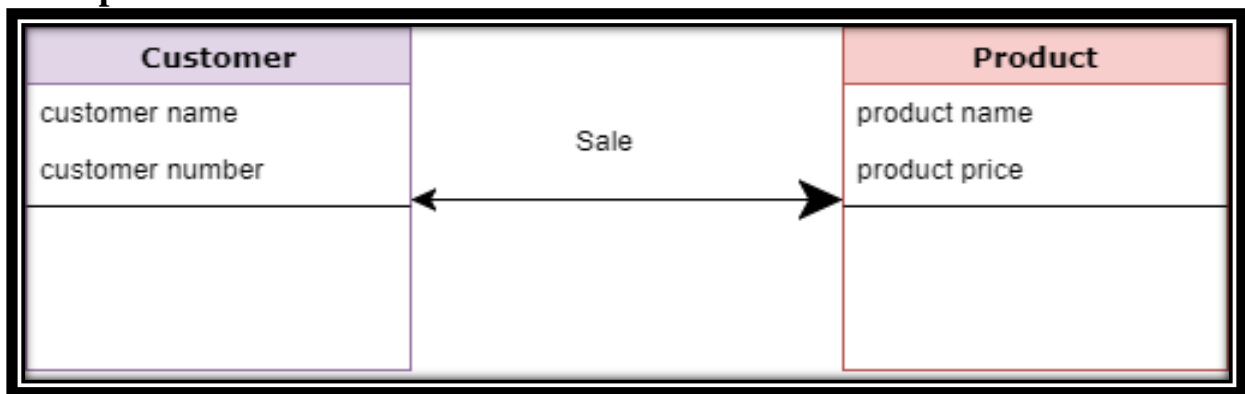


2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

- 2) Attributes: Properties or characteristics of entity. Example: Person name, Car color etc.
- 3) Relationship: Dependency between 2 entities. Example: xyz user uses Mercedes car

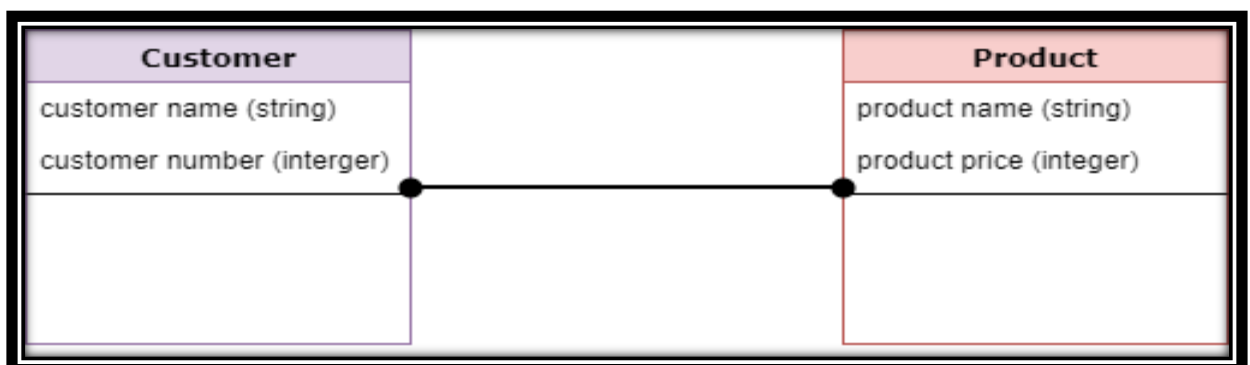
Example:



2) Logical Data Model:

- ❏ The Logical Data Model is used to define the structure of data elements and to set relationships.
- ❏ The logical data model adds further information to the conceptual data model elements. The advantage of using a Logical data model is to provide a foundation to form the base for the Physical model.
- ❏ At this Data Modelling level, no primary or secondary key is defined.

Example:



SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)

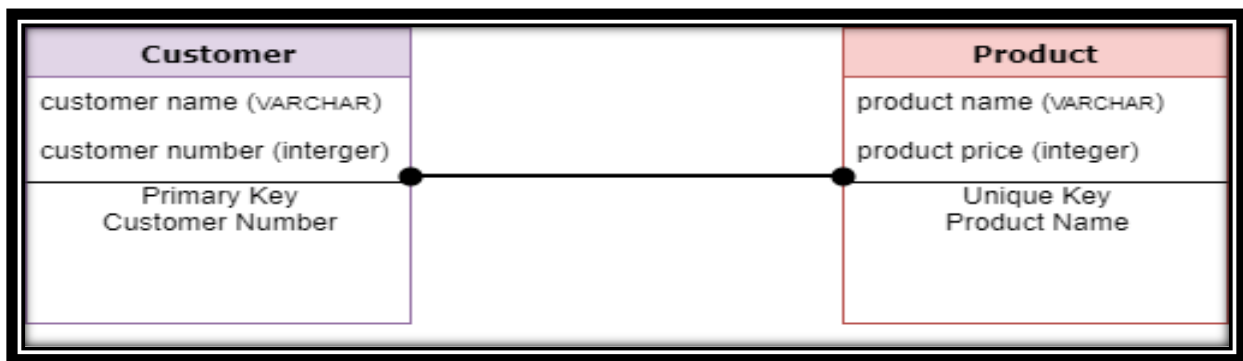


2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

3) Physical Data Model:

- A **Physical Data Model** describes a database-specific implementation of the data model.
- The physical data model also helps in visualizing database structure by replicating database column keys, constraints, indexes, triggers, and other RDBMS



SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

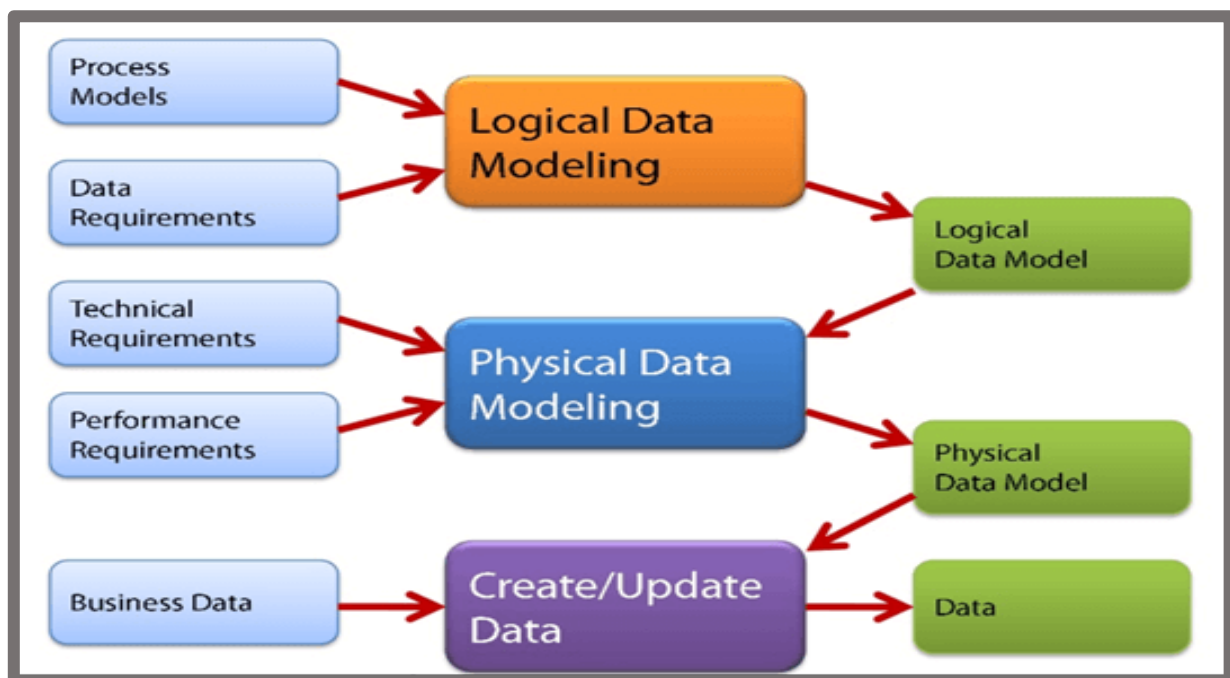
(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

PROCESS SHOWING 3 TYPES OF DATA MODEL



Advantages and Disadvantages of Data Model:

Advantages:

- 1) The main goal of a designing data model is to make certain that data objects offered by the functional team are represented accurately.
- 2) The data model should be detailed enough to be used for building the physical database.
- 3) The information in the data model can be used for defining the relationship between tables, primary and foreign keys, and stored procedures.
- 4) Help to recognize correct sources of data to populate the model

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

Disadvantages:

- 1) Even smaller change made in structure require modification in the entire application.
- 2) To develop Data model one should know physical data stored characteristics.

Sr No.	MCQ	Answer
1model establish entities, their attributes and relationship between different entities.	Conceptual
2	Constraints like primary key or foreign key are given inmodel	Physical

EXTRA QUESTIONS ASKED IN UNIVERSITY EXAMS

1) What is Table Space?

Ans:

- A table space is a storage location where the actual data underlying database objects can be kept
- It provides a layer of abstraction between physical and logical data, and servers to allocate storage for all DBMS managed segments.

2) What is Outtrigger Dimension?

Ans:

- A Dimension may have a reference to another dimension table. These secondary dimensions are known as outrigger dimensions.
- For instance, a bank account dimension can refer a separate dimension representing the date the account was opened.

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

CHAPTER-3:

Creating ETL Solutions with SSIS

Topic: Write a short note on ETL

Ans:

- ETL Stands for Extract, Transform and Load.
- Extract means extract the data from different data sources.
- Transform means transform the data for storing with proper format and structure for analysis purpose.
- Load means loading the data into final target.
- ETL process is generally used for quick execution of request and response the data from database.

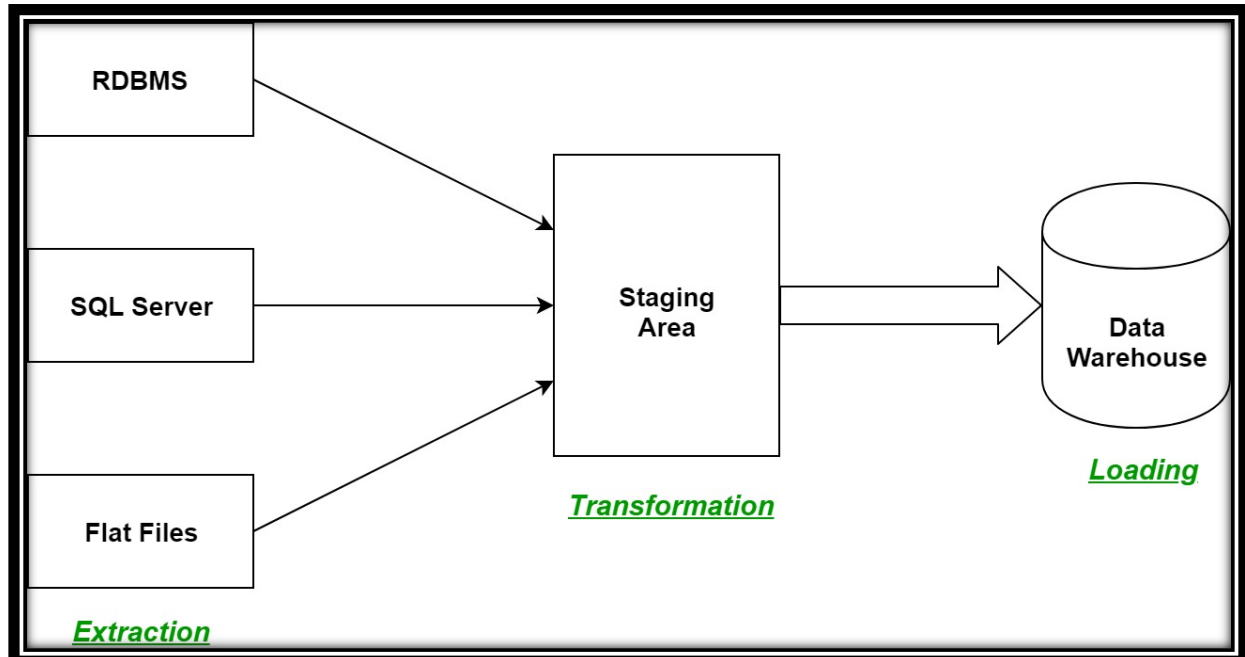
SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645



1) Extraction:

- The first step of the ETL process is extraction.
- In this step, data from various source systems is extracted which can be in various formats like relational databases, No SQL, XML and flat files into the staging area.
- It is important to extract the data from various source systems and store it into the staging area first and not directly into the data warehouse because the extracted data is in various formats and can be corrupted also.
- Hence loading it directly into the data warehouse may damage it and rollback will be much more difficult

2) Transformation:

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

■ The second step of the ETL process is transformation. In this step, a set of rules or functions are applied on the extracted data to convert it into a single standard format.

