Strategic Sponsor

Microsoft

Gold Sponsors

PYRAMID ANALYTICS

KRAJOWY REJESTR DŁUGÓW www.krd.pl

devart Database Interaction Solutions

Silver Sponsors

Profisee®

SQL EXPERT.pl

CREDIT SUISSE

Technical Partners

Microsoft

PTI POLSKIE TOWARZYSTWO INFORMATYCZNE ODDZIAŁ GÓRNOŚLĄSKI

Academic Partners

Wyższa Szkoła Zarządzania i Bankowości w Krakowie

Wyższe Szkoły Bankowe

Media Partners

GEEKCLUB.pl wss.pl

ERP24.pl

4programmers.net

IT professional

IN FACT PR

itnews24

virtual-it.pl

IT SWORD POLSKA

VirtualStudy

MAGAZYN IT PL

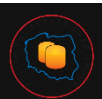
Made in Wro Do IT here!

ERP-view.pl Portal rozwiązań IT w biznesie

Easy ETL with Biml™

Andrzej Kukuła – andrzej.kukula@gmail.com

Marcin Szeliga – marcin@sqlexpert.pl



Agenda

- What's ETL
- SSIS – pros and cons
- Example case
- Traditional approach
- Novel approach
- Biml and BimlScript
- Benefits

ETL

- Extract, Transform, Load
  - Data extraction from an OLTP system, denormalization, conversion, modelling and loading into Data Warehouse, Operational Data Store, Data Mart
  - Also ELT, ETLT

  - Performance reasons
  - Historic data handling (SCD)
  - Better analysis capabilities

Image from: https://namitkabra.wordpress.com/2014/09/16/need-for-etl-tool-as-explained-to-undergraduate-students-part-2/

SQL Server Integration Services

- Well known technology for ETL, integration, interfacing, data movement automation
- First choice for staging/data loading process
- Ability to perform sophisticated data manipulation
- Runtime and Integrated Services Catalog
- Integrated with SQL Server
- Lots of components
- Very good performance

Problems with SSIS

- Package design process in SSDT-BI
  - Manual
  - Not generic
  - Not developer-friendly
    - Doesn't allow code reuse – repeatable work of implementing many similar packages
    - No support for version control
  - Frustrating, slow, boring, painful, error-prone, expensive
- At low level
  - DTSX is almost unreadable and unmodifiable

Problems with SSIS

- Metadata management is difficult
  - substantial development time is spent on solving dreaded metadata issues

  - is this something we should really focus on?...

More problems with SSIS

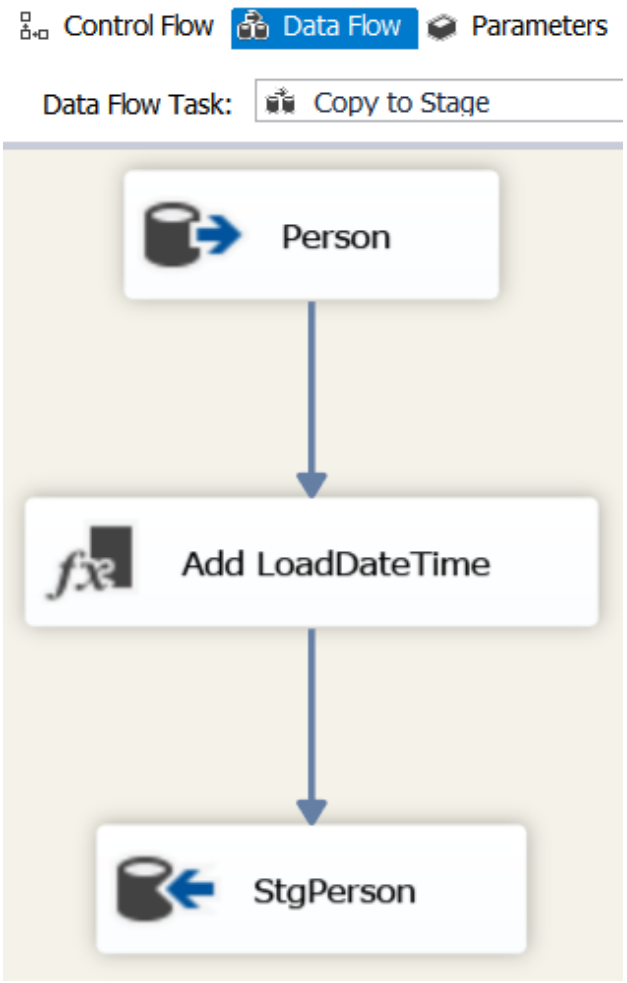
- Automatic generation of packages is difficult
  - DTS API is not trivial
  - EzAPI ditto, also not updated
  - Generation of target DTSX' XML manually is close to impossible in reasonable time
  - Manually generated packages won't run or won't open in SSDT most of the time (the infamous message „`Package Load error 0xC0010014 in CPackage::LoadFromXML`")

