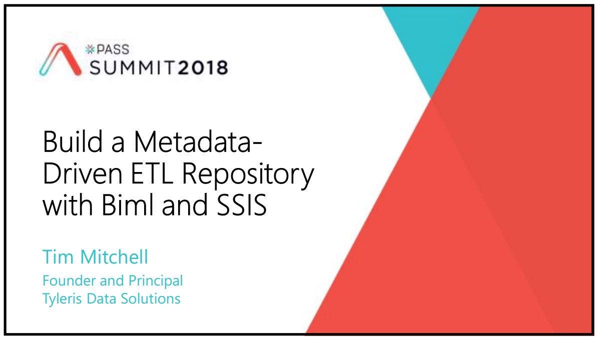
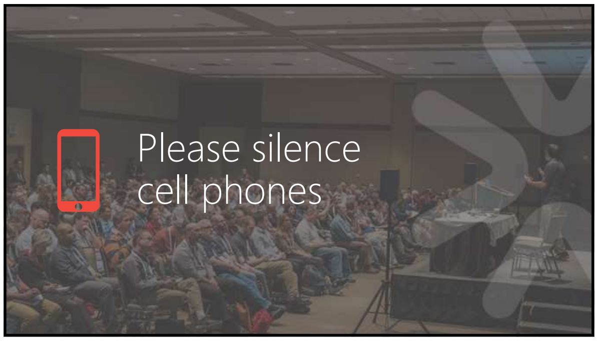


PASS SUMMIT2018

Build a Metadata-Driven ETL Repository with Biml and SSIS

Tim Mitchell
Founder and Principal
Tyleris Data Solutions

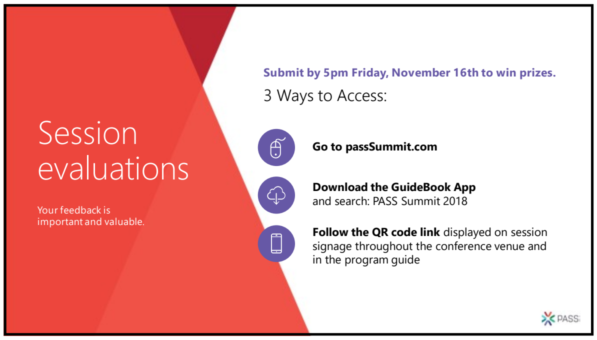


Please silence cell phones

Explore everything PASS has to offer

Free Online Resources
Newsletters
PASS.org

24HOURS of PASS
Free online webinar events
PASS LOCAL GROUPS
Local user groups around the world
PASS SQLSATURDAY
Free 1-day local training events
PASS VIRTUAL GROUPS
Online special interest user groups
PASS MARATHON
Business analytics training
PASS VOLUNTEERS
Get involved



Session evaluations

Your feedback is important and valuable.

Submit by 5pm Friday, November 16th to win prizes.

3 Ways to Access:

Go to passSummit.com

Download the GuideBook App and search: PASS Summit 2018

Follow the QR code link displayed on session signage throughout the conference venue and in the program guide

