# Advanced Integration Services

Stacia Misner
blog.datainspirations.com
smisner@datainspirations.com

pluralsight
hardcore developer training

# Data Warehousing Packages, Part I

Designing Packages for Extract Transform & Load Projects

Stacia Misner
blog.datainspirations.com
smisner@datainspirations.com

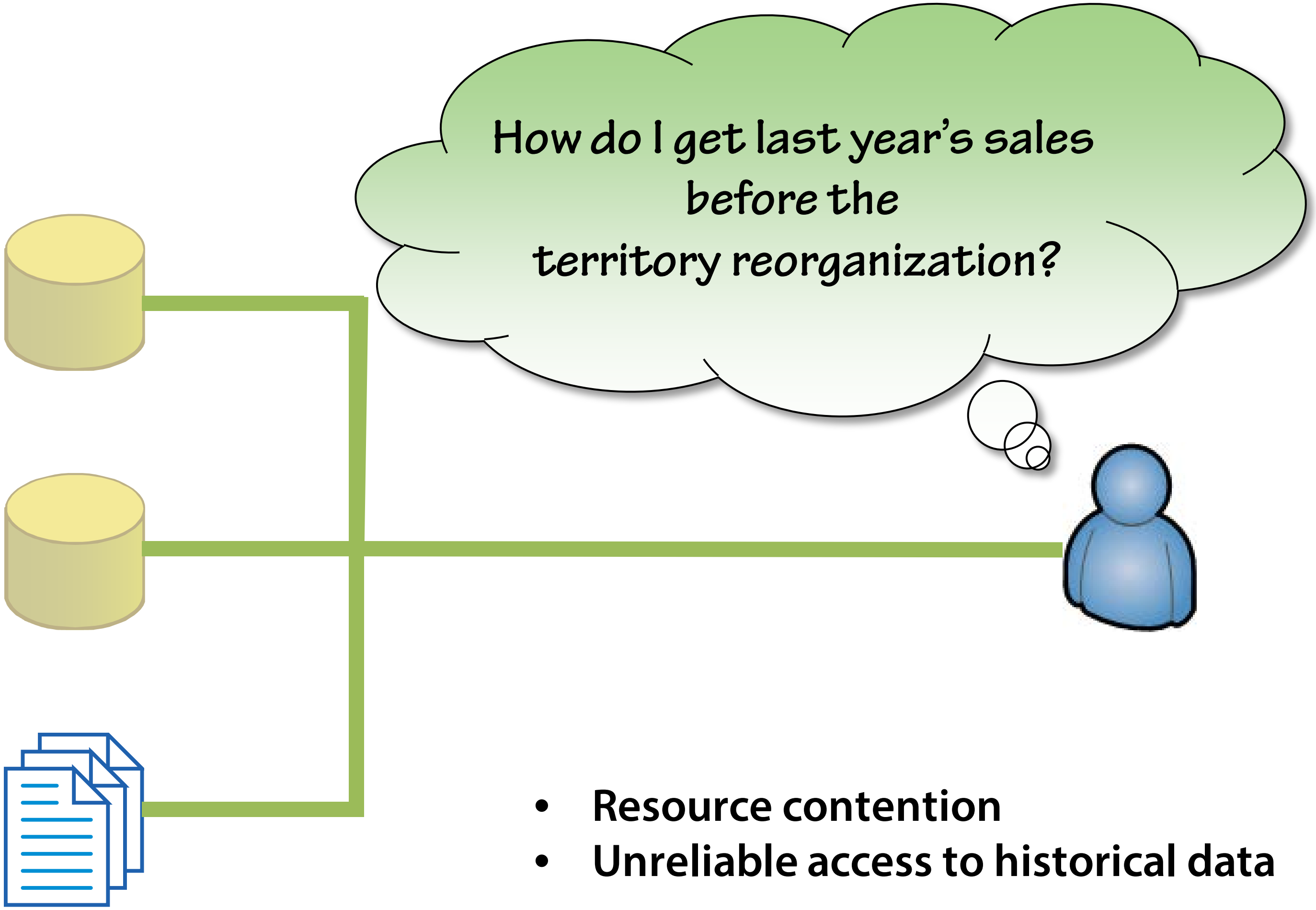
pluralsight
hardcore developer training

# Overview

- Introduction to Data Warehousing
- Dimensional Modeling
- Data Profiling
- ETL Design Patterns

