



—

ЕТЛ/ЕЛТ

ПРЕДРАГ ЈОВАНОВСКИ
DATA CONSULTANT





Агенда

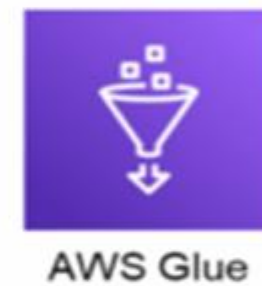
1. УВОД - ШТА ЈЕ И ЧЕМУ СЛУЖИ
2. ПРЕГЛЕД ЕТЛ АЛАТА
3. ЕТЛ VS ЕЛТ
4. MICROSOFT SQL SERVER INTEGRATION (MS SSIS) УПОЗНАВАЊЕ
5. КРЕИРАЊЕ ПРОЈЕКТА, КОНЕКЦИЈА И ЈЕДНОГ ПАКЕТА
6. ПРОМЕНЉИВЕ И ПАРАМЕТРИЗАЦИЈА
7. ОРКЕСТРАЦИЈА ПАКЕТА И ТЕСТИРАЊЕ ПАКЕТА
8. AZURE DATA FACTORY
9. SSIS VS AZURE DATA FACTORY



Увод – шта је и чему служи

- ▶ ЕТЛ се може користити за екстракцију (Е), трансформацију (Т) и учитавање (Л) података из више извора, као што су SQL Server база података, Oracle база података и Excel fajlovi. Користи процесе чишћења и спајања како би подаци били информативнији.
- ▶ Главни задатак ЕТЛ је миграција података из различитих извора на друга одредишта. Такође нуди широк спектар алата и решења. Најчешће употребе ЕТЛ -а укључују:
 1. Архивирање података
 2. Учитавање података (bulk-load data)
 3. Интеграција података из више различитих извора
 4. Чишћење података и трансформација података
 5. Оркестрација активности и пакета

Преглед ЕТЛ алата



ETL vs ELT

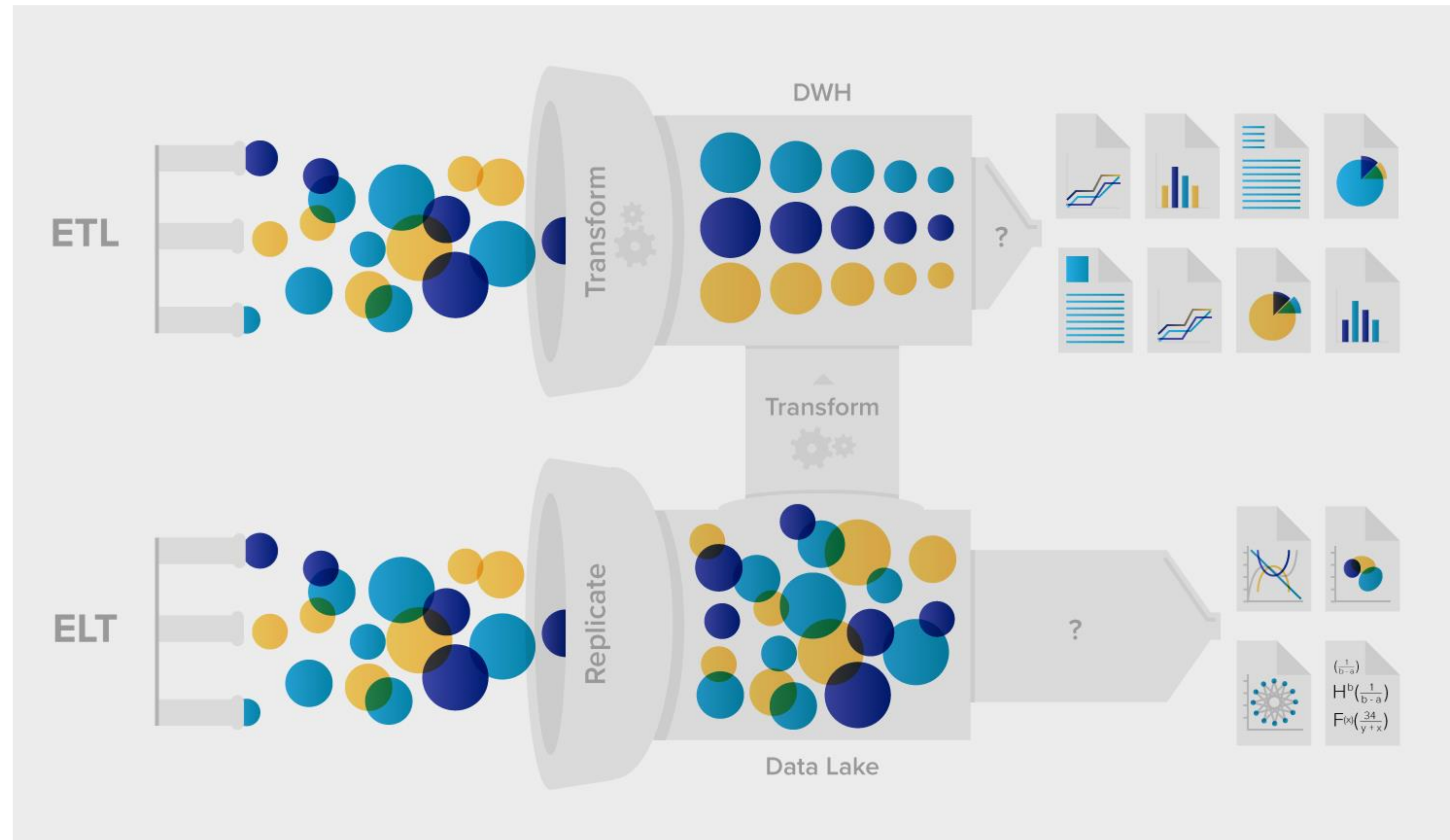
- **ETL УЧИТАВА ПОДАТКЕ ПРВО У STAGE БАЗУ, А ЗАТИМ У ЦИЉНИ СИСТЕМ, ДОК ELT УЧИТАВА ПОДАТКЕ ДИРЕКТНО У ЦИЉНИ СИСТЕМ.**
- **ETL МОДЕЛ СЕ КОРИСТИ ЗА ЛОКАЛНЕ, РЕЛАЦИОНЕ И СТРУКТУРИРАНЕ ПОДАТКЕ, ДОК СЕ ELT КОРИСТИ ЗА СТРУКТУРИРАНЕ И НЕСТРУКТУИРАНЕ ИЗВОРЕ ПОДАТАКА.**
- **УПОРЕЂУЈУЋИ ELT СА ETL, ETL СЕ УГЛАВНОМ КОРИСТИ ЗА МАЛУ КОЛИЧИНУ ПОДАТАКА, ДОК СЕ ELT КОРИСТИ ЗА ВЕЛИКЕ КОЛИЧИНЕ ПОДАТАКА.**

ETL vs EЛТ

Предности ЕЛТ у односу на ЕТЛ

1. **Више флексибилности** - јер је ЕТЛ традиционално намењен за релационе, структуриране податке. Складишта података заснована на облаку омогућавају ЕЛТ за структуриране и неструктуриране податке
2. **Већа приступачност** - пошто ЕТЛ генерално подржавају, одржавају и управљају ИТ одељења организације. ЕЛТ омогућава лакши приступ и коришћење запосленима
3. **Скалабилност** - јер ЕТЛ може бити претерано интензиван ресурсима за нека предузећа. ЕЛТ решења су углавном Саас заснована на облаку, доступна ширем спектру предузећа
4. **Брже време учитавања** - јер ЕТЛ обично траје дуже јер користи Stage area. Са ЕЛТ-ом, постоји само једно оптерећење до одредишног система
5. **Брже време трансформације** - пошто је ЕТЛ обично спорији и зависи од величине скупова података. ЕЛТ трансформација не зависи од величине података
6. **Мање времена потребно за одржавање података** - јер ће подаци можда морати да се пронађу и поново прочитају ако се утврди да трансформација није адекватна за предвиђене сврхе података. Са ЕЛТ-ом, оригинални подаци су нетакнути и већ су учитани ако је потребна додатна трансформација

ETL vs ELT



И ЕТЛ и ЕЛТ омогућавају предузећима да консолидују податке из више база података и других извора у једно спремиште са подацима који су правилно форматирани и квалификовани. Ово обједињено складиште података пружа поједностављен приступ за анализу и додатну обраду. Такође пружа јединствен извор истине, осигуравајући да су сви подаци предузећа доследни и ажурни.

