

**Summer Training Report**

**On**

**“ETL & Dashboard Development”**

Submitted in the partial fulfillment for

The degree of Computer Engineering Submitted by:

**ANAND KUMAR**

Under the guidance of:

**Mr. Rakesh Kumar (Project Manager /supervisor)**

**LiquidHub Analytics pvt. Ltd.**

**(Formerly known as “Annik Technology Services Pvt. Ltd.”)**

**ACKNOWLEDGEMENT**

We take up this occasion of expressing of thankfulness towards all the persons who have been instrumental in the compilation.

We are feeling short of words of expressing ones feeling of gratitude towards my highly respective and esteemed guide,

**ANAND KUMAR**

**DECLARATION**

I certify that

1. The work contained in this report is original and has been done by me under the guidance of my supervisor.
2. The work has not been submitted to any other institute for any degree or diploma.
3. I have followed the guidance provided by the institute in the preparing the report.
4. I have confirmed to the norms and guidelines given in ethical code of conduct of institutions.
5. Whenever I have used materials (data, theoretical, analysis, figures and text) from other sources. I have given credit to them by citing in the report and giving their details in the bibliography. Further I will take permission from the copyright owners of the sources, whenever necessary.

**Name Roll No.**

**ANAND KUMAR 35118002716**

**CONTENTS**

|  |  |  |  |
| --- | --- | --- | --- |
| S.n | **Name Of Topic** | **Subtopic** | **Page NO.** |
| 1 | **Abstract** | ->Feature  ->What effect does it have | 06-07 |
| 2 | **Introduction** | ->Purpose  ->Feature | 08-09 |
| 3. | **Modules with Diagram**  ***A)ETL - (Extract, Transform, Load)*** | 1. Control Flow 2. Data Flow 3. Package Explorer &Execution Result 4. Data on Data warehouse | 010-18 |
|  | **B).*Data on SQL Server After ETL(SSIS),SSMS(SQL SERVER MANAGEMEN STUDIO)*** |  |  |
|  | ***C)Dashboard Result*** |  | 19 |
|  | **Implementation**  *A)*  ***HARDWARE AND SOFTWARE TO BE USED:-*** | 1. Hardware Specification 2. Software Specification | 19-33 |
|  | ***B)Explanation*** | * **1.** PowerBI |  |
|  |  | 2.SQL SERVER DATA TOOLS FOR VISUAL STUDIO(SSDT) |  |
|  |  | 1. SQL Server Express |  |
|  |  | 4.Microsoft SQL Server Management Studio |  |
|  | **c)Implementation**  **procedure** | 1. ETL 2. Output of ETL   3)SSMS(SQL server Management Studio)  **4.**Dashboard  5.Output of Dashboard |  |

**ABSTRACT**

This project work describe in detail, the project work undertaken by us during the final year of degree by us during the final year of degree at **DELHI TECHNICAL CAMPUS.** The contents of this report include a complete description of the **‘ETL & Dashboard Development’**.

This description is empowered with the **ETL & Dashboard Development** is

Scheduled data integration, or **ETL**, is an **important** aspect of warehousing because it consolidates data from multiple sources and transforms it into a useful format. This allows the user to easily access data from one interface, lessening the reliance on your IT team and You can **use** it to pull data from a wide range of systems in the cloud and on premises and create dashboards that track the metrics you care about the most, or drill in and (literally) ask questions about your data.

**FEATURES :** Extract, Transform, and Load (ETL) is a data warehousing process that uses batch processing to help business users analyze and report on data relevant to their business focus. Becoming increasingly popular in a modern data warehouse architecture, the ETL process pulls data out of the source, makes changes according to requirements, and then loads the transformed data into a database or BI platform to provide better business insights. With ETL, business leaders can make data-driven business decisions.

**WHAT IMPACT DOES IT HAVE** Data warehouses(SQL server) are centralized data storage systems that allow your business to integrate data from multiple applications and sources into one location .This provides an environment that is designed for decision support, analytics reporting, and data mining. When you isolate and optimize your data, you can manage it without impacting primary business processes.

In general, the benefits of data warehousing are all based on one central premise: warehousing solves the ongoing problem of analyzing separate data and converting it into actionable information you can use. Warehousing also allows you to process large amounts of complex data in an efficient way. When you successfully implement a data warehouse system, it’s possible to access the benefits associated with the practice— the very benefits that are making data warehousing a common practice for many businesses today.

Introduction

**PURPOSE**

* The purpose of ETL & Dashboard Development is to Extract, Transform, and Load (ETL) is a data warehousing process that uses batch processing to help business users analyze and report on data relevant to their business focus.
* Becoming increasingly popular in a modern data warehouse architecture, the ETL process pulls data out of the source, makes changes according to requirements, and then loads the transformed data into a database or BI platform to provide better business insights.
* With ETL, business leaders can make data-driven business decisions

**FEATURES:**

>Timely access to data

>Enhanced quality and consistency

> High return on investment

**Modules With Diagram**

**A).ETL - (Extract, Transform, Load)**

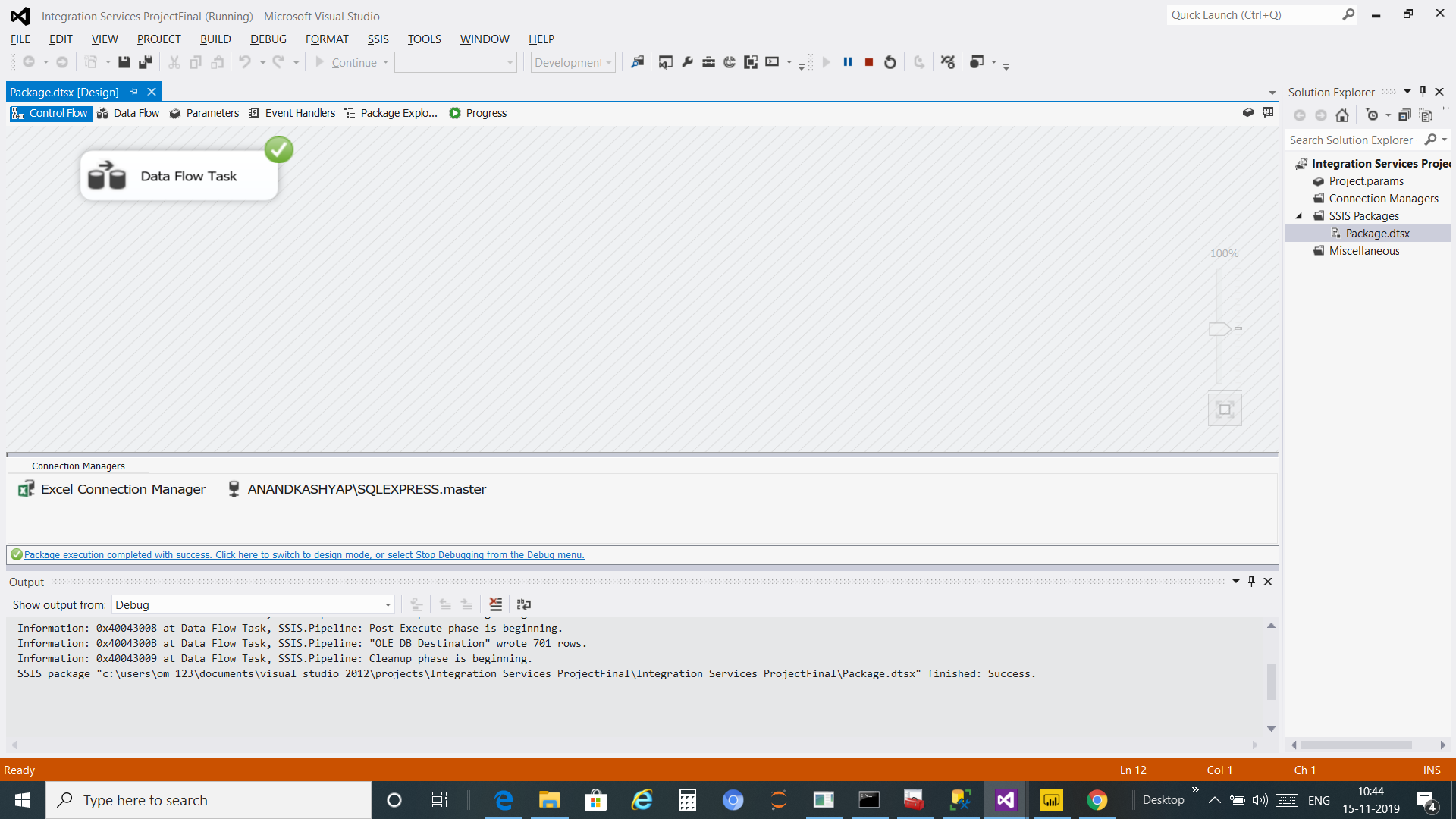
1. Extract is the process of reading data from a database. ...
2. Transform is the process of converting the extracted data from its previous form into the form it needs to be in so that it can be placed into another database. ...
3. Load is the process of writing the data into the target database.

