Unit-1

1)OLTP=Online Transaction Processing

2)OLAP=Online analytical processing

3)DWH=data warehouse

4)EDW=enterprise data warehouse

5)CRM=Customer Relationship Management

6)Explain future trands of datawarehouse.

🡺Data warehousing is very much useful in the future era. For this it use traditional and new improvements because day by day the data will be increased and the market will be grow.

The future trend is the age of the customer and consumer.

So, data warehouse is use with business intelligence for growing and successes the business otherwise you will be fail.

Data warehousing provides you up to date data and analytical information about the customers. So , you can customers using business intelligency(BI).

7)what is normalization data store?

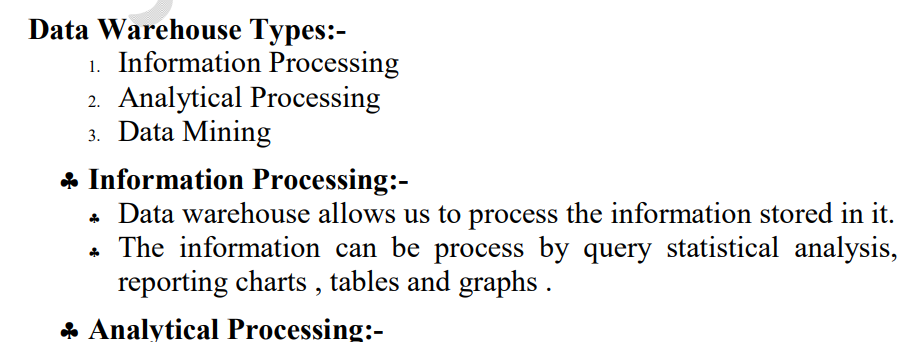
🡺 Other types of data warehouses put the data not in a dimensional data store b store. A normalized data store is one or more relational databases with little or no data redundancy.

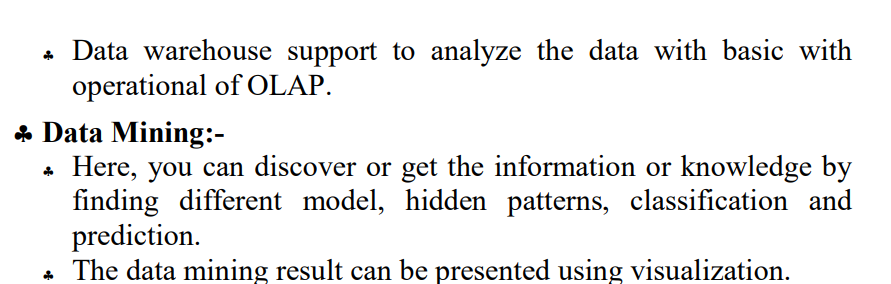
A relational database is a database that consists of entity tables with parent between them. Normalization is a process of removing data redundancy by implementing normalization rules.

There are five degrees of normal forms, from the first normal form to the fifth normal form.

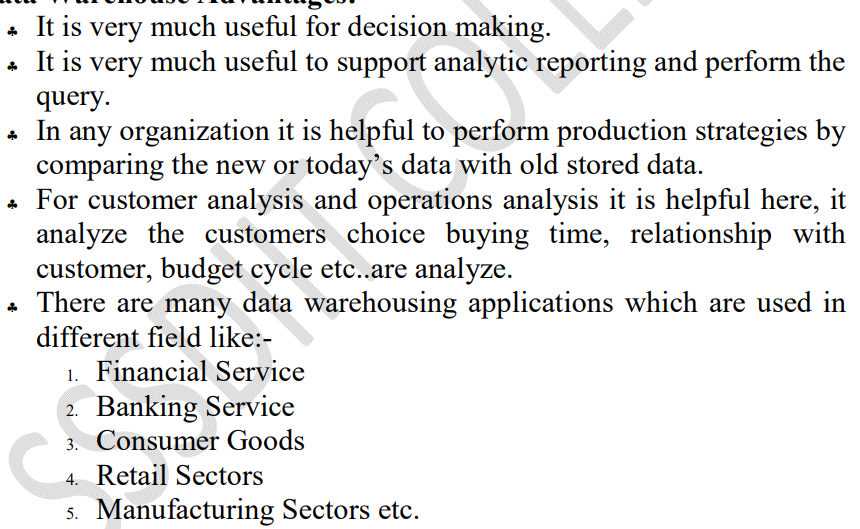
A normalized datastore is usually in third normal form or higher, such as fourth or fifth normal form.

8)Explain data Warehouse types.

🡪

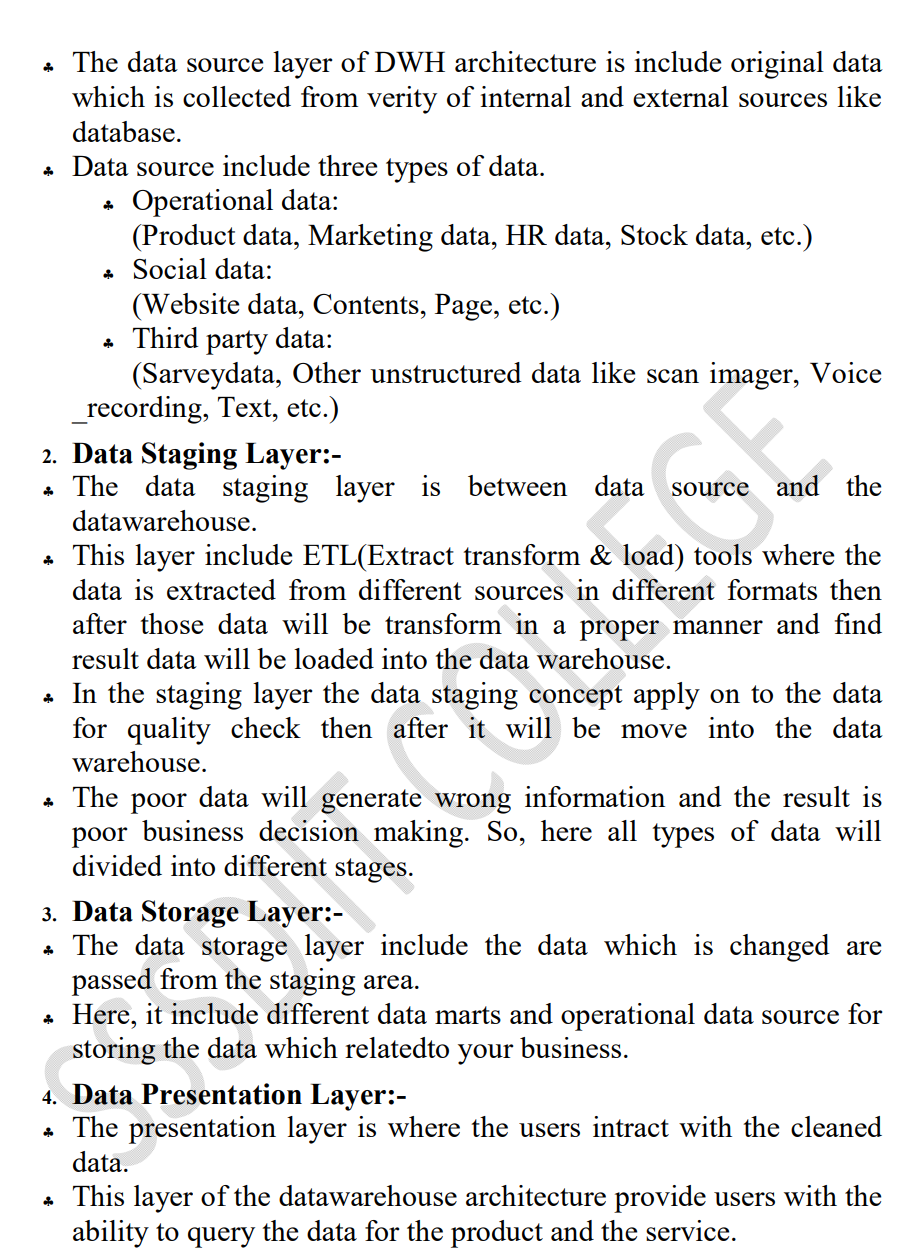


9)Explain Advantages of Datawarehouse.

🡪

10)Explain datawarehouse architecture.

🡪



11)what is datawarehouse ? explain its characteristics.

🡪A **data warehouse** is a type of [data management](https://www.oracle.com/in/database/what-is-data-management/) system that is designed to enable and support business intelligence (BI) activities, especially analytics. Data warehouses are solely intended to perform queries and analysis and often contain large amounts of historical data. The data within a data warehouse is usually derived from a wide range of sources such as application log files and transaction applications.