Microsoft SQL Server Integration Services (MS SSIS) - упознавање

Предуслови за подешавања локалног окружења:

1. SQL Server Development Edition
 - <https://go.microsoft.com/fwlink/p/?linkid=2215158&clcid=0x409&culture=en-us&country=us>
2. SQL Server Management Studio
 - <https://go.microsoft.com/fwlink/p/?linkid=2206714>
3. Azure Data Studio
 - <https://go.microsoft.com/fwlink/p/?linkid=2206560>
4. SSDT for Visual Studio
 - <https://learn.microsoft.com/en-us/sql/ssdt/download-sql-server-data-tools-ssdt?view=sql-server-ver16>
5. Wide World Importers (WWI)
 - <https://learn.microsoft.com/en-us/sql/samples/wide-world-importers-what-is?view=sql-server-ver16>
 - <https://github.com/microsoft/sql-server-samples/tree/master/samples/databases/wide-world-importers>

Пример – Case Study

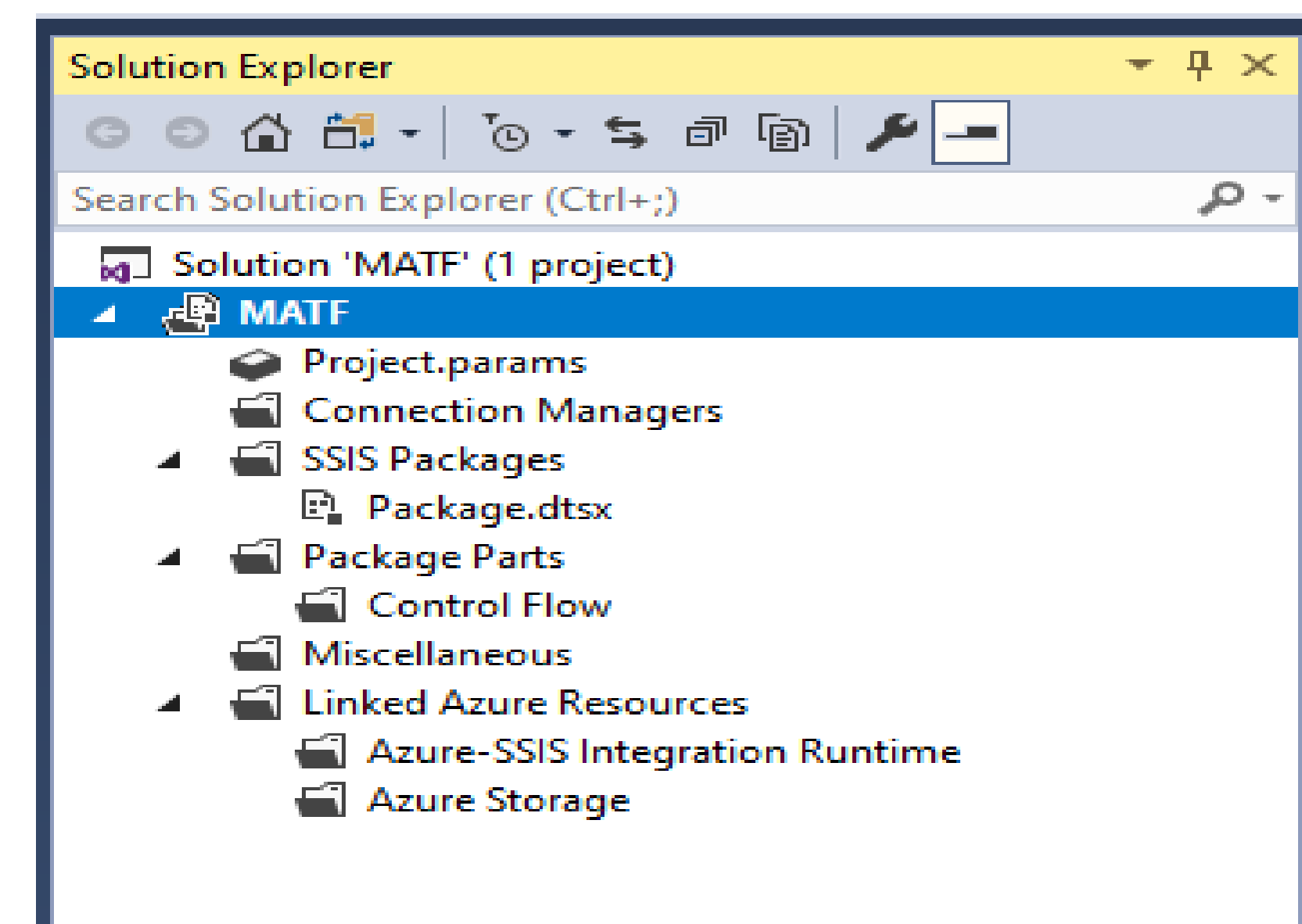
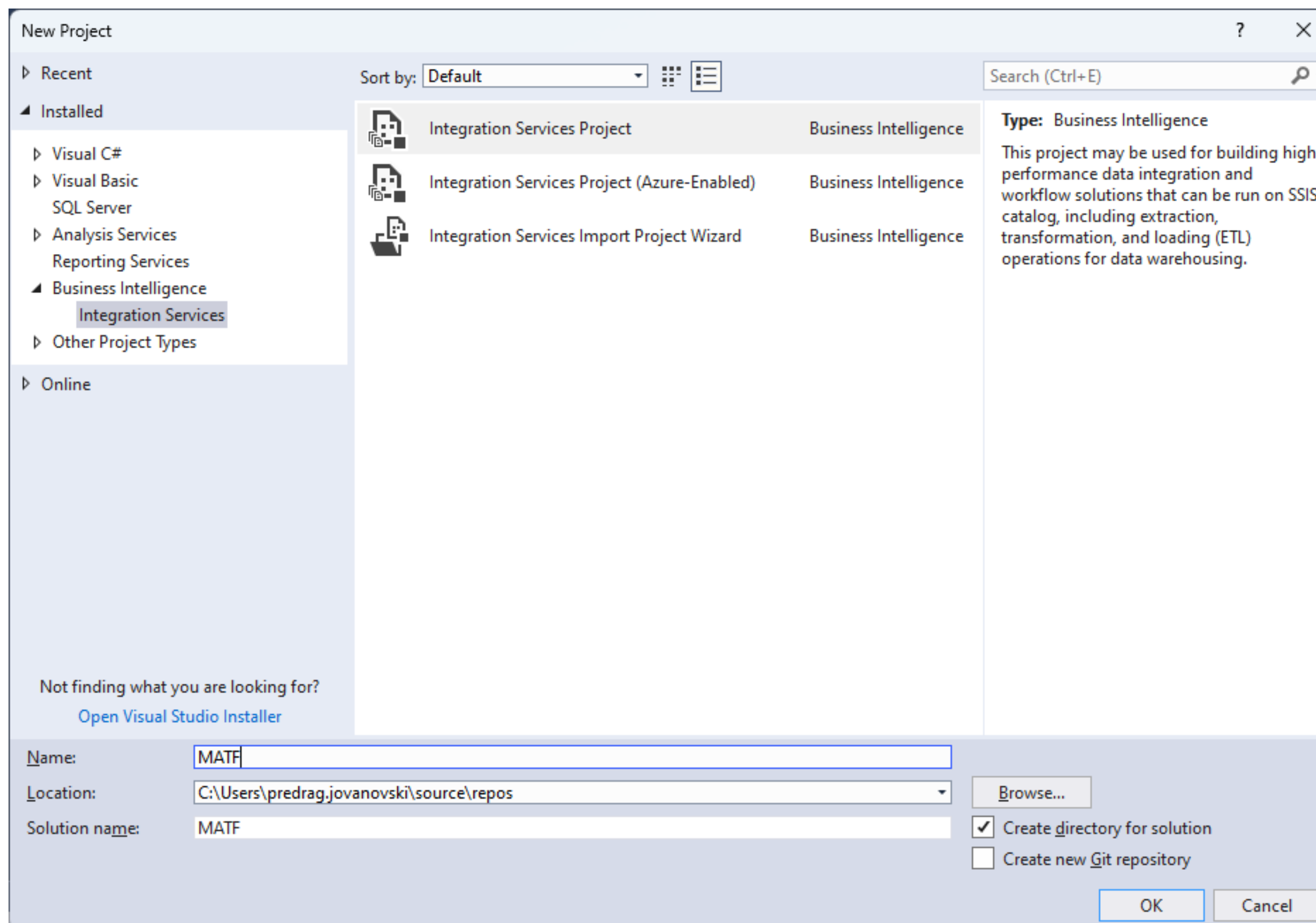
Изворни системи