Executive Problems with SSIS

- ETL process is
  - **Slow**
  - **Expensive**
  - **Difficult to implement**
  - **Difficult to maintain**
  - **Difficult to adapt to changes**
    - One new column introduced in one source table can take **weeks** or even **months** to implement

Image from: http://www.clomedia.com/articles/5605-making-a-case-for-angry-bosses

Example Package



Example DTSX

Is this reasonable amount of code to accomplish the task?…

```
38    <DTS:Executables>
39      <DTS:Executable
40        DTS:CreationName="Microsoft.Pipeline"
41        DTS:DTSID="{33792D2F-53D7-49FA-8AB1-45ADCA7D63CD}"
42        DTS:ExecutableType="Microsoft.Pipeline"
43        DTS:LocaleID="-1"
44        DTS:ObjectName="Copy to Stage"
45        DTS:refId="Package\Copy to Stage"
46        DTS:TaskContact="Performs high-performance data extract
47        <DTS:ObjectData>
48          <pipeline
49            BLOBTempStoragePath=""
50            bufferTempStoragePath=""
51            version="1">
52            <components>
53              <component
54                componentClassID="Microsoft.OLEDBSource"
                  contactInfo="OLE DB Source;Microsoft Corporation; Microsoft SQL Server; (C) Microsoft Cor
56                name="Person"
57                refId="Package\Copy to Stage\Person"
58                usesDispositions="true"
59                rsion="7">
60                <con ctions>
61                  <conne ion
62                    connecti ManagerID="{FE7C9A82-9025-4233-B3
63                    connectionMa erRefId="Package.ConnectionM
64                    description="The E DB runtime connection
65                    name="OleDbConnection
66                    refId="Package\Copy to Sta \Person.Connect
67                </connections>
68                <outputs>
69                  <output
70                    name="Output"
71                    refId="Package\Copy to Stage\Person.Outputs
```

```
1   <?xml version="1.0" encoding="utf-8"
2   <DTS:Executable
3     DTS:CreationDate="04/26/2015 20:46
4     DTS:CreationName="Microsoft.Packag
5     DTS:CreatorComputerName="AK81"
6     DTS:CreatorName="Andrzej"
7     DTS:DTSID="{4510E215-AEBB-4A6B-89C
8     DTS:ExecutableType="Microsoft.Pack
9     DTS:LocaleID="1045"
10    DTS:LastModifiedProductVersion=""
```

```
693                 endId="Package\Copy to
694                 name="Add LoadDateTime
695                 refId="Package\Copy to
696                 startId="Package\Copy
697             </paths>
698           </pipeline>
699         </DTS:ObjectData>
700       </DTS:Executable>
701     </DTS:Executables>
702     <DTS:Property
703       DTS:Name="PackageFormatVersion">
704   </DTS:Executable>
```

What if…?

…we changed rules of the game, and instead of creating SSIS packages by hand, just give a computer **recipes** on how to make the packages for us?

- Recipe language easy to learn and use
  - With smart default values and default behavior
- Be able to use programming language to make recipes more dynamic, easier adapting to changes in databases and business requirements
- **Don't bother with metadata** (most of the time)

What if…?