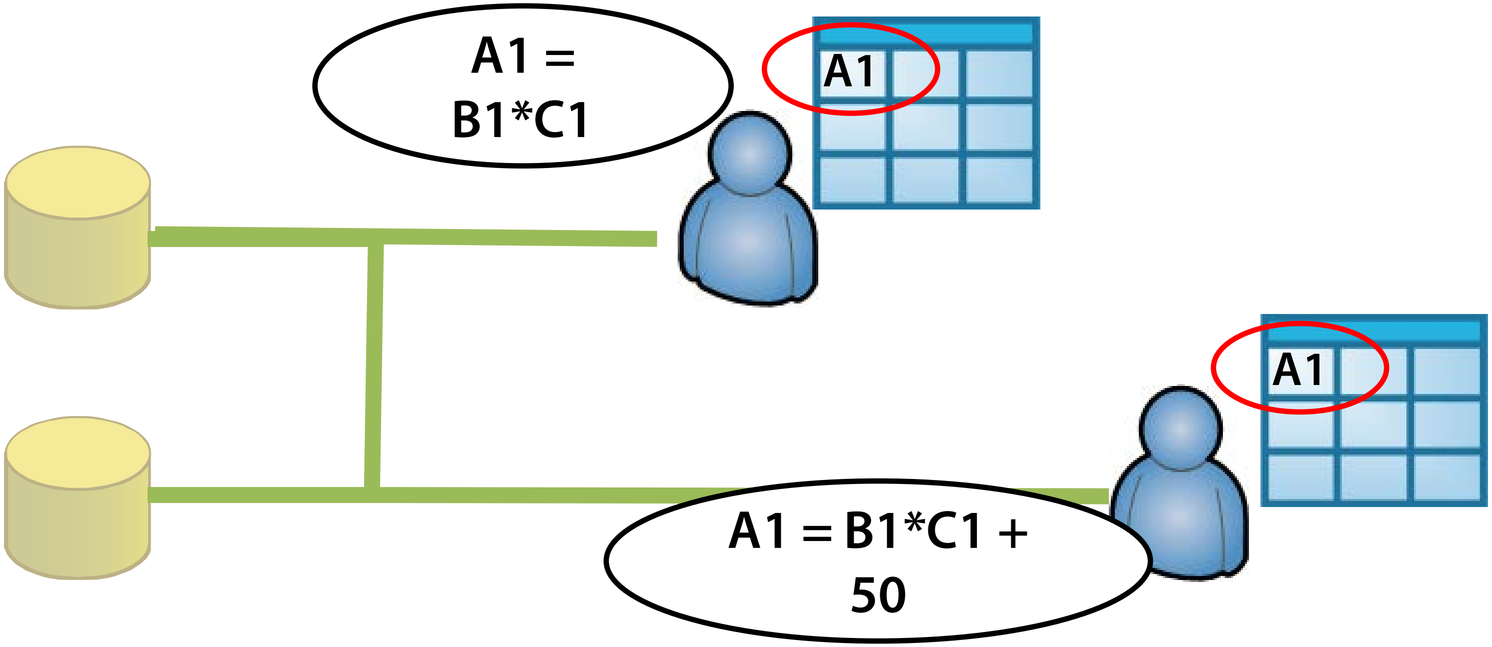
# Introduction to Data Warehousing

# What's the Problem?



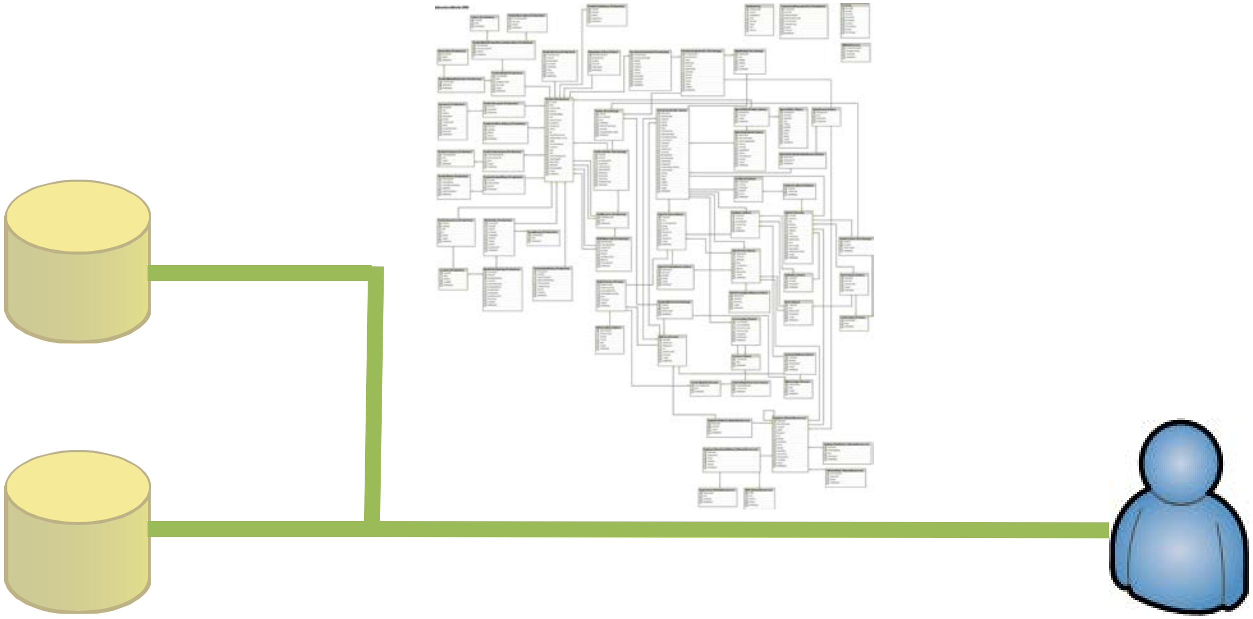
- Resource contention
- Unreliable access to historical data