- Базе података
- API
- Текстуални фајлови – CSV, Excel, Parquet
- Сlike, аудио, видео фајлови
- IoT подаци

Изворни систем	Број Табела/Фајлова
База 1	15
База 2	20
База 3	50
Фајлови	10

Креирање пројекта, конекција и једног пакета

Отворити алат Microsoft SSDT for Visual Studio и направити нови пројекат

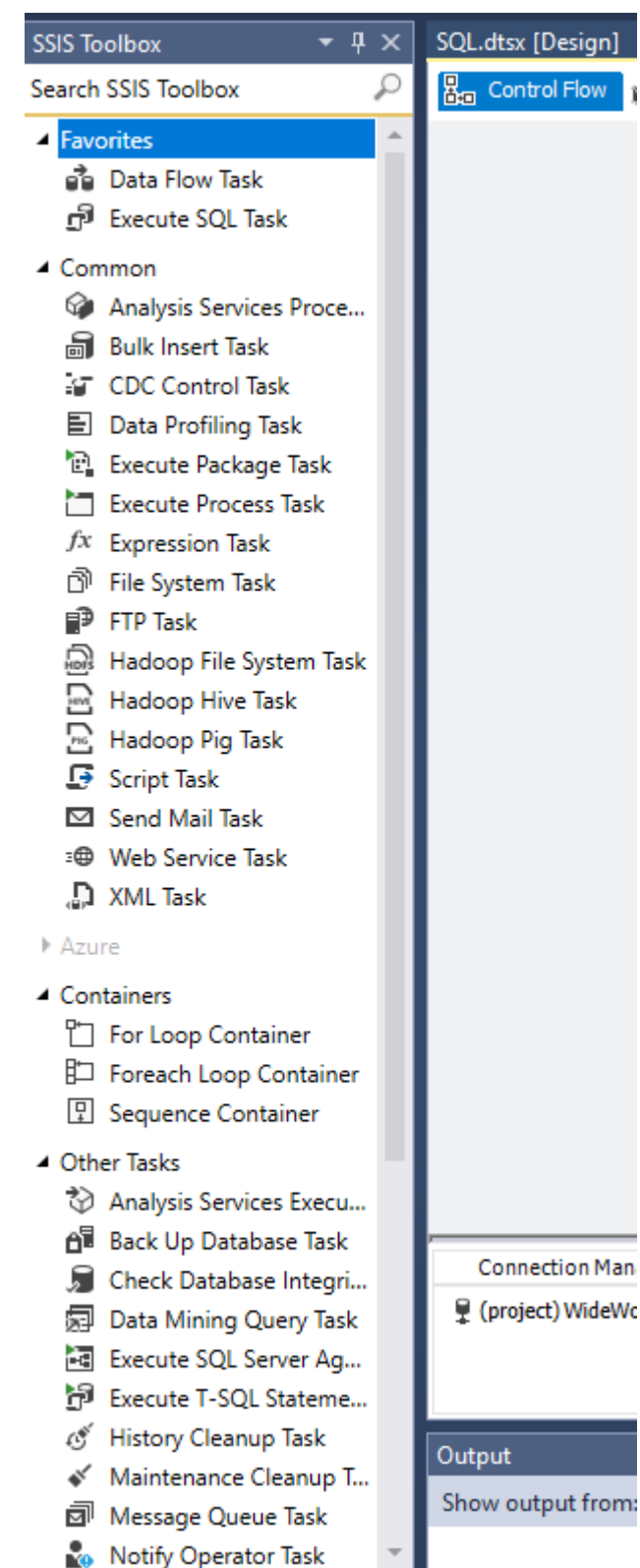


Преглед палета компоненти

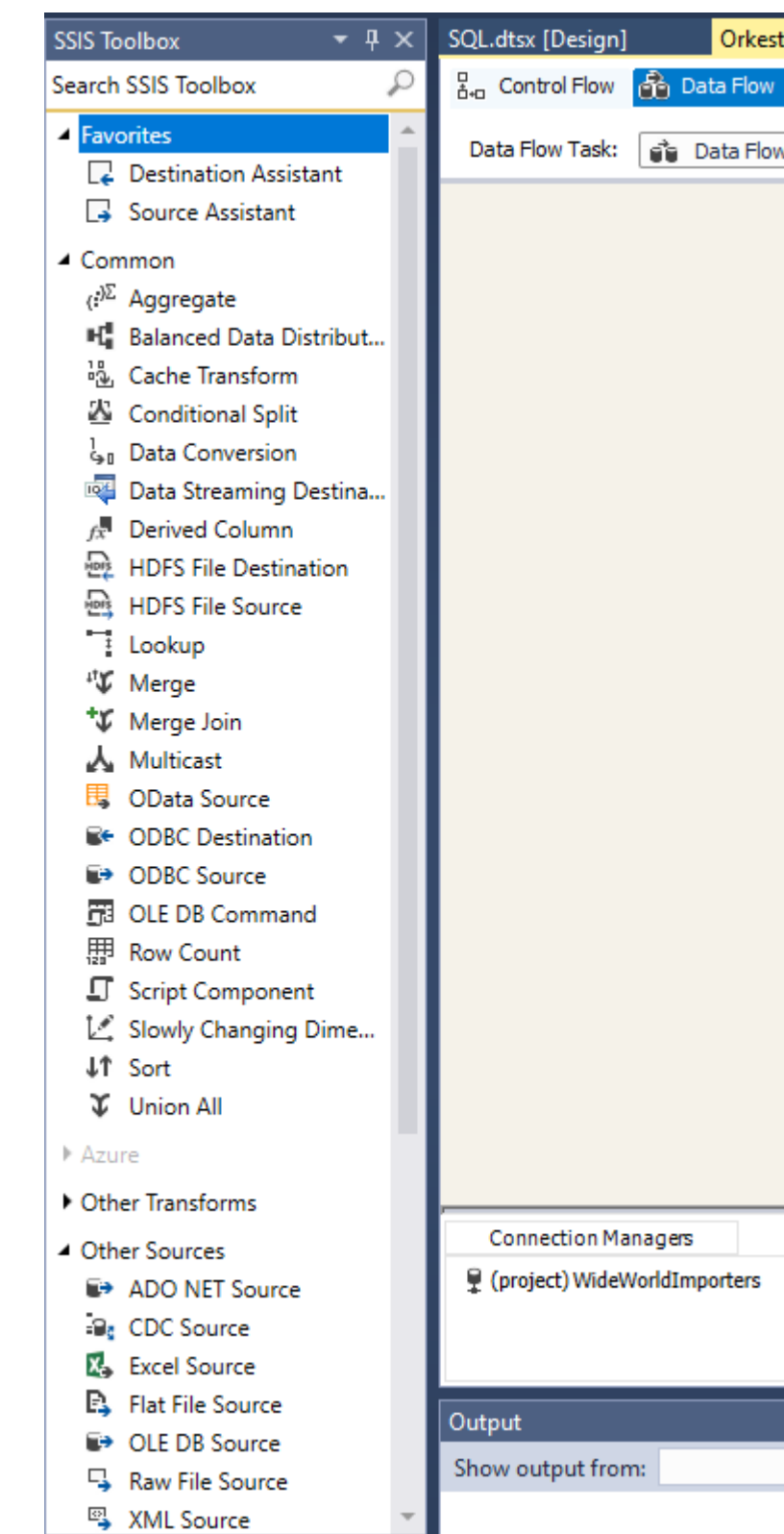
Два типа палета алата:

1. Control Flow и
2. Data Flow

Control Flow



Data Flow



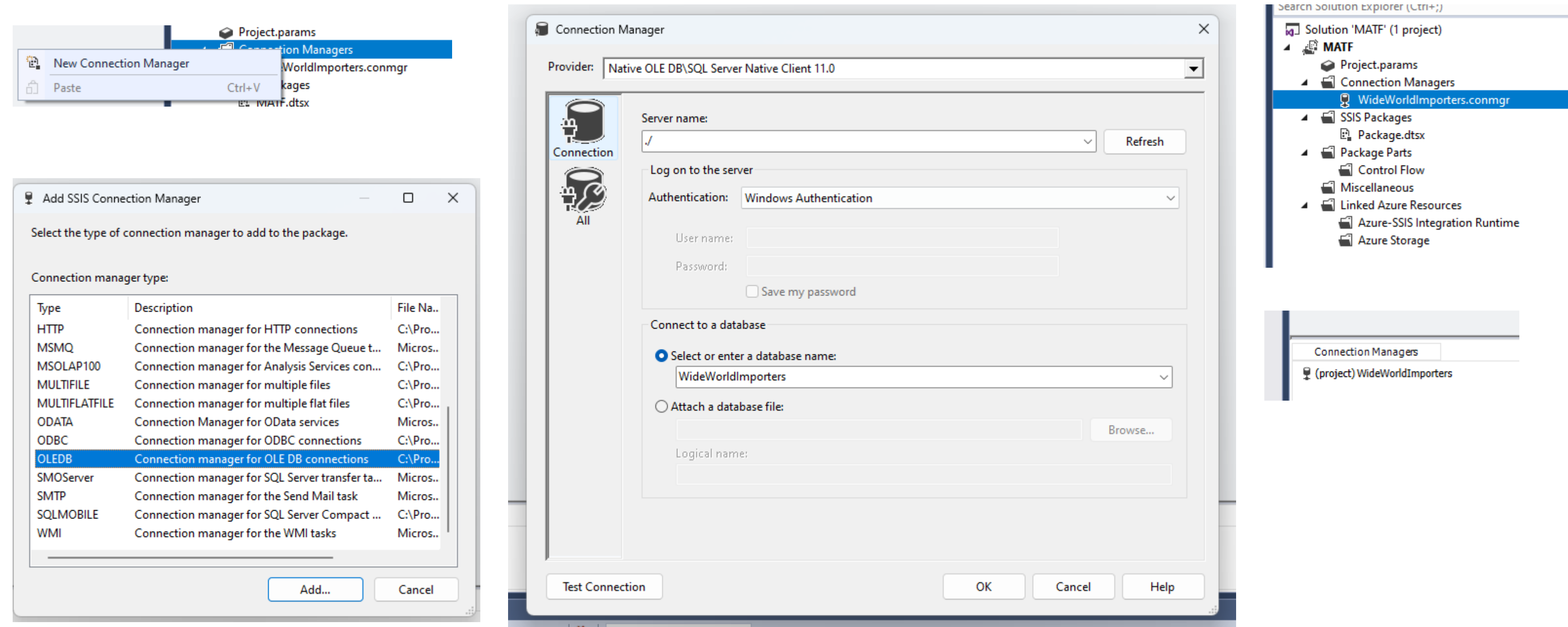
Прављење пакета и конекције према бази

Кораци за конекцију:

1. New Connection Manager
2. OLDB тип
3. Подешавање конекције
4. Преглед направљене конекцији у Solution Explorer-у

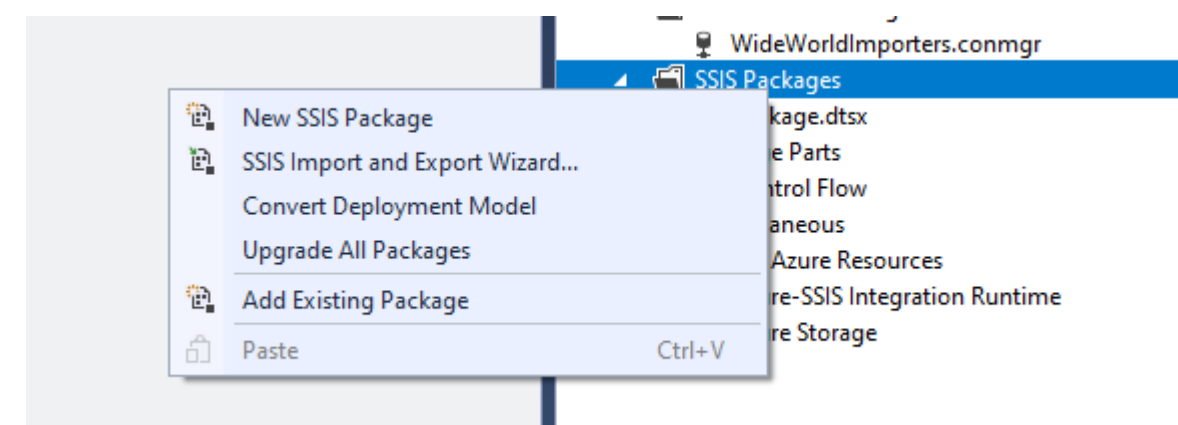
Тип конекције:

1. Project connection
2. Package connection

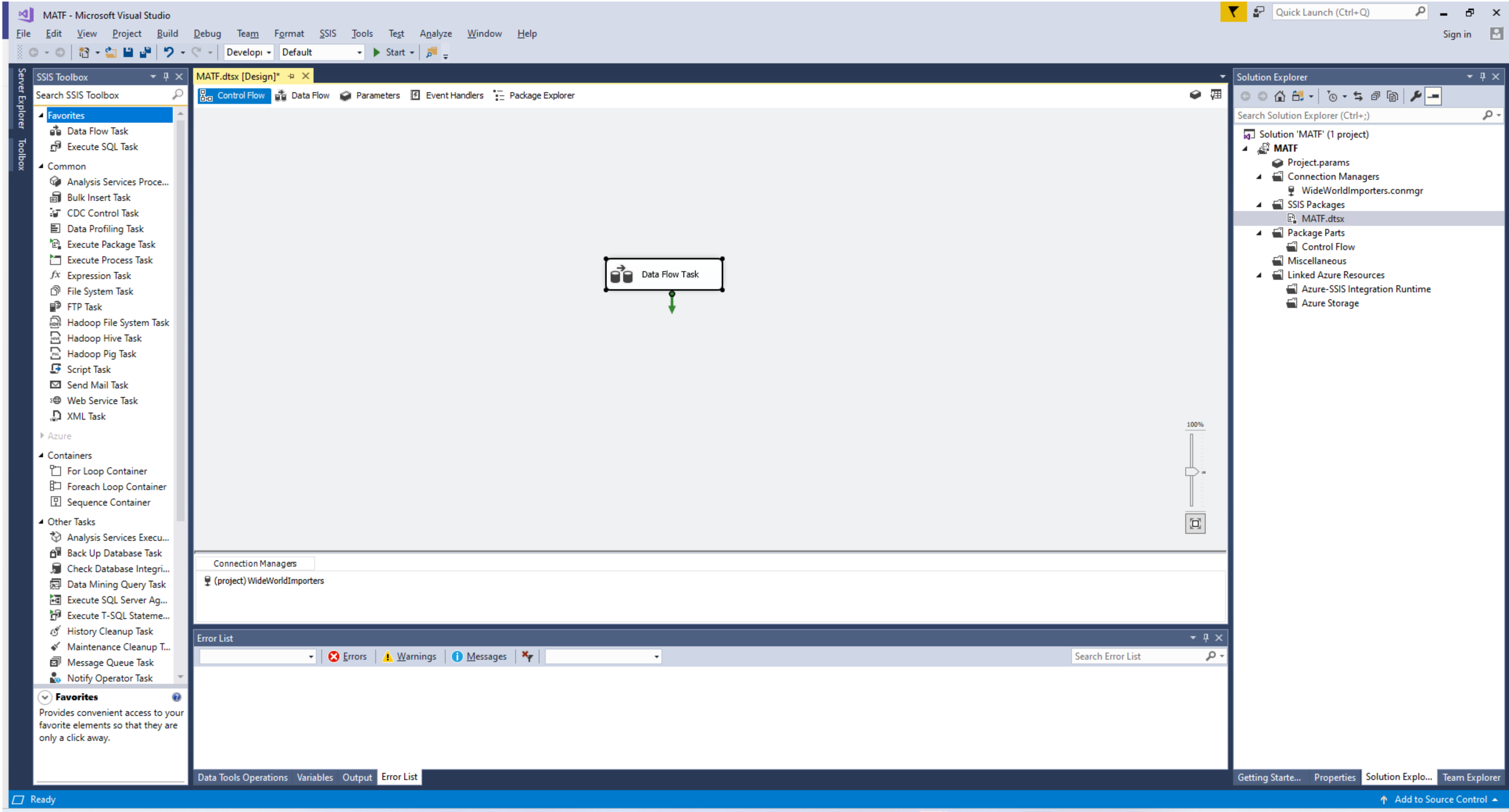


Кораци за пакет:

1. Десни клик на SSIS Package
2. New SSIS Package
3. Променити назив пакет
4. Преглед пакета у Solution Explorer-у