**Characteristics of data wre house**

1. Subject Oriented: Focuses on a specific area or subject such as sales, customers, or inventory.
2. Integrated: Integrates data from multiple sources into a single, consistent format.
3. Read-Optimized: Designed for fast querying and analysis, with indexing and aggregations to support reporting.
4. Summary Data: Data is summarized and aggregated for faster querying and analysis.
5. Historical Data: Stores large amounts of historical data, making it possible to analyze trends and patterns over time.
6. Schema-on-Write: Data is transformed and structured according to a predefined schema before it is loaded into the data warehouse.
7. Query-Driven: Supports ad-hoc querying and reporting by business users, without the need for technical support.

12)write advantages and disadvantages of datawarehouse.

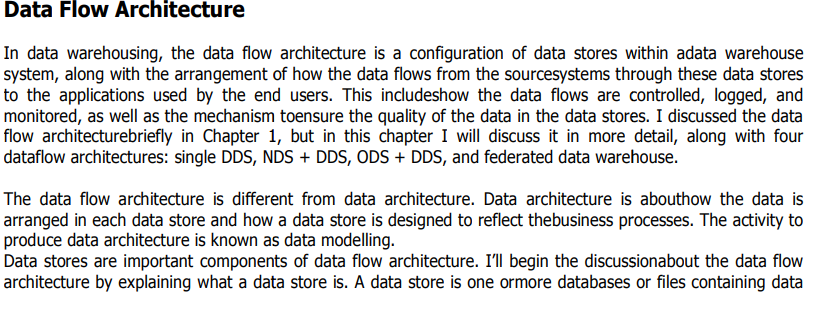
🡪**advantages:**

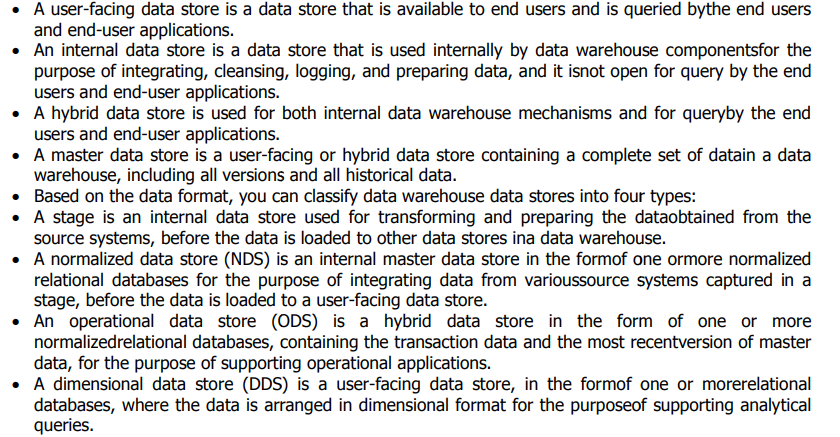
* Data warehouses facilitate end users' access to a variety of data.
* Using numerous data warehouses can increase the operational value of business systems, especially customer relationship management.
* Makes selections with higher quality.
* For the medium and long term, it is especially helpful.
* Storage of analyses and historical search queries is quite beneficial.
* It has a strong capacity for digesting information.
* Access to information is made more flexible and quick because of it.
* Allows for easier corporate decision-making.
* The productivity of businesses rises.
* Gives the company's many departments reliable communication.
* Strengthen connections with customers and suppliers.
* It makes it possible to keep up with business activity and be constantly informed of successful and unsuccessful outcomes.
* transforms information into knowledge and data into information
* You can plan more successfully, thanks to it.
* Cut back on operating expenses and response times.
* The Data Warehouse assists in fusing many data sources, lessening the production system's workload.
* The data warehouse aids in reducing the overall turnaround time for reporting and research.

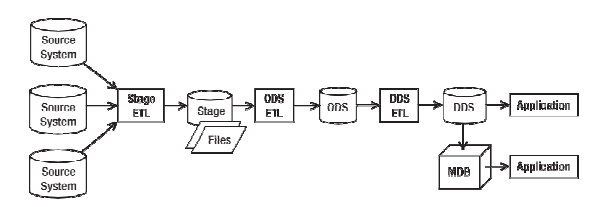
**Disadvantages:**

* The data warehouses may project substantial expenditures throughout his life. The data warehouse is typically not stationary. Costs for maintenance are considerable.
* Data warehouses could soon become outdated.
* They occasionally need to provide complete information before a request for information, which also costs the organization money.
* Regarding the various objectives a company seeks to achieve, challenges may arise during implementation.
* It might be challenging to include new data sources once a system has been implemented.
* They necessitate an examination of the data model, objects, transactions, and storage.
* They were designed in a sophisticated, multidisciplinary manner.
* The operating systems must be reorganized to accommodate them.
* The data warehouse may seem simple, but it is too complex for the typical person to comprehend.
* The scope of the data storage project will start to expand, despite the best efforts of project management.
* At this point, various business regulations may already be in place for warehouse clients.

13)Explain dataflow architecture of data warehouse.

🡪





14)Discuss CRM in details.

🡪Customer Relationship Management (CRM) is the activity of establishing contact and managing communications with customers, analyzing information about customers, campaigning to attract new customers, performing business transactions with customers, servicing customers, and providing support to customers.

There are 3 main types of CRM software: operational CRM systems, collaborative CRM systems, and analytical CRM systems.

Customer Relationship Management (CRM) is the activity of establishing contact and managing communications with customers, analyzing information about customers, campaigning to attract new customers, performing business transactions with customers, servicing customers, and providing support to customers.

**CRM Advantages**

CRM is an integral part of any business’s success. When performing well, it results in higher customer satisfaction, improved customer retention rates, and enhanced customer loyalty.

 It allows the business to nurture relationships with customers, giving them a better experience, and shows that the business is really concerned about their needs.

 CRM also helps in improving the overall sales performance by assisting decision-making processes at all levels of the organization.

 It can also result in better [understanding](https://crm.walkme.com/7-small-business-best-practices-in-microsoft-dynamics-crm-infographic/) customers’ needs, wants, and expectations. This enables you to tailor your products and services according to their demands.

 The importance of CRM has been recognized by most organizations. Implementation of CRM is on the rise and many companies are looking at new ways to utilize this technology.

**CRM Disadvantages**

Customer relationship management (CRM) also has certain disadvantages.

 Although CRM is a great tool for improving customer relationships, it is not without its faults and limitations. There are some aspects of the system that you should be aware of before implementing it in your organization.

 For one thing, they can be quite expensive and [time-consuming](https://www.activecampaign.com/blog/hours-needed-for-crm) to implement. It can take weeks, months, or even several months just to come up with a system that is workable for your organization.

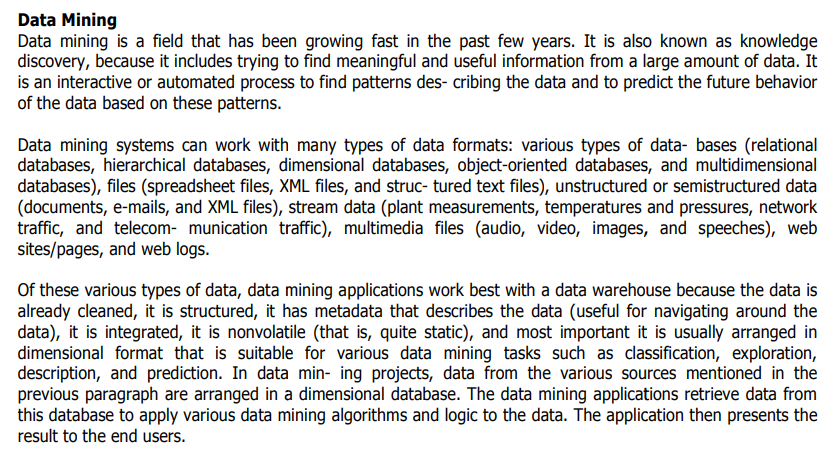
 There are also limitations on data, as CRM requires all your customer data to be in one place.

This means that you will have to go through all your information and decide which information you wish to be stored in the CRM system.

 Furthermore, there are many different CRM systems on the market today and it can be quite challenging to choose the right one for your business.



15)Explain datamining .

🡪

16)

UNIT-2

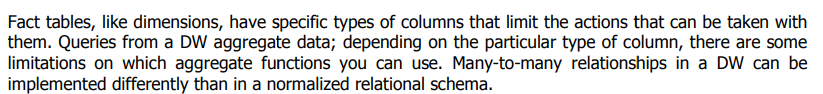
1)List out types of Entity relationship .

🡪

2)what is role playing dimension ?