# What's the Problem?



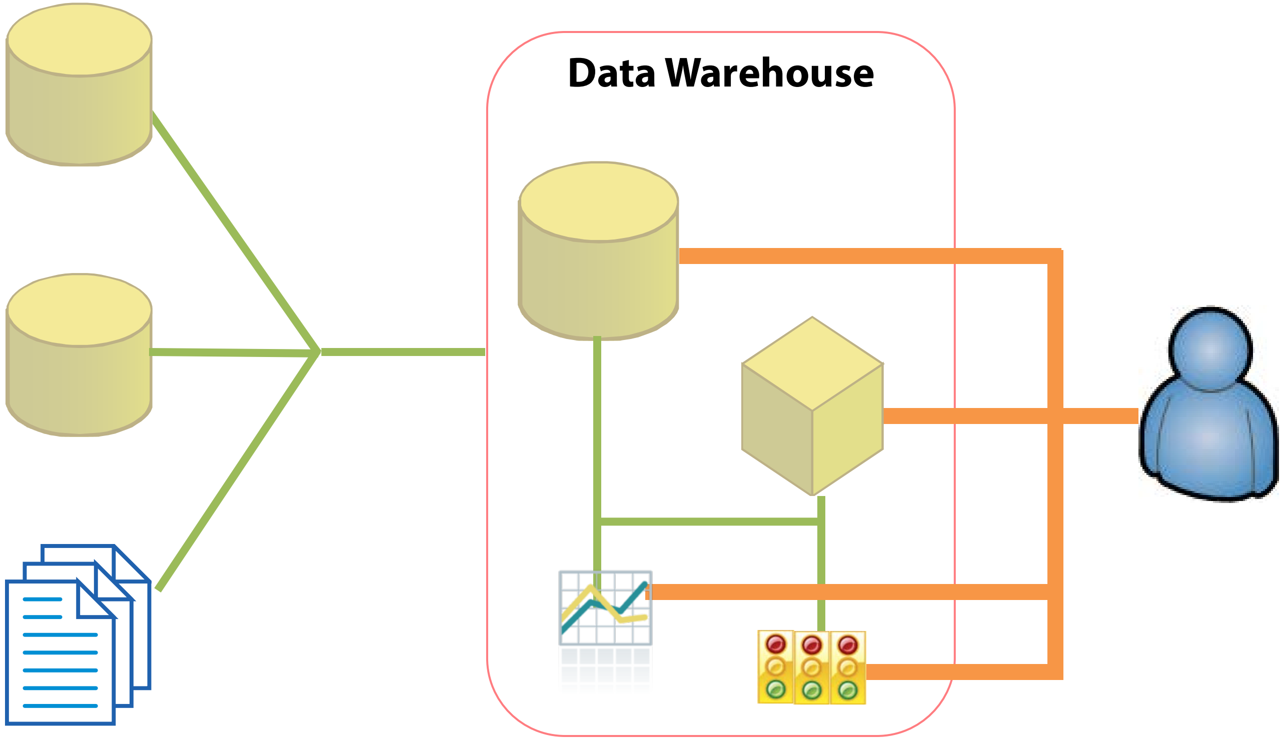
- Resource contention
- Unreliable access to historical data
- Inconsistent application of business rules

