

1

SQL SERVER BUSINESS INTELLIGENCE (BI) - INTRODUCTION

What is BI?

Microsoft SQL Server 2008 provides a scalable Business Intelligence platform optimized for data integration, reporting, and analysis, enabling organizations to deliver intelligence where users want it.

Business intelligence is a method of storing and presenting key enterprise data so that anyone in your company can quickly and easily ask questions of accurate and timely data.

Key Components

- Microsoft has provided a comprehensive Business Intelligence platform for us
 - SQL Server Integration Service (SSIS)
 - SQL Server Reporting Service (SSRS)
 - SQL Server Analysis Service (SSAS)

This presentation provides introductory information on SQL business intelligence components.

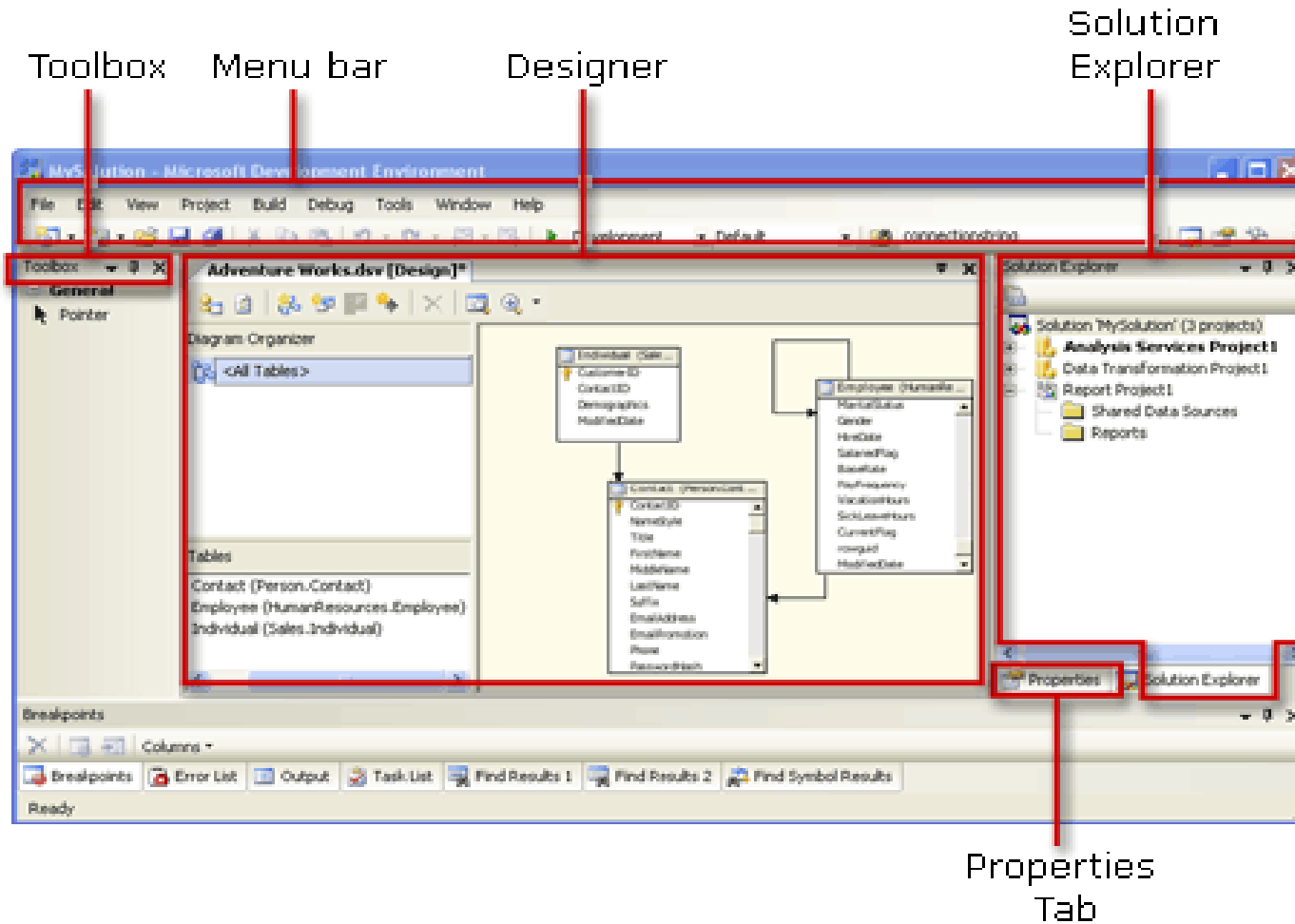
What is BIDS?