🡪 A role-playing dimension is a dimension that can be associated with multiple facts in different roles. For example, a date dimension can play the role of order date, ship date, or delivery date for a sales fact.

3)What is fact table ?

🡺

4) SDC=Slowly changing Dimension

5)what is data wrehoue schema?

🡪 Data warehouse schema is a description, represented by objects such as tables and indexes, of how data relates logically within a data warehouse. Star, galaxy, and snowflake schema are types of warehouse schema that describe different logical arrangements of data.

6)Give steps for logica design of datawarehouse.

🡪

7)Give steps for physical design of data warehouse.

🡺

1. dentify business requirements.
2. Create a conceptual model.
3. Develop a logical model.
4. Define the physical model.
5. Extract, transform, and load (ETL) processes.
6. Develop reporting and analysis tools.
7. Implement data quality and data governance processes.

**8)Explain types of dimension.**

## 🡪 ****Major Types of Dimensions in a Data Warehouse****

1. Slowly Changing Dimension
2. Conformed Dimension
3. Degenerate Dimension
4. Junk Dimension
5. Role-playing Dimension
6. Static Dimension
7. Shrunken Dimension

## 1. Slowly Changing Dimension

A Slowly Changing Dimension (SCD) is a dimension that stores and manages both current and historical data over time in a data warehouse. It is considered and implemented as one of the most critical ETL tasks in tracking the history of dimension records.

Slowly changing dimensions are important in data analytics to track how a record is changing over time. The way the database is designed directly reflects whether historical attributes can be tracked or not, determining different metrics available for the business to use.

**2. Conformed Dimension**

In data warehousing, a conformed dimension is a dimension that has the same meaning to every fact with which it relates. Conformed dimensions allow facts and measures to be categorized and described in the same way across multiple facts or data marts, ensuring consistent reporting across the enterprise.

Conformed dimensions are dimensions that are shared by multiple stars. They are used to compare the measures from each star schema [3]. The reuse of conformed dimensions is very common in order to “support true, cross-business process analysis” [6].

**3.Degenerate Dimension**

It is a type of dimension that resides in a fact table of a data warehouse, helpingin tracking sequence of events and categorizing facts.

It is a type of dimension that resides in a fact table of a data warehouse, helping in tracking sequence of events and categorizing facts

**4.Junk Dimension**

junk Dimension is a dimension table in a data warehouse that combines several low cardinality flags and indicators to improve the efficiency of queries.

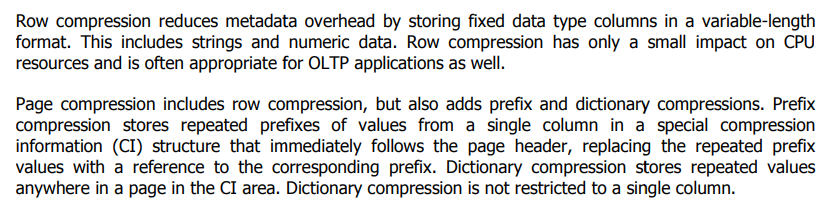
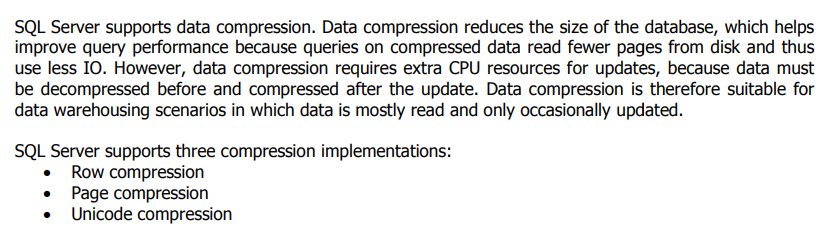
The introduction of the Junk Dimension in a data warehouse setup can bring several benefits to businesses. The most significant advantage being the reduction in query complexity. By consolidating the random, less frequently used fields into a single table, it helps reduce the size and complexity of the fact table.

**5.role playing dimension**

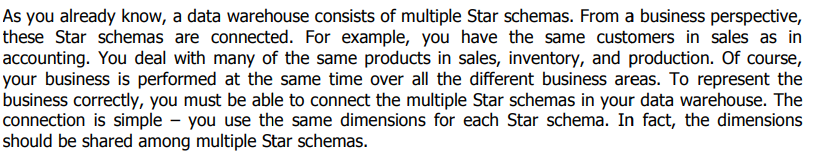
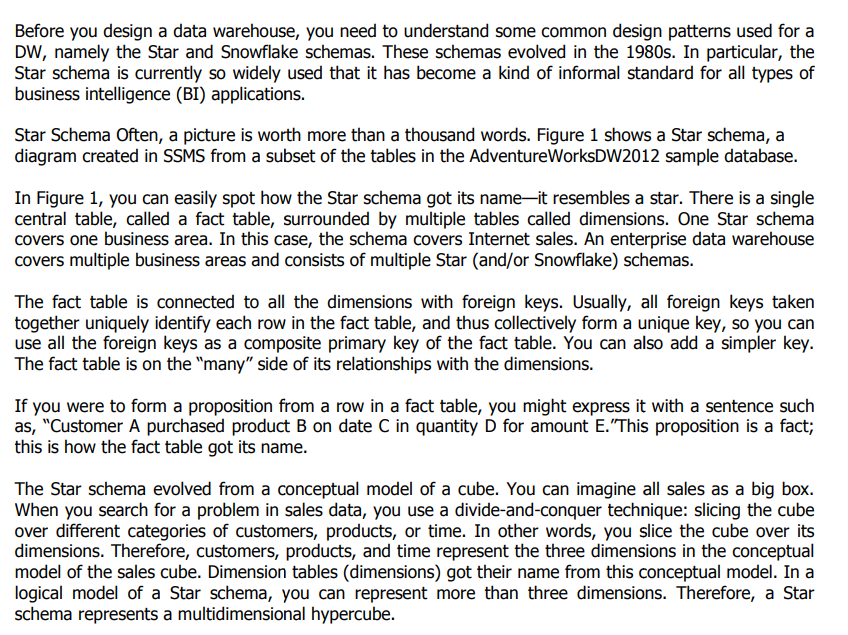
A table with multiple relationships between itself and another table is known as a role-playing dimension. For example, the Sales fact in the following example has multiple relationships to Time on the keys Order Day , Ship Day , and Close Day .

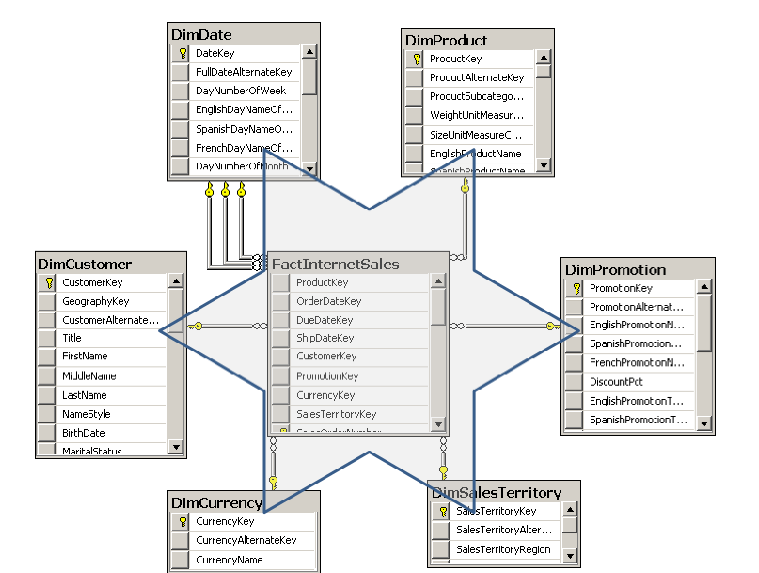
But what is a role-playing dimension and how does it affect the design and usability of a data warehouse? A role-playing dimension is a dimension that can be associated with multiple facts in different roles. For example, a date dimension can play the role of order date, ship date, or delivery date for a sales fact.

