CS –32
Data Warehousing with SQL Server 2012

03. Creating ETL Solutions with SSIS, Implementing Control Flow in SSIS
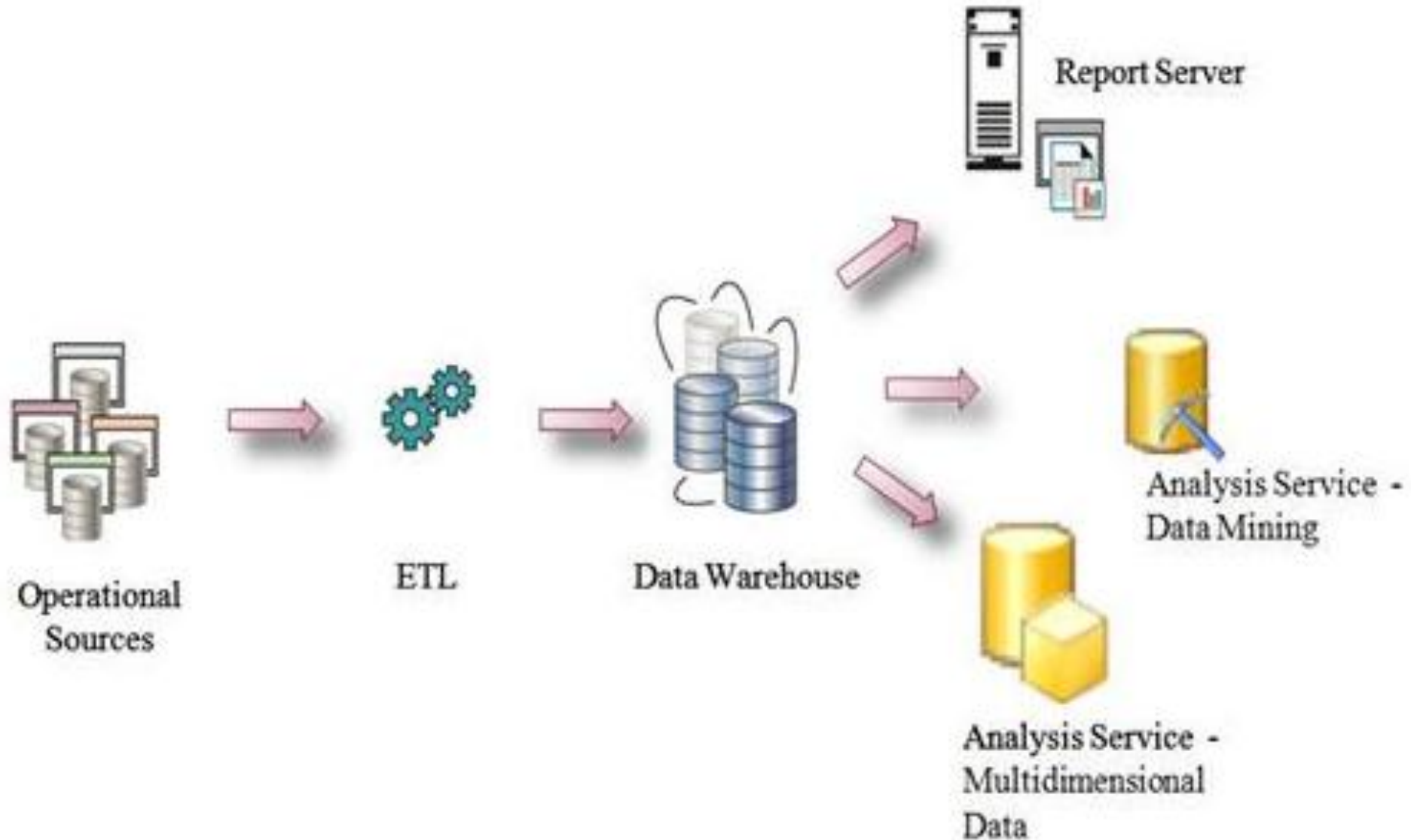
Overview of SSIS :

- SQL Server Integration Services (SSIS) is a platform for building high performance data integration and workflow solutions.