Before understanding SSAS, SSIS, SSRS we need to know what is BIDS - SQL Server Business Intelligence Development Studio.

BIDS is Microsoft Visual Studio 2008 with additional project types that are specific to SQL Server Business Intelligence.

- It's a Development Platform.
- It's a Deployment Platform.

BIDS Preview



SQL SERVER INTEGRATION SERVICES (SSIS)

SSIS is replacement for DTS (Data Transformation Service) in SQL Server 2005 and onwards. It is used to developed enterprise-level data transformations solutions such as...

- Data Migration from Legacy system to new system
- Copying and Downloading files from SQL Server
- Updating Data warehouse
- Bulk data import from one system (e.g. Excel / Flat file/ Oracle / OLEDB source) to SQL Server
- There are also some administrative task can be performed e.g. Data profiling.

Supported Database Service

Can be found at:

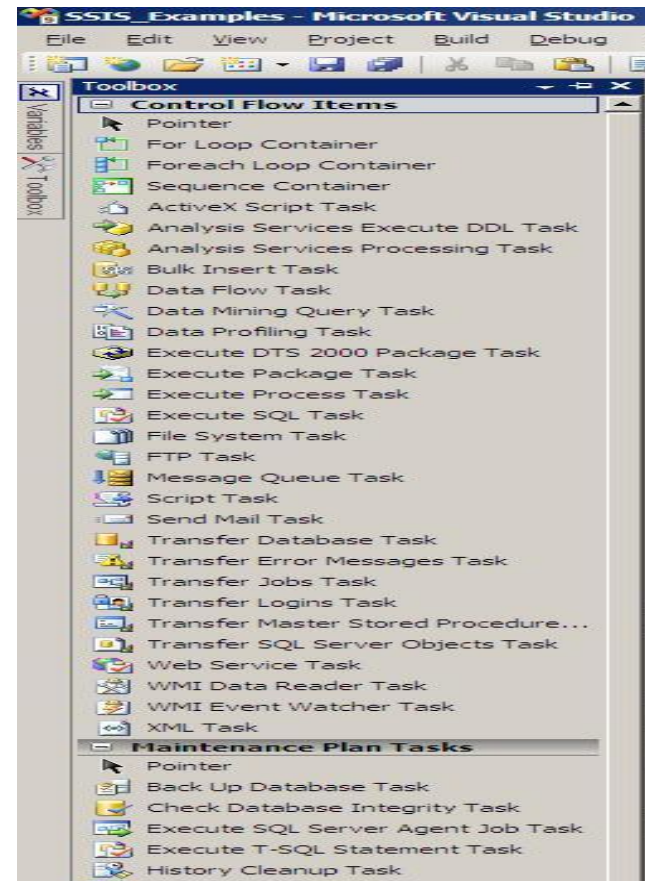
- Program
 - Microsoft SQL Server 2008
 - Configuration Tools
 - SQL Server Configuration Manager
 - *SQL Server Integration Services 10.0*

SSIS Component - Control Flow

Control Flow items – Control the file system level activities.

Examples:

- For Loop Container
- For each Loop Container
- Data Flow Task
- Script Task
- FTP Task
- Bulk Insert Task
- Execute SQL Task
- and many more.....



SSIS Component – Data Flow

One of the most important component of SSIS is Data Flow.

Data Flow can be activated using the “Data Flow Task” from control flow task. Data Flow task supports the ETL process.

Data Flow task comprises of 3 major components

- Data Flow Source – (Extraction)
- Data Flow Transformation – (Transformation)
- Data Flow Destination – (Load)

SSIS Component - Data Flow Transformation

Data flow transformation components are used to process or manipulate data before inserting data into Destination.

For Example, Lookup Transformation – can be used to retrieve some / set of information based on the Lookup Key. There are two types of lookup transformation

1. Lookup Transformation
2. Fuzzy Lookup Transformation

The major diff. between this is LT search for exact match and can be used with any data type, where as FLT searches using tolerance, similarity and can be used only with string data type.