# What's the Problem?

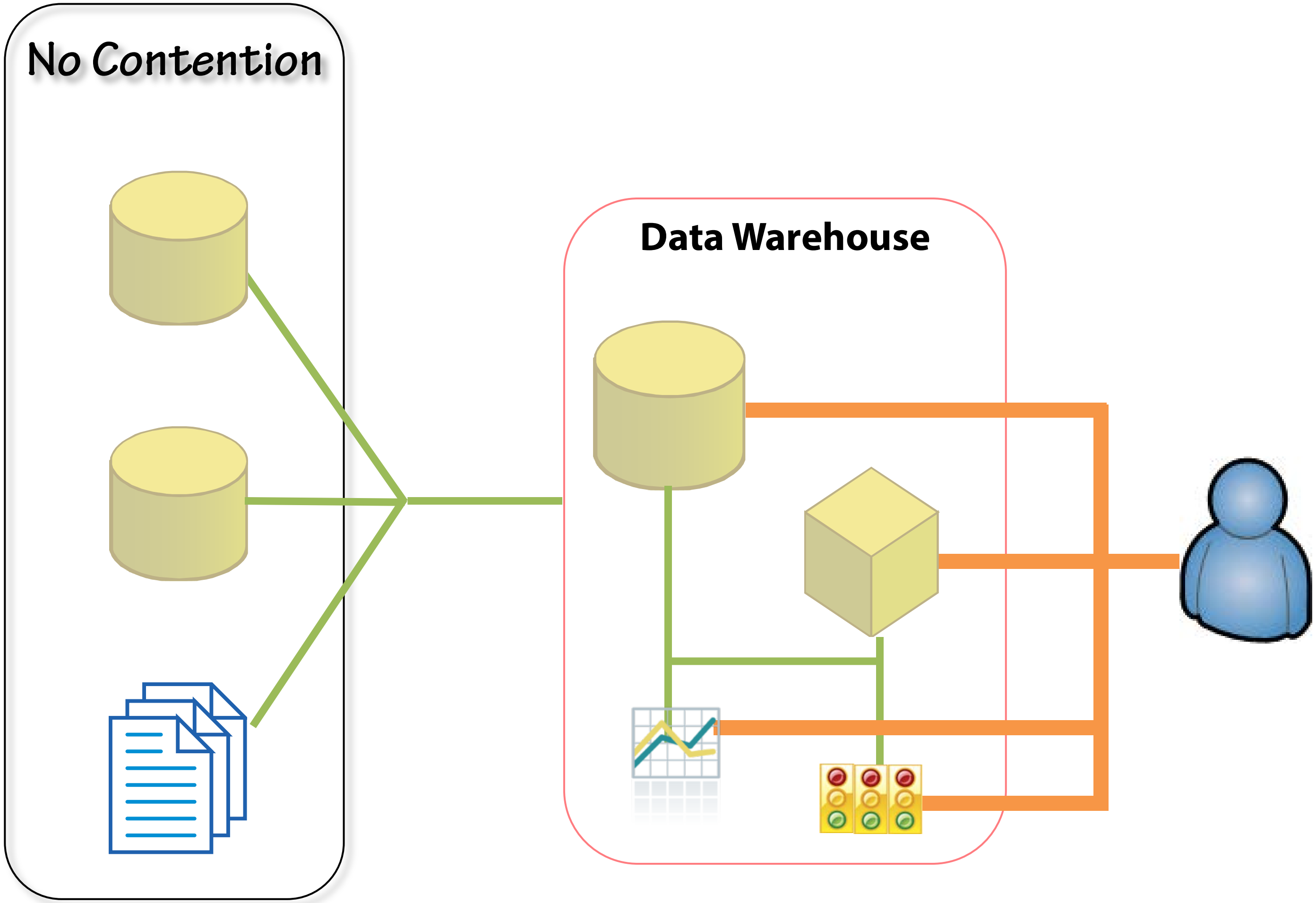


- Resource contention
- Unreliable access to historical data
- Inconsistent application of business rules
- Data structure results in slower, more complex queries