- Have full power of .NET framework available
- Organize recipes into templates and libraries for multiple use
- Build SSIS packages automatically
  - in repeatable way
- Use version control to track code changes
- Use CI and CD to automate deployments

Enter Biml™

```
<Package Name="Load Person" ProtectionLevel="EncryptSensitiveWithUserKey">
    <Tasks>
        <Dataflow Name="Copy to Stage">
            <Transformations>
                <OleDbSource Name="Person" ConnectionName="Source">
                    <TableInput TableName="AdventureWorks2012.Person.Person" />
                </OleDbSource>
                <DerivedColumns Name="Add LoadDateTime">
                    <Columns>
                        <Column Name="LoadDateTime" DataType="DateTime2" Scale="7">
                            @[System::ContainerStartTime]</Column>
                    </Columns>
                </DerivedColumns>
                <OleDbDestination Name="StgPerson" ConnectionName="Stage">
                    <TableOutput TableName="Stage.AW.Person" />
                </OleDbDestination>
            </Transformations>
        </Dataflow>
    </Tasks>
</Package>
```

Is **this** reasonable amount of code to accomplish the task?

What's Biml?

- Business Intelligence Markup Language
- A really easy XML-based language to describe BI assets
  - Connections
  - Tables, Views
  - SSIS Packages, SSIS Projects
  - Dimensions, Measure Groups, Cubes
  - and more…
- Available straight in SSDT-BI for free!
  - All you need to begin is BIDS Helper add-in
  - With Biml Intellisense

Demo!

But wait, there's more!

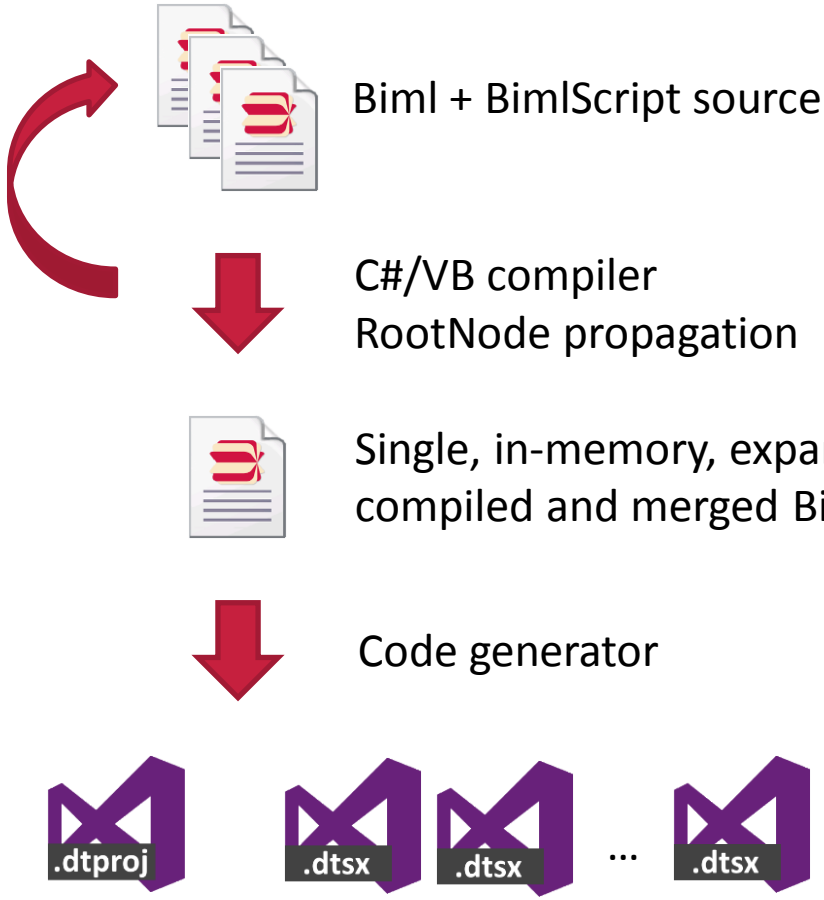
BimlScript

- The ability to instrument Biml with C#/VB.NET code
- Full power of .NET framework and all available libraries at your disposal

```
<Packages>
    <# foreach (var table in listOfTables) { #>
    <Package Name="Load <#= table.Name #>" ... />
    <# } #>
</Packages>
```

- Supports including and calling other Biml files, and referencing external .NET assemblies
- Allows even extending built-in Biml .NET classes
- Available in SSDT with BIDS Helper