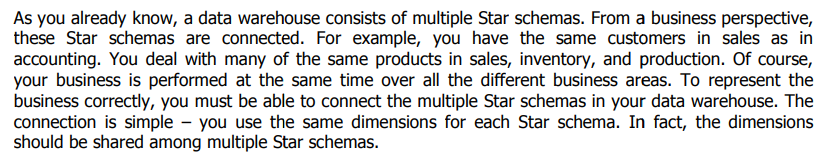
**9)What is data compression?**

**🡪**

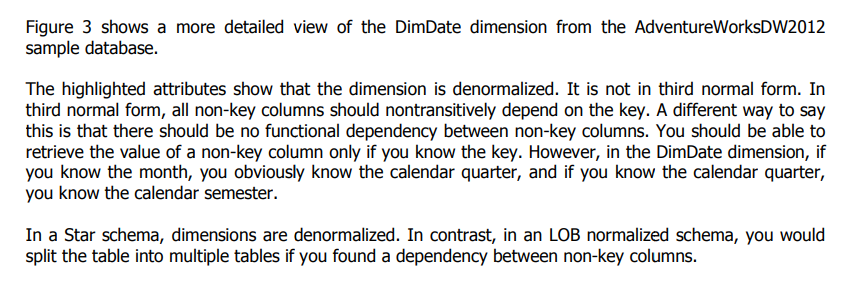
**10)Explain star schema with example.**

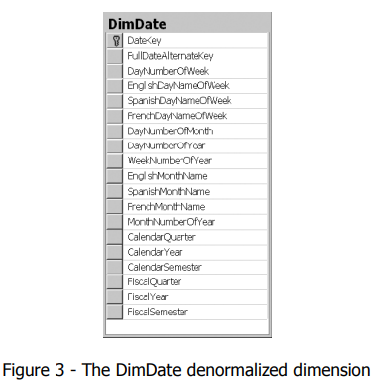
**🡪**

****

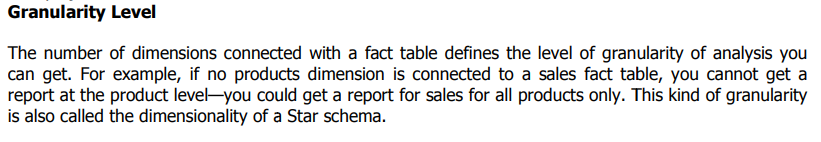
****

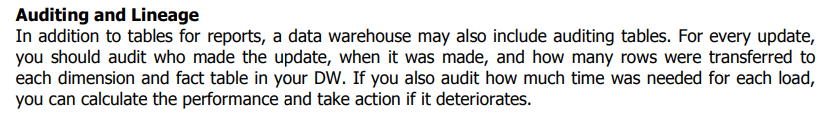
**11)Explain snowflake schema with example.**

****

****

****

****

****

**13)Discuss view in details..**

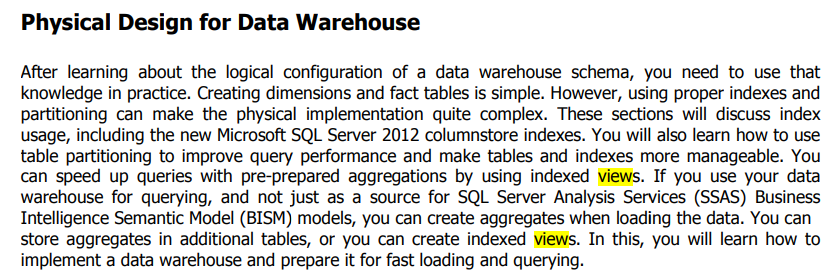
**🡪** the Data Warehouse Views feature is a method of creating new warehoused tables by modifying an existing table, or joining or consolidating multiple tables together by using SQL.

A database view is a subset of a database and is based on a query that runs on one or more database tables. Database views are saved in the database as named queries and can be used to save frequently used, complex queries. There are two types of database views: dynamic views and static views.

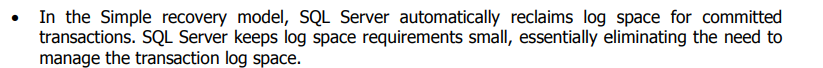
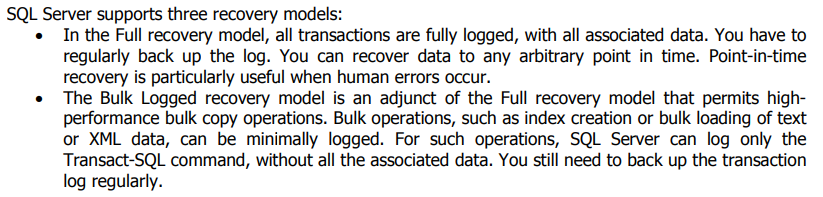
Complex View: A view based on multiple tables, which contain GROUP BY clause and functions. Inline View: A view based on a subquery in FROM Clause, that subquery creates a temporary table and simplifies the complex query. Materialized View: A view that stores the definition as well as data.

Views are used for security purposes because they provide encapsulation of the name of the table. Data is in the virtual table, not stored permanently. Views display only selected data. We can also use Sql Join s in the Select statement in deriving the data for the view.

**14)Discuss phisycal design for datawarehouse.**

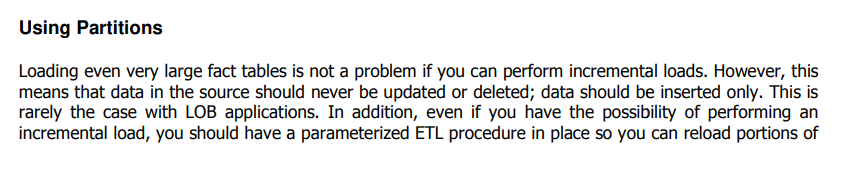
**🡪**

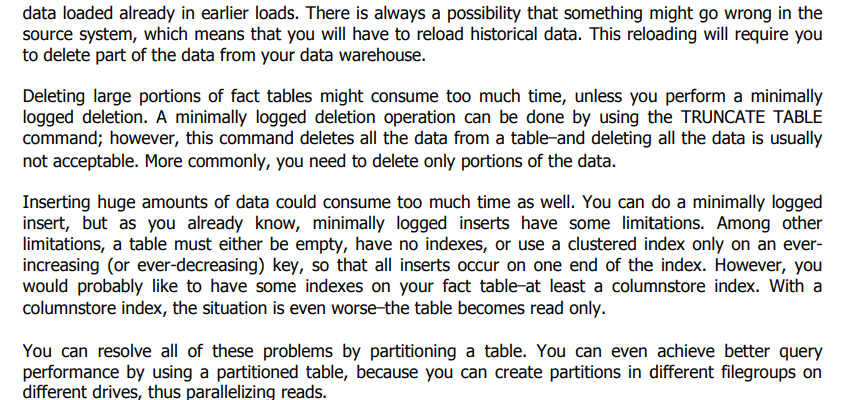
****

****

**15)Define partitioned table . how its use ful.**

**🡪**A partitioned table is divided into segments, called partitions, that make it easier to manage and query your data. By dividing a large table into smaller partitions, you can improve query performance and control costs by reducing the number of bytes read by a query.

****

****

**16)**

**UNIT-3**

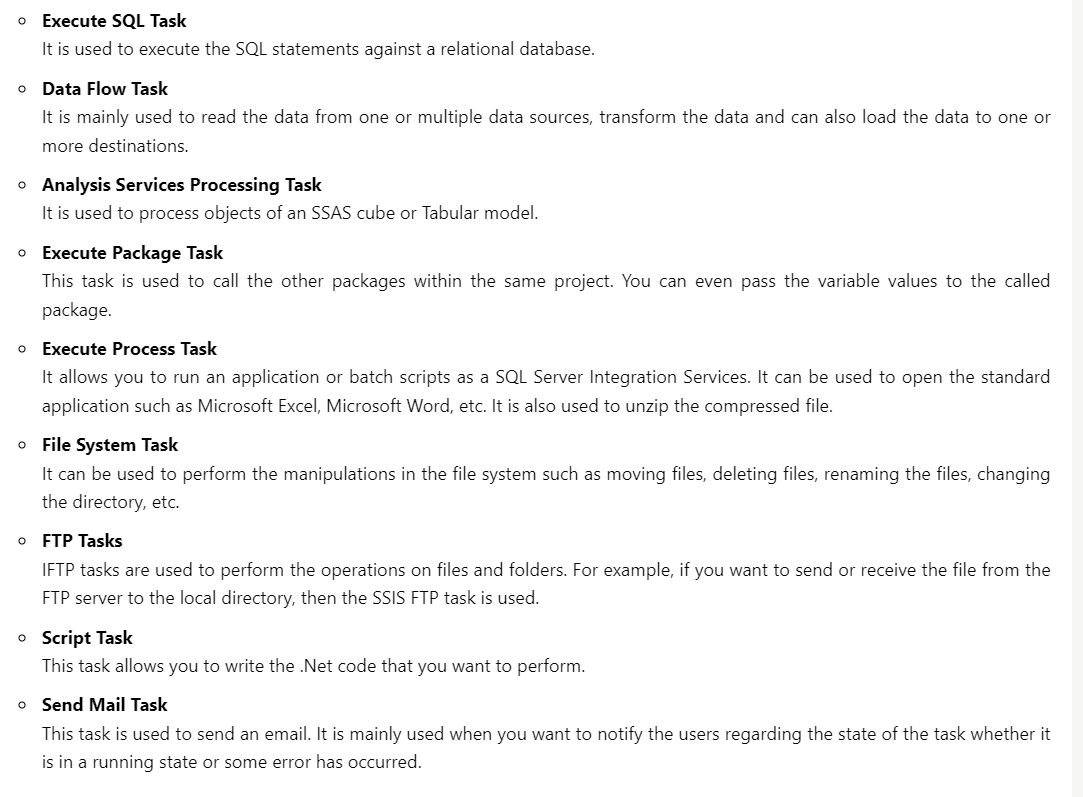
**1)SSIS=** **SQL Server Integration Services**