# The Data Warehouse Solution

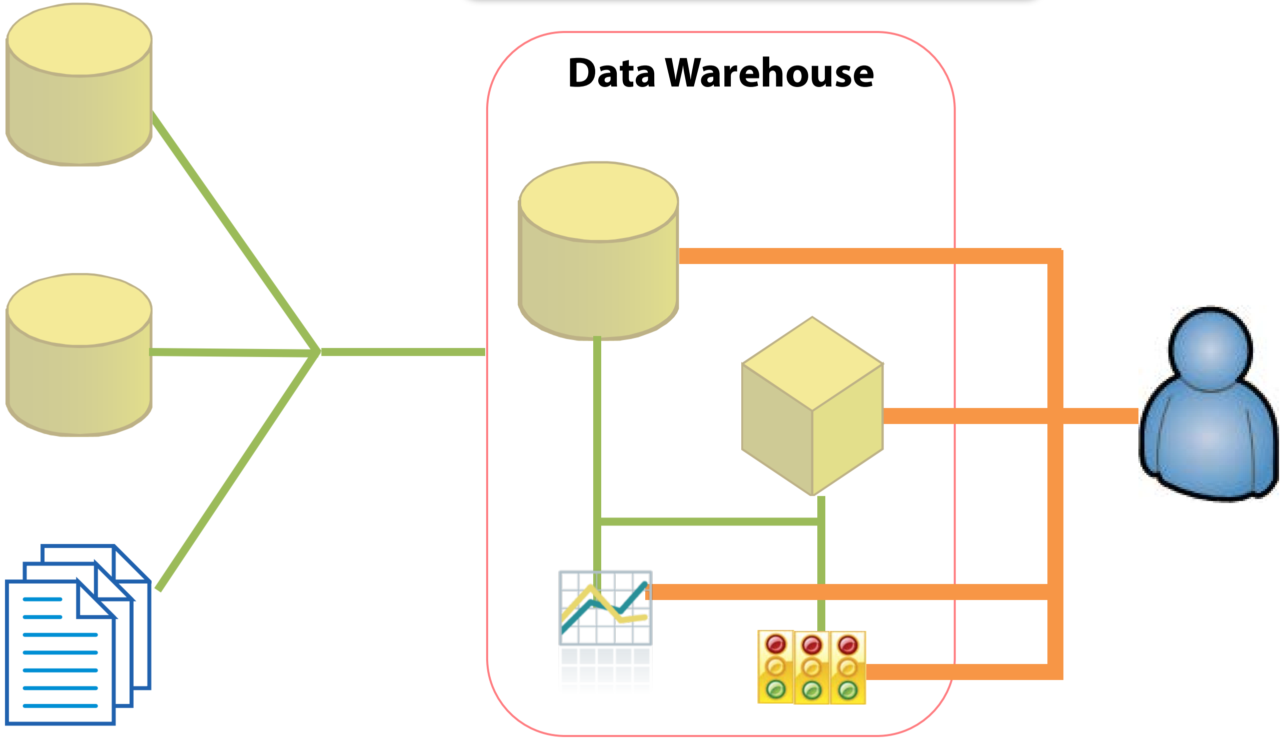


# The Data Warehouse Solution



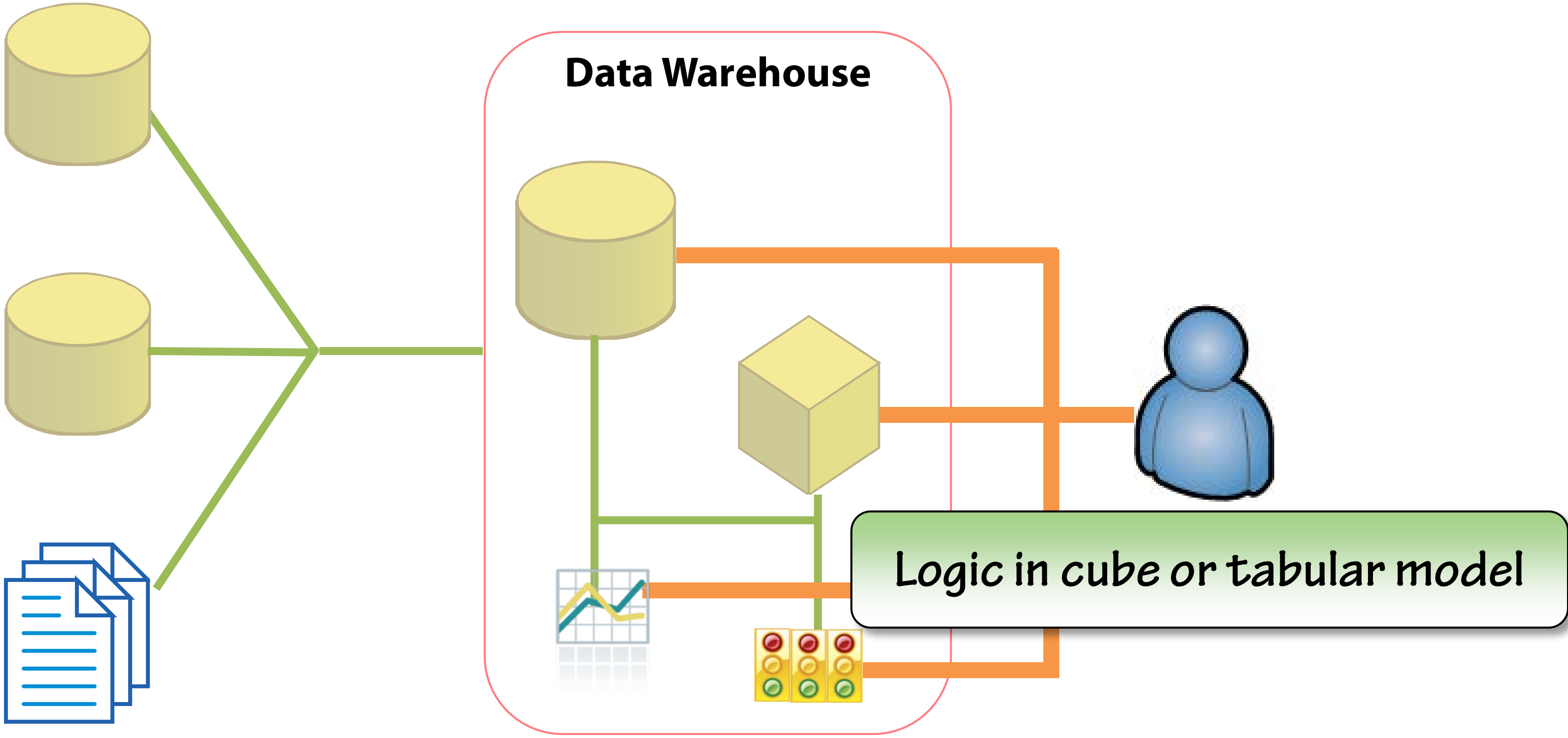