Изглед целог окружења

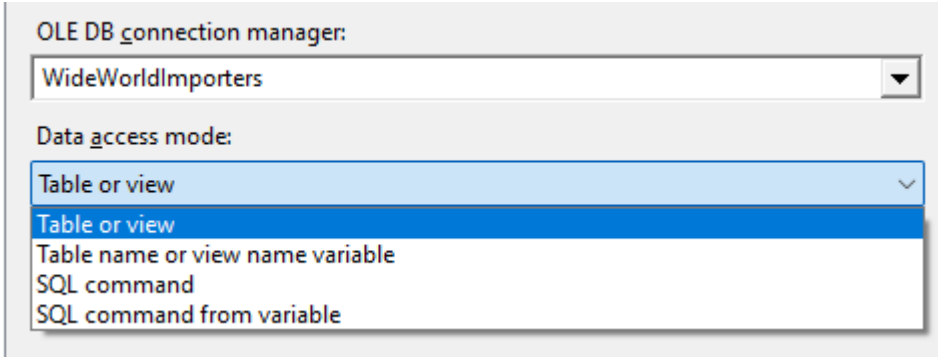


Подешавање Data Source

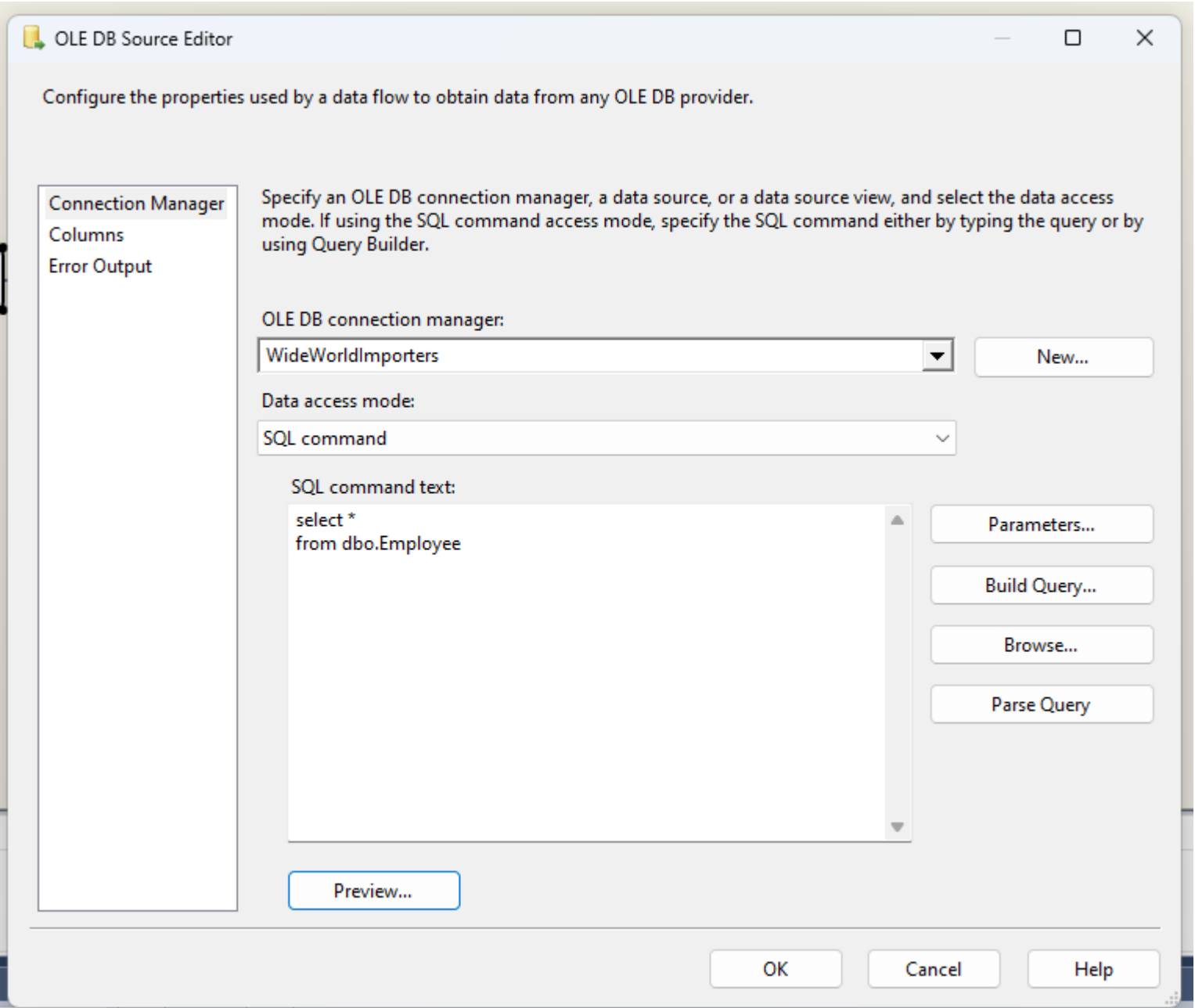
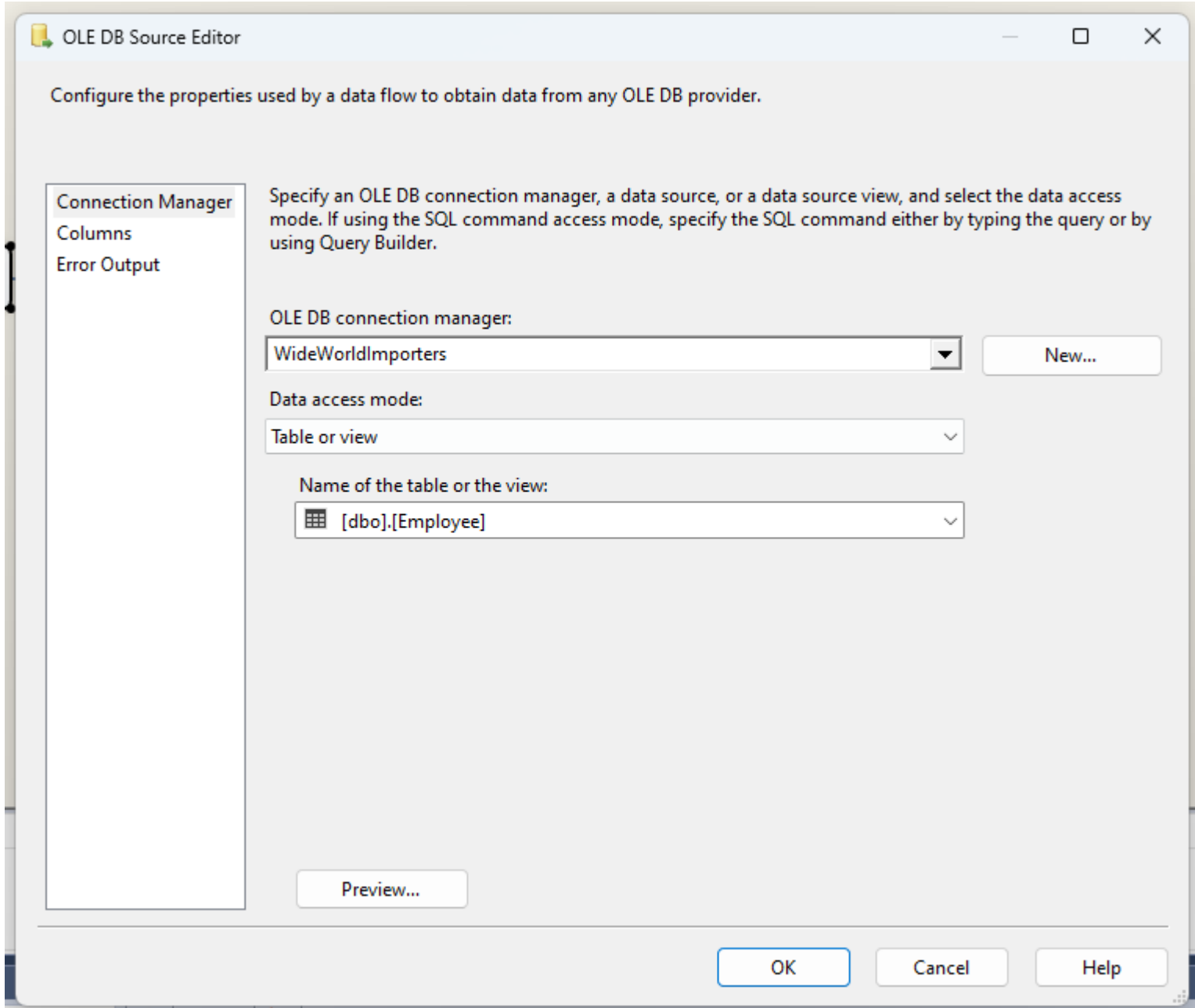
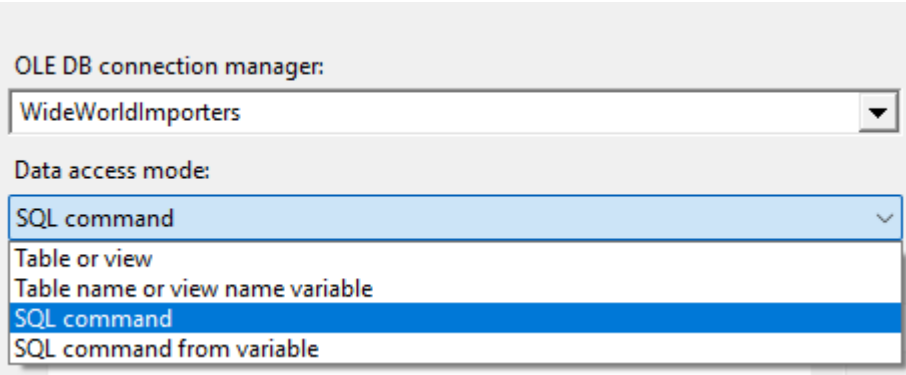
Типови приступа бази:

- 1. Table or View
- 2. SQL Command
- 3. Table name or view name variable
- 4. SQL command from variable

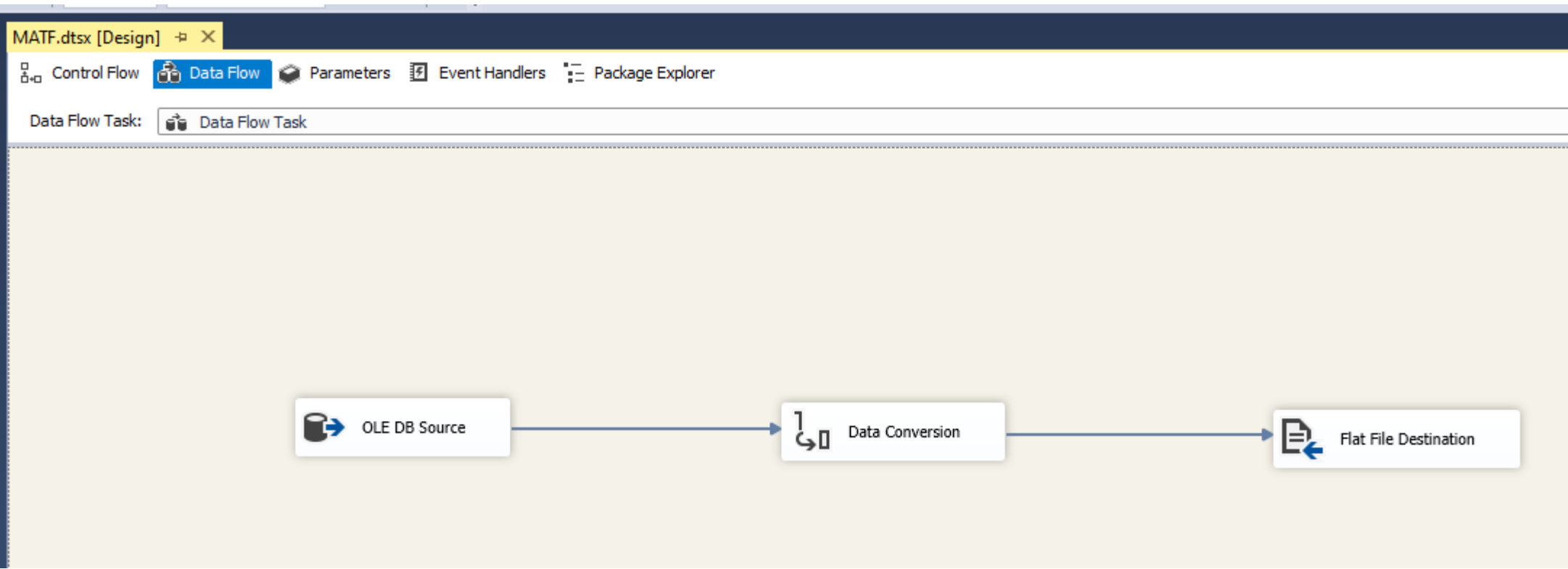
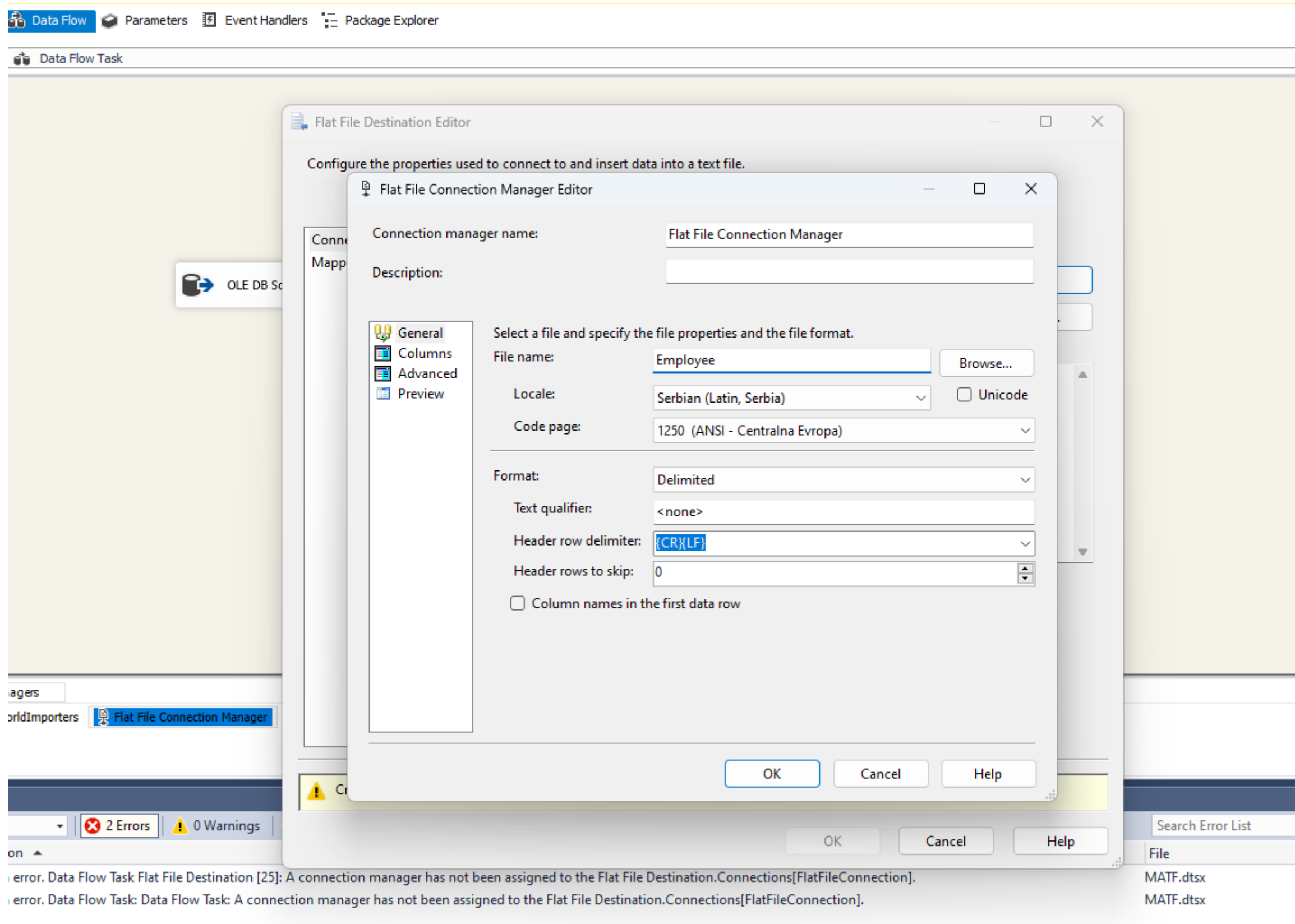
Table or view



SQL command



Прављење конекције према фајлу и додавање Data Conversion трансформације



Data Conversion - Unicode

Променљиве (Variable) и Параметри

Променљива је ентитет који садржи неки податак и који се може мењати током извршавања самог пакета. Дефинисана променљива је видљива само у том пакету у којем је дефинисана. Може се користити више пута.

Параметар је ентитет сличан променљивој али са битном разликом. Вредност параметра се додељује пре извршавања пакета и не може да се мења и тај параметар је видљив свим пакетима у оквиру пројекта.

The screenshot shows the SAP Studio interface with the 'Parameters' and 'Variables' panels. The 'Parameters' panel displays a table with one parameter, and the 'Variables' panel displays a table with one variable.