- It allows creation of packages or SSIS packages. Which and made up of tasks that can move data from source to destination and alter it if required.

- SSIS is basically an ETL tool whose main purpose is to do extraction, transformation and loading of data.

Overview of SSIS :

- SSIS can be used for several other purposes,

- For example,

  - To automate maintenance of SQL Server databases,

  - To update multidimensional cube data or send e-mails detailing the status of the operation as defined by the user.

- SSIS is a component of SQL Server 2005/2008 and is the successor of DTS(Data Transformation Services) which had been in SQL Server 7.0/2000.

Overview of SSIS :

Operational Sources → ETL → Data Warehouse → Report Server / Analysis Service - Data Mining / Analysis Service - Multidimensional Data

Overview of SSIS :

- **Typical Use of Integration Services**
  - Merging Data from Heterogeneous Data Stores
  - Populating Data Warehouses and Data Mart
  - Cleaning and Standardizing Data
  - Building Business Intelligence into a Data Transformation Process
  - Automating Administrative Functions and Data Loading.

Overview of SSIS :

- **What is an ETL process?**

  - ETL stands for Extraction, Transformation and Loading. It is a Process in data warehousing to extract data, transform data and load data to final source.

  - ETL covers a process of how the data are loaded from the source system to the data warehouse. Let us briefly describe each step of the ETL process.

Overview of SSIS :

- **Extraction**

  - Extraction is the first step of ETL process where data from different sources like txt file, XML file, Excel file or various sources collected.

- **Transformation**

  - Transformation is the second step of ETL process where all collected data is been transformed into same format i.e. format can be anything as per our....Cont
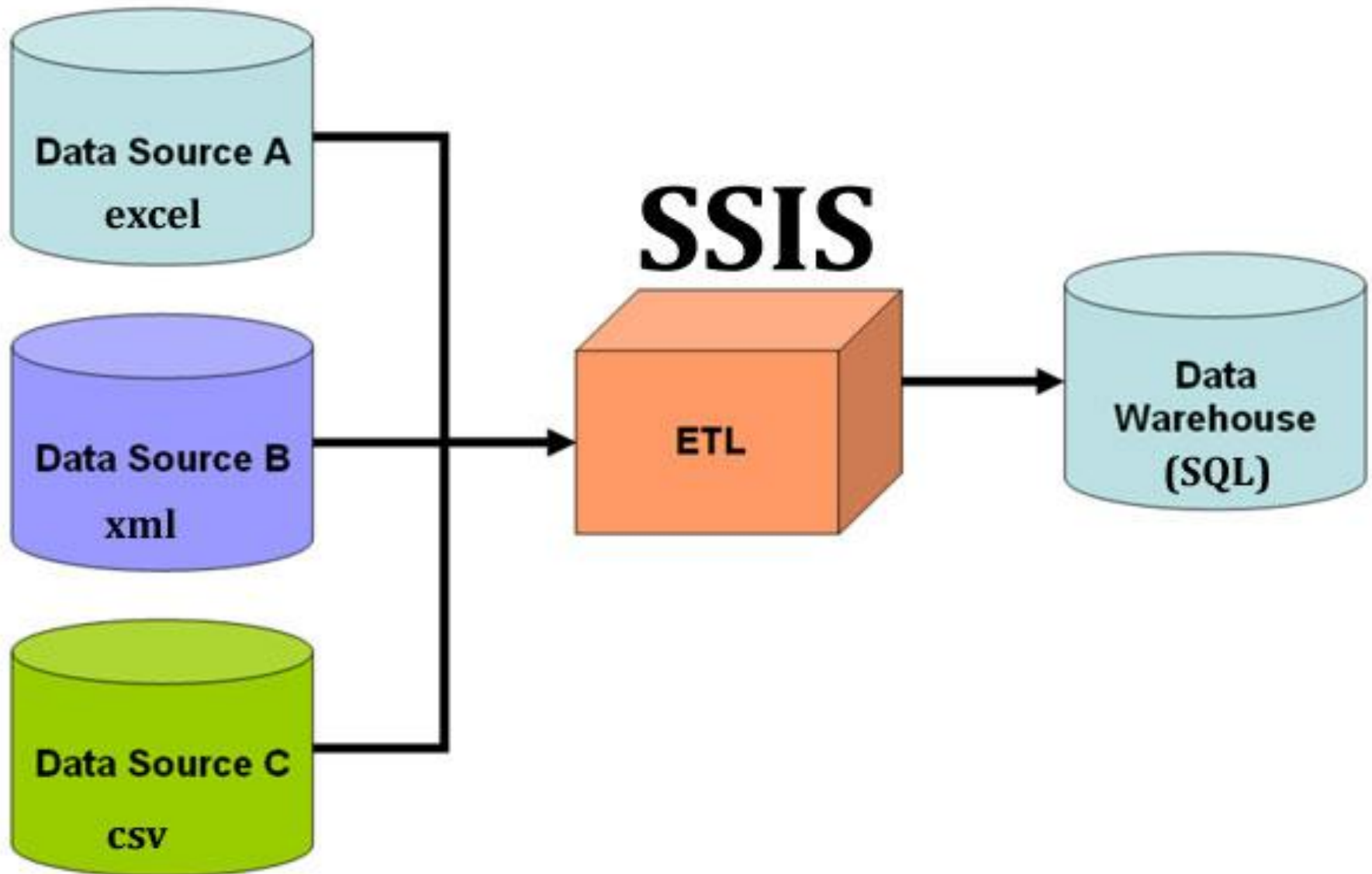
requirement before loading it to data-warehouse i.e. it may be data-type format, data merge format, splitting format, alphabet joining format, currency format etc.