# The Data Warehouse Solution

Historical data preserved



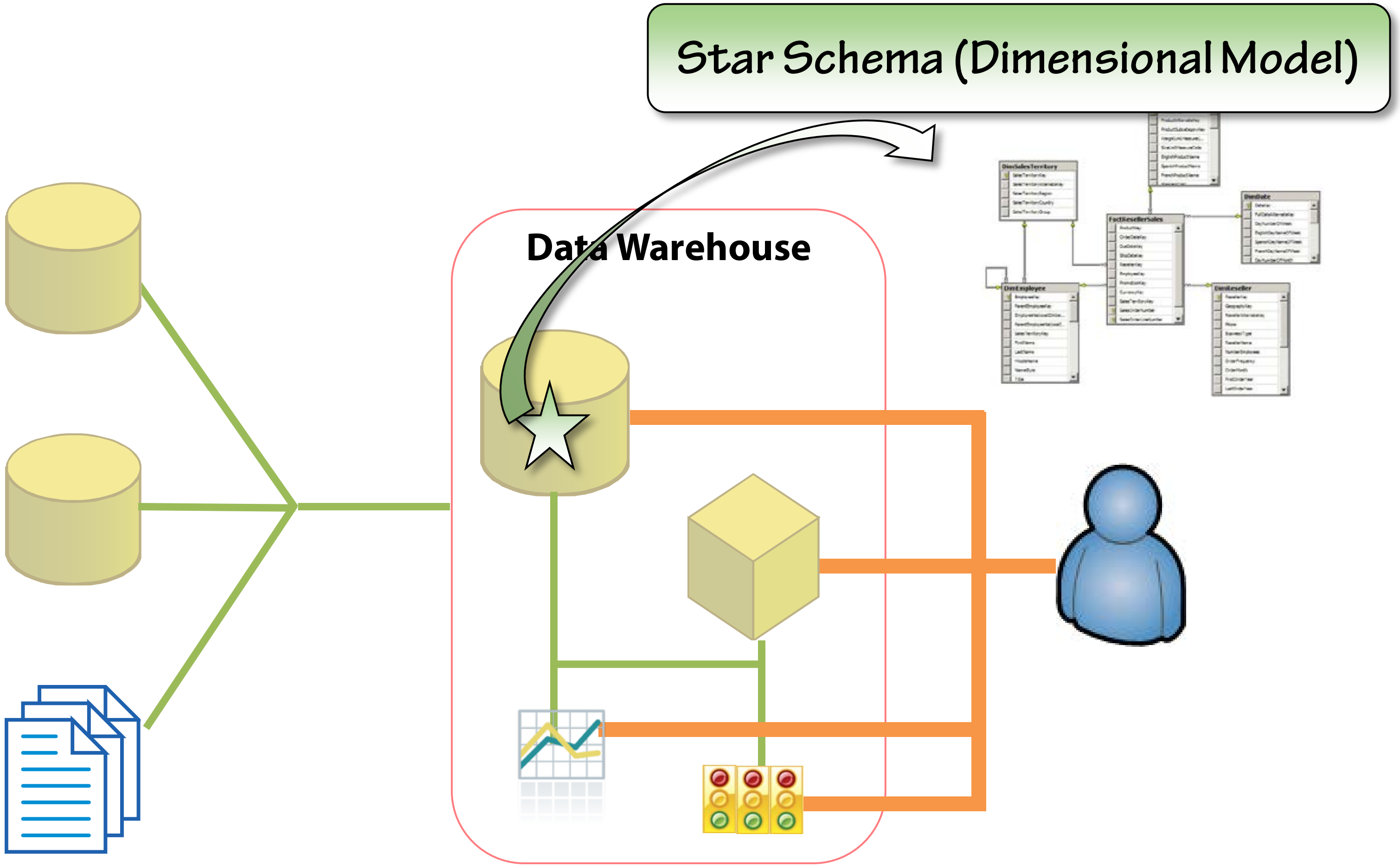
Data Warehouse

# The Data Warehouse