**2)SSDT=** **SQL Server Data Tools**

**3)ETL=** extract, transform, and load

**4)SMO=** **SQL Server Management Objects**

**5)Explain SSIS task.**

**🡪**

**6)Explain Foreach Container.**

**🡪** Foreach Loop container: You map the enumerated value of the container to a user-defined package variable. The container then uses this variable to dynamically modify the ConnectionString property of the Flat File connection manager and iteratively connect to each flat file in the folder

You can use an SSIS Foreach Loop container to define a control flow task to loop through different types of enumerators, such as files, in a specified folder.

**Add a Foreach Loop container**

1. In **SQL Server Data Tools**, select the **Control Flow** tab.
2. In the **SSIS Toolbox**, expand **Containers**, and then drag a **Foreach Loop Container** onto the design surface of the **Control Flow** tab.
3. Right-click the new **Foreach Loop Container** and select **Edit**.
4. In the **Foreach Loop Editor** dialog, on the **General** page, for **Name**, enter **Foreach File in Folder**. Select **OK**.
5. Right-click the Foreach Loop container, select **Properties**, and in the **Properties** window verify that the **LocaleID** property is set to **English (United States)**.

**Configure the enumerator for the Foreach Loop container**

1Double-click Foreach File in Folder to reopen the Foreach Loop Editor.

2Select Collection.

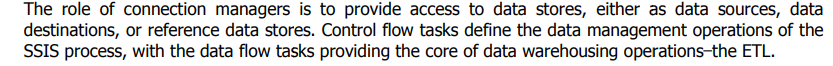
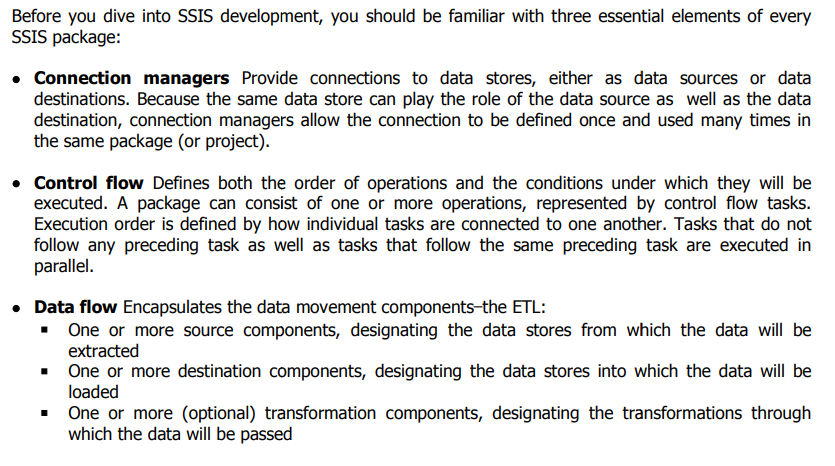
3On the Collection page, select Foreach File Enumerator.

4In the Enumerator configuration group, select Browse.

5In the Browse for Folder dialog box, locate the folder on your machine that contains the Currency\_\*.txt files included with the sample data.

6In the Files box, enter Currency\_\*.txt.

**7)Give the list of creating of flow task.**

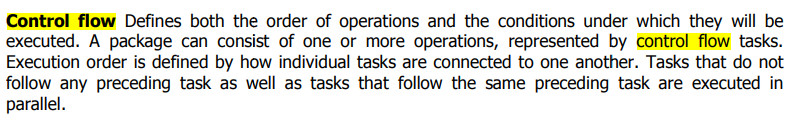
**🡪**

**8)give basic steps for creating package.**

1. **🡪** Defining Business Requirements. ...
2. Setting Up Physical Environments. ...
3. Introducing Data Modelling. ...
4. Choosing Your ETL Solution. ...
5. Online Analytics Processing (OLAP) Cube. ...
6. Creating A Front End. ...
7. Optimising Queries. ...
8. Rolling Out The End Product.

**9)Explain control Flow.**

**🡪**

****

**10)Write a note on ETL Process.**

**🡪**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.

**Extract:** In this phase, the original data is checked. It checks the data, whether it consists of any errors or not. It checks for the errors or inconsistency of data by using some artificial intelligence techniques. In short, it verifies whether the quality of the product is met or not.

**Transform:** It is the third phase in ETL. Transformation is the process in which the original format is converted into a required format that you want. Transformation is modelling or changing the data according to the user requirements. The changes can be either change in the number of columns or rows.

**Load :** The fourth phase is Load and index. It loads the data and validates the number of rows that have been processed. Once the loading of data is completed, the indexing is used. Indexing helps you to track the number of rows that are loaded in the data warehouse. Indexing also helps to identify the data, whether it is in the correct format or not.

**11)Explain Sequence container.**

**🡪** In computing, sequence containers refer to a group of container class templates in the standard library of the C++ programming language that implement storage of data elements. Being templates, they can be used to store arbitrary elements, such as integers or custom classes.

there are many benefits of using a Sequence container:

* Disabling groups of tasks to focus package debugging on one subset of the package control flow.
* Managing properties on multiple tasks in one location by setting properties on a Sequence container instead of on the individual tasks.

For example, you can set the Disable property of the Sequence container to True to disable all the tasks and containers in the Sequence container.

* Providing scope for variables that a group of related tasks and containers use.
* Grouping many tasks so you can more easily managed them by collapsing and expanding the Sequence container.

You can also create task groups, which expand and collapse using the Group box. However, the Group box is a design-time feature that has no properties or run-time behavior. For more information, see [Group or Ungroup Components](https://learn.microsoft.com/en-us/sql/integration-services/group-or-ungroup-components?view=sql-server-ver16)

* Set a transaction attribute on the Sequence container to define a transaction for a subset of the package control flow. In this way, you can manage transactions at a more granular level.

For example, if a Sequence container includes two related tasks, one task that deletes data in a table and another task that inserts data into a table, you can configure a transaction to ensure that the delete action is rolled back if the insert action fails.

**12)What is SSIS ? and Discuss SSIS Package.**

**🡪**SSIS stands for SQL Server Integration Services.

It is a component available in the Microsoft SQL Server database software used to perform a wide range of integration tasks.

It is a data warehousing tool used for data extraction, loading the data into another database, transformations such as cleaning, aggregating, merging data, etc.

SSIS tool also contains the graphical tools and window wizards workflow functions such as sending email messages, ftp operations, data sources.

SSIS is used to perform a wide range of transformation and integration tasks. As a whole, the SSIS tool is used in data migration.

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.

* **Connections**  
  SSIS package will have some connections, and these connections are used to connect to various data sources.
* **Controlflow elements**  
  SSIS package is composed of two elements, i.e., control flow elements and data flow elements. Control flow elements handle workflows. Workflow means that we are performing some tasks in steps, so the sequence is done through control flow.
* **Data flow elements**  
  The data flow elements perform transformations.

**13)write note on flate file.**

**🡪** A flat file consists of a single table of data. It allows the user to specify data attributes, such as columns and data types table by table, and stores those attributes separate from applications. This type of file is commonly used to import data in data warehousing projects.

**flat file example**

A flat file requiring the employee's identification, address, and expertise is one example. For this project, a flat file database would only contain one table. Each entry in the table describes a record. The columns of the table would include the employee's ID, address, and competence.

**Advantages:**

1. **Data Structure:** Relational databases use structured tables with predefined schemas, making them ideal for complex data models and relationships.
2. **Data Integrity:** Relational databases enforce data integrity constraints, ensuring data accuracy and consistency. This includes referential integrity, unique constraints, and data type validation.
3. **Query Capabilities:** Relational databases offer powerful query languages (e.g., SQL) that enable complex data retrieval and analysis, including joins and aggregations.
4. **Scalability:** Modern relational databases support horizontal and vertical scaling, making them suitable for both small and large-scale applications.
5. **Concurrency Control:** Relational databases support concurrent access by multiple users or applications while maintaining data consistency through locking and transaction management.
6. **Security:** Relational databases provide robust security features, including user authentication, authorization, and access control, to protect data from unauthorized access.