■ Following process/tasks are involved:

1. **Filtering** – loading only certain attributes into the data warehouse.
2. **Cleaning** – filling up the NULL values with some default values, mapping U.S.A, United States and America into USA, etc.
3. **Joining** – joining multiple attributes into one.
4. **Splitting** – splitting a single attribute into multiple attributes.
5. **Sorting** – sorting tuples on the basis of some attribute (generally key-attribute).

3) Loading:

■ The third and final step of the ETL process is loading.

■ In this step, the transformed data is finally loaded into the data warehouse.

■ Sometimes the data is updated by loading into the data warehouse very frequently and sometimes it is done after longer but regular intervals. The rate and period of loading solely depends on the requirements and varies from system to system.

Sr No.	MCQ	Answer
1	ETL stands for	Extract Transform Load
2	In which stage of ETL, the data is converted into proper format?	Transformation
3		

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

Topic: Write a short note on SSIS

Ans:

- SSIs stands for SQL Server Integration Services
- It is the component of Microsoft SQL Server database
- SSIS is a fast & flexible data warehousing tool used for data extraction, loading and transformation like cleaning, aggregating, merging data, etc.
- It makes it easy to move data from one database to another database. SSIS can extract data from a wide variety of sources like SQL Server databases, Excel files, Oracle and DB2 databases, etc.
- SSIS also includes graphical tools & wizards for performing workflow functions like sending email messages, FTP operations, data sources, and destinations.

SSIS is a tool mainly used to perform two functionalities:

1) Data Integration

SSIS performs data integration by combining the data from multiple sources and provides unified data to the users.

2) Workflow

Workflow can be used to perform several things. Sometimes we need to execute some specific steps or a particular path which is either based on the time period or the parameter passed to the package or the data queried from the database. It can be used to automate the maintenance of SQL Server databases and provides the update to the multidimensional analytical data.

Sr No.	MCQ	Answer
1	SSIS stands for	Sql Service Integration Service
2is the component of Microsoft SQL Server database	SSIS

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

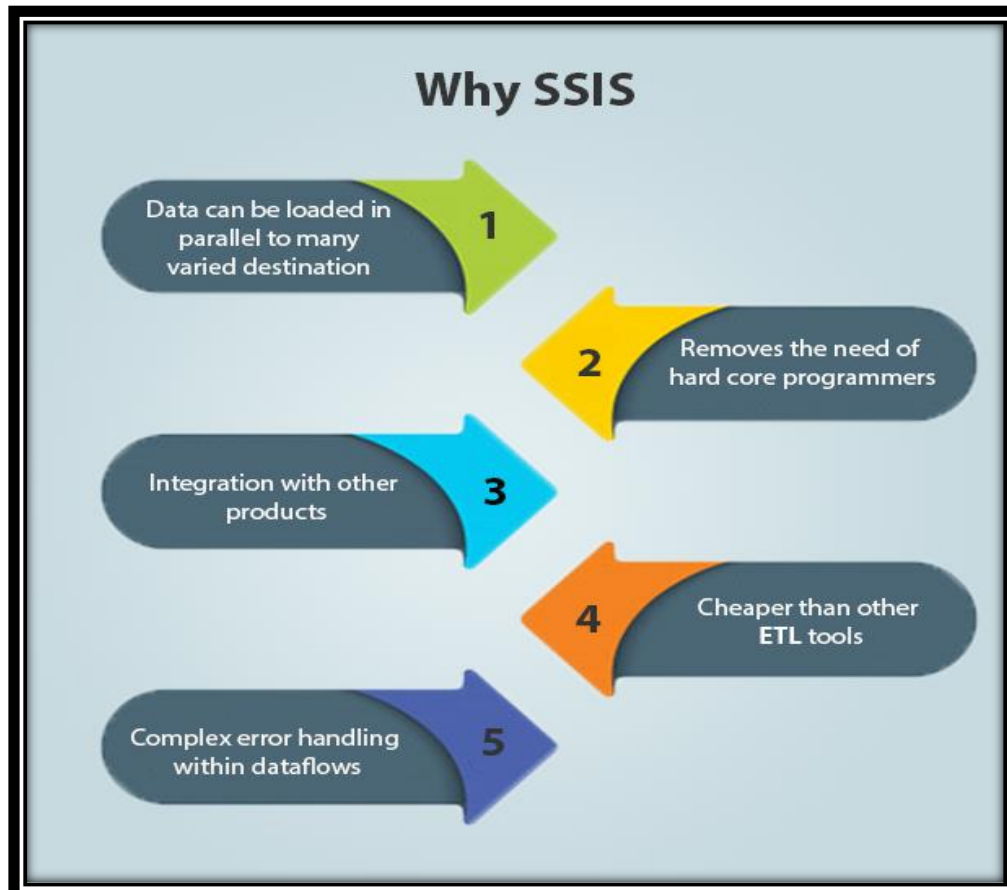
(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

Why SSIS?



- ❏ **Data can be loaded in parallel to many varied destinations**
SSIS is used to combine the data from multiple data sources to generate a single structure in a unified view. Basically, it is responsible for collecting the data, extracting the data from multiple data sources, and merging into a single data source.
- ❏ **Removes the need of hard core programmers**
SSIS is a platform that has the capability to load a large amount of data from excel to a SQL Server database.
- ❏ **Integration with other products**
SSIS tool provides tight integration with other products of Microsoft.

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

Cheaper than other ETL tools

SSSIS tool is cheaper than most of the other tools. It can resist with other base products, their manageability, business intelligence, etc.

Complex error handling within data flows

SSSIS allows you to handle the complex error within a dataflow. You can start and stop the dataflow based on the severity of the error. You can even send an email to admin when some error occurs. When an error is resolved, then you can pick the path in between the workflow.

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

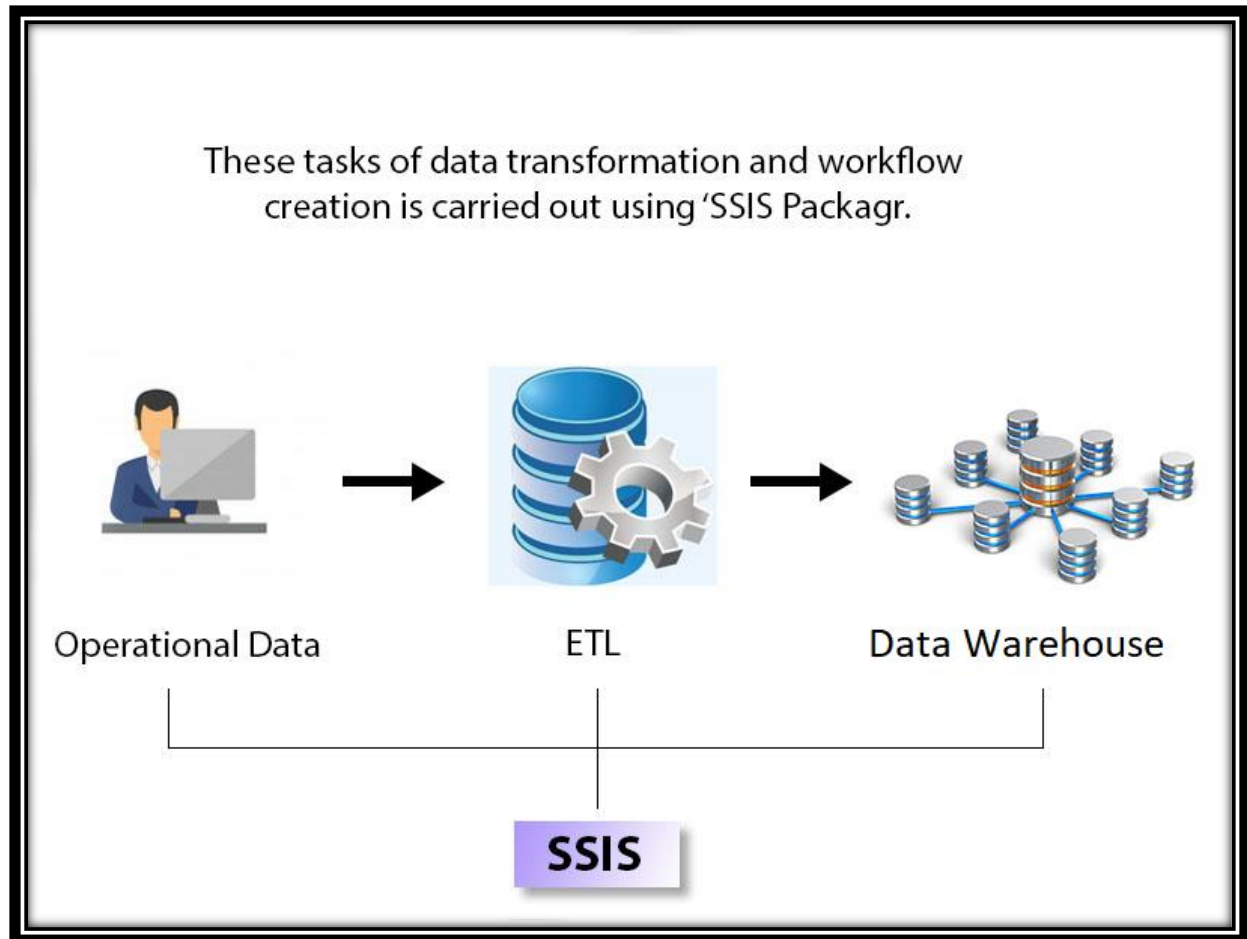
(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

How SSIS works?



Operational data

- Operational data is a database used to integrate the data from multiple data sources to perform additional operations on the data. It is the place where the data is housed for current operation before sending to the data warehouse for storing, reporting, or archiving.

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

ETL:

- ETL is the most important process in SSIS tool. ETL is used to Extract, Transform, and Load the data into a data warehouse.
- ETL is a process responsible for pulling out the data multiple data sources, transforming the data into useful data, and then storing the data into a data warehouse. The data can be in any format xml file, flat file, or any database file.
- It also ensures that the data stored in the data warehouse is relevant, accurate, high quality, and useful to the business users.
- It can be easily accessed so that the data warehouse can be used effectively and efficiently.
- It also helps the organization to make data-driven decisions by retrieving the structured and unstructured data from multiple data sources.

Data warehouse:

- Data warehouse is a single, complete, and consistent store of data which is formulated by combining the data from multiple data sources.

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

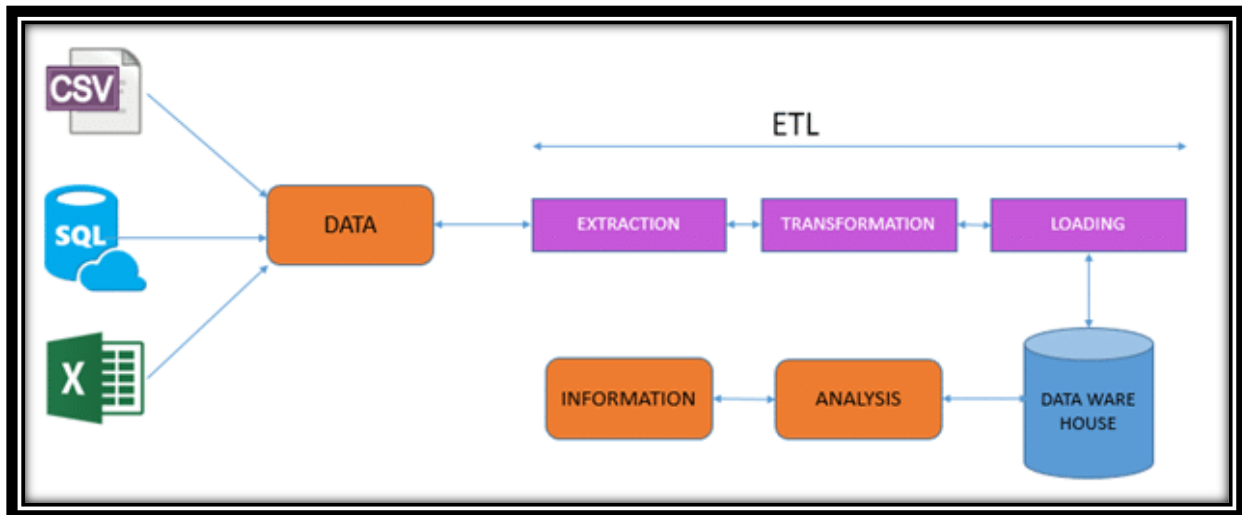
(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

Example showing working of SSIS:



Sr No.	MCQ	Answer
1	Which are the three main elements in working of SSIS?	Operational Data Store, ETL, Data warehouse

Requirement of SSIS:

- To install SQL Server
- To install SSDT (SQL Server Data Tools)

Components of SSIS:

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

Following are the components of SSIS:

- Control Flow
- Data Flow
- Package Explorer
- Parameters
- Event Handler

1) Control Flow (Stores Containers and Task):

- Control flow is a brain of SSIS package. It helps you to arrange the order of execution for all its components.
- The components contain containers and tasks which are managed by precedence constraints.
- Precedence constrain are package component which direct tasks to execute in a predefined order. It also defines the workflow of the entire SSIS package.
- It controls the execution of the two linked tasks by executing the destination tasks based on the result of the earlier task — business rules which are defined using special expressions.