Tim Mitchell

Tyleris Data Solutions

/in/timmitchell/

@tim_mitchell

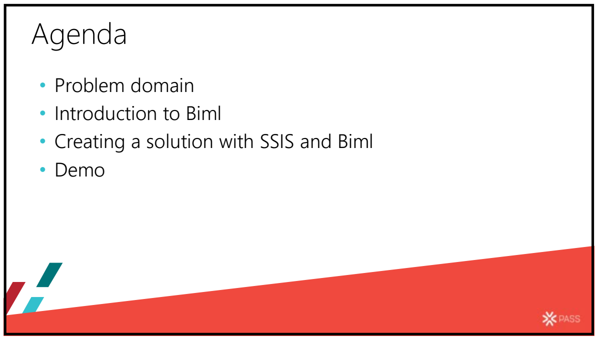
TimMitchell.net

Tyleris.com

Business intelligence consultant

Microsoft Data Platform MVP since 2010

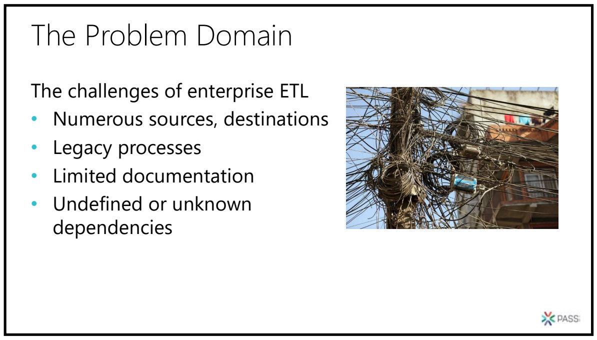
Founder and principal, Tyleris Data Solutions



Agenda

- Problem domain
- Introduction to Biml
- Creating a solution with SSIS and Biml
- Demo

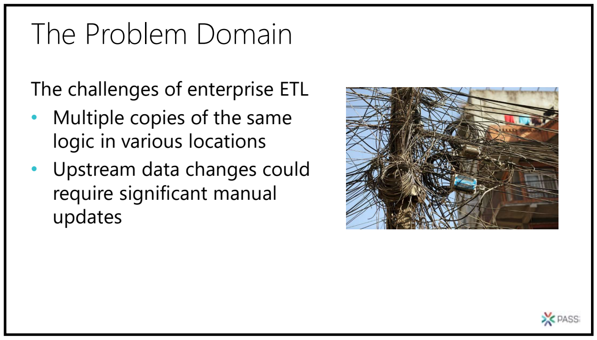
The Problem Domain



The Problem Domain

The challenges of enterprise ETL

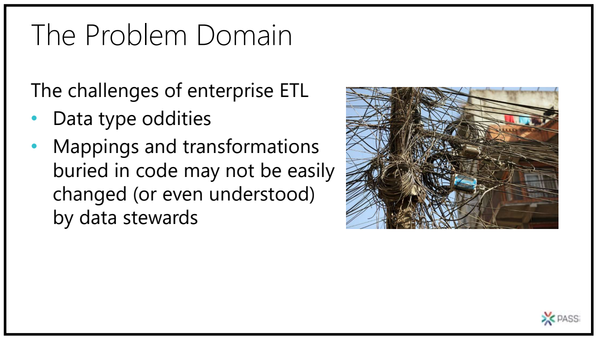
- Numerous sources, destinations
- Legacy processes
- Limited documentation
- Undefined or unknown dependencies



The Problem Domain

The challenges of enterprise ETL

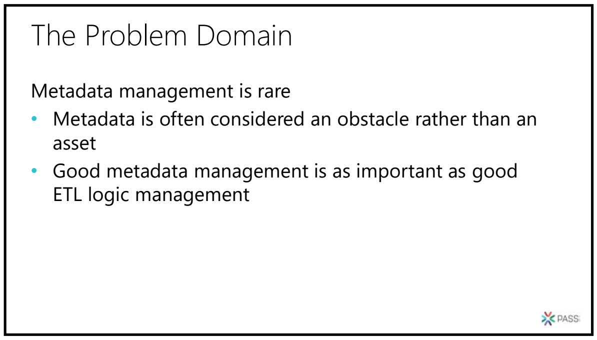
- Multiple copies of the same logic in various locations
- Upstream data changes could require significant manual updates



The Problem Domain

The challenges of enterprise ETL

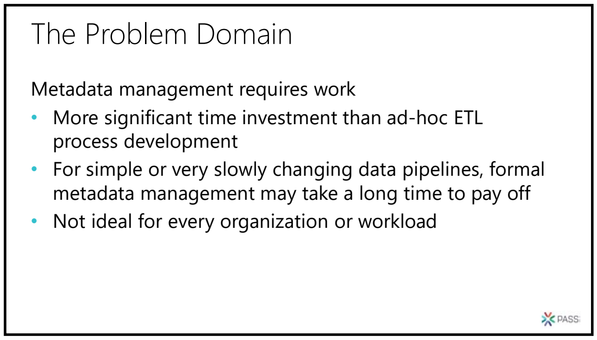
- Data type oddities
- Mappings and transformations buried in code may not be easily changed (or even understood) by data stewards