1. **Control Flow:**

->Control flow element is one that performs any function or provides structure or control the flow of the elements.

->There must be at least one control flow element in the SSIS package. In SSIS a workflow is called a control-flow. A control-flow links together our modular data-flows as a series of operations in order to achieve a desired result.

->A control flow consists of one or more tasks and containers that execute when the package runs. To control order or define the conditions for running the next task or container in the package control flow



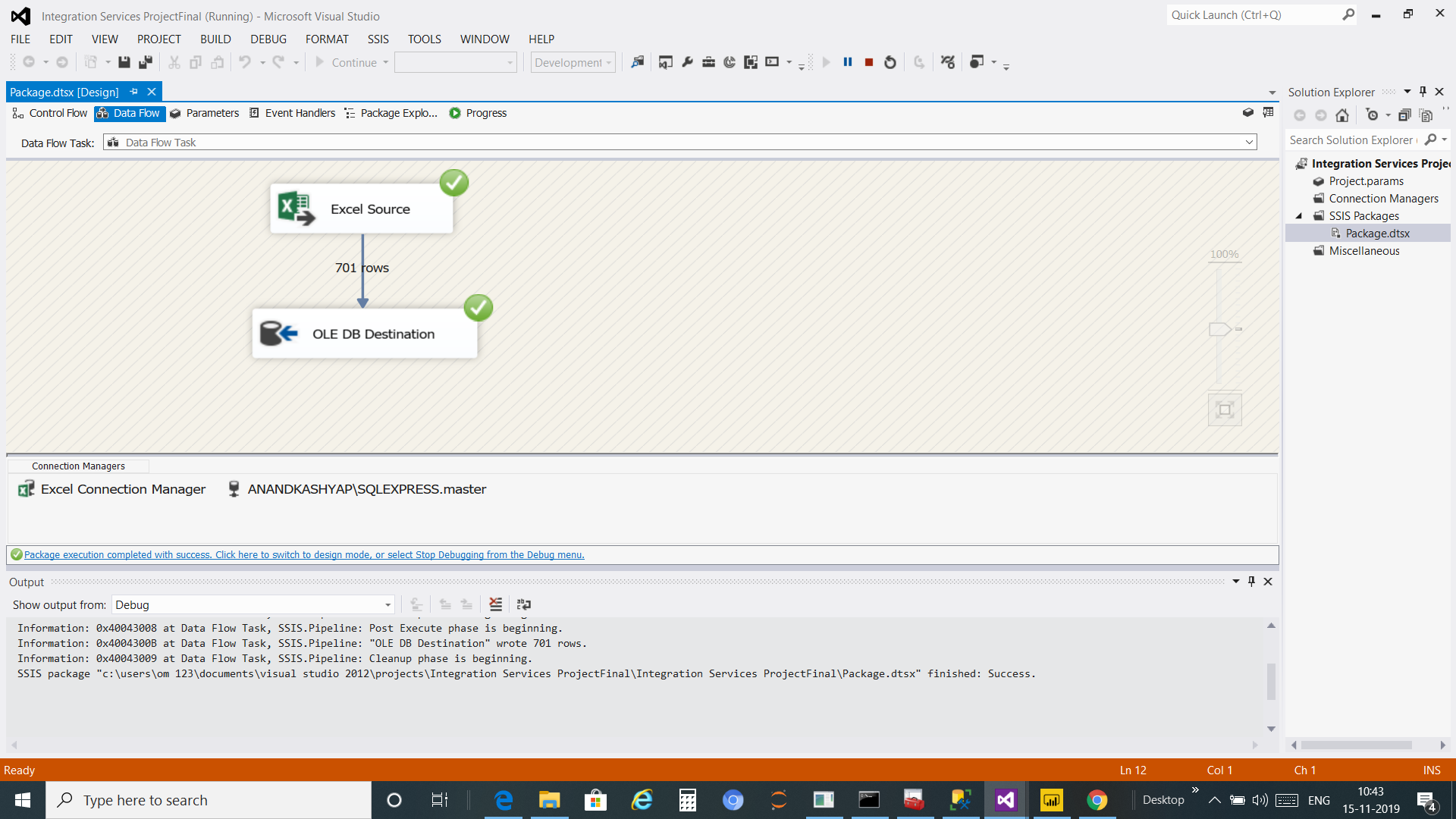
**2.Data Flow**:

->All ETL tasks related to data are done by data flow elements. It is not necessary to have a data flow element in the SSIS package.

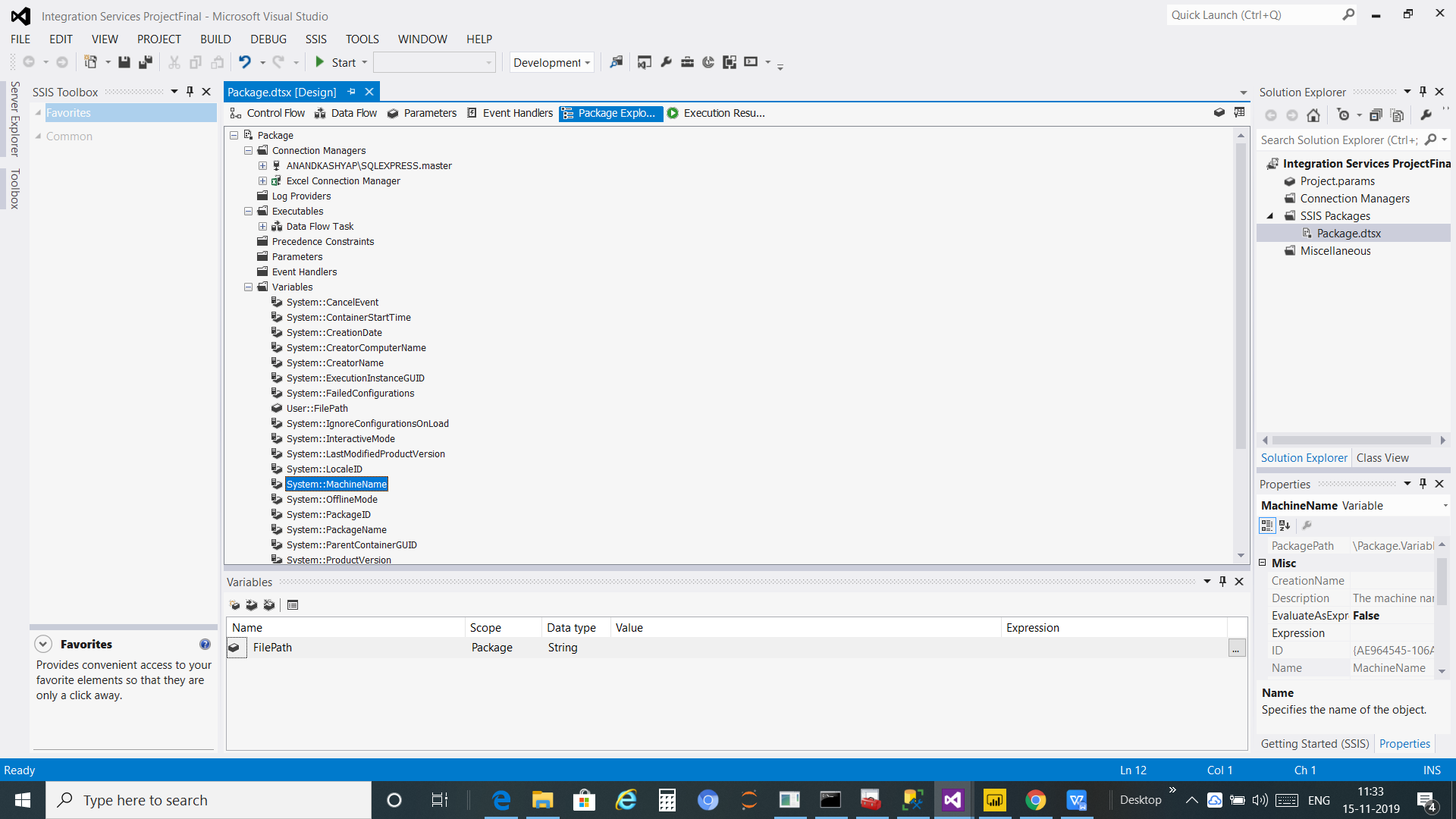
-> A data flow consists of the sources and destinations that extract and load data, the transformations that modify and extend data, and the paths that link sources, transformations, and destinations. Before you can add a data flow to a package, the package control flow must include a Data Flow task.

