BEN WEISSMAN

The SSIS Package that tunes itself

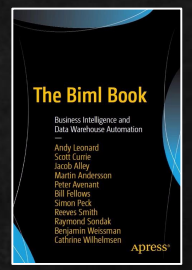
Who am I?

Ben Weissman

Please Talk Data To Me

SOLISYON

@bweissman
b.weissman@solisyon.de
http://biml-blog.de/



BimlHero CERTIFIED EXPERT

Microsoft CERTIFIED
Solutions Associate
Machine Learning

Microsoft CERTIFIED
Solutions Expert
Data Management and Analytics

MVP Microsoft® Most Valuable Professional

I ♥ DATA

FRIEND OF redgate 2019

GENESEE ACADEMY World-Class Training — Certified Data Vault Modeler

Big Data
Data Science
Artificial Intelligence

MICROSOFT PROFESSIONAL PROGRAM

1. A SHORT INTRODUCTION TO BIML

THE 3D PRINTER FOR SSIS AND MORE.

What is Biml?

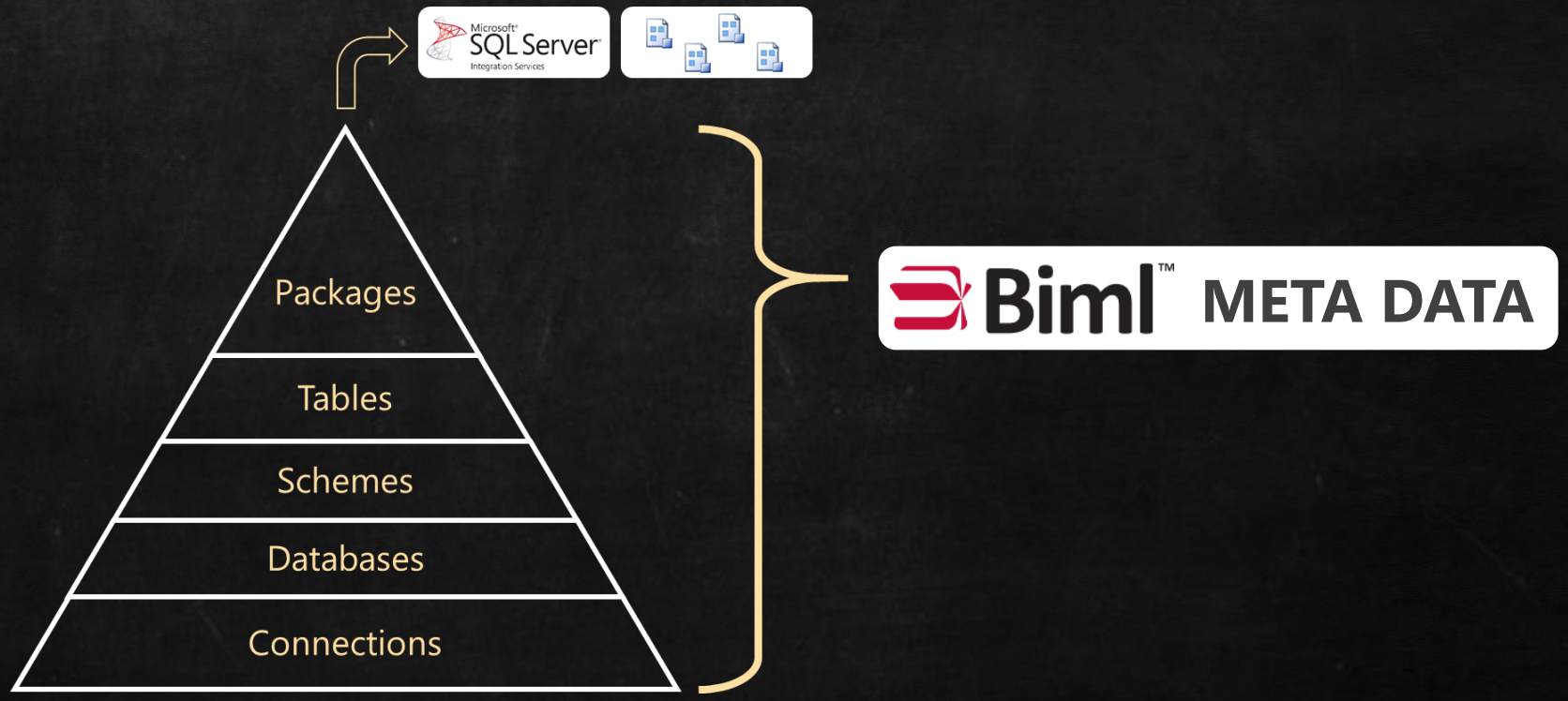
- Biml is the Business Intelligence Markup Language
- Developed by Varigence
- This is Biml:

```
<Biml xmlns="http://schemas.varigence.com/biml.xsd">
    <Packages>
        <Package Name="HelloBiml"/>
    </Packages>
</Biml>
```