The Problem Domain

Metadata management is rare

- Metadata is often considered an obstacle rather than an asset
- Good metadata management is as important as good ETL logic management

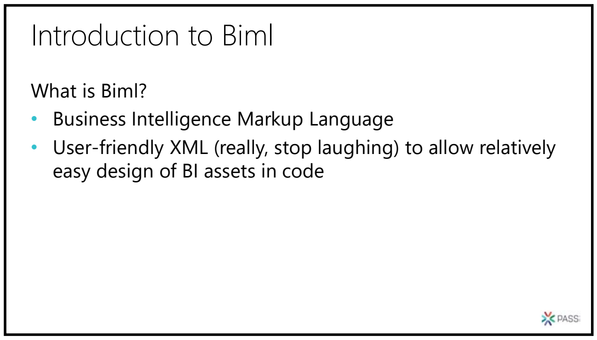


The Problem Domain

Metadata management requires work

- More significant time investment than ad-hoc ETL process development
- For simple or very slowly changing data pipelines, formal metadata management may take a long time to pay off
- Not ideal for every organization or workload

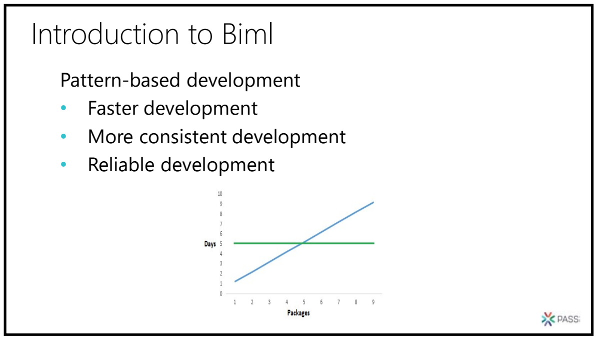
Introduction to Biml



Introduction to Biml

What is Biml?

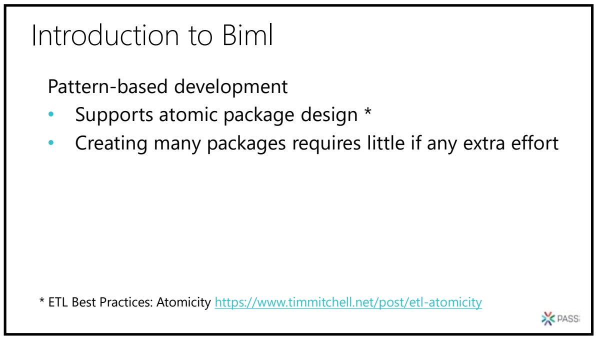
- Business Intelligence Markup Language
- User-friendly XML (really, stop laughing) to allow relatively easy design of BI assets in code



Introduction to Biml

Pattern-based development

- Faster development
- More consistent development
- Reliable development

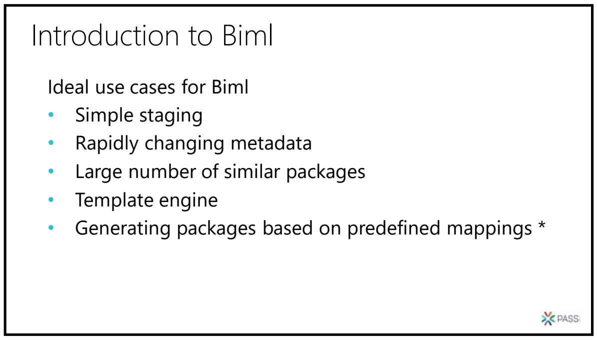


Introduction to Biml

Pattern-based development

- Supports atomic package design *
- Creating many packages requires little if any extra effort