**Disadvantages:**

1. **Complexity:** Relational databases can be complex to set up and maintain, especially for users without prior database experience.
2. **Overhead:** Relational databases have higher overhead in terms of storage and processing compared to flat files.
3. **Learning Curve:** Learning SQL and understanding database design concepts can be challenging for beginners.
4. **Cost:** Relational database systems may involve licensing and operational costs, especially for commercial offerings.

**14)write note on control flow in SSIS**

**🡪** Control flow is a significant cornerstone of each SSIS package and defines how the integration works. In general, control flow includes:

* Operations (eg Data flow task, Execute SQL task, Containers)
* Order of operations and relationships between them (Precedence constraints)

**15)**

**UNIT -4**

**1)DQS=Data Quality Services**

**2)DQAF=Data Quality Assessment Framework**

**3)What is data profiling?**

🡪Data profiling is the process of examining, analyzing, and creating useful summaries of data. The process yields a high-level overview which aids in the discovery of [data quality](https://www.talend.com/resources/what-is-data-quality/) issues, risks, and overall trends. Data profiling produces critical insights into data that companies can then leverage to their advantage.

**4)DQS**

**5)DQKB=Data Quality Knowledge Base**

**6)DQM=**Data quality management

**7)Explain Featurs of DQS.**

**🡪**SQL Server Data Quality Services (DQS) is a knowledge-driven data quality product.

DQS enables you to build a knowledge base and use it to perform a variety of critical data quality tasks, including correction, enrichment, standardization, and de-duplication of your data.

DQS enables you to perform data cleansing by using cloud-based reference data services provided by reference data providers.

DQS also provides you with profiling that is integrated into its data-quality tasks, enabling you to analyze the integrity of your data.

DQS consists of Data Quality Server and Data Quality Client, both of which are installed as part of SQL Server.

Data Quality Server is a SQL Server instance feature that consists of three SQL Server catalogs with data-quality functionality and storage.

Data Quality Client is a SQL Server shared feature that business users, information workers, and IT professionals can use to perform computer-assisted data quality analyses and manage their data quality interactively.

You can also perform data quality processes by using the DQS Cleansing component in Integration Services and Master Data Services (MDS) data quality functionality, both of which are based on DQS.

**8)write note on Data Cleansing.**

**🡪** Data cleaning is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset. When combining multiple data sources, there are many opportunities for data to be duplicated or mislabeled. If data is incorrect, outcomes and algorithms are unreliable, even though they may look correct. There is no one absolute way to prescribe the exact steps in the data cleaning process because the processes will vary from dataset to dataset. But it is crucial to establish a template for your data cleaning process so you know you are doing it the right way every time.

Advantages and benefits of data cleaning

Having clean data will ultimately increase overall productivity and allow for the highest quality information in your decision-making. Benefits include:

* Removal of errors when multiple sources of data are at play.
* Fewer errors make for happier clients and less-frustrated employees.
* Ability to map the different functions and what your data is intended to do.
* Monitoring errors and better reporting to see where errors are coming from, making it easier to fix incorrect or corrupt data for future applications.
* Using tools for data cleaning will make for more efficient business practices and quicker decision-making

**How to clean data**

1. Step 1: Remove duplicate or irrelevant observations. Remove unwanted observations from your dataset, including duplicate observations or irrelevant observations. ...
2. Step 2: Fix structural errors. ...
3. Step 3: Filter unwanted outliers. ...
4. Step 4: Handle missing data. ...
5. Step 5: Validate and QA.

**9)Explain importance of data quality.**

**🡪 Data quality is defined as *the degree to which data meets a company’s expectations of accuracy, validity, completeness, and consistency.***

It is a critical aspect of [data management](https://www.alation.com/blog/data-catalog-and-master-data-management/), ensuring that the data used for analysis, reporting, and decision-making is reliable and trustworthy.

By tracking data quality, a business can pinpoint potential issues harming quality, and ensure that shared data is fit to be used for a given purpose.

When collected data fails to meet the company expectations of accuracy, validity, completeness, and consistency, it can have massive negative impacts on customer service, employee productivity, and key strategies.

**important**

Low-quality data can have significant business consequences for an organization. Bad data is often the culprit behind operational snafus, inaccurate analytics and ill-conceived business strategies. For example, it can potentially cause any of the following problems:

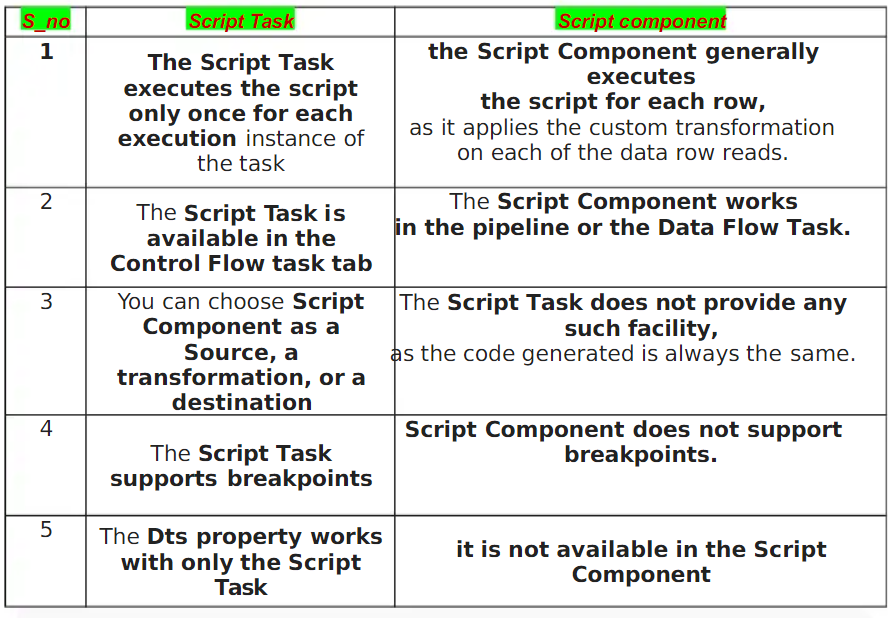
* Shipping products to the wrong customer addresses.
* Missing sales opportunities because of erroneous or incomplete customer records.
* Being fined for improper financial or regulatory compliance reporting.

In 2021, consulting firm Gartner stated that bad data quality costs organizations an average of $12.9 million per year. Another figure that's still often cited comes from IBM, which estimated that data quality issues in the U.S. cost $3.1 trillion in 2016. And in an article he wrote for the MIT Sloan Management Review in 2017, data quality consultant Thomas Redman estimated that correcting data errors and dealing with the business problems caused by bad data costs companies an average of 15% to 25% of their annual revenue.

**10)Explain Invalid DQS.**

**🡪**

**11)difference between script task v/s component task.**

**🡪**

**12)Explain data qulity services with match data.**

**🡪**The Data Quality Services (DQS) data matching process enables you to reduce data duplication and improve data accuracy in a data source. Matching analyzes 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 matches and take the appropriate action on the source data.

The DQS matching process has the following benefits:

Matching enables you to eliminate differences between data values that should be equal, determining the correct value and reducing the errors that data differences can cause. For example, names and addresses are often the identifying data for a data source, particularly customer data, but the data can become dirty and deteriorate over time. Performing matching to identify and correct these errors can make data use and maintenance much easier.

Matching enables you to ensure that values that are equivalent, but were entered in a different format or style, are rendered uniform.

Matching identifies exact and approximate 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. You define which fields are assessed for matching, and which are not.

DQS enables you to create a matching policy using a computer-assisted process, modify it interactively based upon matching results, and add it to a knowledge base that is reusable.