Task:

- It is the individual unit of work.
- It is same as method/function used in the programming language

Container:

- Grouping task together into units of work
- Allows you to declare variable and event handlers.

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

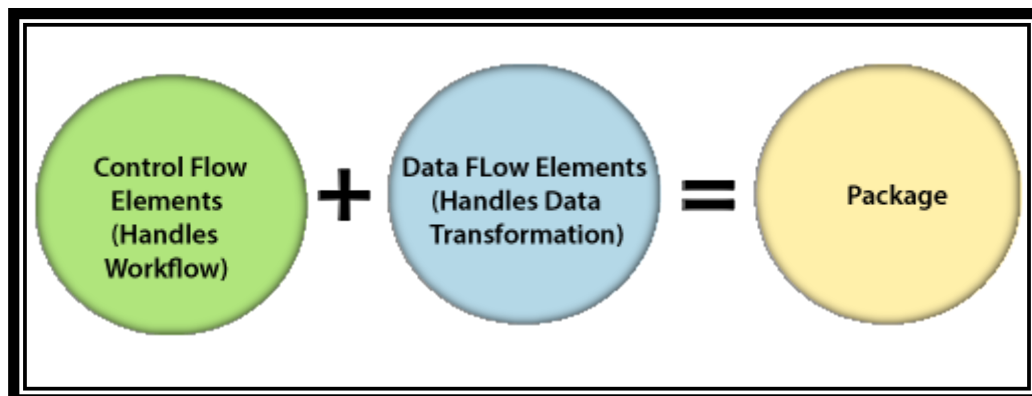
3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

2) Data Flow (Source, Destination and Transformations):

- The main use of the SSIS tool is to extract data into the server's memory, transform it, and write it to another destination. If Control Flow is the brain, Data Flow is the heart of SSIS

3) SSIS Package:

- The Package is a fundamental block where you code in SSIS. Here, code does not mean that you are coding in some programming language; it means the development that you do.
- The development is done in the SSIS package. SSIS is mainly used for the ETL process, and the ETL process is performed inside the SSIS package.
- SSIS package is composed of two parts:**



4) Parameters:

- Parameters behave much like a variable but with a few main exceptions. It can be set outside the package easily. It can be designated as values that must be passed in for the package to start.

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

5) Event Handler:

-  Allows you declare event handlers which should be in the scope of that specific container

Sr No.	MCQ	Answer
1	SSDT stands for	Sql Server Data Tools
2	Control flow contains	Container and Task
3	Data Flow contains	Source and Destination
4	Control Flow + Data Flow is known as	Package
5	What is the extension of package?	.dtsx
6	dtsx stands for	Data Transformation Service
7	What are the requirement for SSIS?	1) Install SQL Server 2) Install SSDT

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

PRACTICAL: 1 Example demonstrating Data Flow Task (Create the database in excel and import it in text file)

Step 1: Create one excel sheet

stuid	stuname	marks	gender
1	abc	100	Female
2	xyz	100	Male

Excel Sheet

Step 2: Create one blank text file (Name it as: stu.txt)



SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

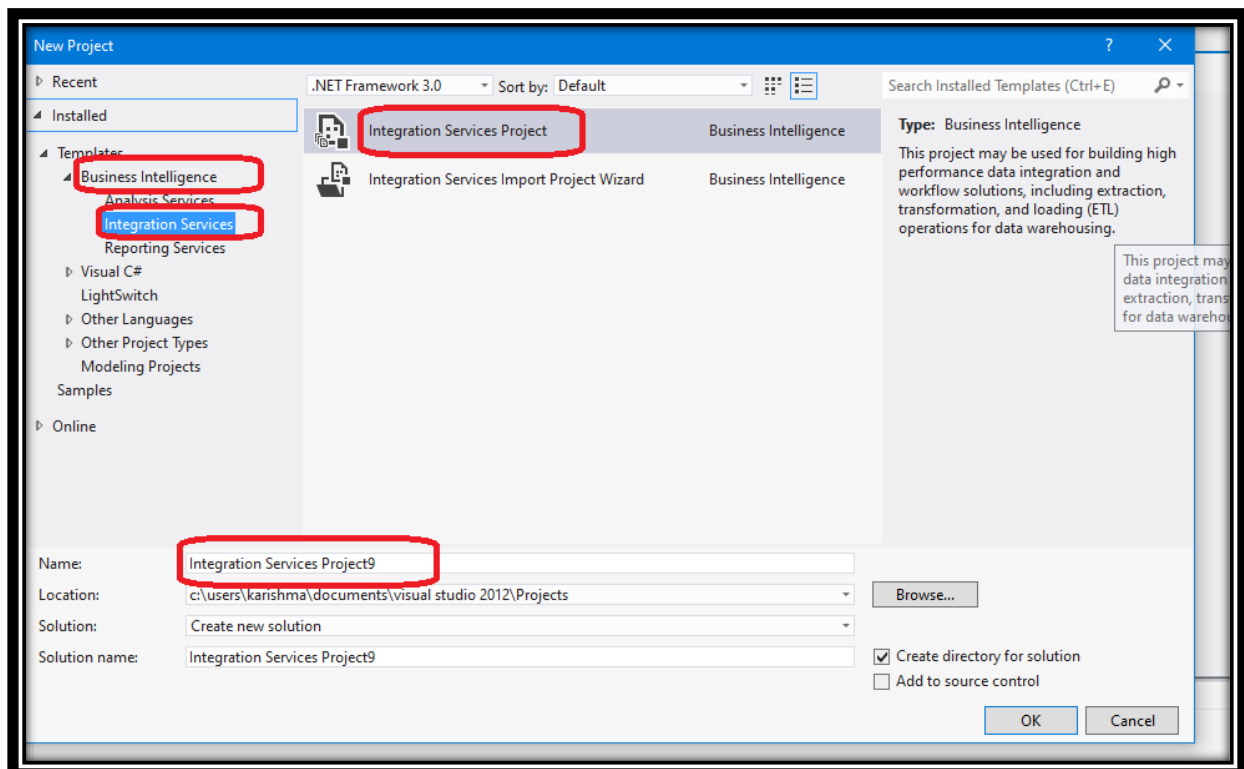
(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

Step 3: Start Visual Studio – New Project – Integration service project



Step 4: Click on Control Flow Tab and then add Data Flow Task

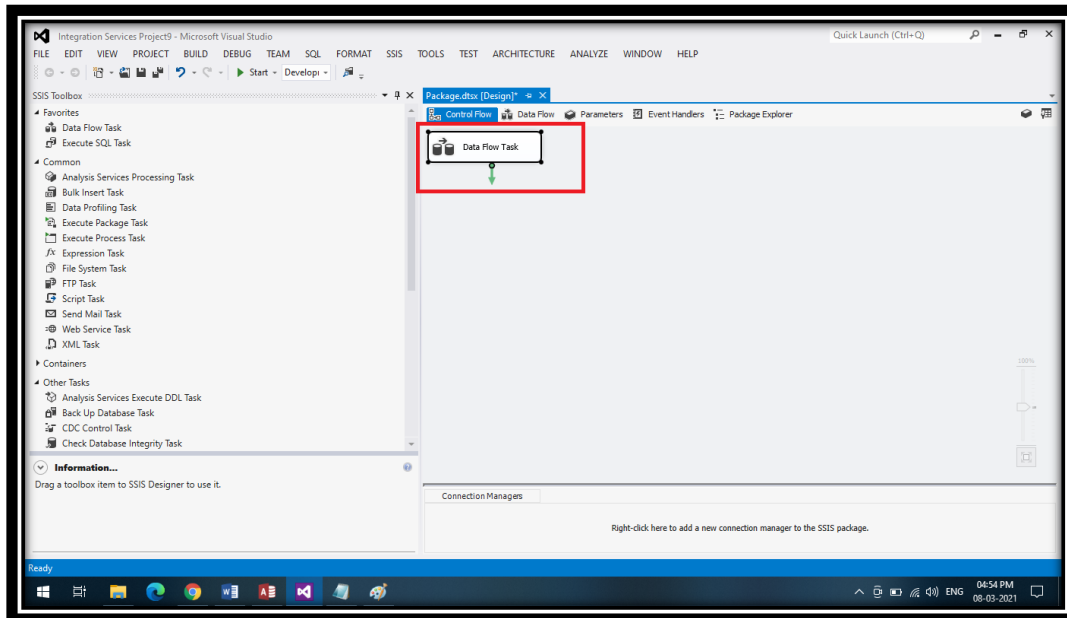
SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)

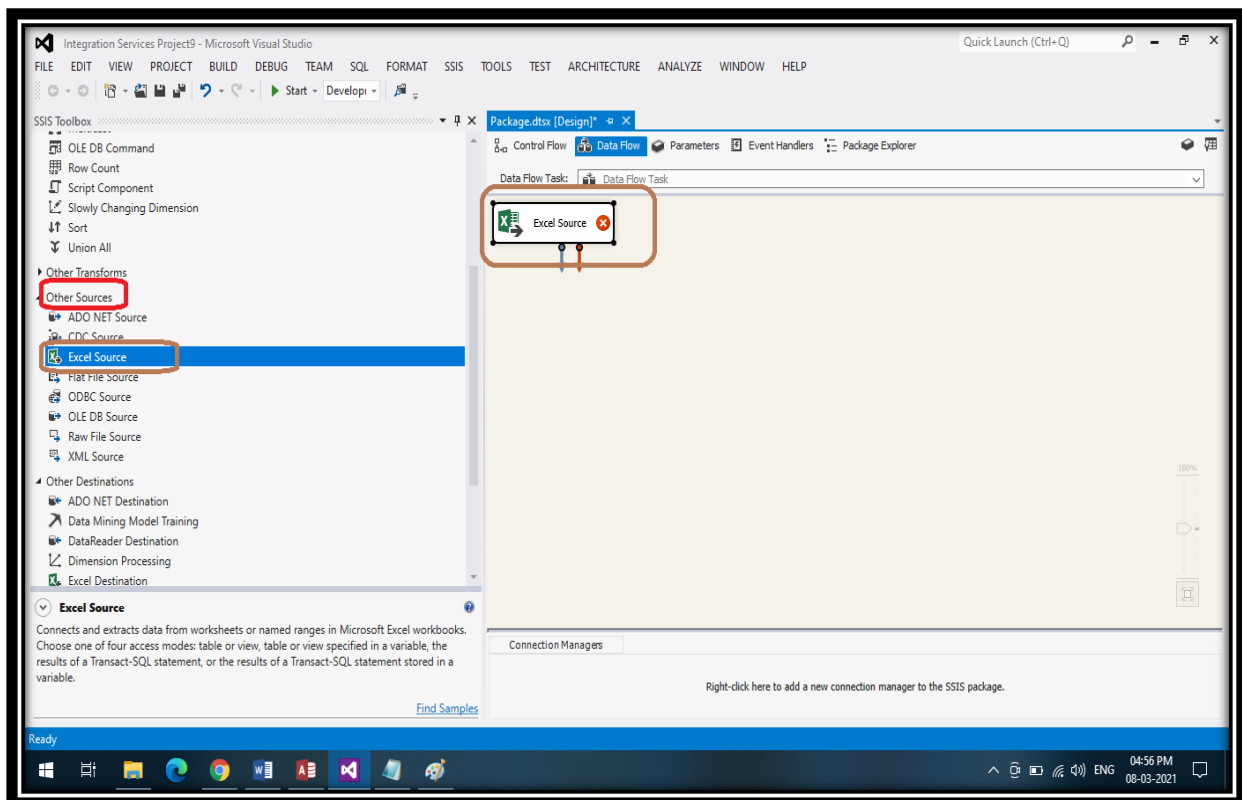


2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645



Step



5: Goto SSIS Toolbox and select excel data source

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

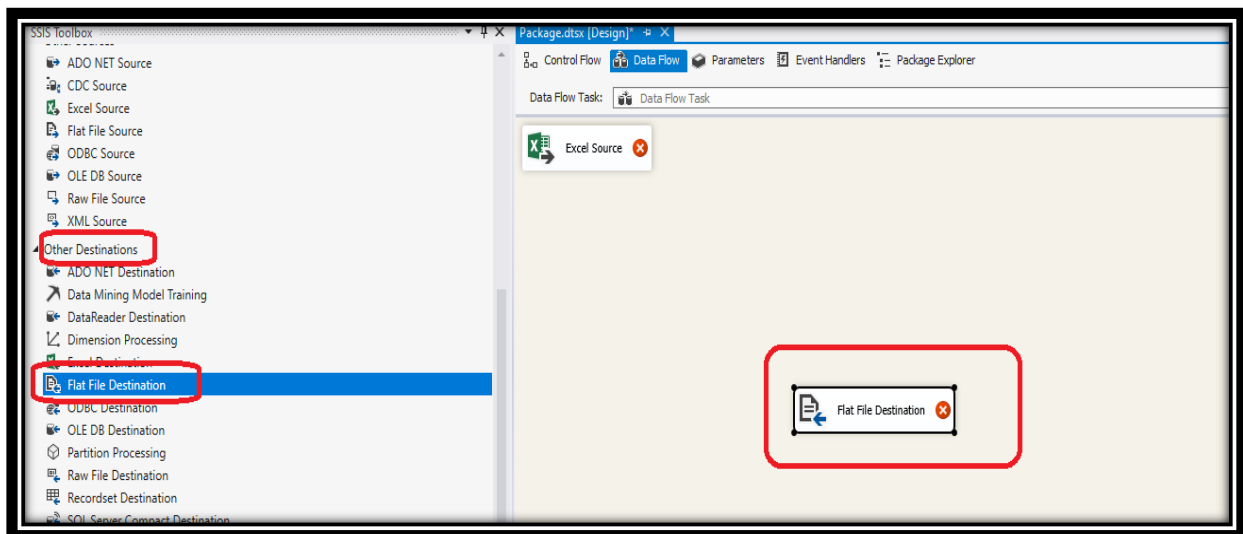
(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