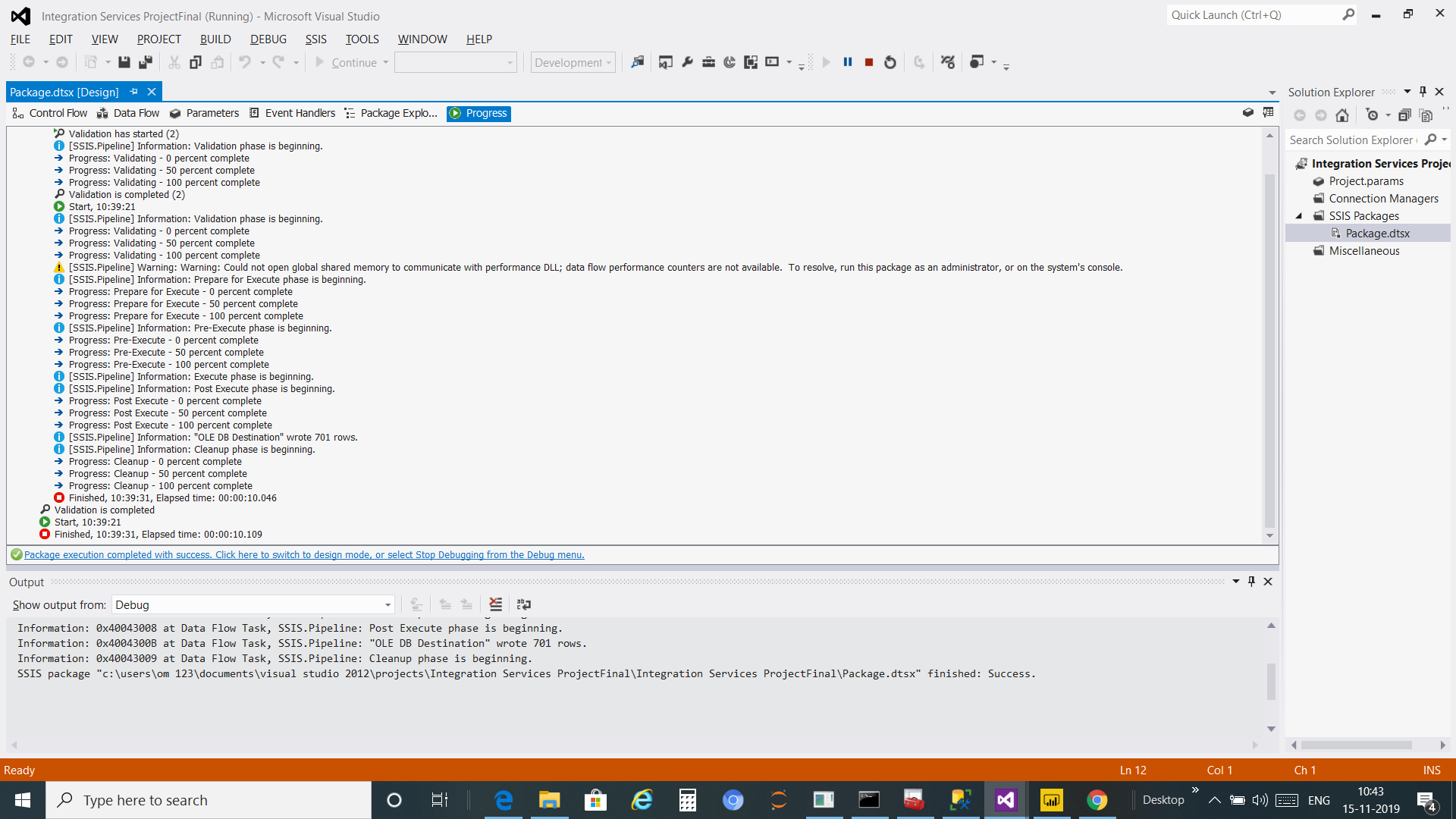
->The Data Flow task is the executable within the SSIS

package that creates, orders, and runs the data flow. A separate instance of the data flow engine is opened for each Data Flow task in a package



**3.Package Explorer and Execution Result**



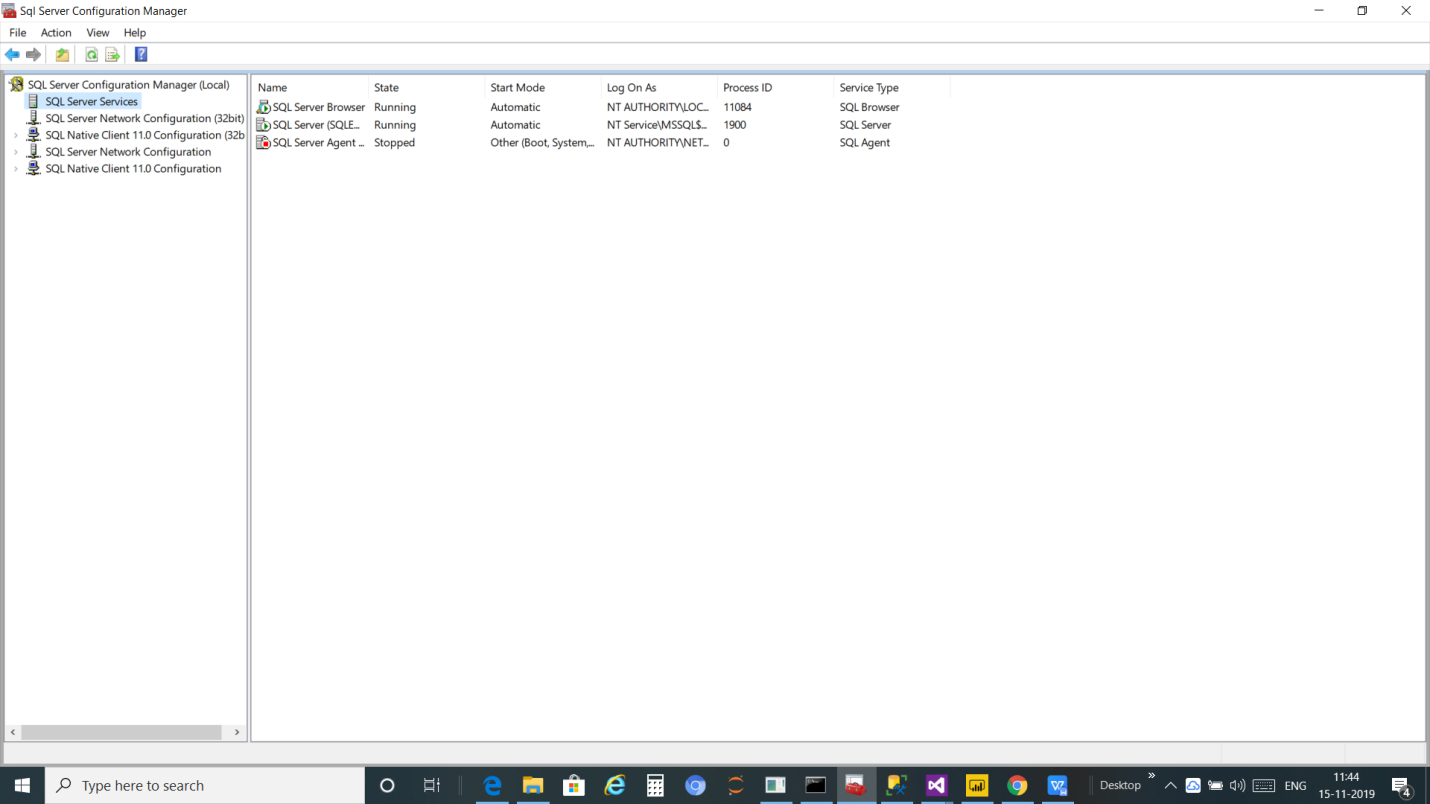


**4.Data on Data warehouses**

->Data warehouses(SQL server) are centralized data storage systems that allow your business to integrate data from multiple applications and sources into one location .