Solution



Logic in cube or tabular model

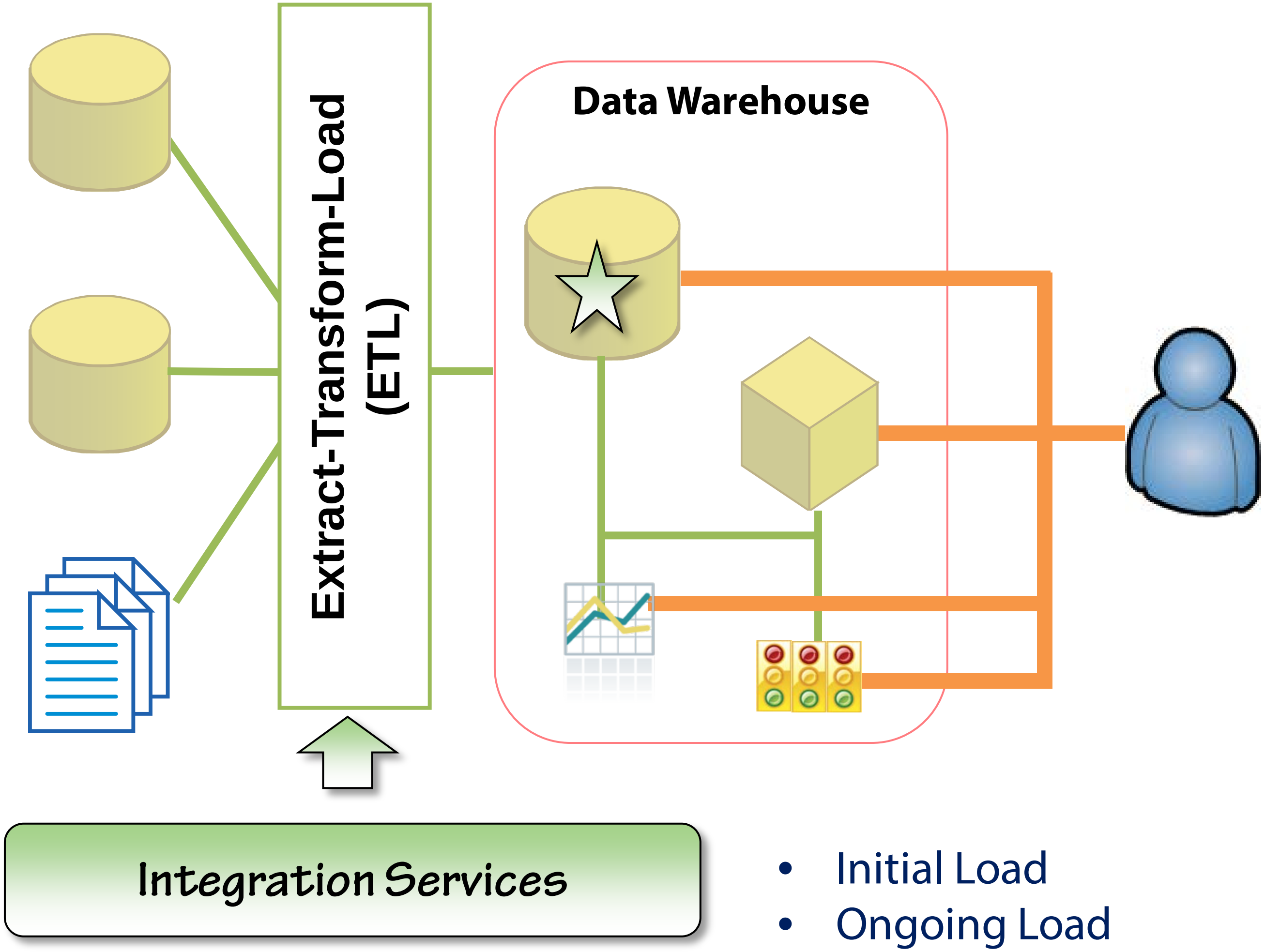
# The Data Warehouse Solution



Star Schema (Dimensional Model)