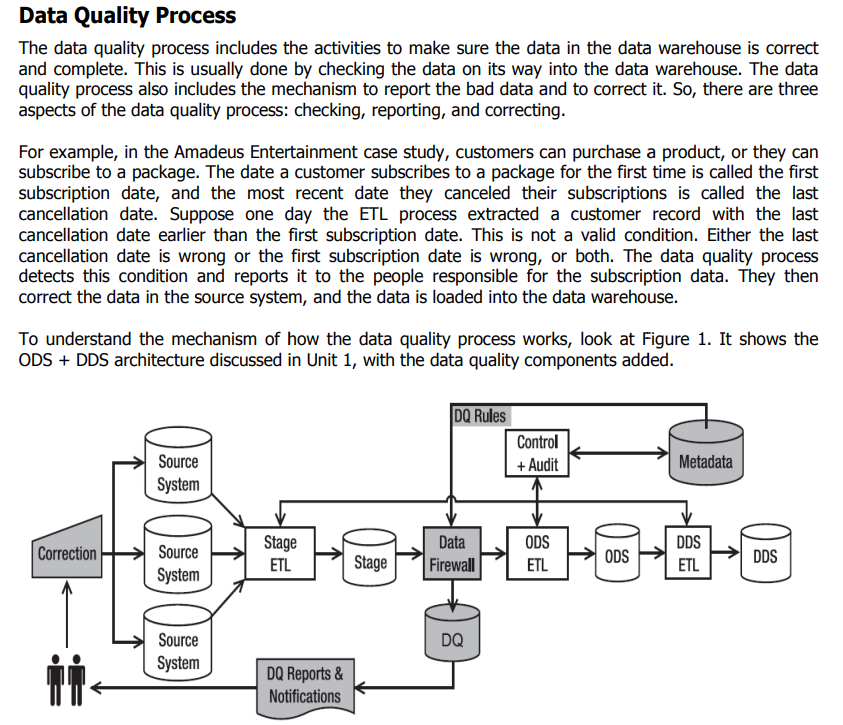
You can re-index data copied from the source to the staging table, or not re-index, depending on the state of the matching policy and the source data. Not re-indexing can improve performance.

HOW tO MATCH DATA:

Create a matching policy in the knowledge base

Perform a de-duplication process in a matching activity that is part of a data quality project.

**13)Explain data qulity process with diagram.**

**🡪**

**14)**

**UNIT-5**

**1)SSAS-**SQL Server Analysis Services

**2)SSRS-**SQL Server Reporting Services

**3)BIML=**Business Intelligence Markup Language

**4)What is data analysis?**

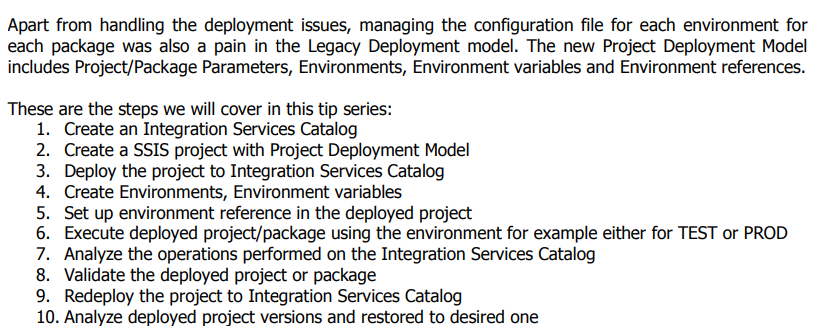
**🡺**a process of inspecting, cleansing, transforming, and modeling data with the goal of discovering useful information; suggesting conclusions; and supporting decision-making

**5)What is deployeement?**

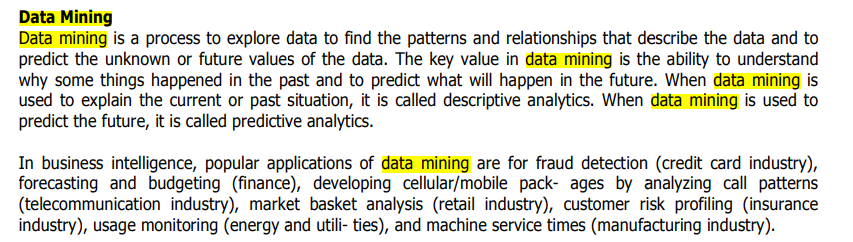
**🡺**  Deployment is the next phase after construction. In the deployment phase, you attend to the last few details, turn the data warehouse on, and let the users reap the benefits. By the time you reach the deployment phase, the majority of the functions are completed. The main concerns in the deployment phase relate to the users getting the training, support, and the hardware and tools they need to get into the warehouse.

To find our place in the whole life cycle of data warehouse development, let us summarize the functions and operations that have been completed up to this point. Here is the list of major activities already completed in the construction phase:

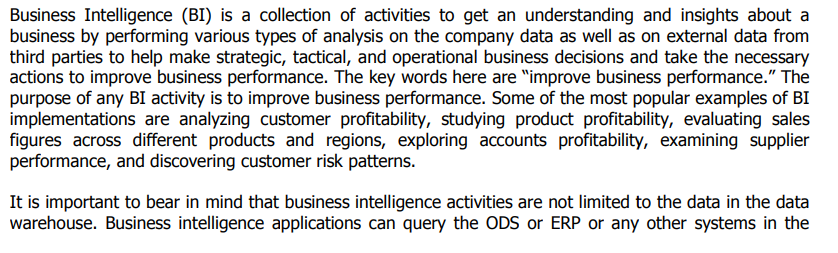
* The infrastructure is in place with the components fully tested.
* The validity of the architecture is already verified.
* The database is defined. Space allocation for the various tables is completed.
* The staging area is fully set up with file allocations.
* The extract, transformation, and all other staging area jobs are tested.



**7)What is data mining?**

**🡪** ****

**8)Explain Bussiness Intaligence**

**🡪**

****

**Adavantages**

* Reduced complexity. ...
* Improved data access. ...
* Enhanced productivity. ...
* Right time, right data. ...
* Increased data quality. ...
* Empowered decision-making. ...
* Lower costs. ...
* Trend insights

**disadvantages**:

* Initial cost
* User resistance
* Data skills gap

benefit:

* Data clarity
* Increased efficiency
* Better customer experience
* Improved employee satisfaction

**methods**

* [**Data mining**](https://www.tableau.com/learn/articles/what-is-data-mining)**:** Using databases, statistics, and [machine learning (ML)](https://www.tableau.com/learn/articles/define-machine-learning) to uncover trends in large datasets
* **Reporting:**Sharing data analysis to stakeholders so they can draw conclusions and make decisions
* **Performance metrics and benchmarking:** Comparing current performance data to historical data to track performance against goals, typically using customized dashboards
* **Descriptive analytics:** Using preliminary data analysis to find out what happened
* **Querying:** Asking the data-specific questions, BI pulling the answers from the data sets
* **Statistical analysis:** Taking the results from descriptive analytics and further exploring the data using statistics such as how this trend happened and why
* [**Data visualization**](https://www.tableau.com/learn/articles/data-visualization)**:** Turning data analysis into visual representations such as charts, graphs, and histograms to more easily consume data
* [**Visual analysis**](https://www.tableau.com/data-insights/reference-library/visual-analytics)**:** Exploring data through visual storytelling to communicate insights on the fly and stay in the flow of analysis
* **Data preparation:** Compiling multiple data sources, identifying the dimensions and measurements, and preparing it for data analysis

**9)Discuus control flow and data flow.**

**🡪**The data flow in the Data Warehouse describes which objects are needed at design time and which objects are needed at runtime to transfer data from a source to BI and cleanse, consolidate and integrate the data so that it can be used for analysis, reporting and possibly for planning.

the metadata description of the source data is modeled with **DataSources**

 A DataSource is a set of fields that are used to extract data of a business unit from a source system and transfer it to the entry layer of the BI system or provide it for direct access.

There is a new object concept available for DataSources in BI. In BI, the DataSource is edited or created independently of 3.x objects on a unified user interface. When the DataSource is activated, the system creates a **PSA** table in the Persistent Staging Area (PSA), the entry layer of BI

**InfoPackage**, you specify the selection parameters for transferring data into the PSA. In the new data flow, InfoPackages are only used to load data into the PSA.

Using the **transformation,** data is copied from a source format to a target format in BI. The transformation process thus allows you to consolidate, cleanse, and integrate data.

**InfoObjects** are the smallest units of BI. You map the information in a structured form that is required for constructing InfoProviders.

**InfoProviders** are persistent data repositories that are used in the layer architecture of the Data Warehouse or in views on data

Use:

Use of the new DataSource permits real-time data acquisition as well as direct access to source systems of type File and DB Connect.

The data transfer process (DTP) makes the transfer processes in the data warehousing layers more transparent. The performance of the transfer processes increases when you optimize parallelization. With the DTP, delta processes can be separated for different targets and filtering options can be used for the persistent objects on different levels. Error handling can also be defined for DataStore objects with the DTP.

The use of transformations simplifies the maintenance of rules for cleansing and consolidating data. Instead of two rules (transfer rules and update rules), as in the past, only the transformation rules are still needed.

## Control flow:

A control flow is a graphical model that sequences data flows and mining flows, integrates external commands, programs, and stored procedures, and provides conditional processing logic for a warehousing application. Deployed warehousing applications consist of control flows that you can run and monitor from the Administration Console.