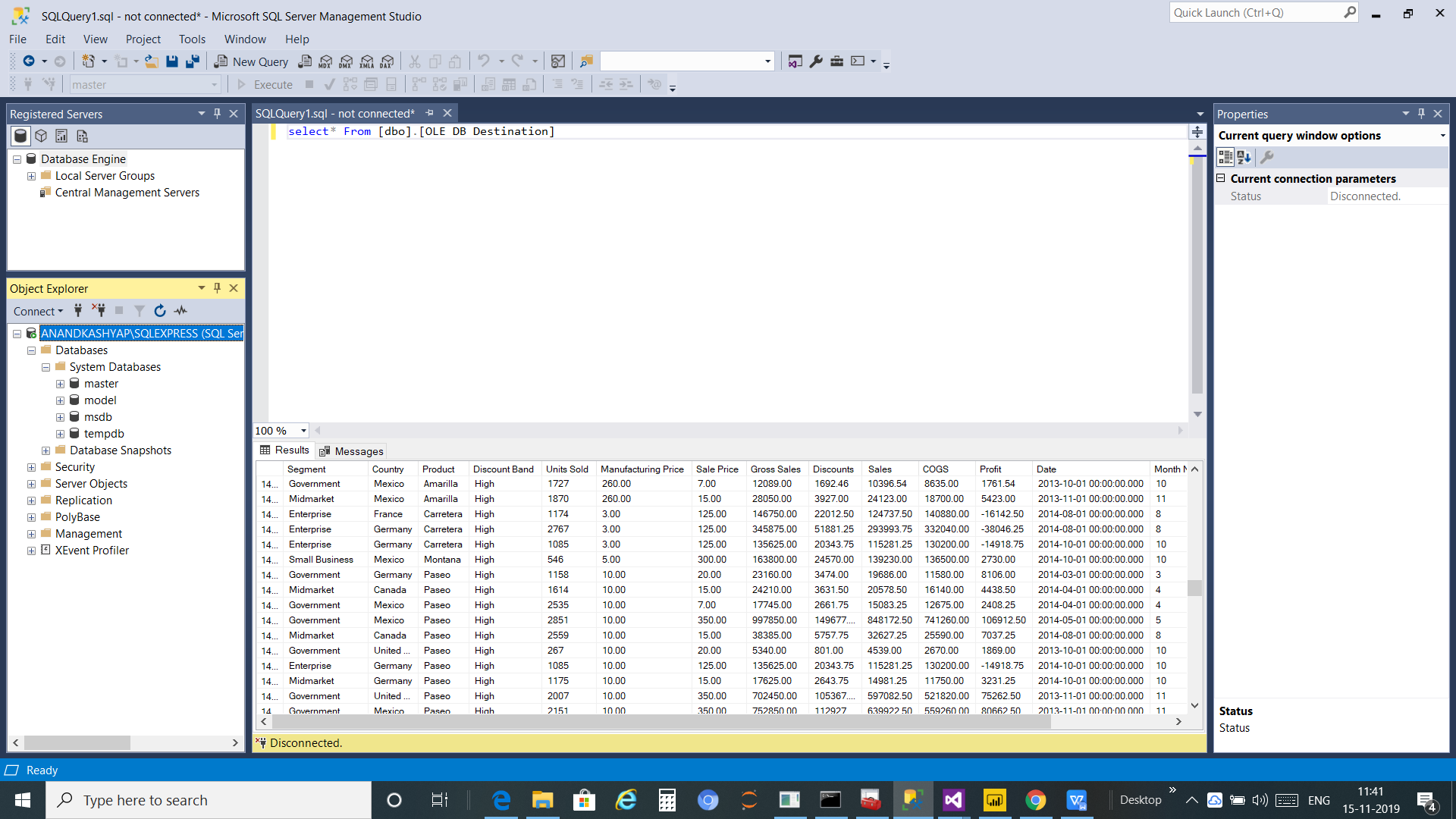
->This provides an environment that is designed for decision support, analytics reporting, and data mining. When you isolate and optimize your data, you can manage it without impacting primary business processes.

**SQL configuration manager:**



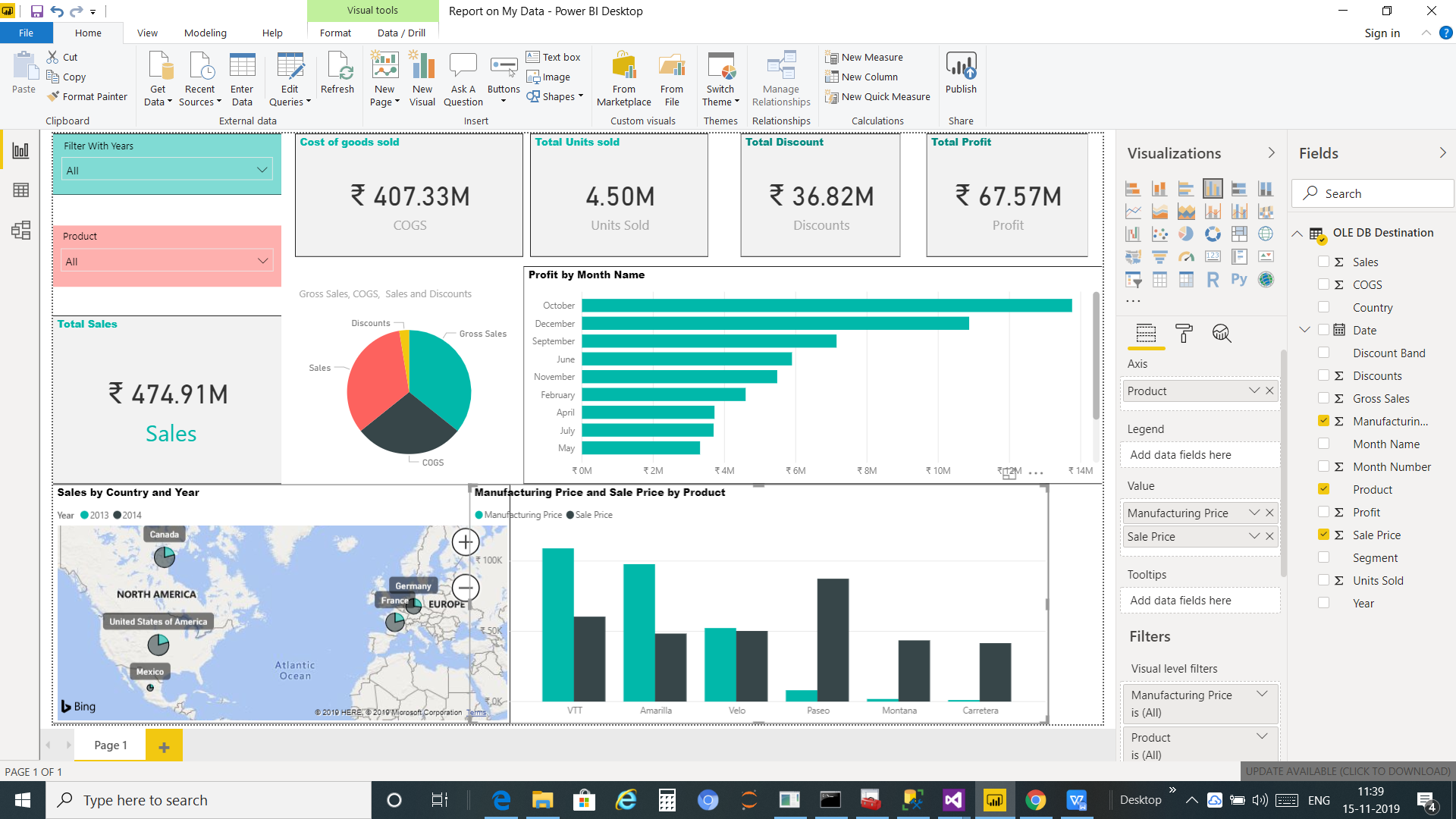
**B).Data on SQL Server After ETL(SSIS)**

**SSMS(SQL SERVER MANAGEMEN STUDIO)**



**C).Dashboard Result:**

A **dashboard** is a tool used for information management and business intelligence. Much like **the dashboard** of a **car**, data **dashboards** organize, store, and display important information from multiple data sources into one, easy-to-access place.