Few Concepts

- Precedence – Success, Failure, Completion
- Connection String Manager
- Containers
- Expressions
- Package Execution
- Package Deployment
- Execution Utility – dtexecui, dtexec

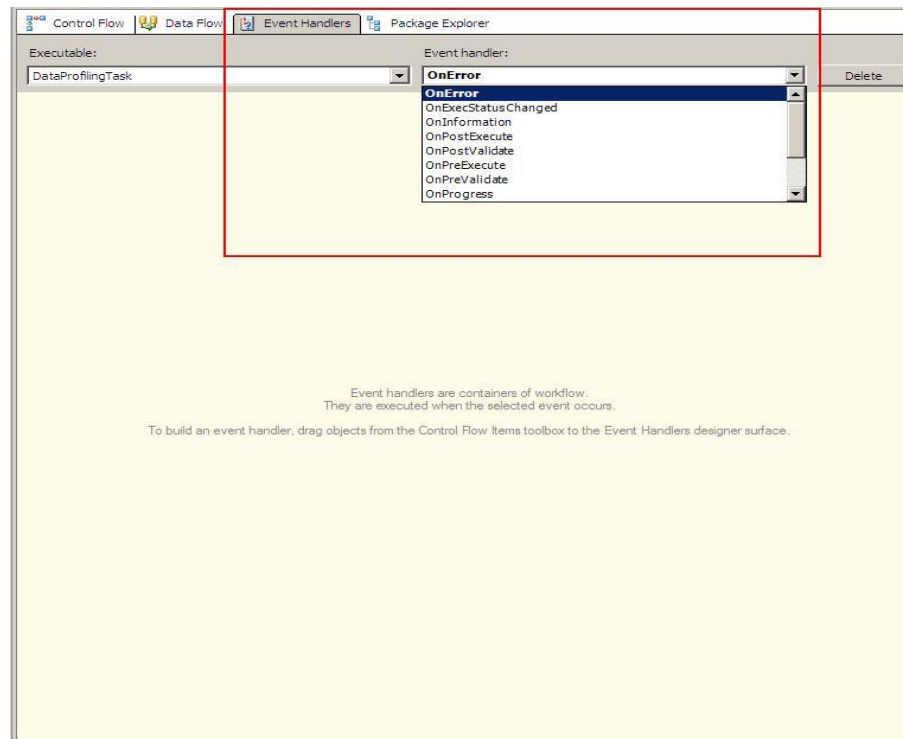
Checkpoint

Checkpoint is basically a file that helps in restarting failed packages. Instead of re-running the whole package if checkpoint is available the package will load from point of failure. Checkpoints can be configure at package level (control flow) only using below properties.

- *CheckpointFileName* - Specifies the name of the checkpoint file.
- *CheckpointUsage* - Specifies whether checkpoints are used.
- *SaveCheckpoints* - Indicates whether the package saves checkpoints. This property must be set to True to restart a package from a point of failure.

Event Handling in SSIS

SSIS got its own Event Handler for specified events, which we can leverage to put the event handler at package level or container level.



Screen Shot 1

The screenshot displays the Microsoft Visual Studio interface for an SSIS package. The main design area shows a control flow diagram with the following tasks:

- Execute SQL Task
- Extract, Transform, Load (Data from Input File)
- Extract, Transform, Load (Data from Input File)