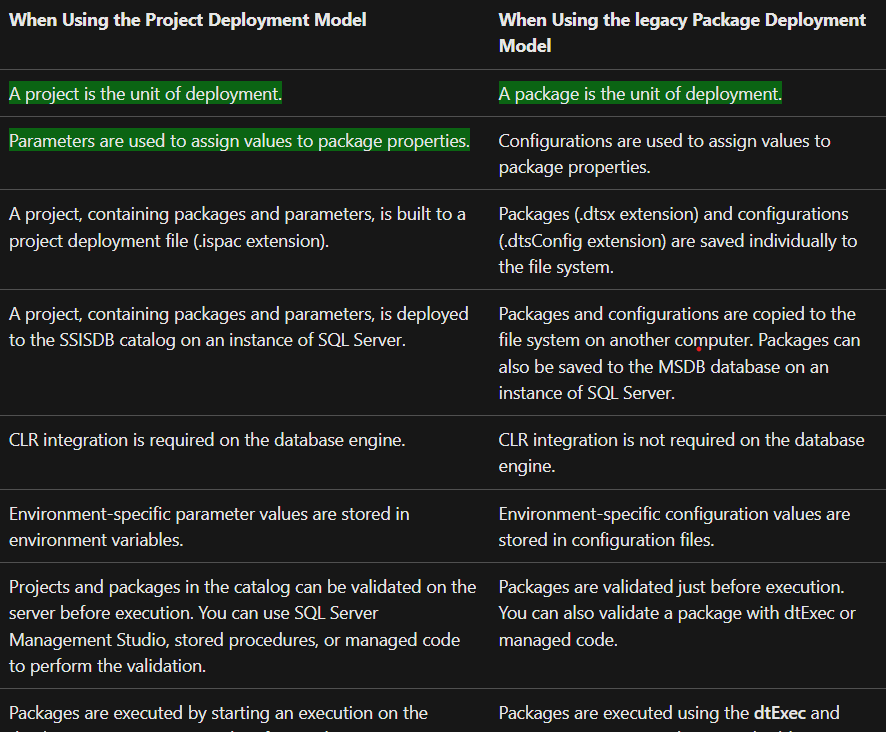
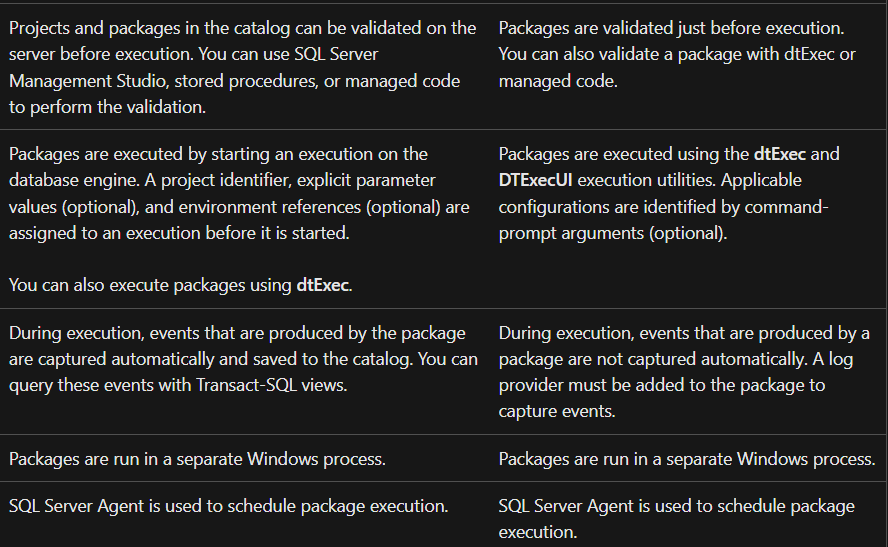
An individual control flow in the runtime environment is equivalent to a control flow in the design-time environment. When you schedule or start a control flow, you are running all of the activities within a particular control flow.

To run or schedule a control flow, you use the Manage Control Flows page of the console.

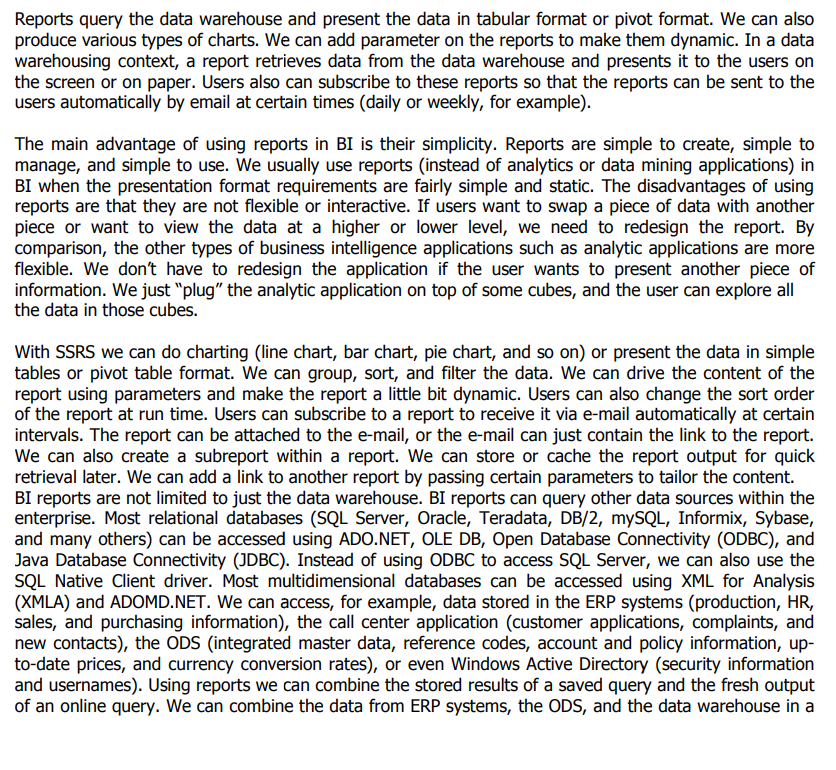
* [**Control flow instances**](https://www.ibm.com/docs/en/SSEPGG_9.7.0/com.ibm.dwe.admin.doc/cf.instances.html)  
  Control flow instances are specific jobs that are created when you start or schedule control flows from the console. Each instance has a specific name, either user-defined or system-generated, so you can monitor multiple instances of the same control flow when they run.
* [**Instance states**](https://www.ibm.com/docs/en/SSEPGG_9.7.0/com.ibm.dwe.admin.doc/instance.states.html)  
  When you schedule a control flow, each instance goes through a cycle of states: ideally from Scheduled to Running to Successful. To troubleshoot problems, you need to understand these states and what actions you can take for each state.
* [**Missed control flow instance**](https://www.ibm.com/docs/en/SSEPGG_9.7.0/com.ibm.dwe.admin.doc/missed_control_flow_instance.html)  
  You can manage the way in which the Administration Console runs the control flow instances that were missed at the scheduled times.
* [**Control flows and activities**](https://www.ibm.com/docs/en/SSEPGG_9.7.0/com.ibm.dwe.admin.doc/cfs.activities.html)  
  Warehousing applications consist of control flows and activities. Control flows are runtime data warehousing processes.
* [**Activity attributes**](https://www.ibm.com/docs/en/SSEPGG_9.7.0/com.ibm.dwe.admin.doc/activity.attributes.html)  
  Activity attributes are the set of properties that are associated with a particular type of activity. These attributes are treated as system-defined variables with the phase automatically set to run time.
* [**Variables in control flows**](https://www.ibm.com/docs/en/SSEPGG_9.7.0/com.ibm.dwe.admin.doc/variables.cf.html)  
  A variable is a user-defined name that represents data that can be changed within an activity property in a data flow or control flow. By setting variables for certain resources, you can defer the definition of critical properties until a later phase in the life cycle of the warehousing application.
* [**Control flow profiles**](https://www.ibm.com/docs/en/SSEPGG_9.7.0/com.ibm.dwe.admin.doc/copcflprofiles.html)  
  Control flow profiles store variable values that you can use during runtime or execution time of the control flow. You can create multiple control flow profiles in the Administration Console.

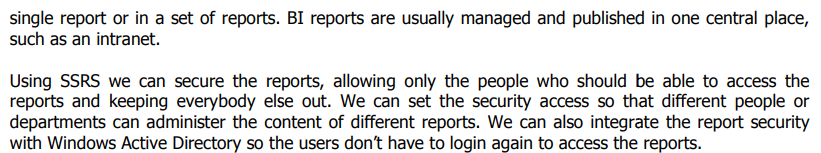
**10)Differentiate project deployment model v/s package deployment package.**

**🡺**

**** ****

**11)write note on busssiness intelligence report.**

**🡪** ****

****