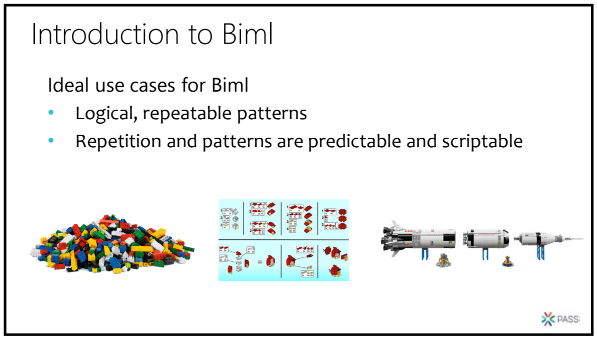
* ETL Best Practices: Atomicity https://www.timmitchell.net/post/etl-atomicity



Introduction to Biml

Ideal use cases for Biml

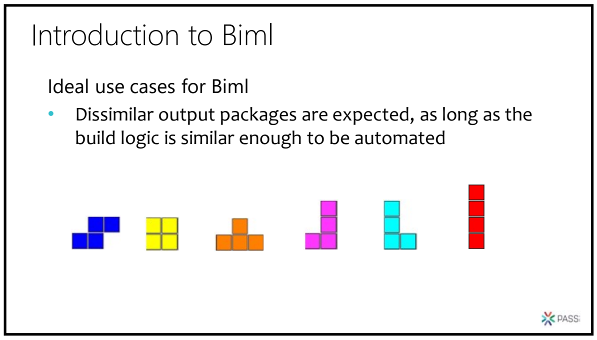
- Simple staging
- Rapidly changing metadata
- Large number of similar packages
- Template engine
- Generating packages based on predefined mappings *



Introduction to Biml

Ideal use cases for Biml

- Logical, repeatable patterns
- Repetition and patterns are predictable and scriptable



Introduction to Biml

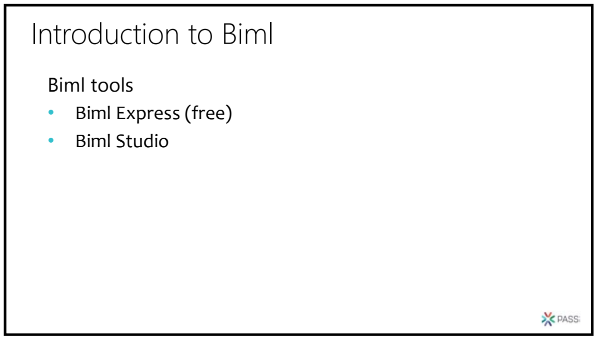
Ideal use cases for Biml

- Dissimilar output packages are expected, as long as the build logic is similar enough to be automated

Introduction to Biml

Ideal use cases for Biml

- Dissimilar patterns or unpredictable logic may not be a good fit



Introduction to Biml

Biml tools

- Biml Express (free)
- Biml Studio

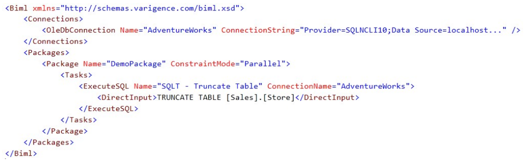
Introduction to Biml

The Biml language

```
<Packages>
    <Package Name="BimlPackage">
        <Tasks>
            <Dataflow Name="DFT Orders">
            …
            </Dataflow>
        </Tasks>
    </Package>
</Packages>
```

Introduction to Biml

The Biml language



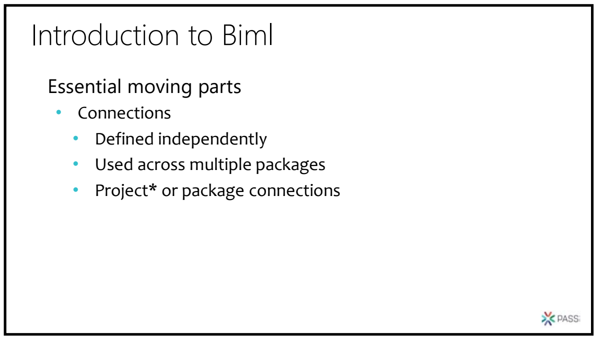
```
<Biml xmlns="http://schemas.varigence.com/biml.xsd">
    <Connections>
        <OleDbConnection Name="AdventureWorks" ConnectionString="Provider=SQLNCLI10;Data Source=localhost..." />
    </Connections>
    <Packages>
        <Package Name="DemoPackage" ConstraintMode="Parallel">
            <Tasks>
                <ExecuteSQL Name="SQLT - Truncate Table" ConnectionName="AdventureWorks">
                    <DirectInput>TRUNCATE TABLE [Sales].[Store]</DirectInput>
                </ExecuteSQL>
            </Tasks>
        </Package>
    </Packages>
</Biml>
```