Step 5: Goto SSIS Toolbox and select flat file destination



SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

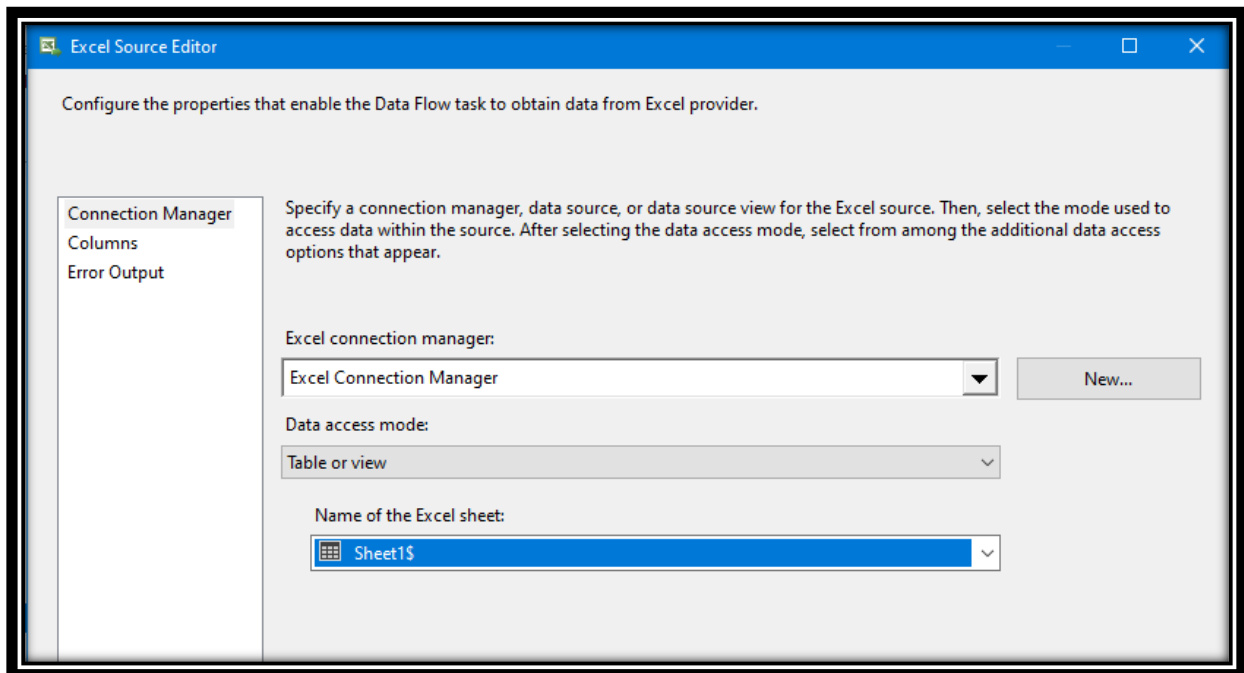
(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

Step 6: Double click on source and select source file by giving proper path



Step 7: Drag the connection line from source to destination.

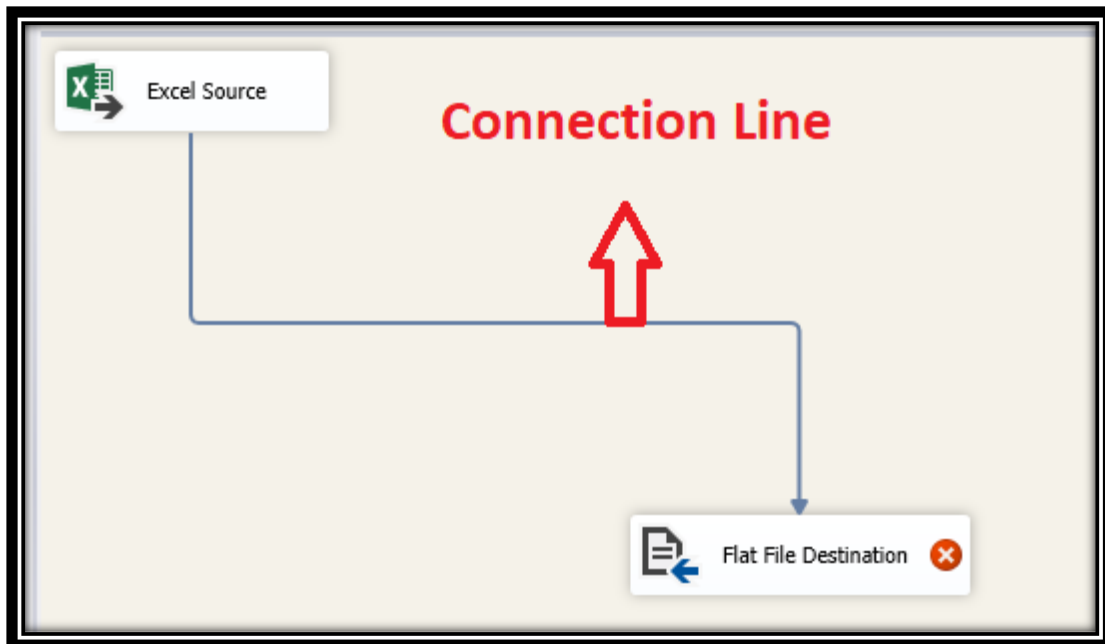
SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645



Step 8: Double click on destination file and give the name of empty text file

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

Flat File Connection Manager Editor

Connection manager name: Flat File Connection Manager 1

Description:

Select a file and specify the file properties and the file format.

File name: F:\Stu.txt

Locale: English (India) Unicode

Code page: 1252 (ANSI - Latin I)

Format: Delimited

Text qualifier: <none>

Header row delimiter: {CR}\n

Header rows to skip: 0

Column names in the first data row

Step 9: Click on mapping and then click ok

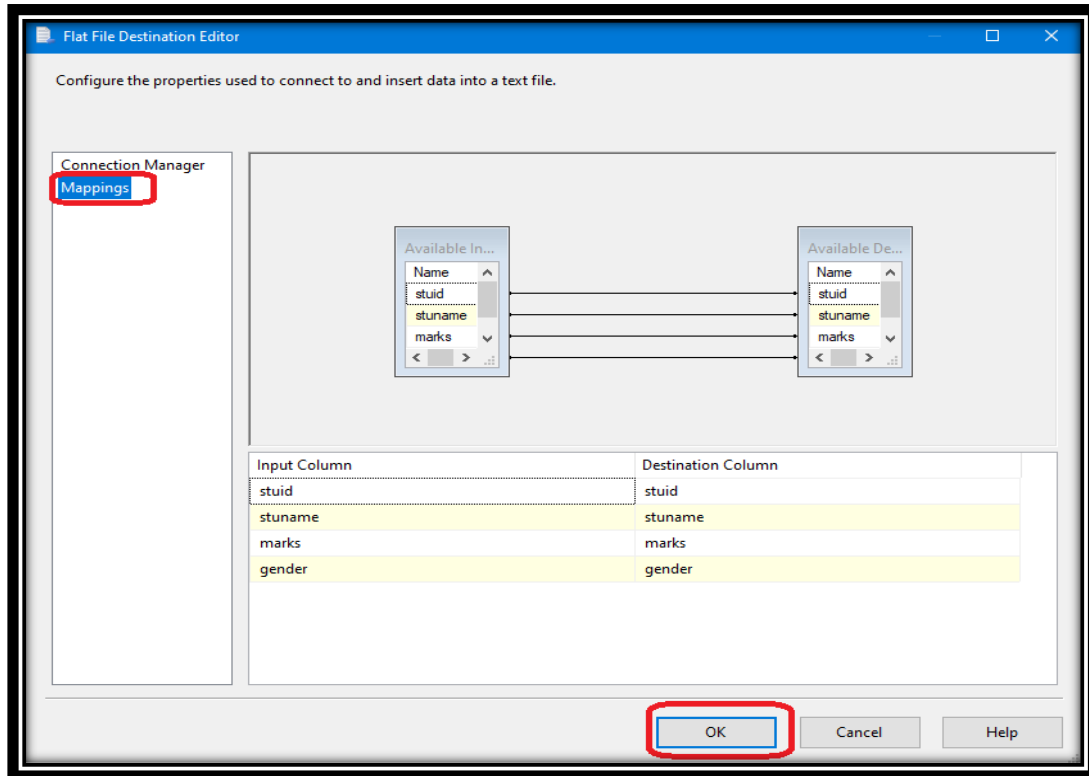
SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645



SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

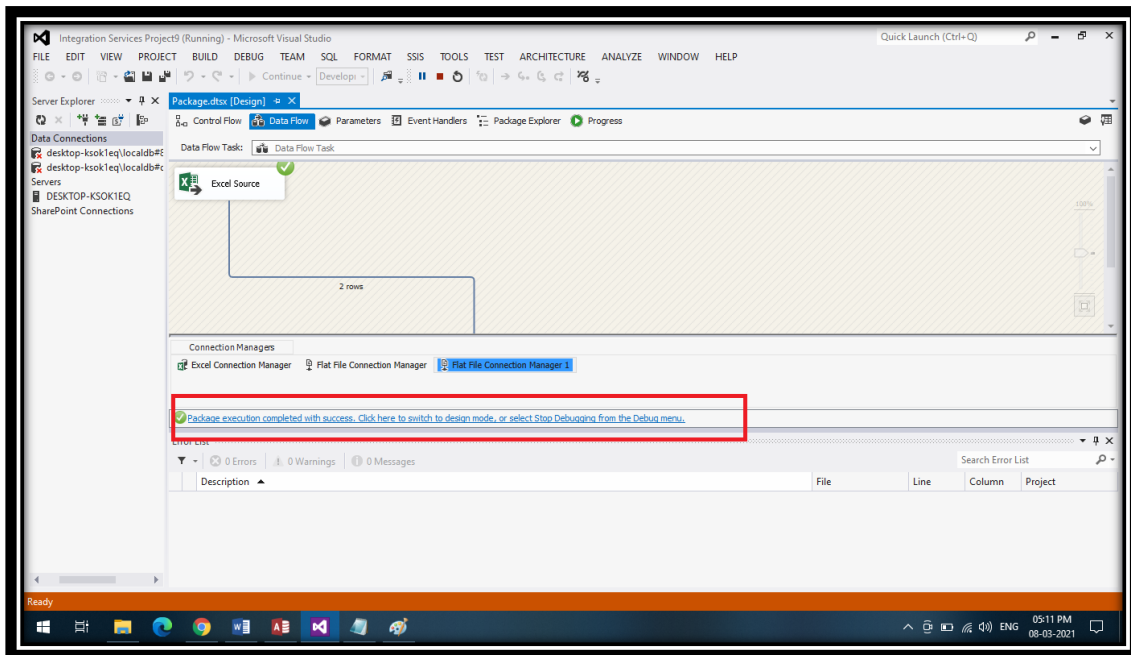
(AFFILIATED TO SAURASHTRA UNIVERSITY)