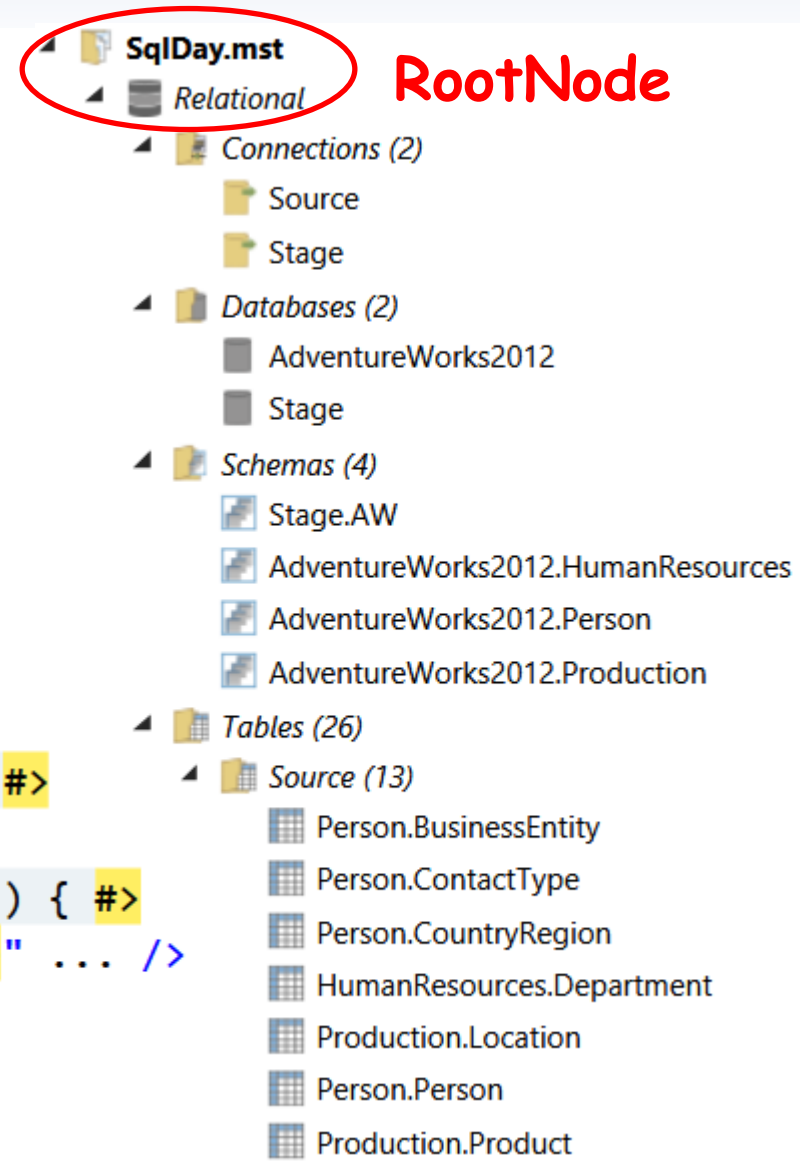
How does it work?



RootNode

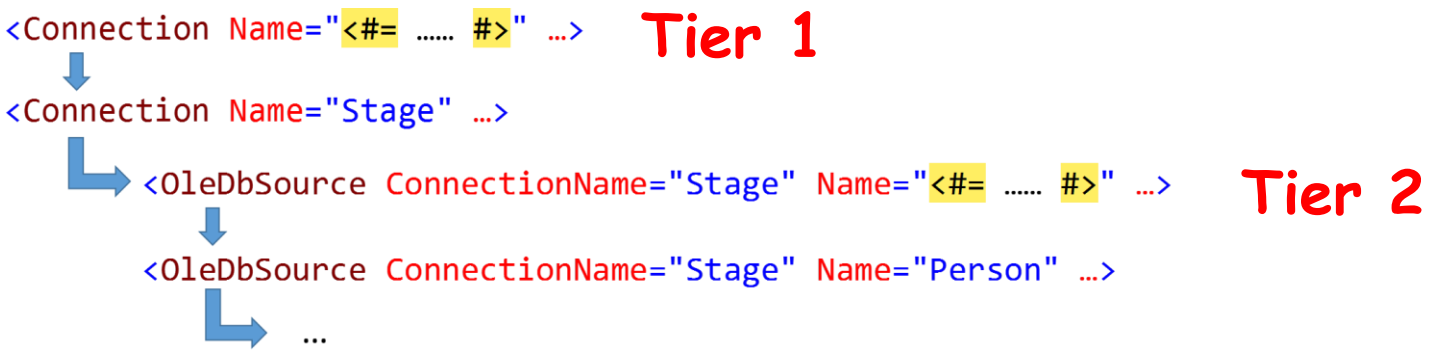
- The **model** of all assets in the project
- Used within C#/VB code
- Read/write

```
<# var listOfTables = RootNode.Tables
        .Where(t => t.Database.Name
                == "AdventureWorks2012"); #>
<Packages>
    <# foreach (var table in listOfTables) { #>
    <Package Name="Load <#= table.Name #>" ... />
    <# } #>
</Packages>
```