Language reference: https://www.varigence.com/Documentation/Language/Index

Introduction to Biml

Essential moving parts

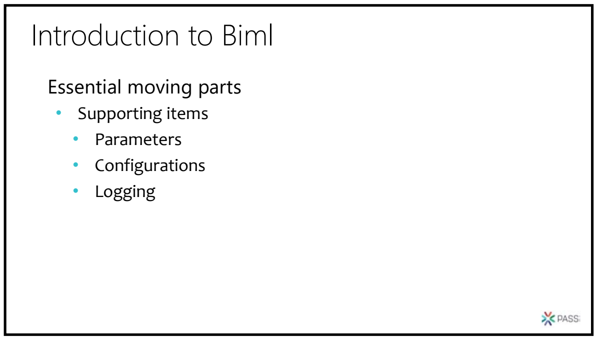
- Package
  - Central component of Biml development
  - All native tasks and components are supported
  - Data flow components



Introduction to Biml

Essential moving parts

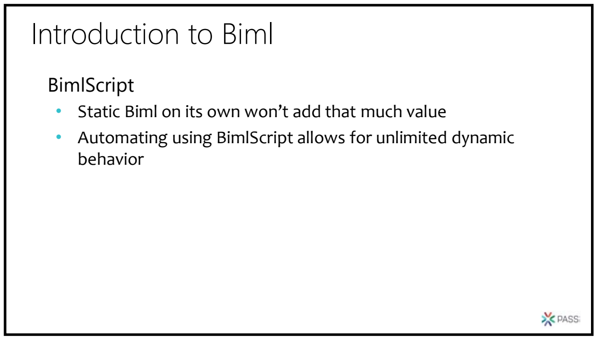
- Connections
  - Defined independently
  - Used across multiple packages
  - Project* or package connections



Introduction to Biml

Essential moving parts

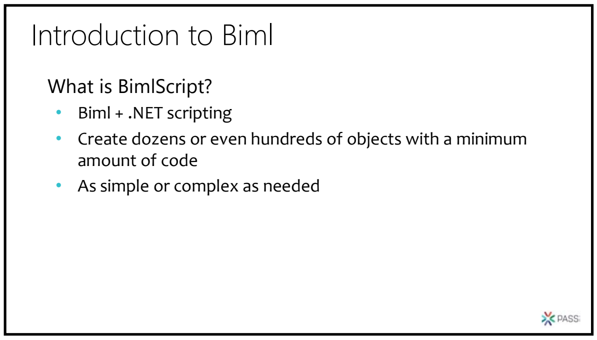
- Supporting items
  - Parameters
  - Configurations
  - Logging



Introduction to Biml

BimlScript

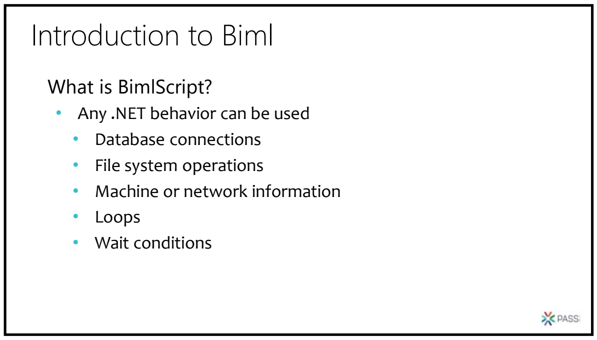
- Static Biml on its own won’t add that much value
- Automating using BimlScript allows for unlimited dynamic behavior



Introduction to Biml

What is BimlScript?