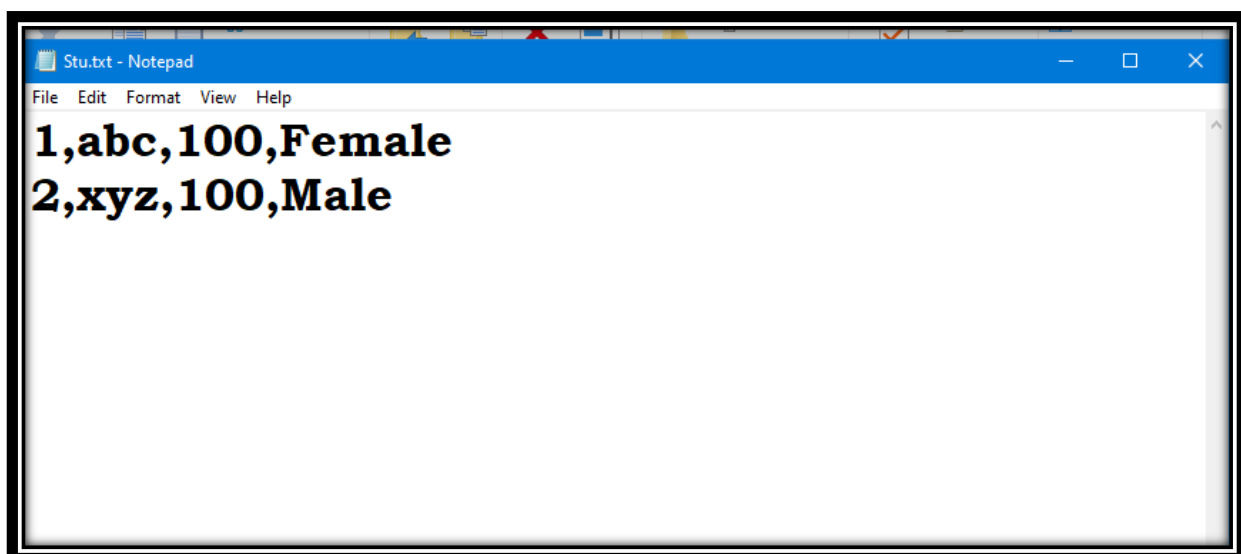
**2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590**

**3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645**

Step 10: Run your integration project by clicking on start button or pressing f5



Step 11: Check your output in the text file



SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

SSIS Task:

- In SSIS package, we can add the tasks. A task is a unit of work and we will have different types of tasks to perform different kinds of work.
- There are various types of tasks, but we will discuss the most common tasks used in SSIS:

1) Execute SQL Task:

It is used to execute the SQL statements against a relational database.

2) Data Flow Task:

It is mainly used to read the data from one or multiple data sources, transform the data and can also load the data to one or more destinations.

3) Analysis Services Processing Task:

It is used to process objects of an SSAS cube or Tabular model.

4) Execute Package Task:

This task is used to call the other packages within the same project. You can even pass the variable values to the called package.

5) Execute Process Task:

It allows you to run an application or batch scripts as a SQL Server Integration Services. It can be used to open the standard application such as Microsoft Excel, Microsoft Word, etc. It is also used to unzip the compressed file.

6) File System Task:

It can be used to perform the manipulations in the file system such as moving files, deleting files, renaming the files, changing the directory, etc.

7) FTP Tasks:

FTP tasks are used to perform the operations on files and folders. For example, if you want to send or receive the file from the FTP server to the local directory, then the SSIS FTP task is used.

8) Script Task:

This task allows you to write the .Net code that you want to perform.

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645



9) Send Mail Task:

This task is used to send an email. It is mainly used when you want to notify the users regarding the state of the task whether it is in a running state or some error has occurred.

Sr No.	MCQ	Answer
1	Which task in SSIS is used to write code?	Script Task
2task is used to move the data from source to destination	Data Flow
3	To send email which task of SSIS is used?	Send Email Task
4	If you want to send or receive the file from the FTP server to the local directory, then the SSIS task is used.	FTP
5	Which task is used to execute SQL Statements against relational database?	ExecuteSQL

Topic: Write a short note on Control Flow

Ans:

-  A package consists of a control flow and, optionally, one or more data flows.
-  SQL Server Integration Services provides three different types of control flow elements:
 - 1) Containers: They provide structures in packages
 - 2) Task: It provide functionality
 - 3) Precedence Constraints: It connect the executables, containers and tasks in an ordered control flow.

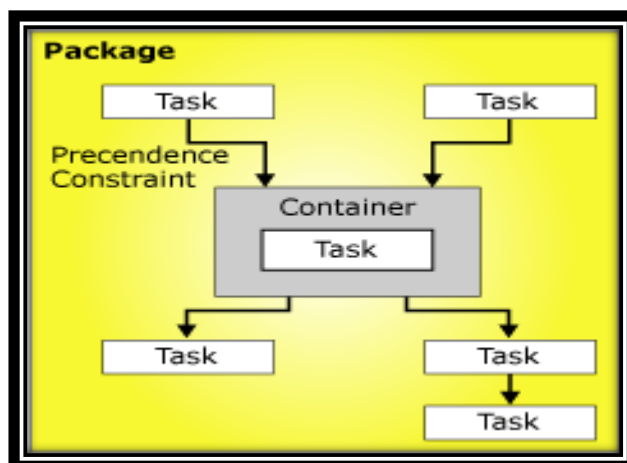
SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645



- The above diagram shows a control flow that has one container and six tasks. Five of the tasks are defined at the package level, and one task is defined at the container level. The task is inside a container.
- You can create control flow in a package by using control flow tab in SSIS designer.

Sr No.	MCQ	Answer
1	What is precedence Constraint?	It connects executables, containers and tasks in an ordered control flow

Topic: Write a short note on Data Flow

Ans:

- Data flow in SSIS defines the flow of data from a source to a destination.
- The Data Flow Task is the most frequently used task in SSIS, which contains the data transformation logic (ETL processes).
- It moves the data from source to destination and adds transforms in them to merge, update or split data.

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

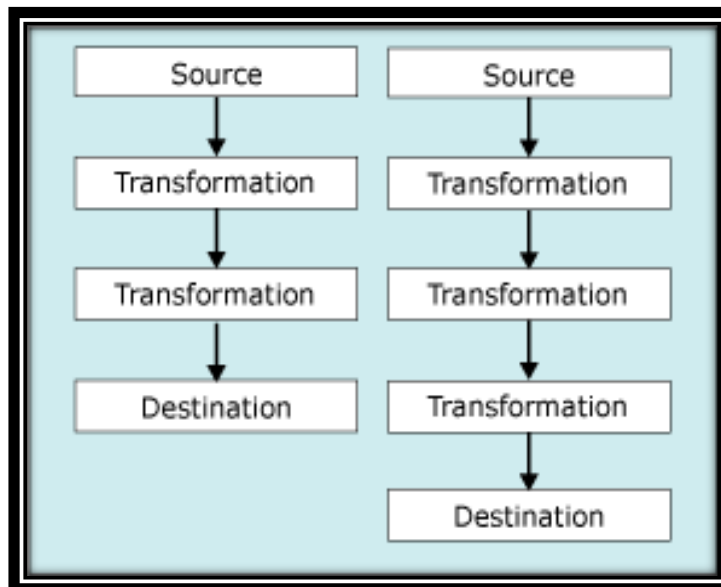
(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

- When you want multiple flows then a data flow task can include multiple data flows
- Data flow engine determines order of execution from one data flow task which has multiple data flows
- To insert bulk data from text files into a SQL Server database, you can use bulk insert task instead of Data flow task and a data flow
- However, bulk insert task cannot transform data.



Sr No.	MCQ	Answer
1	To insert bulk data from text files into a SQL Server database, you can useinstead of data flow task and a data flow	Bulk insert task

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

Topic: What is Container OR Explain Containers of Integration of services?

Ans:

- Containers are objects in SSIS
- It provides structure to packages and services to tasks
- They support repeating controls flows in packages and group task and containers into meaningful unit of work
- Containers can include other container in editions to task.
- There are 4 types of containers:

1) For each loop Container:

It runs a control flow repeatedly by using enumerator. The foreach loop container repeats the control flow for each member of a specified enumerator.

Example: For each file enumerator loops through a collection of files within a folder and makes it possible to execute. (individual student par operation perform karwu)

2) For loop Container:

It runs control flow by testing a condition. The for loop container evaluates an expression and repeats its workflow until the expression evaluates to false.

For loop container uses the following elements to define the loop:

- 1) An optional initialization expression that assigns the value to loop counters
- 2) An evaluation expression that contains the expression used to test whether the loop should stop or continue
- 3) An optional iteration expression that increments or decrements the loop counter

Example: to work in project till error gets solved.

3) Sequence Container:

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

This container simply groups tasks together. The task will execute together. This container is useful to split your control flow into logical units of work.



Here you can see a couple of tasks inside a sequence container. When you would execute the sequence container, all three tasks will execute at the same time.

The Sequence Container defines a control flow that is a subset of the package control flow. Sequence container groups the package into multiple separate control flows, each containing one or more tasks and containers that run within the overall package control flow.

Benefits of Sequence Container:

- 1) Disabling the group of tasks to focus package debugging on one subset of the package control flow
- 2) Managing the properties on multiple tasks in one location by setting properties on a sequence container instead of an individual tasks
- 3) Providing scope of variables that a group of related tasks and containers use

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

4) Grouping many tasks

4) Task Host Container:

It provides services to a single task. The task host container is the default container under which single tasks fall and is used only behind the scenes for SSIS. This type of container is not available in the toolbox in visual studio and is implicitly (automatically) assigned to each task

Purpose of using Container:

- 1) It is used to repeat tasks for each element in a collection
- 2) It repeats the tasks until a specified expression evaluates to false
- 3) Group tasks and containers can be executed
- 4) Container includes control flows with different constraints. It include executable, event handlers, variables etc.

Properties of Container

Property	Description
Delay Validation	Indicates about the validation of container is delayed until runtime or not
Disable	Disables the container
Disable Event Handler	Boolean value that indicate the container runs or the event handler runs with container or not
ID	Container GUID that is assign when the package is created
Name	Name of the container

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

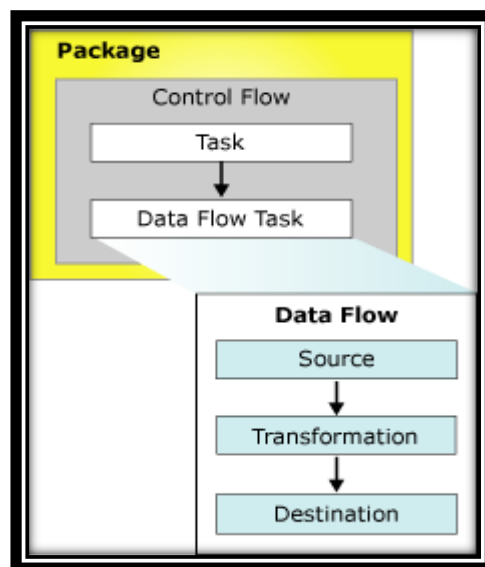
3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

Sr No.	MCQ	Answer
1are the objects in SSIS	Container
2	for each container runs a control flow repeatedly by using	Enumerator
3container runs control flow by testing a condition	for loop
4	Which container is used to provide service to a single task?	Task Host Container
5	Which is the default container?	Task Host Container
6	Which container is not available in toolbox of Visual studio?	Task Host Container

Topic: What is Package? Explain the steps to create package

Ans:

- Package is an organized collection of connections, control flow elements, data flow elements, event handlers, variables, parameters and configurations.
- The package is collection of control flow and data flow.
- The package is the unit of work that is retrieved, executed and saved.



SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

Steps to create package:

- 1) Create new integration service project
- 2) Start the control flow with control flow tab
- 3) Click on data flow task to create data flow
- 4) In the data flow interface, you have to put the source and destination from SSIS toolbox
- 5) Drag the connection line from source to destination
- 6) Double click on source or right click and go in edit option and select your source file
- 7) Double click on destination or right click and go in edit option and select your destination file
- 8) Click on Mapping and then click on OK
- 9) Run your integration project by clicking f5
- 10) Check your output in the destination file.

Topic: What is dynamic package?

Ans:

- Dynamic package means creating the package or executing integration services outside the development environment
- If you need to create package dynamically or to manage and execute integration services packages outside the development environment, you can build package's programmatically
- You can load and execute the existing package or you can create the new packages or you can extend the properties of packages like:
 - Restarting the package by using check point property
 - You can secure a package with digital signature and encryption
 - You can support the transaction that is for managing different task and different elements.
- SSIS factory is a .net program which read the template package, modify the parts and create the dynamic package for the target table which is defined in the Meta data.

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

Sr No.	MCQ	Answer
1	What is dynamic Package?	Creating the package or executing Integration service outside the development environment.

EXTRA QUESTION ASKED IN UNI. PAPER

1) What is SSMS?

Ans:

- **SSMS stands for SQL Server Management Studio**
- SQL Server Management Studio (SSMS) provides Import and Export Wizard tasks which you can use to copy data from one data source to another.
- You can choose from a variety of source and destination data source types, select tables to copy or specify your own query to extract data, and save your work as an **SSIS** package.

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

CHAPTER-4:

Enforcing Data Quality, Extending SQL Server Integration Service

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

Topic: What is data Quality?