**IMPLEMENTATION**

**A).HARDWARE AND SOFTWARE TO BE USED:-**

1) **Hardware Specifications**

|  |  |
| --- | --- |
| **HARDWARE** | **SPECIFICATIONS** |
| CPU | Intel Core 2 Duo CPU |
| Processor speed | 2.10 GHz |
| RAM | 1.00 GB |
| Hard Disk | 120 GB |

2). **Software Specifications**

|  |  |
| --- | --- |
| FOR DASHBOARD | **PoweBI** |
| ETL or SSIS | **SQL SERVER DATA TOOLS FOR VISUAL STUDIO(SSDT)** |
| SERVER  (DATA WAREHOUSE) | **SQL SERVER express** |
| To manage SQL server Database | **Microsoft SQL SERVER Management   Studio(SSMS)** |

**Explanation :-**

**1.PowerBI:**

->Power BI is a business analytic service by Microsoft.

->It aims to provide interactive visualizations and business intelligence capabilities with an interface simple enough for end users to create their own reports and dashboards.

**2.SQL SERVER DATA TOOLS FOR VISUAL STUDIO(SSDT):**

****->SQL Server Data Tools**** is a modern development tool for building SQL Server relational databases, Azure SQL databases, Analysis Services (AS) data models, Integration Services (IS) packages, and Reporting Services (RS) reports. With SSDT, you can design and deploy any SQL Server content type with the same ease as you would develop an application in Visual Studio.

**3.SQL server Express:**

->SQL Server Express is a free version of Microsoft’s primary relational database management system (RDBMS) – the SQL Server.

->Essentially, the SQL Server is a database management system that can be used to store and access the information stored in many different databases.

->SQL Server comes with an impressive range of features like business intelligence, reporting, and in-depth advanced analytics.

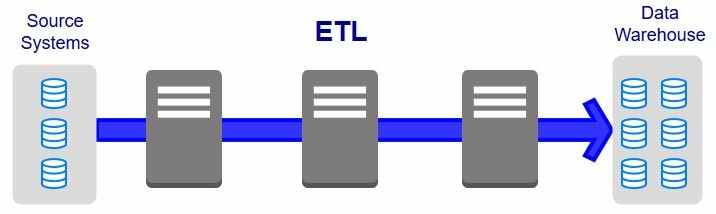
**4.Microsoft SQL Server Management Studio:**

SQL Server Management Studio (SSMS) is an integrated environment for managing any SQL infrastructure, from SQL Server to Azure SQL Database. SSMS provides tools to configure, monitor, and administer instances of SQL Server and databases. Use SSMS to deploy, monitor, and upgrade the data-tier components used by your applications, and build queries and scripts.

**Implementation Procedure:**

**1). ETL**

Data from one or more sources is extracted and then copied to the [data warehouse](https://www.webopedia.com/TERM/D/data_warehouse.html). When dealing with large volumes of data and multiple source systems, the data is consolidated. ETL is used to [migrate](https://www.webopedia.com/TERM/D/data_migration.html) data from one database to another, and is often the specific process required to load data to and from [data marts](https://www.webopedia.com/TERM/D/data_mart.html) and data warehouses, but is a process that is also used to to large convert (transform) databases from one format or type to another.



## ***ETL***

->ETL is an important part of today's [business intelligence](https://www.webopedia.com/TERM/B/Business_Intelligence.html) (BI) processes and systems.

It is the IT process from which data from disparate sources can be put in one place to programmatically analyze and discover business insights.

->Databases are the most traditional kind of data source in BI. There are many different kinds of databases, and many vendors providing databases with different architectures and different features. Common databases used today include MS Access, Oracle, DB2, Informix, SQL, MySQL, Amazon SimpleDB and a host of others.

->Traditionally, transactional databases—namely the ones that record the company’s daily transactions, such as CRM, HRM and ERP—are not considered optimal for business intelligence.

->This is for a variety of reasons, including the fact that a) data is not optimized for reporting and analysis and b) querying directly against these databases may slow down the system and prevent the databases from recording transactions in real time.

->In some cases, companies use an ETL tool to collect data from their transactional databases, transform them to be optimized for BI and load them into a data warehouse or other data mart.

**->**The main downside of this approach is that a data warehouse is a complex and expensive architecture, which is why many other companies opt to report directly against their transactional databases.

Steps-1

Sample Of data ,It may be any company data or another organization and the data may present in the form of Excel,ADO NET,CDC,FlAT FILE,ODBC,OLE DB,RAW FILE,XML etc

So, I have my source sample data is Excel.

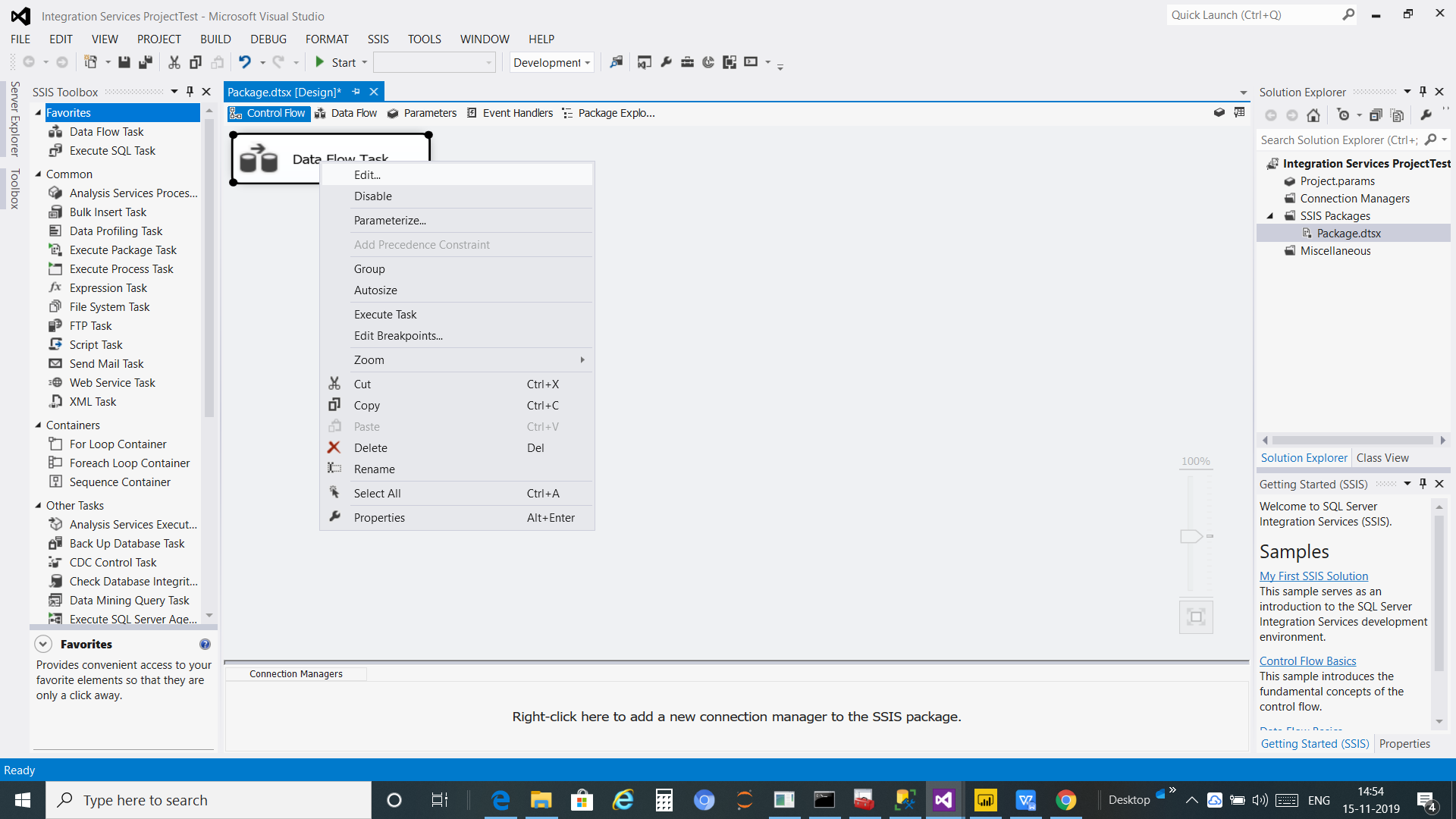
We need t SQL SERVER DATA TOOLS FOR VISUAL STUDIO(SSDT) which is provides SSIS (SQL server integration services) or ETL tools

Steps-2

Create new project on SSTD select Integration Services

Steps-3

Select Data Flow Task locked into SSIS Toolbox and on Data Flow Task double click and select the EDIT.