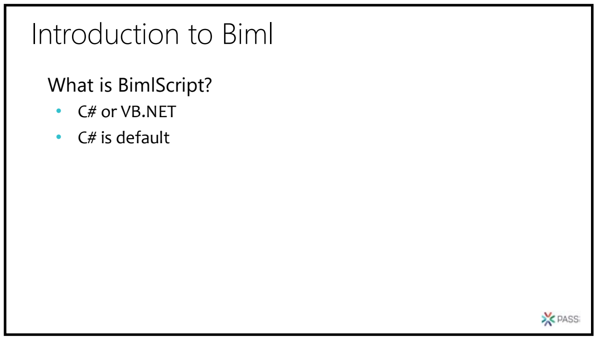
- Biml + .NET scripting
- Create dozens or even hundreds of objects with a minimum amount of code
- As simple or complex as needed



Introduction to Biml

What is BimlScript?

- Any .NET behavior can be used
  - Database connections
  - File system operations
  - Machine or network information
  - Loops
  - Wait conditions



Introduction to Biml

What is BimlScript?

- C# or VB.NET
- C# is default

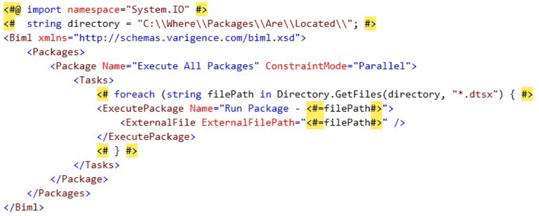
Introduction to Biml

BimlScript Syntax

- **<# Code Block #>**
  - Any .NET Code
- **<#= Inline Code Block #>**
  - Calls the .NET ToString() method on the expression
- **<#@ Directive #>**
  - Set configuration options
- **<#+ Module Level Code Block #>**
  - Functions that are callable from within code block

Introduction to Biml

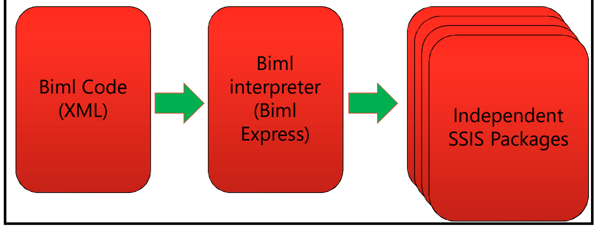
BimlScript Syntax



```
<#@ import namespace="System.IO" #>
<#  string directory = "C:\\Where\\Packages\\Are\\Located\\"; #>
<Biml xmlns="http://schemas.varigence.com/biml.xsd">
    <Packages>
        <Package Name="Execute All Packages" ConstraintMode="Parallel">
            <Tasks>
                <# foreach (string filePath in Directory.GetFiles(directory, "*.dtsx") { #>
                <ExecutePackage Name="Run Package - <#=filePath#>">
                    <ExternalFile ExternalFilePath="<#=filePath#>" />
                </ExecutePackage>
                <# } #>
            </Tasks>
        </Package>
    </Packages>
</Biml>
```