Ans:

- Data Quality is the integral part of data governance (Process that ensure high quality of data) that ensures your organization data is fit for the purpose
- **It is an assessment of data's fitness to serve its purpose.**
- Data quality is usually defined as a metric on how well a product meets user requirements.
- Data quality does not mean the data has to be completely or almost error-free—it depends on the users' requirements.
- **Following are the data quality dimensions or measurements.**

1) Completeness:

Is all the data required available and accessible? Are all sources needed available and loaded? Was data lost between stages?

2) Consistency:

Is there erroneous/conflicting/inconsistent data? For example, the termination date of a contract in a “Terminated” state must contain a valid date higher than or equal to the start date of the contract

3) Uniqueness:

Are there any duplicates?

4) Integrity:

Is all data linked correctly? For example, are there orders linking to non-existent customer IDs (a classic referential integrity problem)?

5) Accuracy:

The data reflects the real-world objects and/or events it is intended to model. Accuracy is often measured by how the values agree with an information source that is known to be correct.

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

The DWH Quality Management focuses on:

- Delivers end-to-end quality solutions
- Enforces Data Quality and Data Profiling as important processes during implementation of data warehouse
- Keeps a check on the metadata and its storage repository to ensure
- Generates mappings for data correction based on business rules and ethics.

Sr No.	MCQ	Answer
1	DQ stands for.....	Data Quality
2	List out the measurements of data quality.	Completeness Consistency Uniqueness Integrity Accuracy
3	Data Quality does not mean that data has to be completely or almost error free-it depends upon user requirement (T/F)	True

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)

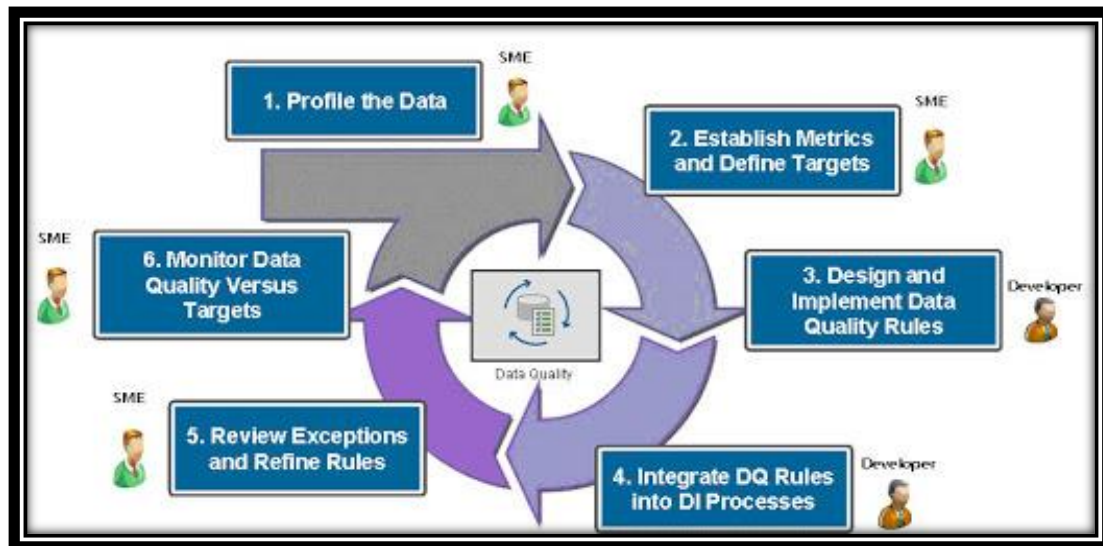


2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

Topic: Write the steps for the cycle of data quality management.

Ans:



1) Data profiling: (Data Analysis):

It is a key element in scoping overall data quality initiatives—it enables you to determine the content, structure, and quality of highly complex data structures and to discover hidden inconsistencies and incompatibilities between data sources and target applications.

2) Establishing metrics and defining targets:

Metrics are the measurements that help estimate the progress, quality of data.

3) Design and implementation of data quality rules:

It help define and measure targets and criteria for data quality.

4) Integrating:

Integrating data quality rules and activities (profiling, cleansing/matching, automated remediation, and management) with data integration processes.

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

5) Reviewing exceptions and refining rules:

It is best accomplished as a joint effort involving core team members and BI stakeholders. In many cases, BI stakeholders have limited control over the business processes and operational systems that cause poor data quality, which is why it is important to involve key stakeholders and executives throughout an organization in documenting data quality issues and launching a formal data quality program.

6) Data quality monitoring :

It is a continuous measurement. Data quality must be tracked, managed and monitored if it is to improve business efficiency and transparency.

Sr No.	MCQ	Answer
1	Data Profiling means?	Data Analysis
2	What is data quality monitoring?	It is a continuous measurement. Data Quality must be tracked, and monitored continuously

Topic: How Data Quality Services is used to for cleansing the data?

Ans:

- ❑ Data Cleansing is the process of analysing the quality of data in a data source, manually approving or rejecting the suggestions by the system, and thereby making changes to the data.
- ❑ Data Cleansing includes computer-assisted process and an interactive process.
- ❑ Data cleansing can also be done in Integration Service Packaging Process. In this case, DQS Cleansing component in integration service is used that automatically performs data cleansing by using existing knowledge base.
- ❑ Data Cleansing feature in DQS has following features:

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

- Identifies incomplete or incorrect data in your data source (Excel file or Sql Server Database) and then corrects or alerts you about the invalid data
- Provides twostep process to cleanse the data: computer assisted and interactive.
- Standardizes and enriches customer data by using domain values, domain rules (condition that is used to validate, correct, and standardize **domain values**)

Types of Data Cleansing:

1) Computer-assisted Cleansing:

- DQS data cleansing process applies the knowledge base to the data to be cleansed and proposes the changes to the data.
- To perform data cleansing process, the data steward proceeds as follows:
 - 1) Create a data quality project, select a knowledge base against which you want to analyse and cleanse your source data, and select the cleansing activity.
 - 2) Specify the database table or an excel file that contains the source data to be cleansed. Excel file can be the same one that was used for knowledge base or it can be different database or excel file.
 - 3) Map the data fields to be cleansed to appropriate domains or composite domains in knowledge base. If you map a field to a composite domain, the mapping happens between the field and composite domain, and not with individual domains in the composite domain.
 - 4) Run the computer assisted cleansing process by clicking start on the cleanse page.

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

2) Interactive Cleansing:

Based on the computer-assisted cleansing process, DQS provides the data steward with information that they need to make a decision about changing the data.

DQS categorizes the data under following 5 tabs:

- 1) Suggested: Values for which DQS found suggestions that have a confidence level higher than the auto-suggestion threshold value.
- 2) New : Valid value for which DQS does not have enough information and therefore cannot be mapped to any other tab
- 3) Invalid: Values that were marked as invalid in the domain in the knowledge base. This tab also contains the values that are rejected by the user in any of the other four tabs during the interactive cleansing process
- 4) Corrected: Values that are corrected by DQS during the automated cleansing process.
- 5) Correct: Values that were found correct. For example, the value matched a domain value. This tab will also contain values that were approved by the user during interactive cleansing.

DQS will merge any changes that the data steward has made with the results of computer-assisted data cleansing.

These changes will stay with the project, however they will not be added to the knowledge base.

During data cleansing, the associated knowledge base is read only.

When the data cleaning process has completed, you can choose to export the processed data to new table in SQL Server Database.

The source data on which cleansing is performed is maintained in its original state.

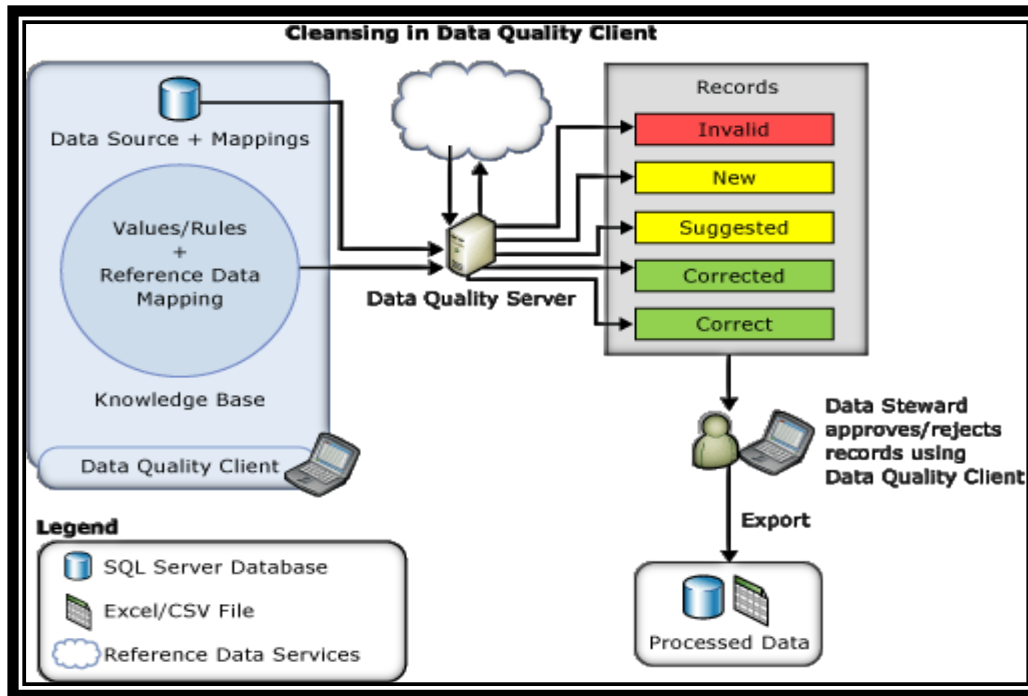
SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645



Sr No.	MCQ	Answer
1	Which are the types of data cleansing?	Computer-assisted Interactive
2	Which type of cleansing process involves user?	Interactive
3	List out the 5 different tabs of interactive cleansing process?	Suggested New Invalid Corrected Correct

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

Topic: How data quality service is used to match the data?

Ans:

- The Data Quality Service (DQS) data matching process enables you to reduce the data duplication and improve data accuracy in a data source
- Matching analyses the degree of duplication in all records of a single data source, returning weighted probabilities of a match between each set of records compared. You can then decide which records are matched and then take appropriate action on the source data.

Advantages of DQS matching process:

- 1) Matching enables you to eliminate the differences between data values that should be equal determining the correct value and reducing the errors that differences can cause
- 2) Matching enables you to ensure that values that are equivalent, but were entered in a different format are rendered uniform
- 3) Matching identifies the exact and appropriate matches, enabling you to remove duplicate data as you define it. You define the point at which an approximate match is in fact a match
- 4) DQS enables you to create matching policy using computer assisted process, modify it interactively based upon matching results and add it to a knowledge base that is reusable.
- 5) You can re-index data copied from the source to the staging table.
- 6) You can perform matching process in conjunction with other data cleansing process to improve overall data quality.

How to perform Data Matching

- 1) Create a matching policy in the knowledge base
- 2) Perform a de-duplication process in a matching activity that is part of data quality project.

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

Building a Matching Policy:

- 1) You prepare the knowledge base for performing matching by creating a matching policy in the knowledge base to define how DQS assigns the matching probability
- 2) A matching policy consists of one or more matching rules that identify which domains will be used when DQS assesses how well one record matches to another.
- 3) You can specify in the rule whether the domain values have to be an exact match or can just be similar and to what degree of similarity
- 4) The matching policy in the knowledge base management wizard analyses sample data by applying each matching rule to compare two records at time throughout the range of records.
- 5) These matching results are not added to knowledge base.

Running a Matching Project:

- 1) Before running a matching process, you can export the results of the cleansing the project into a data table and then create a matching project in which you map the cleansing results to domains in the matching project
- 2) A data matching project consists of a computer-assisted process and an interactive process. The matching project applies the matching rules in the matching policy to the data source to be assessed
- 3) When DQS performs the matching analysis, it creates clusters of records that DQS considers matches. DQS randomly identifies one of the records in each cluster as pivot, or leading record.
- 4) The data steward verifies the matching results, and rejects any record that is not an appropriate match for cluster.

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

Sr No.	MCQ	Answer
1is the process which reduces data duplication and maintains accuracy of data	Data Matching
2	DQS stands for	Data Quality Service
3	Re-indexing is the benefit of Data Matching?	True

Topic: What is Script task in SSIS?

Ans:

- The SSIS script task is one of the task of SSIS
- It is used to add functionality to your SSIS package.
- Script task is used to generate log messages as well as many other logics related to packages
- With the script task, you can program new functionality using C# or VB
- Example: Display hello message using script task

Sr No.	MCQ	Answer
1	Script task is one of task of	SSIS
2	With script task, you can program new functionality using Andlanguage	C#, VB

Topic: Write a short note on SSIS custom components

Ans:

- The SSIS is the Microsoft platform for building enterprise level data integration and data transformation solutions
- It supports variety of sources like XML, Data Files etc.
- SSIS custom components provide a new functionality not provided by standard.