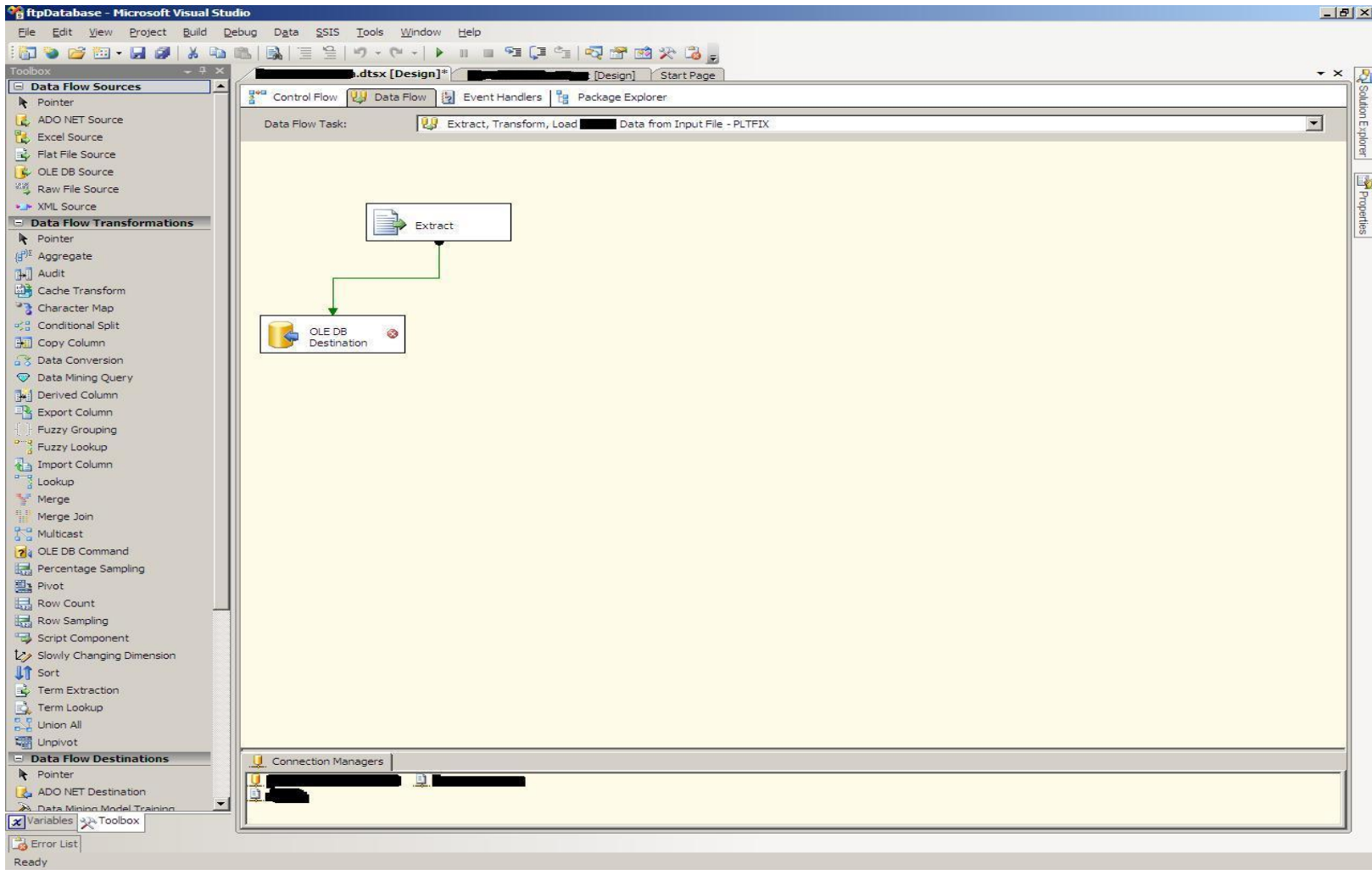
The Solution Explorer on the right shows the project structure:

- Solution 'ftpDatabase' (2 projects)
- ftpDatabase
- Data Sources
- Data Source Views
- SSIS Packages
- [Redacted] .dtsx
- Miscellaneous

The Properties window shows the following details for the selected .dtsx file:

Property	Value
File Name	[Redacted]
Full Path	D:\Shared\Di

Screen Shot 2



SQL SERVER ANALYSIS SERVICE (SSIS)

SSAS Multidimensional Data allows you to design, create, and manage multidimensional structures that contain detail and aggregated data from multiple data sources, such as relational databases, in a single unified logical model supported by built-in calculations.

SSAS provides fast, intuitive and high level analysis for the OLTP data, as it calculates, aggregates and stores the data in Data warehouse. Data warehouse database are mostly used for reporting purposes (SSRS) high level reports

e.g. Total Sales of the Product by Year, Month, Week and Location wise, Country wise, State wise

Supported Database Service

Can be found at:

- Program
 - Microsoft SQL Server 2008
 - Configuration Tools
 - SQL Server Configuration Manager
 - *SQL Server Analysis Services*

Key Concepts

- Staging Database – A database for data cleansing activity. It clean the data transform, combine and prepare the source data to the data mart or data ware house
- OLTP – Online transactional processing. A database contains tables the stores a specific set of structured data.
- MDX – Multidimensional Expression. OLAP query language.

Key Components

- Data Source – A data source is a connection reference that you create outside a package.
- Data Source View – A subset of the data that populates a large data warehouse.
- Measures – Measurable part of the data. E.g. SalesAmount attribute from SalesDetails can be used to measure the TOTAL SALES.
- Dimensions – How we can measure the data from different dimension. E.g. by Year, Location, Product, Product Category.