Introduction to Biml

Biml workflow

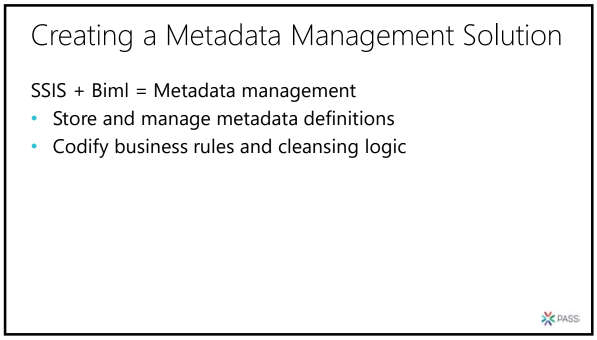


Demo:
Introduction to Biml

Creating a Biml Metadata Management Solution

Creating a Metadata Management Solution

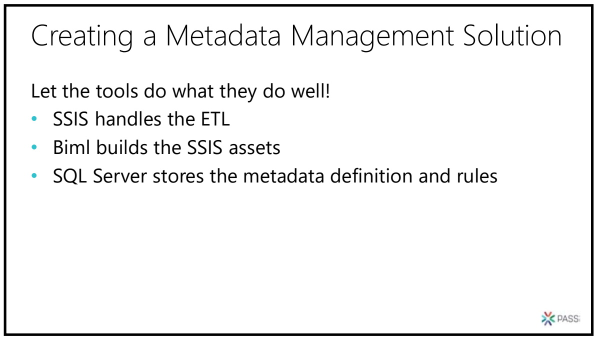
SSIS + Biml = Metadata management



Creating a Metadata Management Solution

SSIS + Biml = Metadata management

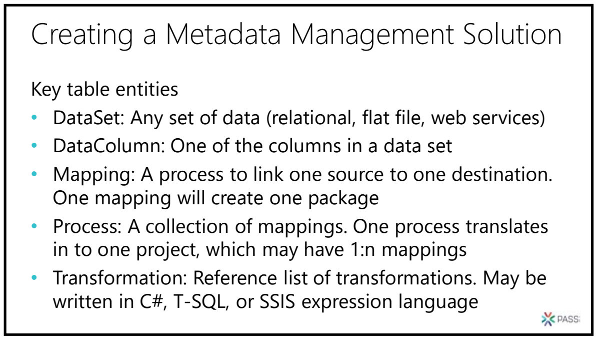
- Store and manage metadata definitions
- Codify business rules and cleansing logic



Creating a Metadata Management Solution

Let the tools do what they do well!

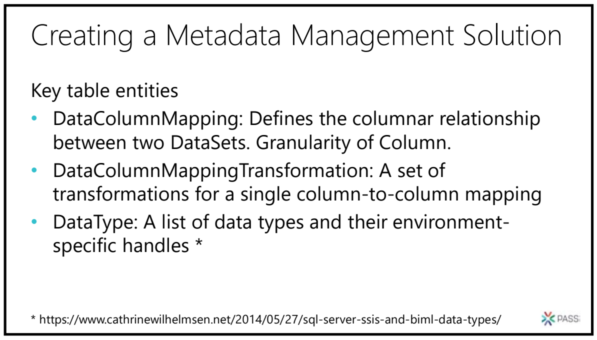
- SSIS handles the ETL
- Biml builds the SSIS assets
- SQL Server stores the metadata definition and rules



Creating a Metadata Management Solution

Key table entities

- DataSet: Any set of data (relational, flat file, web services)
- DataColumn: One of the columns in a data set
- Mapping: A process to link one source to one destination. One mapping will create one package
- Process: A collection of mappings. One process translates in to one project, which may have 1:n mappings
- Transformation: Reference list of transformations. May be written in C#, T-SQL, or SSIS expression language



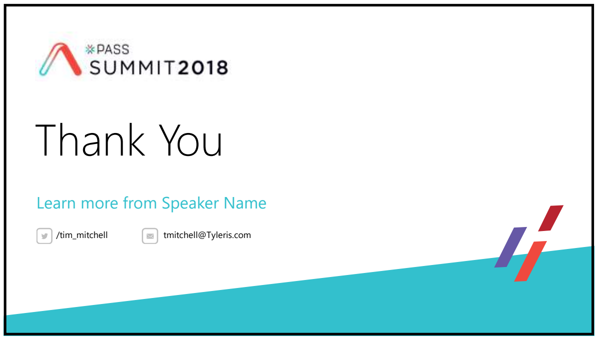
Creating a Metadata Management Solution

Key table entities

- DataColumnMapping: Defines the columnar relationship between two DataSets. Granularity of Column.
- DataColumnMappingTransformation: A set of transformations for a single column-to-column mapping
- DataType: A list of data types and their environment-specific handles *

* https://www.cathrinewilhelmsen.net/2014/05/27/sql-server-ssis-and-biml-data-types/

Demo: Creating a Metadata Management



PASS SUMMIT2018

Thank You

Learn more from Speaker Name

/tim_mitchell

tmitchell@Tyleris.com