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

Advantages:

- 1) Reusability:
 - a. Component is a DLL
 - b. Single Code base
 - c. Can be used multiple times in a project
 - d. Can be shared across multiple projects
- 2) Performance:
 - a. Faster than scripting
- 3) Well documented:
 - a. Documentation is good

Types of Components:

- 1) Data Connections
- 2) Log Providers
- 3) For each loop
- 4) Control Flow task
- 5) Data flow pipeline component
- 6) Custom user interface

Data Flow Pipeline Component:

- A **data pipeline** is a set of tools and activities for moving data from one system with its method of data storage and processing to another system in which it can be stored and managed differently.
- Moreover, pipelines allow for automatically getting information from many disparate sources, then transforming and consolidating it in one high-performing data storage.

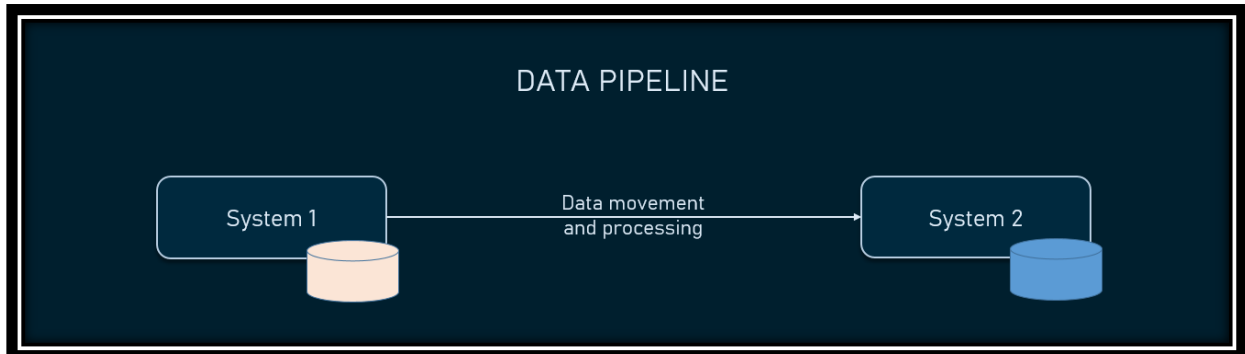
SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)

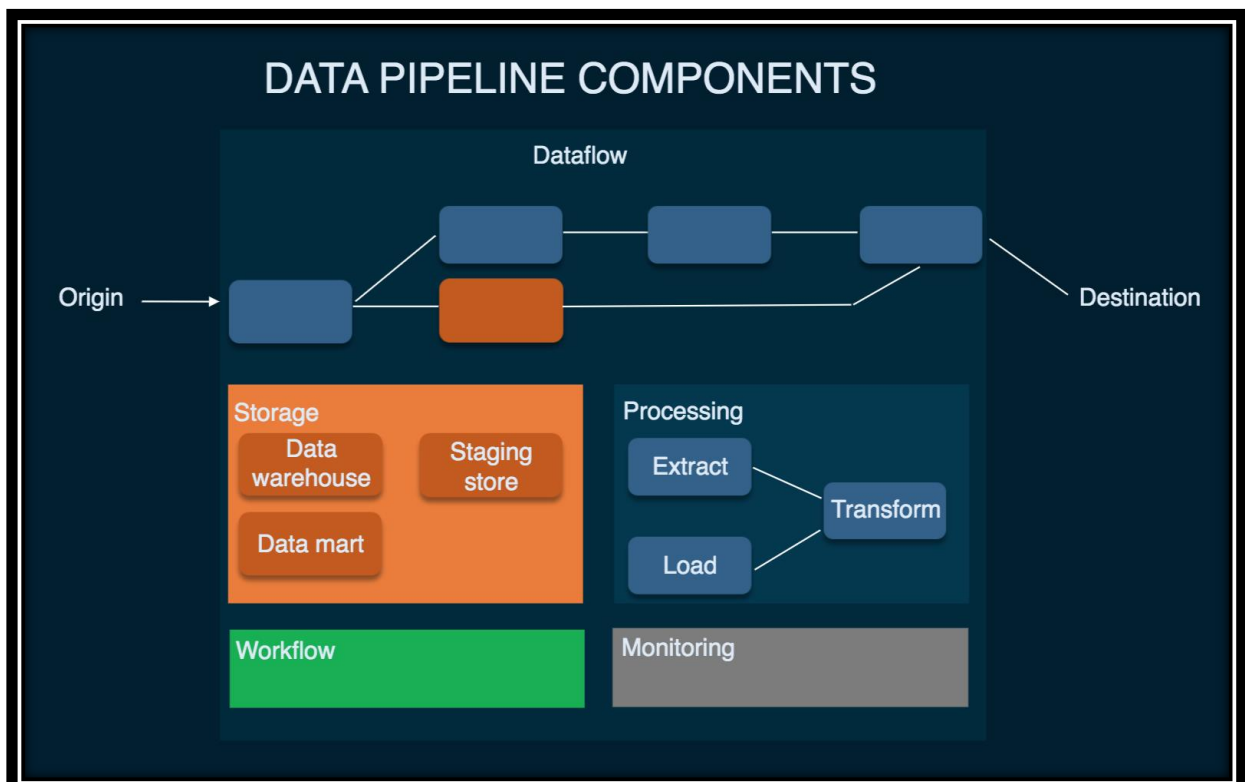


2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645



List out types of data pipeline components?



Origin:

🌱 Origin is the point of data entry in a data pipeline.

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

Destination

- The final point to which data is transferred is called a destination. Destination depends on a use case:

Dataflow

- That's the movement of data from origin to destination, including the changes it undergoes along the way as well as data stores it goes through. One of the approaches to dataflow is called [ETL](#), which stands for extract, transform, and load:

Storage

- Storage refers to systems where data is preserved at different stages as it moves through the pipeline.
- Data storage choices depend on various factors, for example, volume of data and frequency and volume of queries to a storage system, uses of data, etc.
- Example: Online bookstore

Processing

- Processing includes activities and steps for ingesting data from sources, storing it, transforming, and delivering to a destination.
- While data processing is related to dataflow, it focuses on *how* to implement this movement.
- For instance, one can ingest data by extracting it from source systems, copying from one database to another one (database replication), or by streaming data.
- We mention just three options, but there are more of them.

Workflow

- Workflow defines a sequence of processes (tasks) and their dependence on each other in a data pipeline

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

Monitoring

- The goal of monitoring is to check how the data pipeline and its stages are working: whether it remains efficient with growing data load, data remains accurate and consistent as it goes through processing stages, or whether no data is lost along the way.

Sr No.	MCQ	Answer
1is the main advantage of SSIS custom components	Reusability

Topic: Write a short note on Multicast component

Ans:

- Multicast Component is one component of SSIS which is used to create multiple copies of source or input data
- Suppose we want to create multiple copies of input data in .txt, .xls etc.

Example:

- Step 1: Go to File – New Project – Select integration service project
- Step 2: Start Data Flow task
- Step 3: Select Source from SSIS toolbox (Excel data source)
- Step 4: Drag and drop multicast component from SSIS toolbox
- Step 5: Select 3 different destinations like flat file, excel etc.
- Step 6: Map the connection lines from source to multicast component
- Step 7: Map the connection line from multicast component to all the destinations
- Step 8: Double click on source and give the path of excel data
- Step 9: Double click on first destination and select the flat file path
- Step 10: Double click on second destination and select excel file path
- Step 11: Now map all the destinations with multicast component to fetch the data from source
- Step 12: Run the program.

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

CHAPTER-5:

Deploying and Configuring

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

Topic: Write a short note on overview of SSIS Development

Ans:

- SSIS (SQL Server Integration Service) is a component of Microsoft SQL Server Database software that can be used to execute wide range of data migration tasks.
- SSIS is fast and flexible data warehousing tool used for data extraction, loading and transformation like cleaning, aggregating, merging etc.
- It makes it easy to move data from one database to another database. SSIS can extract data from a wide variety of sources like SQL Server databases, Excel files, Oracle and DB2 databases, etc.
- To perform integration services, there are two packages available:
 - 1) SQL Server Data Tools
 - 2) SQL Server Management Studio

1) SSDT:

- SQL Server data tools are used for developing the integration services packages that business solution required.
- You can create packages with SSDT
- With SSDT, you can perform following task:
 - Run the SQL Server import and export wizard to create basic packages that copy the data from source to destination
 - Create the packages that include complex control flow, data flow, event logics and logging.
 - Test and debug the packages by using the troubleshooting and monitoring in SSIS designer and the debugging features.
 - Create configuration that update the properties and the package object at run time
 - Create the deployment utility that can install packages and there dependency on other computer.

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

2) SQL Server Management Studio:

SQL Server Management studio provide the integration services that you can use to manage the packages, monitor the running packages and provide the data for integration service and SQL Server object.

With SSMS, following task can be performed.

- 1) Create the folders to organized packages in a way that is meaningful to your organization
- 2) Run the packages that are stored on the local computer by using the execute package utility.
- 3) Run the execute package utility to generate a command line to use when you run the dtexec command prompt utility (dtexec.exe)
- 4) Import and export packages to the SQL Server database of from SQL Server database.

Sr No.	MCQ	Answer
1	SSDT stands for	SQL Server Data Tools
2	SSMS stands for.....	Sql Server Management Studio

Topic: Write a short note on deploying SSIS Projects

Ans:

- Deployment means hosting the website or application on web server
- SIS Package deployment model in SQL Server is useful for deploying the project so it is known as project deployment model
- This model creates deployment package that includes package, different parameters, environment, variables etc.
- When you create the project in this model then it will create a single file with .ispac extension (which specifies the format of sql Server Integration Service Project file deployment)

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

- Before creating deployment project, we first need to create the integration service catalog. This catalog is like a container for all the deployed project, their settings and historical versions.
- There is only one catalog per sql server instance and it is represented by a separate SQL server database called SSISDB that include all the settings and performance data.

Steps:

- 1) First, create the Integration Service Catalog
- 2) Integration service catalog must be represented by separate SQL Server database called SSISDB
- 3) This catalog contains all the settings, historical versions and performance data
- 4) Create SSISDB catalog with blank folder
- 5) In the blank folder, create two sub folder with name projects and environments
- 6) The project folder will contain the deployed SIS project and the environment folder will contain different sets of environment settings
- 7) Deploy the project by build the solution and run the project.

Sr No.	MCQ	Answer
1	.ispac extension is for	Integration Service Project file Deployment
2	What is Catalog?	It is the container for all the deployed projects, their settings and historical versions

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

Topic: Write a short note on BI

Ans:

- BI stands for Business intelligence
- It is a broad category of application programs and technology for gathering, storing, analysing and providing the data from different data sources and from different users.
- SQL Server Business intelligence is a series of tools for transferring raw data into information that can be used for decision making process in any business
- BI tools allows developer to perform ETL on data from data warehouse.
- BI is also helpful for decision support, data mining, online analytical processing, querying and knowledge management



- 1) Measurement: Creating KPI (Key Performance Indicator) based on historic data

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

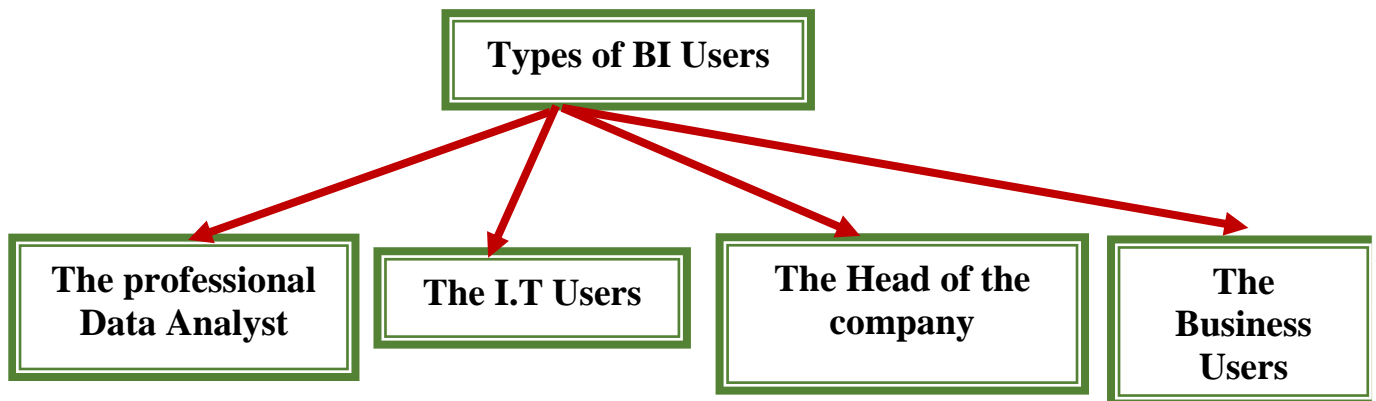
(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

- 2) Identify and set benchmarks (standard or point of reference against which things may be compared) for varied processes
- 3) With BI Systems, organizations can identify market trends and spot business problems that need to be addressed
- 4) BI helps on data visualization that enhances that data quality and thereby the quality of decision making



1) The Professional Data Analyst:

The data analyst is the one who always needs to drill deep down into data. BI System helps them to get fresh insights to develop unique business strategies.