****

**Step-4**

Select the Sources File within one

ADO NET source

CDC Source

Excel Source

Flat File Source

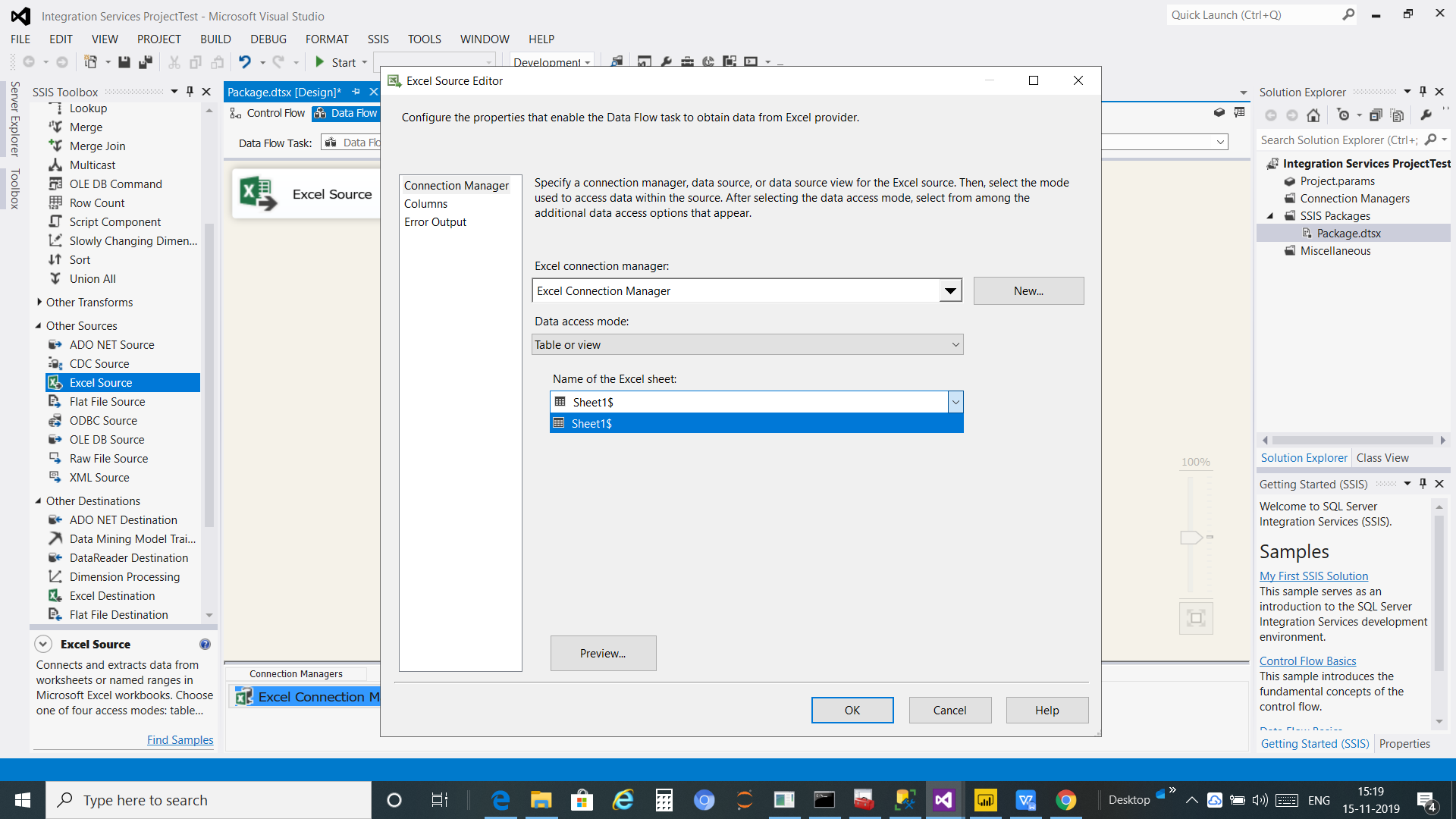
ODBC source

OLE DB source

RAW File Source

I have selected Excel Source and Edit it and configure the properties that enable the Data Flow task to obtain data from Excel provider.

It will need to established connection to the source file like that

****

**Step 5:**

so the above step identify the connection established to the source file and now we need to make connection to the destination