Key Components (Continue)

- Fact – The table from where we can get the measurable data.
E.g. SalesDetails
- Cube – A logical collection of Dimensions and Measures.
- Dimension Table – A table from where we can get the Dimension data.
- Fact Table – A table from where can get the Fact data.

Storage structures

- MOLAP – Measure group data and aggregation are store in a multidimensional format (OLAP db).
- ROLAP - Measure group data and aggregation are store in relational format (OLTP db).
- HOLAP - Measure group data is maintained in relational format (OLTP db) and aggregation are store in a multidimensional format (OLAP db).

Schema

- Snow fake schema – Creating dimension with more than 3 dimension table.

E.g. Consider database AdventureWorksDW2008

For getting the TotalSales from DimInternetSales Productwise, Subcategorywise, Categorywise, we require 3 dimensions.

- Star schema – Creating dimension with One dimension table

E.g. In above example we can get the TotalSales by creating a view which contains Product, Subcategory, Category information in one table.

Key points to remember

- By default any attribute in dimension has type – Regular.
- Reference Dimension – One dimension linked with fact table using other dimension.
- KPI – Key Performance Indicator
- Cube structure
 - Calculations
 - Aggregates – Pre-calculated values that creates a separate file “Aggregate File”.
 - Partitions
 - Translations
 - Browser

Practice more on SSAS

SSAS is gigantic subject and can't be covered in just few slides. I have shared some introductory articles and books reference at the end of this presentation, which give you the further in-depth information on SQL Business Intelligence.

SQL SERVER REPORTING SERVICE (SSRS)

SSRS provides a full range of ready-to-use tools and services to help you create, deploy, and manage reports for your organization, as well as programming features that enable you to extend and customize your reporting functionality.

With Reporting Services, you can create interactive, tabular, graphical, or free-form reports from relational, multidimensional, or XML-based data sources. You can publish reports, schedule report processing, or access reports on-demand. Reporting Services also enables you to create ad hoc reports based on predefined models, and to interactively explore data within the model

SQL Server 2008 Reporting Services (SSRS) is a server-based reporting platform that provides comprehensive reporting functionality for a variety of data sources.

Supported Database Service

Can be found at:

- Program
 - Microsoft SQL Server 2008
 - Configuration Tools
 - SQL Server Configuration Manager
 - *SQL Server Reporting Services*

Key Components

- Report Server – A key component of SSRS, which provides data & report processing and report delivery.
- Report Designer - A graphical interface which uses Business Intelligence Development Studio for report design. Report Designer also provides a wizard that steps you to design a report and to produce a simple tabular /chart report.
- Report Builder – An interface to create the reports. It allows the user to explore the business information using report model without having to understand the underlying data structure. No pre-defined template used.
- Report Manager – It's a web based report access and management tool.

CONCEPTS

- Report definition - A report definition is a file that you create in Report Designer or Report Builder, which contains report connection string, command text/query, layout.
- Report Model - A report model is a metadata description of a data source and its relationships. It defines only design no layout is defined when you use Report Model.
- Report Designer – An interface to design the complete report.
- Shared Data Source - A *shared data source* is a separate item stored on the report server that describes a data source connection. It is reusable across multiple reports.

Types of Reports

- Parameterized reports - A parameterized report uses input values to complete report or data processing.
- Linked reports – A report which links to existing report in Report Server.
- Snapshot reports - A report snapshot is a report that contains layout information and query results that were retrieved at a specific point in time.
- Cached reports - A cached report is a saved copy of a processed report.

Types of Reports (Continue)

- Ad hoc reports - An ad hoc report can be created from an existing report model by using Report Builder.
- Clickthrough reports - A clickthrough report is a report that displays related data from a report model when you click the interactive data contained within your model-based report.
- Drillthrough reports - Drillthrough reports are standard reports that are accessed through a hyperlink on a text box in the original report.
- Subreports - A subreport is a report that displays another report inside the body of a main report.

Dig into it practically more ...



Questions



If you have any questions/comments related this presentations you can comment it on my blog <http://abhijitmore.wordpress.com/>

Reference Books & Articles

- ✓ MCTS 70-448 Training Kit MS SQL Server 2008 BI Development and Maintenance
- ✓ <http://www.microsoft.com/sqlserver/2008/en/us/business-intelligence.aspx>