- Loading

  ❑ Final step of ETL process, The big chunk of data which is collected from various sources and transformed then finally load to our data warehouse.

Overview of SSIS :

Overview of SSIS :

- **Control Flow :**

  - Defines both the order of operations and the conditions under which they will be executed.

  - A package can consist of one or more operations, represented by control flow task. Execution order is defined by how individual tasks are connected to one another.

  - Tasks that do not follow any preceding task as well as tasks that follow the same preceding task are executed in parallel.

- Encapsulates the data movement components-the ETL :

  - One or more source components, designating the data stores from which the data will be extracted.

  - One or more destination components, designating the a data stores into which the data will be loaded.

  - One or more (optional) transformation components, designating the transformations through which the data will be passed.
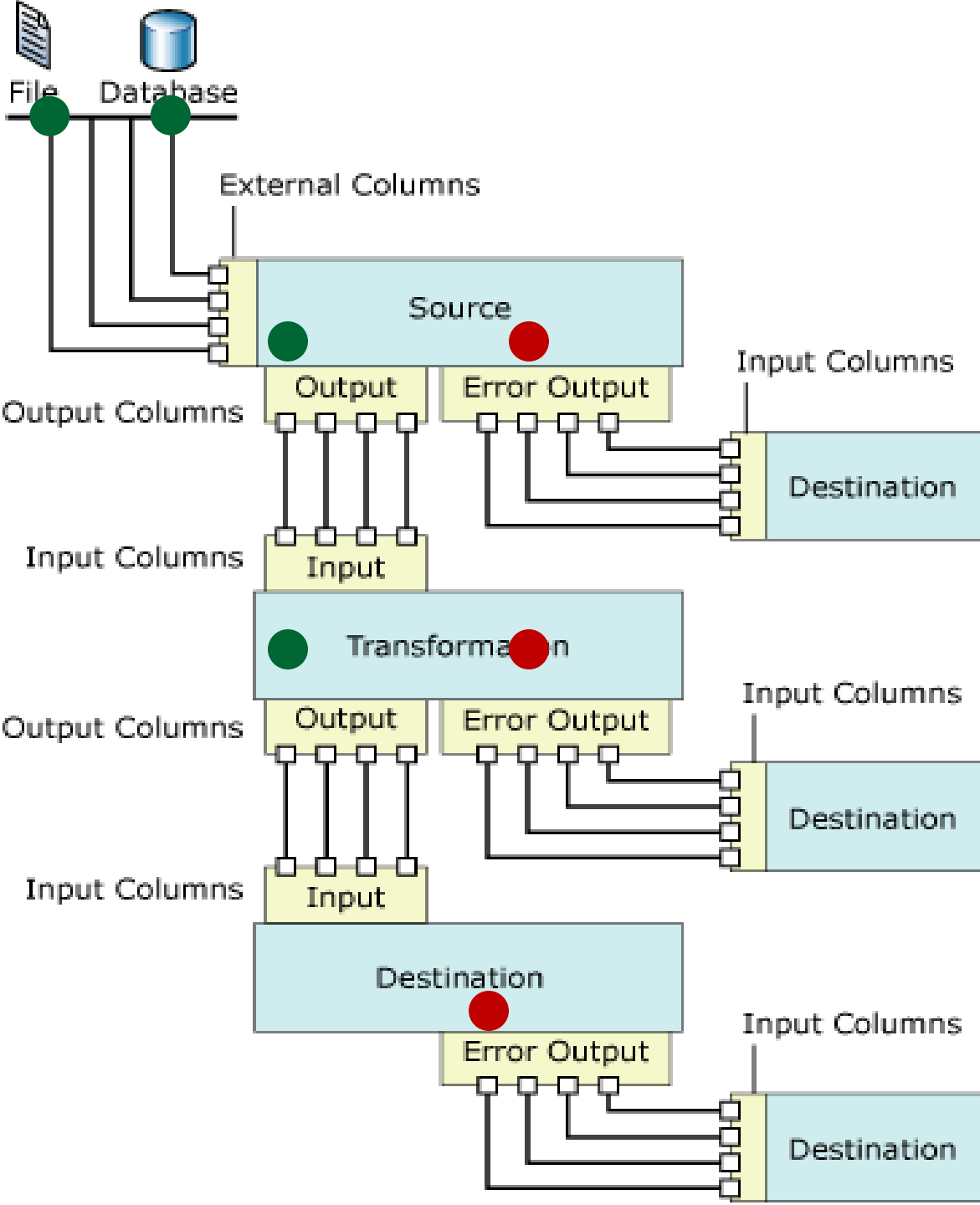
Overview of SSIS

- Connection Manager :

  - The role of connection managers is to provide access to data stores, either as data sources, data destinations, or reference data stores.

  - Control flow tasks define the data management operation of the SSIS process, with the data flow tasks providing the core of data warehousing operations – the ETL.

Data FLOW

- Elements of Data Flow include Elements of Data Flow are categorized into three parts :

- Data Flow Sources :

  - These elements are used to read data from different type of sources like (SQL Server, Excel sheet, Etc.)

Data FLOW

- Data Flow Transformations :

  - These elements are used to do process on data like (cleaning adding new columns, ect.)

- Data Flow Destinations :

  - These elements are used to save processed data into desired destination. (SQL Server, Excelsheet, etc.)

Container :

- Containers provide structure in packages and services to tasks in the control flow, Integration Services include the following container types, for grouping tasks and implementing repeating control flows:

  - The Foreach Loop container

  - For Loop Container

  - Sequence Container

Container :

- The Foreach Loop container :

  - It enumerates a collection and repeats its control flow for each member of the collection. The Foreach Loop Container is for situations where you have a collection of items and wish to use each item within it as some kind of input into the downstream flow.

Container :

- **For Loop Container:**

  - It's a basic container that provides looping functionality. A for loop contains a counter that usually increments (though it sometimes decrements), at which point a comparison is made with a constant value. If the condition evaluates to True, then the loop execution continues.

Container :

- Sequence Container :

  - One special kind of container both conceptually and physically can hold any other type of container or Control Flow component. It is also called "container of container", or super container.