Name	Data type	Value
Parameter1	Int32	0

Name	Scope	Data type	Value
Variable1	MATF	Int32	0

Примена променљивих и параметара

Table name or view name variable

SQL command from variable

OLE DB connection manager:
WideWorldImporters

Data access mode:
Table name or view name variable

Table or view
Table name or view name variable
SQL command
SQL command from variable

OLE DB connection manager:
WideWorldImporters

Data access mode:
SQL command from variable

Table or view
Table name or view name variable
SQL command
SQL command from variable

OLE DB Source Editor

Configure the properties used by a data flow to obtain data from any OLE DB provider.

Connection Manager
Columns
Error Output

Specify an OLE DB connection manager, a data source, or a data source view, and select the data access mode. If using the SQL command access mode, specify the SQL command either by typing the query or by using Query Builder.

OLE DB connection manager:
WideWorldImporters

Data access mode:
Table name or view name variable

Variable name:
User::NazivTabele

Preview...

OK Cancel Help

Name	Scope	Data type	Value
NazivTabele	MATF	String	[dbo].[Employee]

Expression Builder

Specify the expression:

Variables and Parameters
System Variables
\$Package::NazivFajla
User::NazivTabele
User::SQL

Mathematical Functions
String Functions
Date/Time Functions
NULL Functions
Type Casts
Operators

Description:

Expression:
"SELECT * from " + @[User::NazivTabele]

Evaluated value:
SELECT * from [dbo].[Employee]

Evaluate Expression

OK Cancel

Name	Scope	Data type	Value
NazivTabele	MATF	String	[dbo].[Employee]
SQL	MATF	String	SELECT * from [dbo].[Employee]

OLE DB Source Editor

Configure the properties used by a data flow to obtain data from any OLE DB provider.

Connection Manager
Columns
Error Output

Specify an OLE DB connection manager, a data source, or a data source view, and select the data access mode. If using the SQL command access mode, specify the SQL command either by typing the query or by using Query Builder.

OLE DB connection manager:
WideWorldImporters

Data access mode:
Table name or view name variable

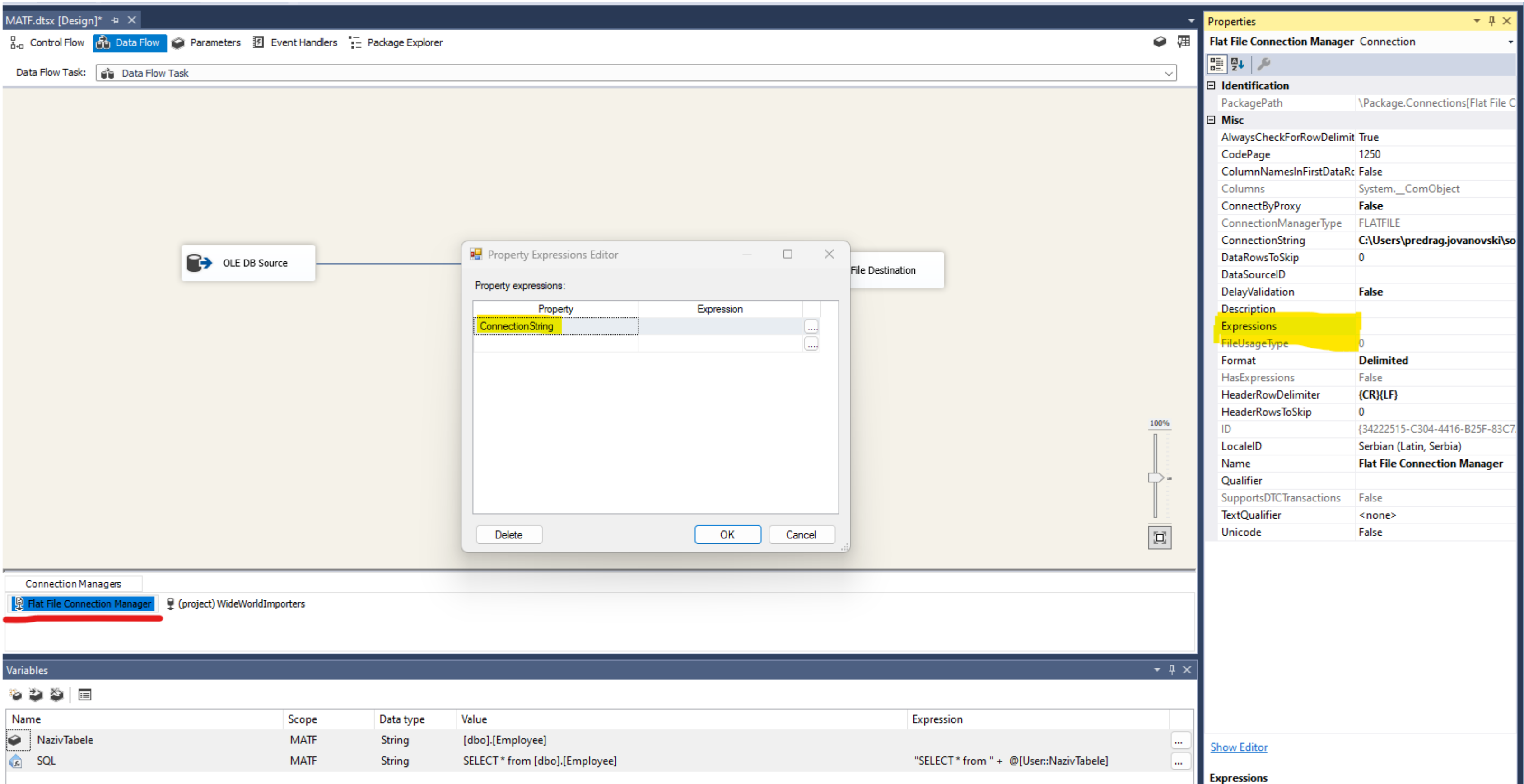
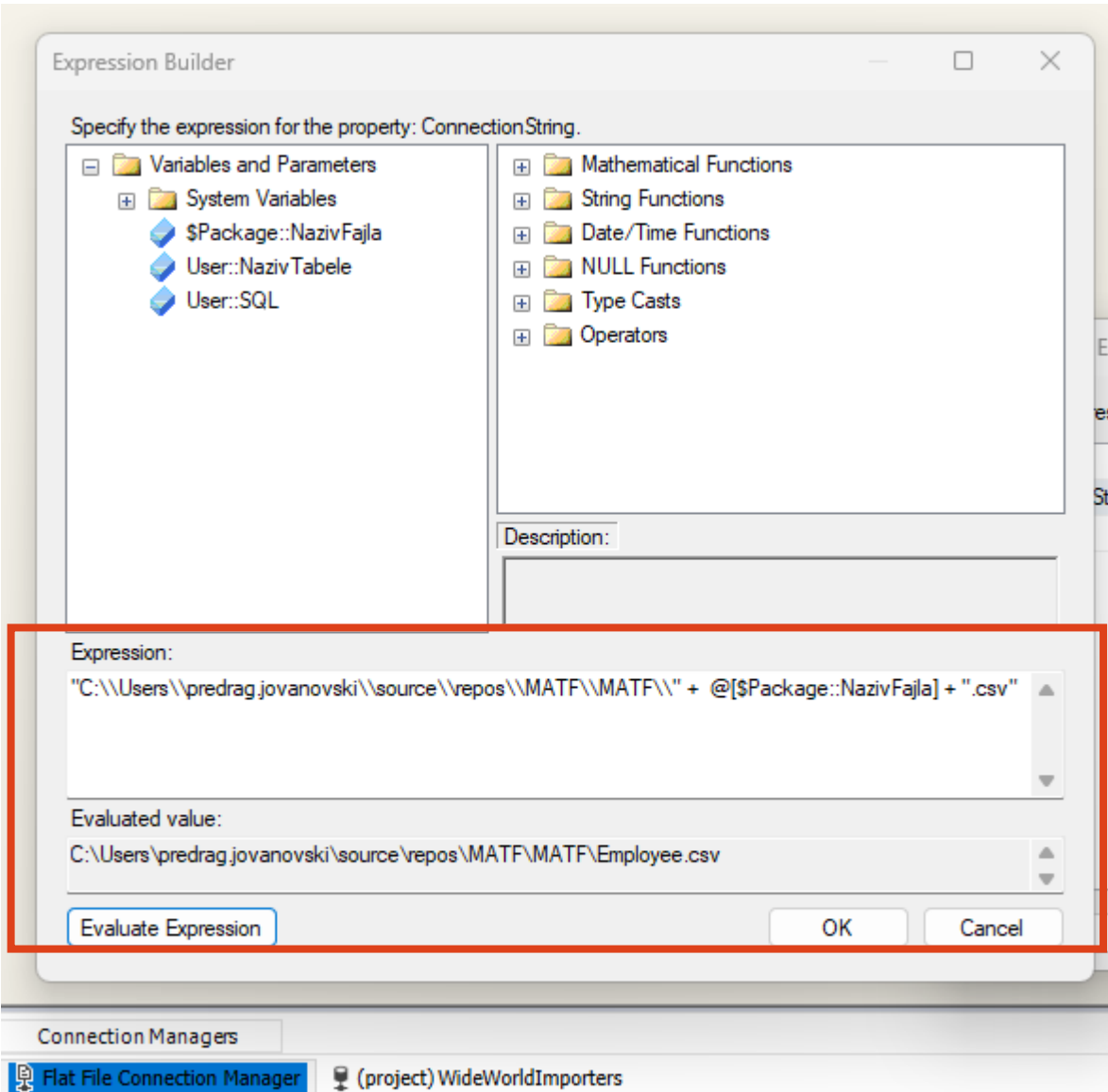
Variable name:
User::SQL

Preview...