BimlScript Features

- „Layered” expansion based on „tiers”



- Makes it easy to prepare resources before using them (e.g. fetch from metadata store)
- Tier *n*+1 sees everything in tier *n* in already expanded form
- No limit on number of tiers

BimlScript Features

- Just to name a few
  - Ability to dynamically fetch database schema during generation of packages
  - Ease of implementation of custom metadata-based processing logic
  - Automatic data type conversions during Data Flow Task
  - SCD handling during loading of dimensions
  - SQL Server's built-in extended properties proven to be good metadata storage
- Can be easily adapted to new/changed requirements
- Spectacular effects need just a few lines of C# code

Mist IDE for Biml

Mist™

- Dedicated development environment just for Biml
- Compatible with SQL Server/SSIS/SSAS 2008-2014
- Makes it super easy to work with large, advanced BimlScript projects
  - Full Biml/C#/VB Intellisense
  - Live preview of expanded BimlScript
  - Graphical designers
- It can even completely replace SSDT-BI for SSIS development!
  - Integration Services is just a runtime
- The fastest ETL development
- Our further demos will use Mist

Demo!

More features

- Transformers and „Frameworks”
  - Ability to change the way how code is generated
- DDL generation

- MSBuild integration
  - CI with TFS Build, TeamCity immediately possible

How to benefit it?

- BI Project Decision Makers
  - Faster initial delivery
  - Lower cost of change
  - Immediate Biml/Mist ROI
- BI Architects
  - Reusable Design Patterns with Biml/BimlScript
  - One project compatible with SQL Server 2005-2014
  - Easily manage large BI code base, tasks, and issues using TFS
  - Plan for BI Continuous Integration/Continuous Delivery

How to benefit it?

- ETL developers
  - Fast Biml learning curve
  - Generate your DTSX faster with Biml instead of drag'n'drop
  - Embrace DRY in BI development
  - Use proper version control to manage your sources
- BI consultants
  - Increased productivity
  - Build your Biml/BimlScript code library and reuse it in different projects
  - No runtime license costs for customers

Eric Heilman
@EricHeilman

Obserwuj

Only with #biml can you add 5 lines of code to a single file, press a button and spit out 55 new #SSIS packages. I am a total nerd #fb

Q&A

Strategic Sponsor

Microsoft

Gold Sponsors

PYRAMID ANALYTICS

KRAJOWY REJESTR DŁUGÓW www.krd.pl

devart Database Interaction Solutions

Silver Sponsors

Profisee®

SQL EXPERT.pl

CREDIT SUISSE

Technical Partners

Microsoft

PTI POLSKIE TOWARZYSTWO INFORMATYCZNE ODDZIAŁ GÓRNOŚLĄSKI

Academic Partners

Wyższa Szkoła Zarządzania i Bankowości w Krakowie

Wyższe Szkoły Bankowe

Media Partners

GEEKCLUB.pl wss.pl

ERP24.pl

4programmers.net

IT professional

IN FACT PR

itnews24

virtual-it.pl

IT SWORD POLSKA

VirtualStudy

MAGAZYN IT PL

Made in Wro Do IT here!

ERP-view.pl Portal rozwiązań IT w biznesie