And the destination may be

ADO NET Destination

Data Mining Model

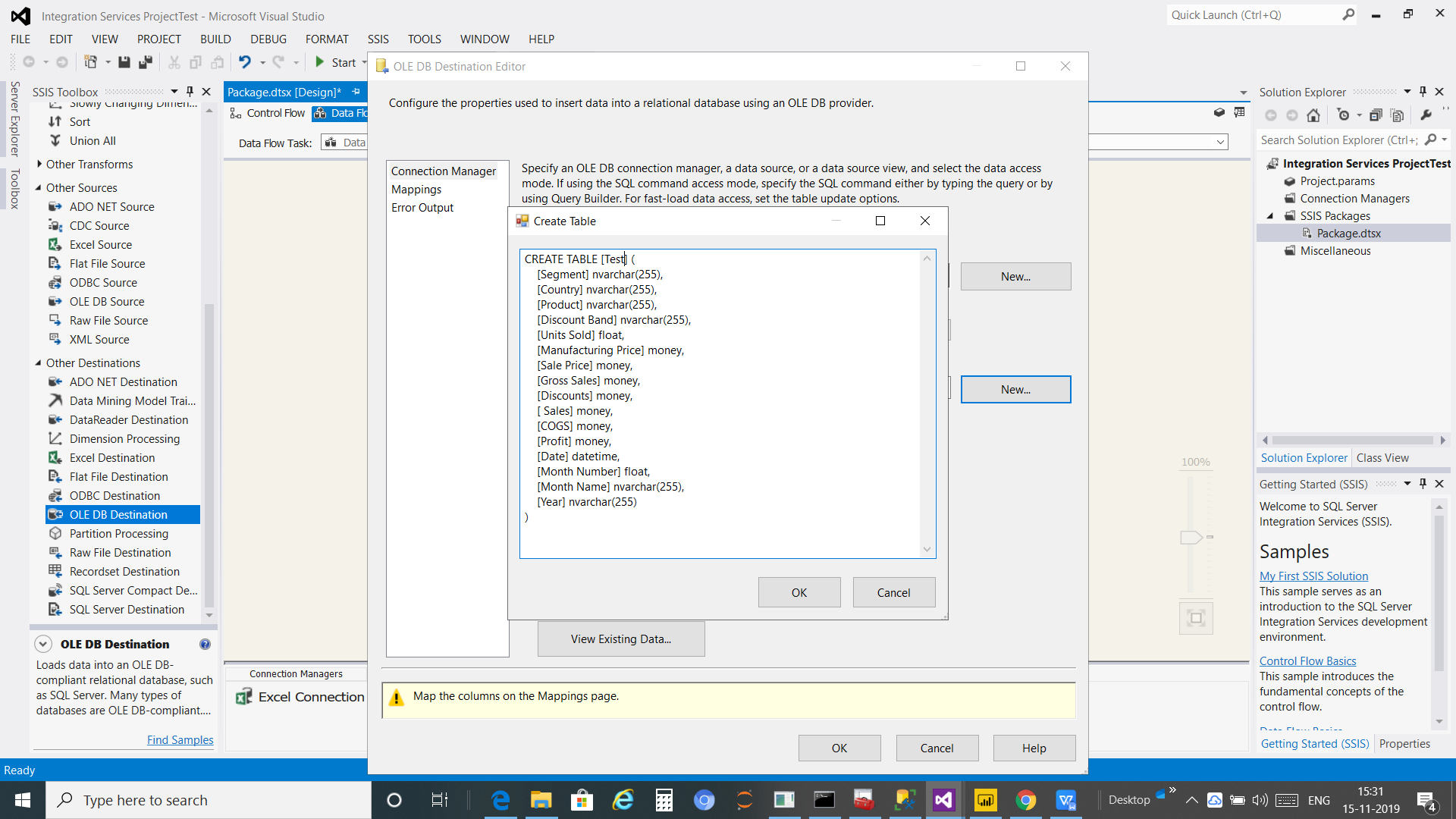
Data Reader Destination

Excel Destination

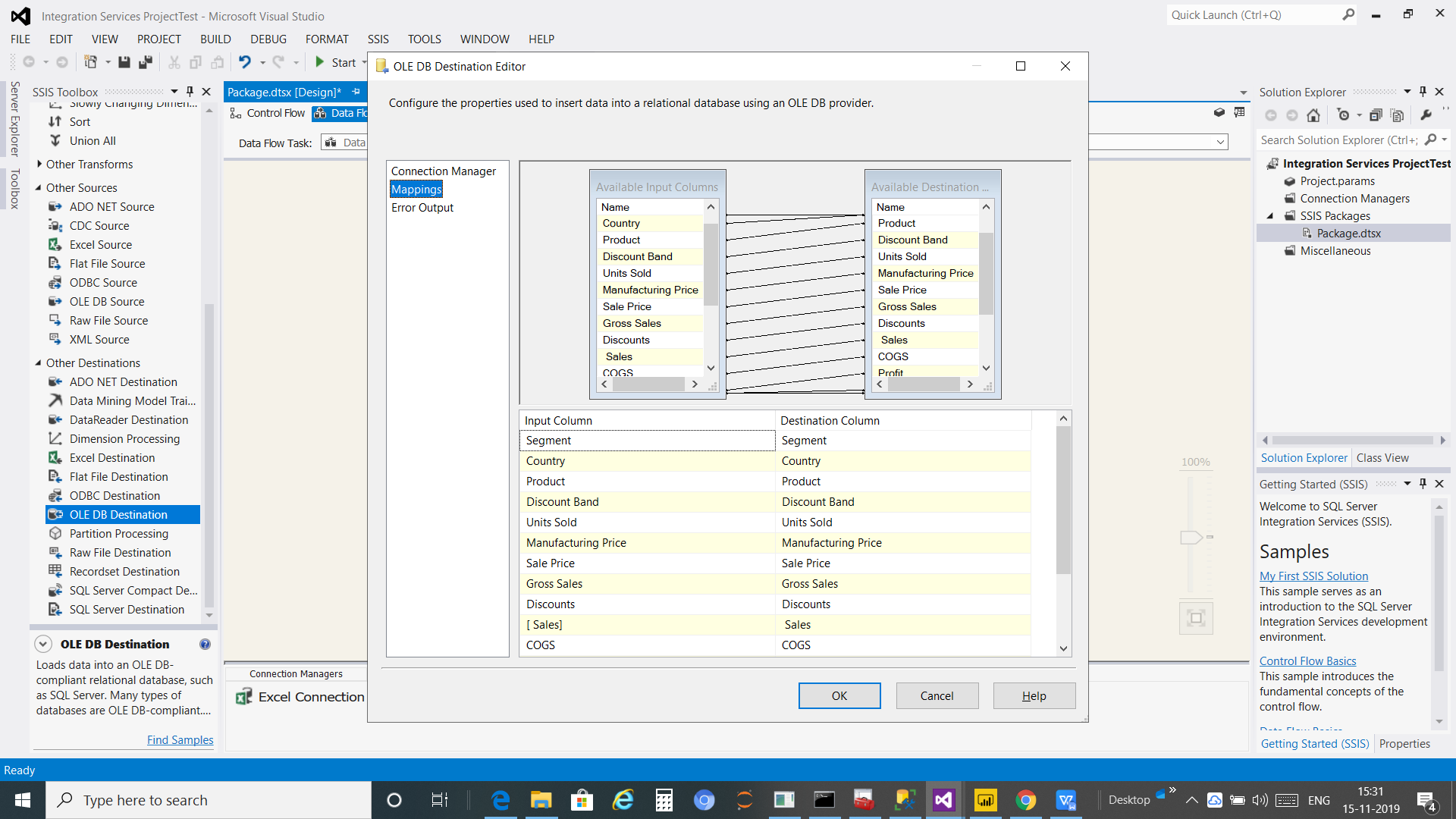
ODBC destination

OLE DB Destination

I have selected OLE DB destination ,Now It need to established connection to configure the properties use to insert data into a relation database using an OLE DB provide