OK Cancel Help

Name	Scope	Data type	Value
NazivTabele	MATF	String	[dbo].[Employee]
SQL	MATF	String	SELECT * from [dbo].[Employee]

Примена променљиве и параметра у конекцији за фајл



Оркестрација пакета и тестирање пакета

Solution 'MATF' (1 project)

- MATF
 - Project.params
 - Connection Managers
 - WideWorldImporters.conmgr
 - SSIS Packages
 - SQL.dtsx
 - MATF.dtsx
 - Orkestracija.dtsx
 - Package Parts
 - Control Flow
 - Miscellaneous
 - Linked Azure Resources
 - Azure-SSIS Integration Runtime
 - Azure Storage

SQL.dtsx [Design] Orkestracija.dtsx [Design] MATF.dtsx [Design]

Control Flow Data Flow Parameters Event Handlers Package Explorer

Execute SQL Task

Execute SQL Task Editor

Configure the properties required to run SQL statements and stored procedures using the selected connection.

General

Parameter Mapping

Result Set

Expressions

General

Options

Result Set

SQL Statement

Name

Specifies the name of the task.

Browse...

Build Query...

Parse Query

OK

Cancel

Help

SQL.dtsx [Design] Orkestracija.dtsx [Design] MATF.dtsx [Design]

Control Flow Data Flow Parameters Event Handlers Package Explorer

SQL

MATF

Направљена су три пакета:

- 1. MATF
- 2. SQL
- 3. Orkestracija

Search Solution Explorer (Ctrl+;)

Solution 'MATF' (1 project)

- MATF
 - Project.params
 - Connection Managers
 - WideWorldImporters.conmgr
 - SSIS Packages
 - SQL.dtsx
 - MATF.dtsx
 - Orkestracija.dtsx
 - Package Parts
 - Control Flow
 - Miscellaneous
 - Linked Azure Resources
 - Azure-SSIS Integration Runtime
 - Azure Storage

Deploy

Build

Rebuild

SSIS in Azure Data Factory

Scope to This

New Solution Explorer View

Add

Manage NuGet Packages...

Set as StartUp Project

Debug

Convert to Package Deployment Model

Cut

Paste

Remove

Rename

Unload Project

Properties

Azure Data Factory

- Azure Data Factory је Azure ЕТЛ сервис у облаку за интеграцију података без сервера и трансформацију података. Такође, можете подићи и пребацити постојеће SSIS пакете на Azure и покренути их са потпуном компатибилношћу у ADF-у. SSIS Integration Runtime нуди услугу којом се у потпуности управља, тако да не морате да бринете о управљању инфраструктуром.



SSIS vs Azure Data Factory

SSIS	Azure Data Factory
Package	Pipeline
Task	Activity
Connection Manger	Linked Services
Source/Destination	Source/Sink
SQL Agent Job	Trigger
Data Flow	Data Flow
Variable/Parameter	Variable/Parameter

Azure ресурси

- 1. SQL Server
- 2. SQL Database
- 3. Storage Account
- 4. Azure Data Factory

Azure Data Factory Linked Services

»

Data Factory

Validate all

Publish all

Home

Factory settings

Connections

Linked services

Integration runtimes

Microsoft Purview

Source control

Git configuration

Linked services

Linked service defines the connection information to a data store or compute. [Learn more](#)

+ New

Filter by name

Annotations : Any

Showing 1 - 2 of 2 items

Name	Type
ls_abs_matf	Azure Blob Storage
ls_adb_matf	Azure SQL Database

Edit linked service

Azure Blob Storage [Learn more](#)

Name *

ls_abs_matf

Description

Connect via integration runtime *

AutoResolveIntegrationRuntime

Authentication type

Account key

Connection string

Azure Key Vault

Account selection method

From Azure subscription

Enter manually

Storage account name *

sgmatf

Storage account key

Azure Key Vault

Storage account key *

.....

Partioned DNS enabled

Endpoint suffix

core.windows.net

Additional connection properties

+ New

Test connection

To linked service

To file path

Annotations

Apply

Cancel

Test connection

Edit linked service

Azure SQL Database [Learn more](#)

Name *

ls_adb_matf

Description

Connect via integration runtime *

AutoResolveIntegrationRuntime

Connection string

Azure Key Vault

Account selection method

From Azure subscription

Enter manually

Fully qualified domain name *

srvmatf.database.windows.net

Database name *

matf

Authentication type *

SQL authentication

User name *

matfadmin

Password

Azure Key Vault

Password *

.....

Always encrypted

Additional connection properties

+ New

Annotations

Azure Data Factory Data Sets

Data Factory

Validate all

Publish all

Factory Resources

Filter resources by name

Pipelines

Change Data Capture (preview)

Datasets

Data flows

Power Query

0

0

2

0

0

ds_adb_matf

ds_csv_matf

ds_adb_matf

ds_csv_matf

SQL

Azure SQL Database

ds_adb_matf

Connection

Schema

Parameters

Linked service *

Is_adb_matf

Test connection

Edit

New

Learn more

Table

dbo.Employee

Refresh

Preview data

Edit

ds_adb_matf

ds_csv_matf

DelimitedText

ds_csv_matf

Connection

Schema

Parameters

Linked service *

Is_abs_matf

Test connection

Edit

New

Learn more

File path *

matf

Directory

@dataset().pFileName

Browse

Preview data

Detect format

Compression type

Select...

Column delimiter ①

Comma (,)

Row delimiter ①

Default (\r,\n, or \r\n)

Encoding ①

Default(UTF-8)

Quote character ①

Double quote (")

Escape character ①

Backslash (\)

First row as header ①

☒

Null value ①

Azure Data Factory Pipeline

Data Factory

Validate all

Factory Resources

Filter resources by name

Pipelines3

pOrkestracija

pIMATF

pISQL

Change Data Capture (preview)0

Datasets...

Data flows0

Power Query0

MATF Pipeline

Copy data

cdMATF

SQL Pipeline

Script

scrSQL

pIMATF

pISQL

pOrkestracija

Activities

Search activities

Move and transform

Synapse

Azure Data Explorer

Azure Function

Batch Service

Databricks

Data Lake Analytics

General

HDInsight

Iteration & conditionals

Machine Learning

Power Query

Validate

Debug

Add trigger

Execute Pipeline

exeMATF
pIMATF

Execute Pipeline

exeSQL
pISQL

Parameters

Variables

Settings

Output

New



Предраг Јовановски

DATA CONSULTANT

SOFTWARE ENGINEER WITH SEVEN YEARS OF EXPERIENCE WORKING AS A DWH/BI CONSULTANT, BI DEVELOPER, TEAM LEAD OF BI TEAM. GOAL ORIENTED, INTERESTED IN DATA MANAGEMENT, DATA METHODOLOGY, DATA ANALYTICS. FAST SOLVING PROBLEMS AND DEVELOPING PLANS FOR IMPLEMENTATION AND OPERATION OF CLIENT SERVICES, PRODUCT SUPPORT AND QUALITY ASSURANCE. DETAIL ORIENTED TEAM PLAYER FOLLOW EVERY SITUATION IN TEAM AND INNOVATE MANY SOLUTIONS FOR BUSINESS NEEDS. DURING HIS CARRIER HE MOVED QUICKLY AND EFFORTLESSLY ACROSS VARIOUS TECHNOLOGIES AND DIFFERENT VENDORS. HIS EXPERIENCE IS VERY MUCH HANDS ON AND HE INVESTS A LOT OF EFFORT TO KEEP IT UP TO DATE WITH LATEST TECHNOLOGIES AND DATA SCIENCE PRACTICE.

predrag.jovanovski@endava.com