- Enhanced by APIs and BimlScript, it becomes very powerful
- Can be used for T-SQL, SSIS and SSAS

Biml Frontends

Basic Concept of Biml



How does it work?

Biml Code

XML, BimlScript, C#, VB
…

Biml Engine

Result

SQL Server Tables,
DTSX Packages etc.

But what does that mean?

- Care less about compatibility between different SQL Server Versions
- Increase your productivity by decreasing the time it takes to build stupid data load packages… and to handle changes in meta data
- In my opinion: Biggest advantage comes with standardization so building a staging area will see the highest improvement rates – still, there are 100s of other use cases

DEMO

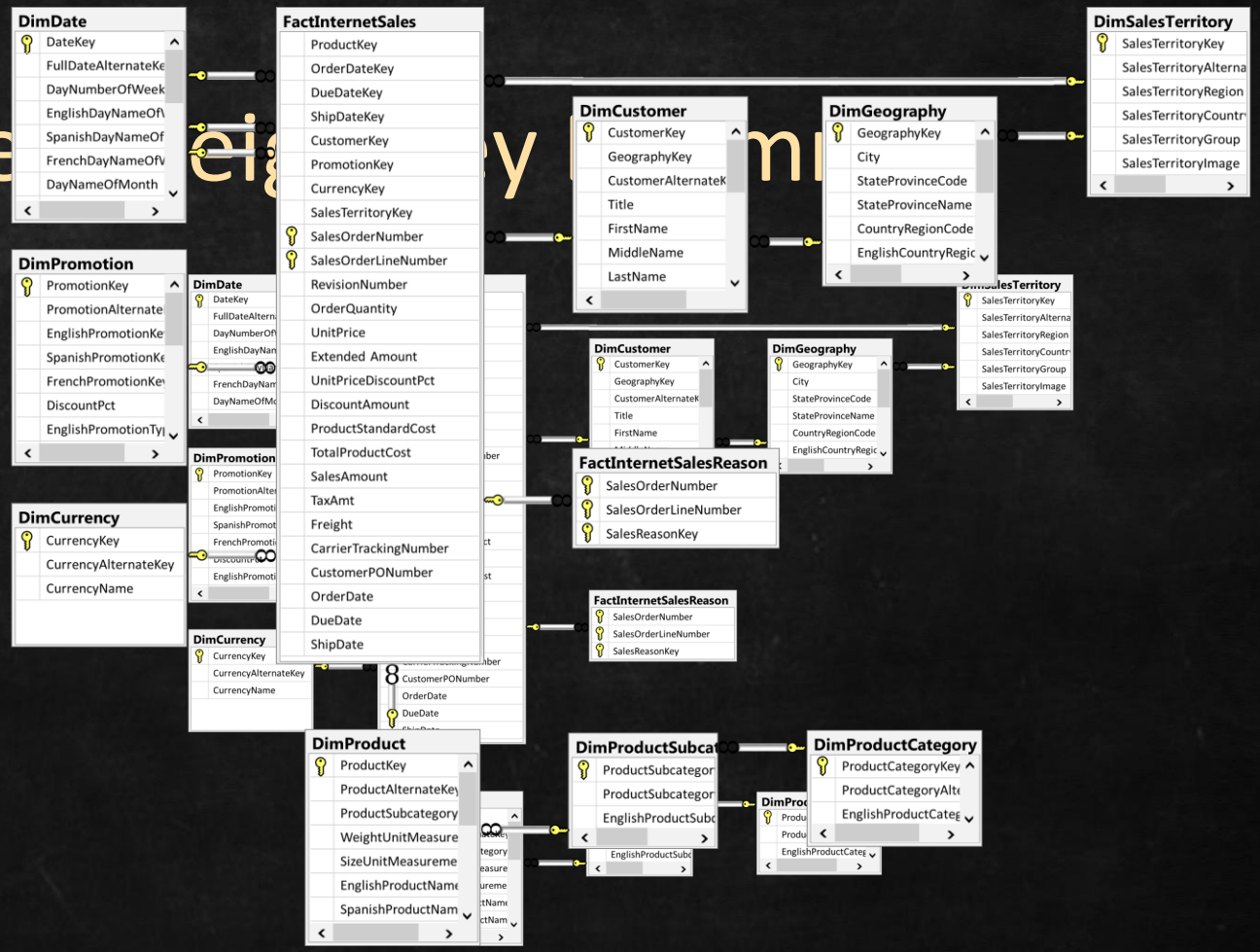
AN INTRODUCTION TO BIML
THE 3D PRINTER FOR SSIS AND MORE.