2) The I.T Users:

They are responsible for the infrastructure and available tools in the company and they manage the rights and roles of employees. But they will also try to close the gap between IT and business operations to accelerate the adoption of BI.

When selecting, implementing, managing and using Business Intelligence in your equipment dealership, you should take all these types of users into account. On the one hand, the tool must be accessible and user-friendly, on the other hand it must enable in-depth and extensive analyses

3) The head of the company:

CEO or head increase the profit of their business by improving operational efficiency in their business

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

4) The Business Users:

- * Casual BI Users: He will use dashboards to evaluate predefined sets of data.
- * Power User: He has the capability to work with complex data sets

Advantages:

- Business folks can jump in and start analysing data themselves, rather than wait for IT specialist to run custom and complex reports.
- Takes less time to develop a dashboard of information than what a report will take when it has to be programmed by a developer.
- Make key performance indicators easy to understand.
- Information can be easily isolated on small-related reports rather than getting all data at once.
- Help to increase productivity by assisting with relevant information.
- Can increase profit by addressing inefficient activities revealed by a dashboard of information.

Disadvantages:

- User resistance is one big barrier (Involvement of Users) to Business Intelligence and Information as a Service.
- Sometimes is necessary to analyse voluminous amounts of irrelevant and poor quality data.
- The clean-up of irrelevant data is often time consuming.
 - Most companies don't understand their business processes well enough to determine how to improve them. If the process does not have a direct impact, the entire BI effort could disintegrate.

Sr No.	MCQ	Answer
1	BI stands for	Business Intelligence
2	KPI stands for?	Key Performance Indicator
3	How many types of BI users are there? List out.	4 types

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

		<ol style="list-style-type: none">1) The professional Data Analyst2) I.T Users3) Head of Company4) Business Users
--	--	--

Topic: Explain SSDT as BI

Ans:

- ❖ Business Intelligence Development Studio (BIDS) is available in Microsoft Visual Studio 2008 with additional project types that are specific to SQL server business intelligence.
- ❖ SSDT-BI replaces BIDS for visual studio 2010, 2012,2013, 2015
- ❖ It serve the same purpose as BIDS (Business Intelligence Development Studio), providing a development environment for developing business solutions.
- ❖ Both BIDS and SSDT-BI are client based applications that include additional project type's specific to SQL Server Business Intelligence.

Sr No.	MCQ	Answer
1	BIDS stands for.....	Business Intelligence Development Studio
2	SSDT-BI stands for...	Sql Server Data tools – Business Intelligence

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

Topic: Explain Reporting in SSDT

Ans:

- In SSDT, reporting service is used to create the report.
- SSDT features the Report Designer authoring environment, where you can open, modify, preview, save and deploy reporting concept.
- It also include report definition, shared data sources, shared data sets and report parts.
- It provides a graphical interface where you can set data source, data set, query, layouts, design, different fields, positions etc.
- Paginated reports are the files those are stored in report project. It can be stand as a report or sub report.

Steps to create the report/steps to add report to the project:

- 1) File – New – Project
- 2) Go to Business Intelligence Tab—Reporting services—Report Server Project
- 3) Click Ok to create the project.

Creating new report definition file:

- 1) Go to view—Solution Explorer—right click on reports—add –new item
- 2) Click report
- 3) Report designer opens and displays the new file (.rdl) file in design view.
- 4) Report designer is a reporting service component that runs in SQL Server data tools. It has 2 views: Design and Preview
- 5) You can define your data in Report Data Pane. You define your layout in Design View and run the report and see what it looks in Preview View.
- 6) Configure the data source for the report
- 7) Create a dataset from the datasource
- 8) Select the type of the report which you want. It may be either tabular or matrix format.
 - a. A tabular format have a fixed number of column, while a matrix report have variable number of columns

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

9) Design and format the report layout.

Benefits of Report Projects:

- 1) Report project act as a container for report definition and resources
- 2) Project can be used for below purpose:
 - a. Organized report and related items in one container
 - b. Test report solution that include reports and related item locally
 - c. Deploy related items together. Use project properties and configuration management to deploy to multiple environments
 - d. Preserve a set of master copies for reports and related items. After deployment, reports can be modified.

Sr No.	MCQ	Answer
1	What is Paginated Report?	They are the files stored in the report project. It can be entire report or subreport

Topic: Explain Data Analysis

Ans:

It is the process of inspecting, cleansing, transforming and modeling the data with the goal of discovering useful information, conclusions and supporting decision making process.

The purpose of data analysis is to extract useful information from the data and taking the decision based upon the data analysis.

Data analysis can be performed by different techniques and is very useful in business, different organizations

Data Analysis Process:

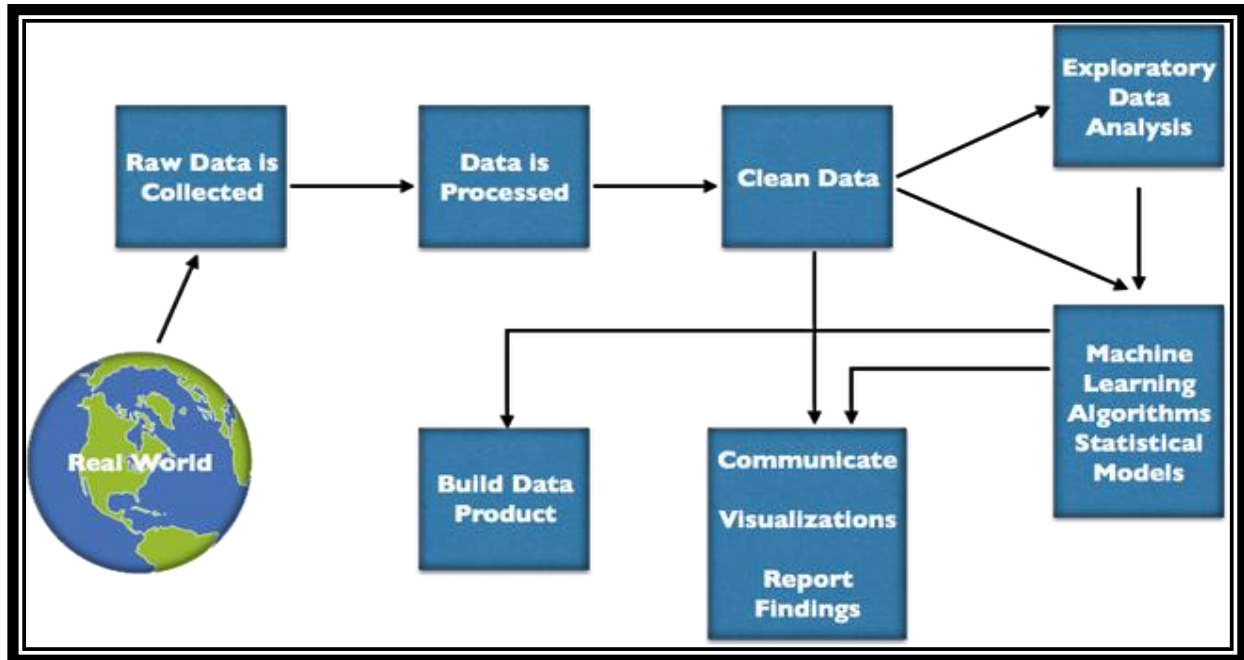
SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645



Types of Data Analysis:

There are several **types of Data Analysis** techniques that exist based on business and technology.

- 1) Text Analysis
- 2) Statistical Analysis
- 3) Diagnostic Analysis
- 4) Predictive Analysis
- 5) Prescriptive Analysis

1) Text Analysis:

Text Analysis is also referred to as Data Mining. It is one of the methods of data analysis to discover a pattern in large data sets using databases or [data mining tools](#). It used to transform raw data into business information.

2) Statistical Analysis:

Statistical Analysis shows "What happen?" by using past data in the form of dashboards.

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

Statistical Analysis includes collection, Analysis, interpretation, presentation, and modelling of data. It analyses a set of data or a sample of data.

There are two categories of this type of Analysis - Descriptive Analysis and Inferential Analysis.

Descriptive Analysis: analyses complete data or a sample of summarized numerical data.

Inferential Analysis: analyses sample from complete data.

3) Diagnostic Analysis:

Diagnostic Analysis shows "Why did it happen?" by finding the cause from the insight found in Statistical Analysis.

This Analysis is useful to identify behaviour patterns of data.

If a new problem arrives in your business process, then you can look into this Analysis to find similar patterns of that problem.

4) Predictive Analysis:

Predictive Analysis shows "what is likely to happen" by using previous data. The simplest data analysis example is like if last year I bought two dresses based on my savings and if this year my salary is increasing double then I can buy four dresses. But of course it's not easy like this because you have to think about other circumstances like chances of prices of clothes is increased this year or maybe instead of dresses you want to buy a new bike, or you need to buy a house!

So here, this Analysis makes predictions about future outcomes based on current or past data.

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

5) Prescriptive Analysis:

Prescriptive Analysis combines the insight from all previous Analysis to determine which action to take in a current problem or decision.

Most data-driven companies are utilizing Prescriptive Analysis because predictive and descriptive Analysis are not enough to improve data performance.

Sr No.	MCQ	Answer
1is the process of inspecting, cleansing, transforming and modelling the data	Data Analysis
2	Descriptive and inferential analysis are the types ofanalysis	Statistical

Topic: Write a short note on Analysis Service Project

Ans:

- Analysis service projects provide project templates and design surfaces for building professional data models hosted in SQL Server Analysis service.
- Following are the steps to create analysis service project in SSDT
 - Go to File – New Project-- BI—Analysis service project
 - Give the name of project. The name of the project given will be used as your database name
 - In the location dropdown list, type or select the folder in which you want to store the files for the project
 - To add new project to the existing solution, in the solution dropdown list, select Add to the solution

Contents of Analysis Service Project Folder:

- 1) Data Sources:
Contains data sources for an Analysis Services project.
- 2) Data Sources View:

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

Data source view means create logical view of only tables involved in your data source

3) Cubes:

A cube is a basic unit of storage. It is a collection of data which has been aggregated to allow queries to return data quickly.

4) Dimensions:

Dimension offers the context surrounding a business process event. In simple terms, they give who, what, where of a fact. In the Sales business process, for the fact sales number, dimensions would be who customer Names.

Where – Location, What – Product Name

In other words, you can say that a dimension is a window to view the information in the facts.

5) Mining Structures:

A **mining structure** is just a metadata container until it is processed. When you process a **mining structure**, **Analysis Services** creates a cache that stores statistics about the data, information about how any continuous attributes are discretized, and other information that is later used by **mining** models

6) Roles:

Roles are used in **Analysis Services** to manage security for **Analysis Services** objects and data. In basic terms, a **role** associates the security identifiers (SIDs) of Microsoft Windows users and groups that have specific access rights and permissions defined for objects managed by an instance of **Analysis Services**.

Sr No.	MCQ	Answer
1	What is Cube in content of analysis service project folder	It is a basic unit of storage.
2	What is Mining Structure in content of analysis service project folder?	Metadata container until it is processed

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

Topic: Write a short note on Deployment process for ssis package

Ans:

- Integration services support two deployment models.
 - The Project Deployment Model
 - The legacy package deployment model
- SQL Server integration services includes tolls and wizards that make to deploy packages process simple.
- Steps for the deployment process package:
 - 1) Configuring the package and create the package using update property. The configuration automatically include when you deploy the packages.
 - 2) Build the service project to create a package deployment utility.
 - 3) Copy the deployment folder that was created when you create the project to target computer
 - 4) Run on the target computer the package installation wizard to install the packages.

Topic: Write a short note on Configuration of SSIS package

Ans:

- Configuration of package is required to update the values of property at runtime
- It is a layer to move the package's from development environment to production environment using configuration
- To deploy the package's to many server's configuration is useful
- You can use dtexec.exe command prompt utility to run the deployed package
- There are different package configurations types like:
 - XML Configuration

SHREE H. N. SHUKLA COLLEGE OF I.T. & MGMT.

(AFFILIATED TO SAURASHTRA UNIVERSITY)



2 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot – 360001
Ph.No–(0281)2440478,2472590

3 – Vaishalinagar
Nr. Amrapali Under Bridge
Raiya Road
Rajkot - 360001
Ph.No–(0281)2471645

- Environment Variable
- Register Entry
- Package variable
- SQL Server tables

■ You can use stored procedure to deploy the package and project and you can use package deployment model that allows you to convert the package-to-package deployment model.