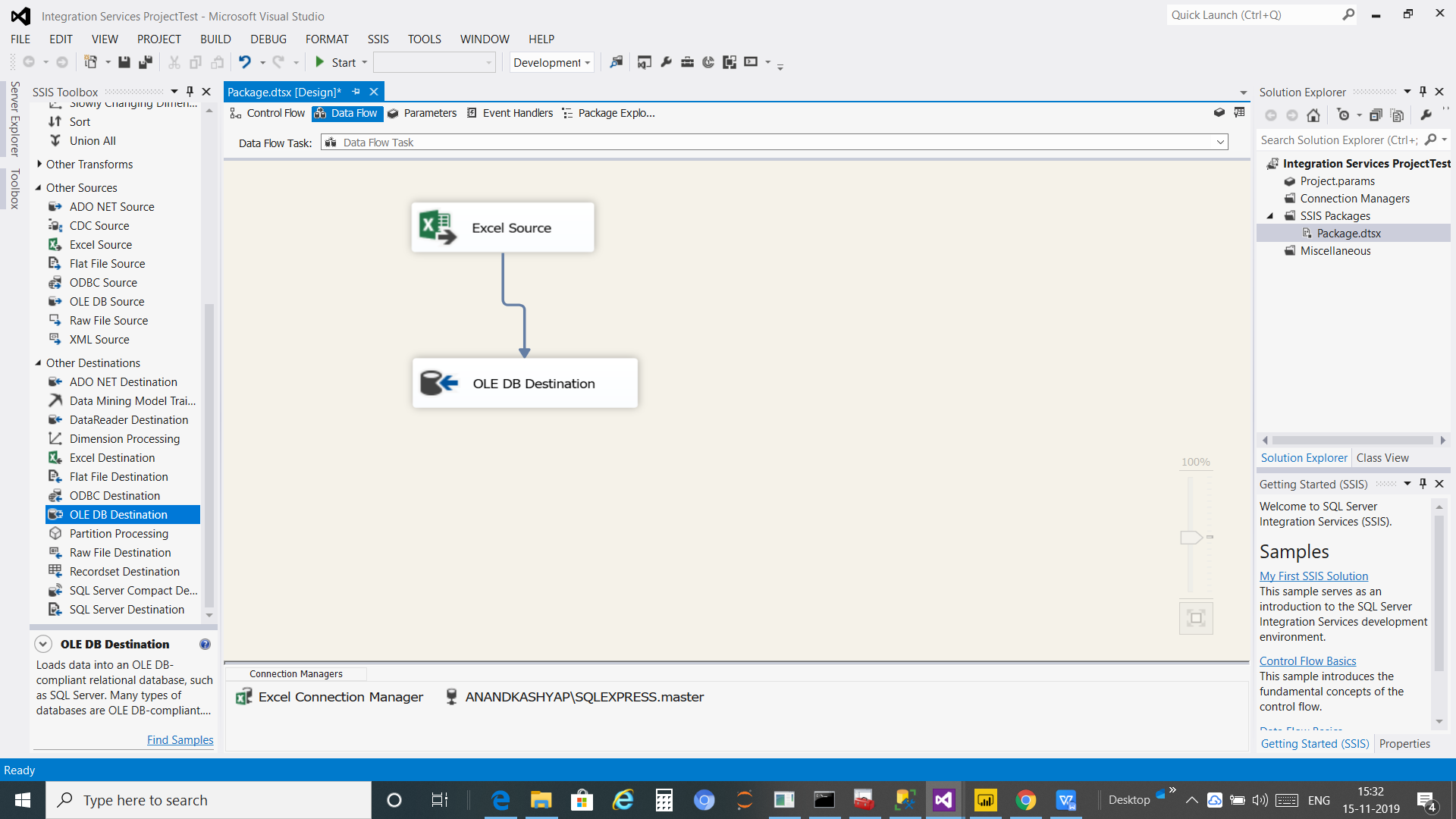
****

Also need to mapping the source data table to the destination data table

****

**Step 5**

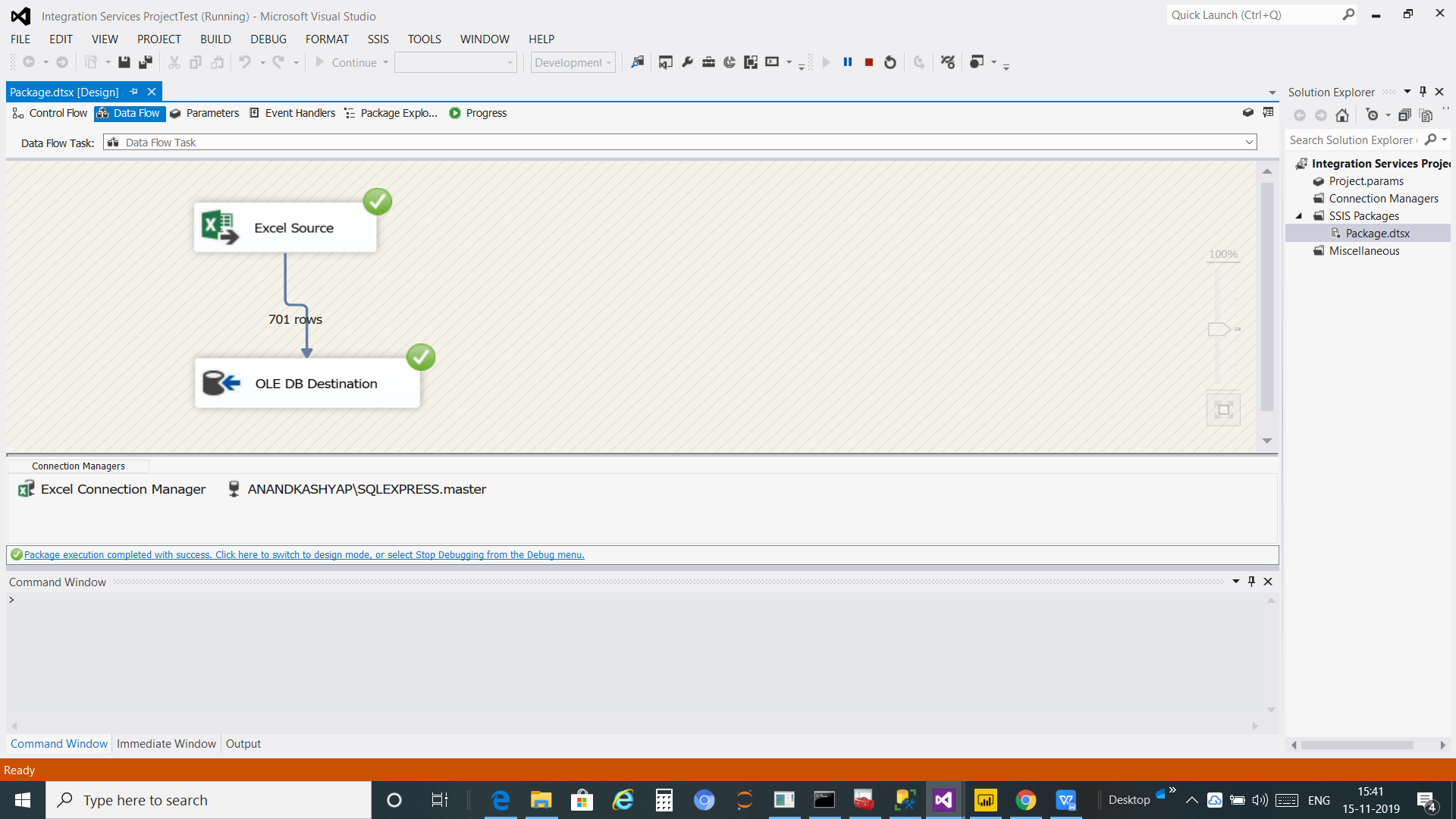
**Connect the Excel Source TO the OLE DB Destination**

****

**Step 6**

Start the SSIS Tools and then It will Extract ,transform and Load the data on destination data warehouse says OLE DB

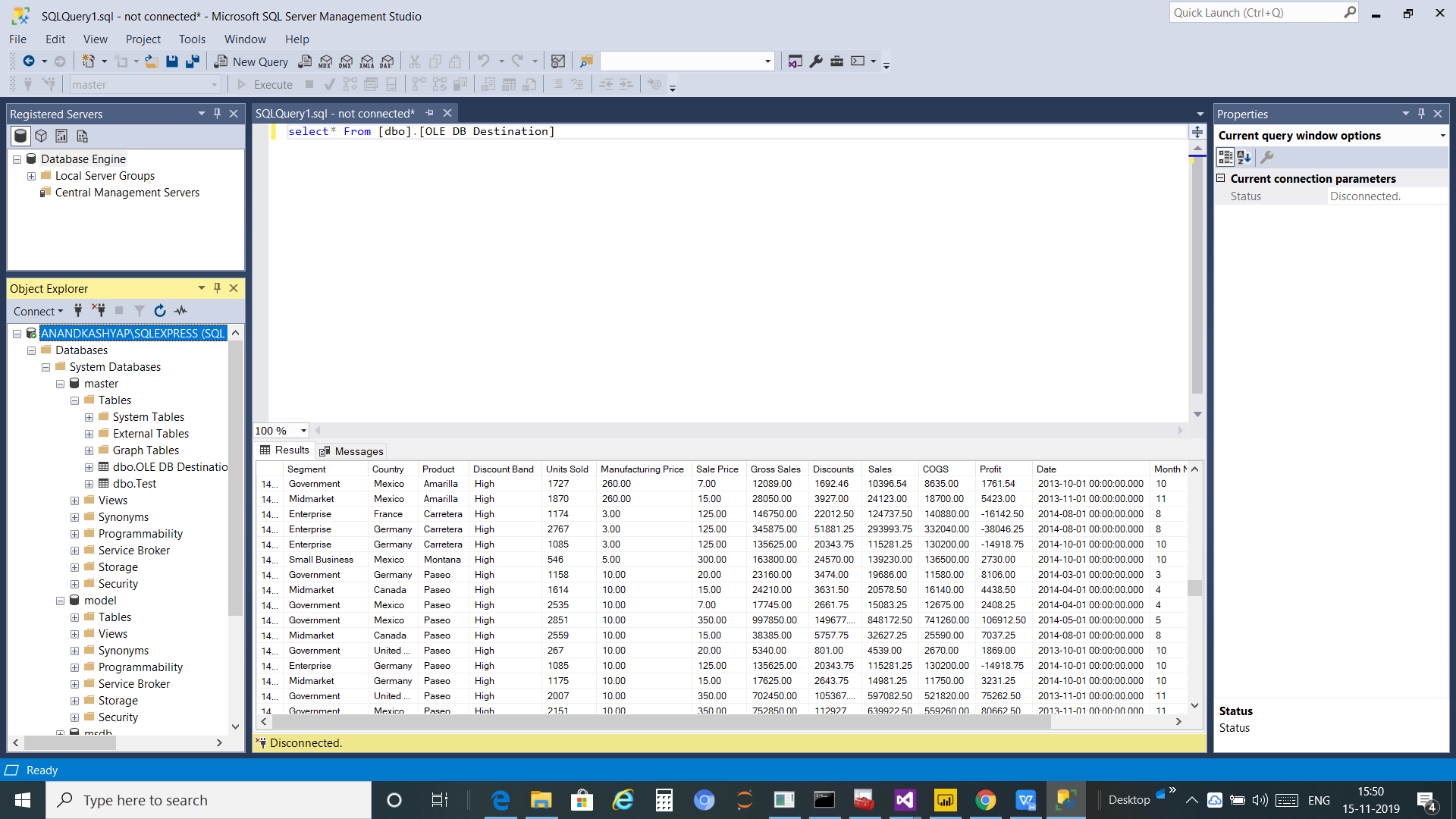
**2.)Output of ETL**

****

**3).SSMS(SQL server Management Studio):**

SQL Server is the tool which is used to modified the data in data warehouse(SQL server)

we can remove and modified ,implement on and many more on our data For business Intelligence. By use SQL query.

****

**4).Dashboard:**

It will need tools says PowerBI ,Power BI is a business Analytic service by Microsoft. It aims to provide interactive visualizations and business intelligence capabilities with an interface simple enough for end users to create their own reports and dashboards

Step-1

Get the data from data warehouse(OLE DB) in the powerBI

Step -2 connect to the sql server to get the data for interacive visualization and Business intelligence and create report and dashboards

**5).Outputs: Output of Dashboard**