The Business Intelligence Markup Language

- Basic empty packages
- Automating empty packages with BimlScript
- Automating staging patterns

2. TOPOLOGY SORT

The


DimDate
DateKey
FullDateAlternateKe
DayNumberOfWeek
EnglishDayNameOf
SpanishDayNameOf
FrenchDayNameOf
DayNameOfMonth
FactInternetSales
ProductKey
OrderDateKey
DueDateKey
ShipDateKey
CustomerKey
PromotionKey
CurrencyKey
SalesTerritoryKey
SalesOrderNumber
SalesOrderLineNumber
RevisionNumber
OrderQuantity
UnitPrice
Extended Amount
UnitPriceDiscountPct
DiscountAmount
ProductStandardCost
TotalProductCost
SalesAmount
TaxAmt
Freight
CarrierTrackingNumber
CustomerPONumber
OrderDate
DueDate
ShipDate
DimSalesTerritory
SalesTerritoryKey
SalesTerritoryAlterna
SalesTerritoryRegion
SalesTerritoryCountr
SalesTerritoryGroup
SalesTerritoryImage
DimCustomer
CustomerKey
GeographyKey
CustomerAlternateK
Title
FirstName
MiddleName
LastName
DimGeography
GeographyKey
City
StateProvinceCode
StateProvinceName
CountryRegionCode
EnglishCountryRegic
DimPromotion
PromotionKey
PromotionAlternate
EnglishPromotionKe
SpanishPromotionKe
FrenchPromotionKe
DiscountPct
EnglishPromotionTy
DimCurrency
CurrencyKey
CurrencyAlternateKey
CurrencyName
FactInternetSalesReason
SalesOrderNumber
SalesOrderLineNumber
SalesReasonKey
DimProduct
ProductKey
ProductAlternateKey
ProductSubcategory
WeightUnitMeasure
SizeUnitMeasureme
EnglishProductName
SpanishProductNam
DimProductSubcat
ProductSubcategor
ProductSubcategor
EnglishProductSubc
DimProductCategory
ProductCategoryKey
ProductCategoryAlte
EnglishProductCateg

Topology Sort

- Your tables have foreign key constraints
- There may be more tables required over time or tables becoming obsolete
- Therefore, we need to maintain a certain orchestration of our dataflows in SSIS, starting with tables that do not have a foreign key constraint, followed by the descendants

Topology Sort

- Run through a list of all tables
- Mark each table as „visited“ and add them to a list
- Check for any tables pointing to the table that has not been „visited“ yet and add them behind this table so they get loaded afterwards

DEMO

TOPOLOGY SORT

Linear

Does that work in parallel as well?!

- Load all tables without any dependencies and mark them as loaded
- Load all immediate decendants
- Run all remaining tables in groups until they are all done

DEMO

TOPOLOGY SORT

Parallel

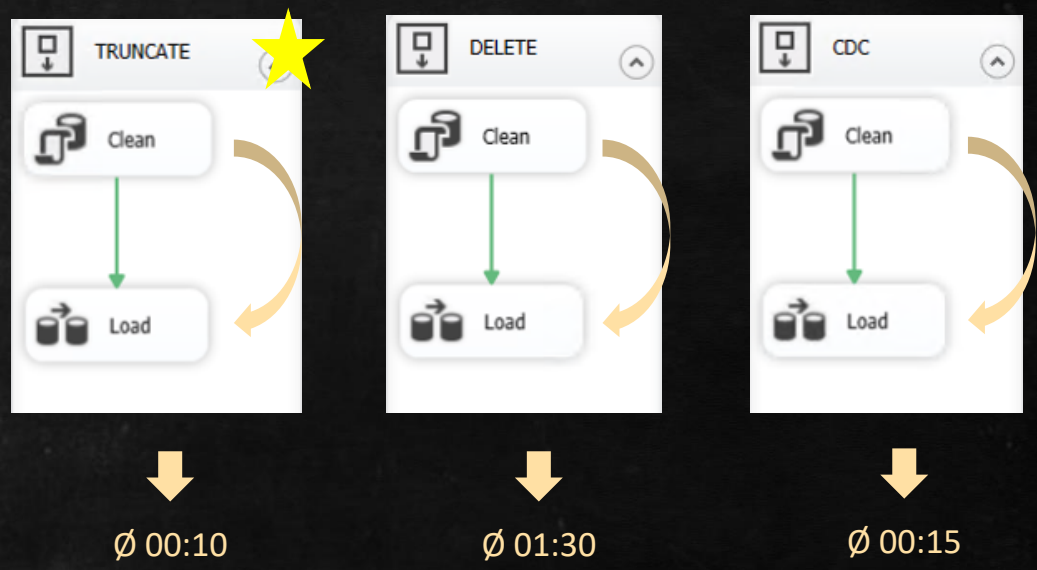
3. PATTERN EVALUATION

Pattern Evaluation

- Most tables can be loaded using different patterns (DELETE, TRUNCATE, CDC etc.) and/or connectors
- This patterns will usually show different performance results
- Their performance and therefore the “best pattern” may change over time
- We want to identify the best performing pattern and use it in our regular loads

Pattern Evaluation

- Execute each pattern n times
- Get the average execution time
- Mark the „best“ pattern



DEMO

PATTERN EVALUATION

4. PERFECT CURRENCY

Perfect Concurrency

- My Source and/or Target only allow a certain (reasonable) number of connections at a time
- I want all my dataflows to complete in total as fast as possible
- MaxConcurrent will not necessarily bring them in a good order
- I want small Dataflows to run in linear groups while larger tasks should run individually in parallel

Perfect Concurrency

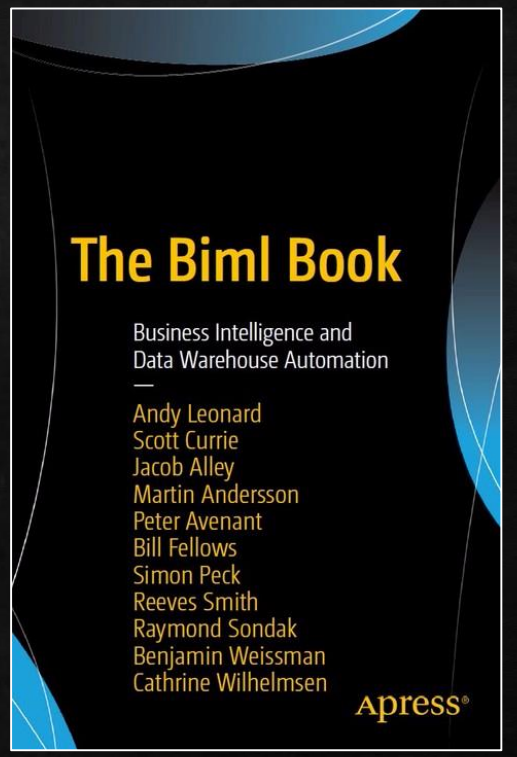
- Define the number of
  which results in SSIS c
- Get the average execu ataflow,
  ordered by time desc
- Put each table in the „sm lles container at
  that time and recalculate t e co tainer sizes
- This results in a more or less equ
  distribution of tables along the cont ers

DEMO

PERFECT CURRENCY

Want to learn more about Biml?

- http://biml.blog/
- http://www.bimlscript.com/
- http://www.cathrinewilhelmsen.net/biml/



QUESTIONS?

BEN WEISSMAN

@BWEISSMAN

If not now, feel free to reach out anytime!

b.weissman@solisyon.de

http://biml.blog/

FEEDBACK FORMS

PLEASE FILL OUT AND PASS TO YOUR ROOM
HELPER BEFORE YOU LEAVE THE SESSION

FRIDAY NIGHT, SQLBITS

PROHIBITION PARTY

THANK YOU!

BEN WEISSMAN

@BWEISSMAN