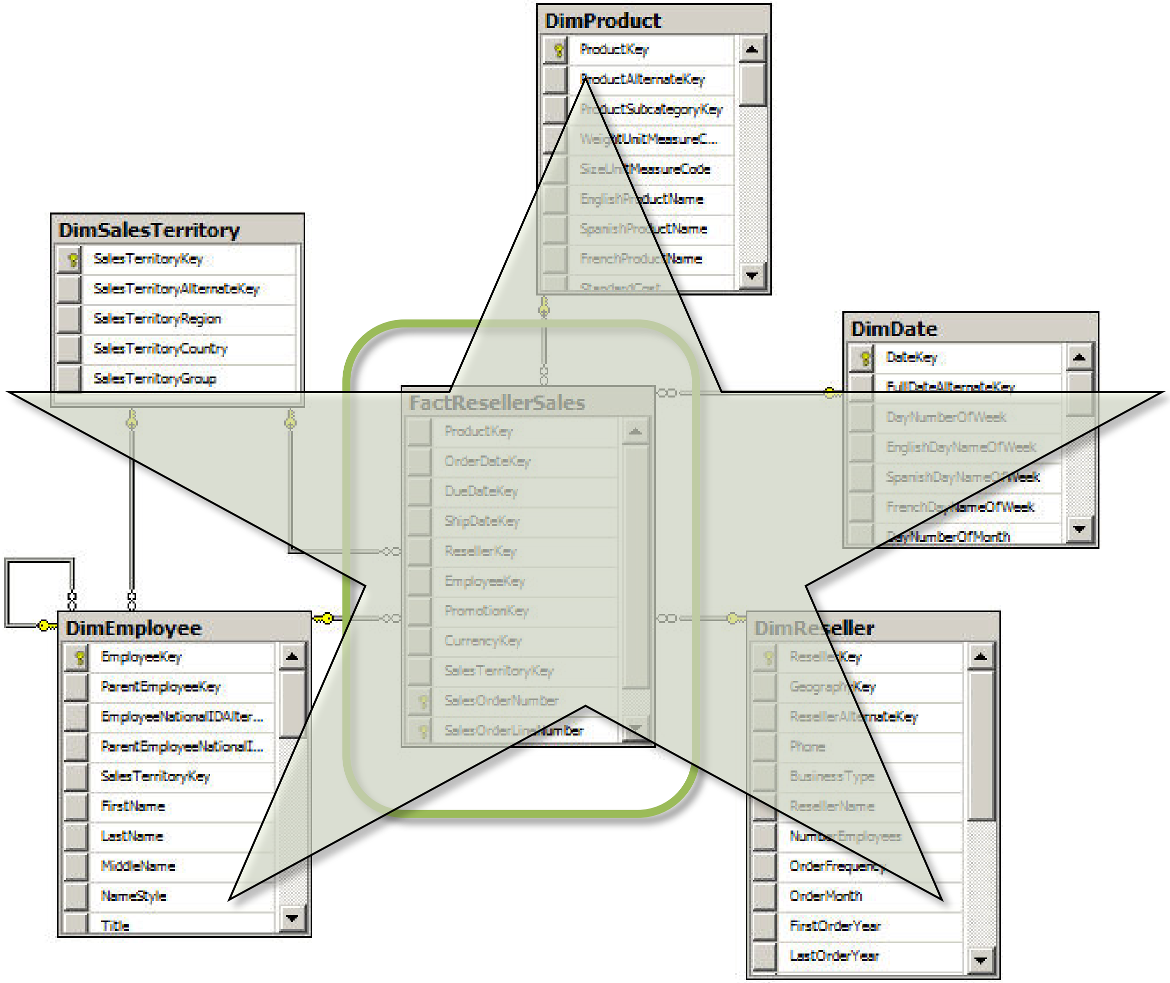
Data Warehouse

# The Data Warehouse Solution



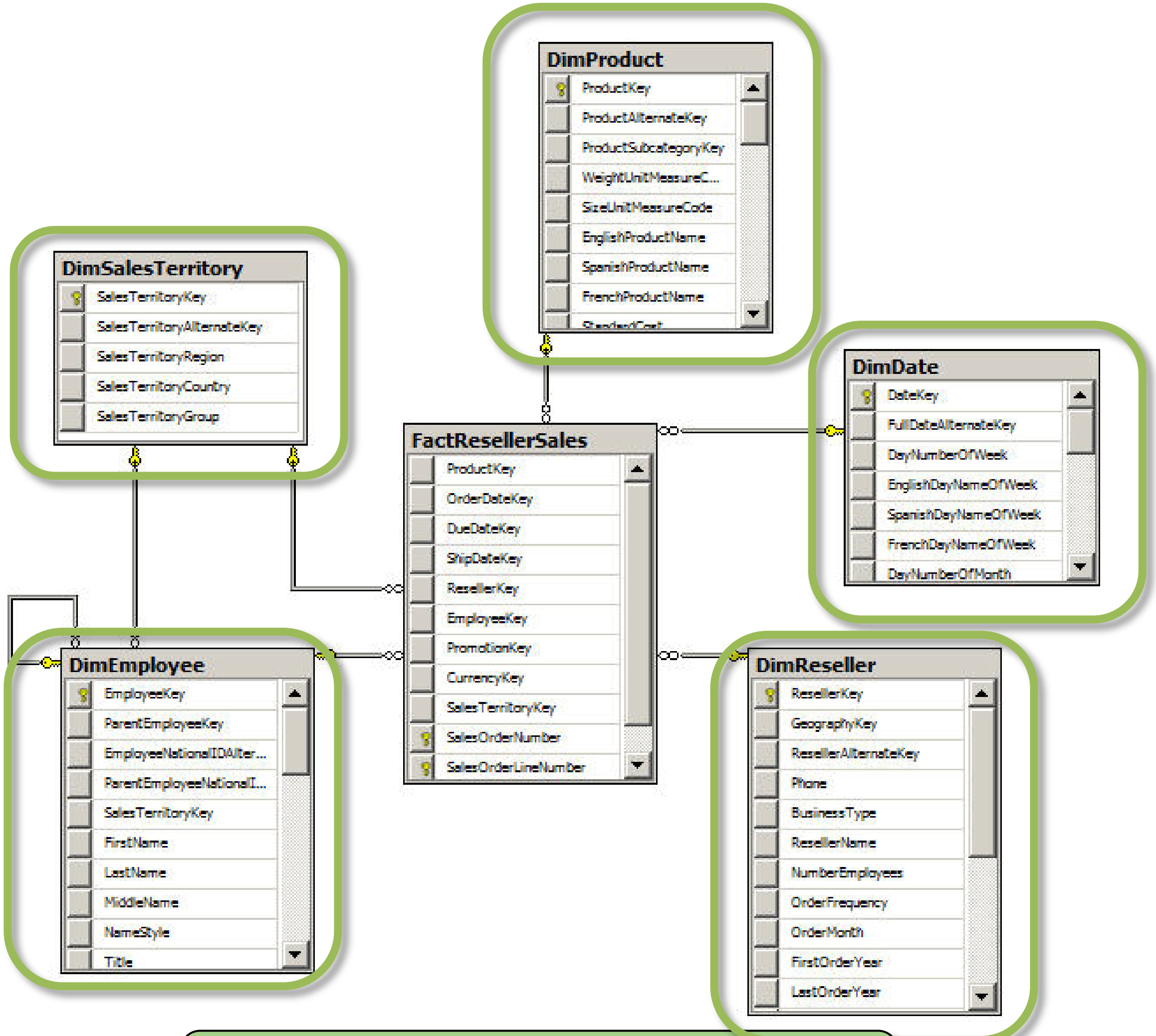
- Initial Load
- Ongoing Load

# Dimensional Modeling



Fact Table

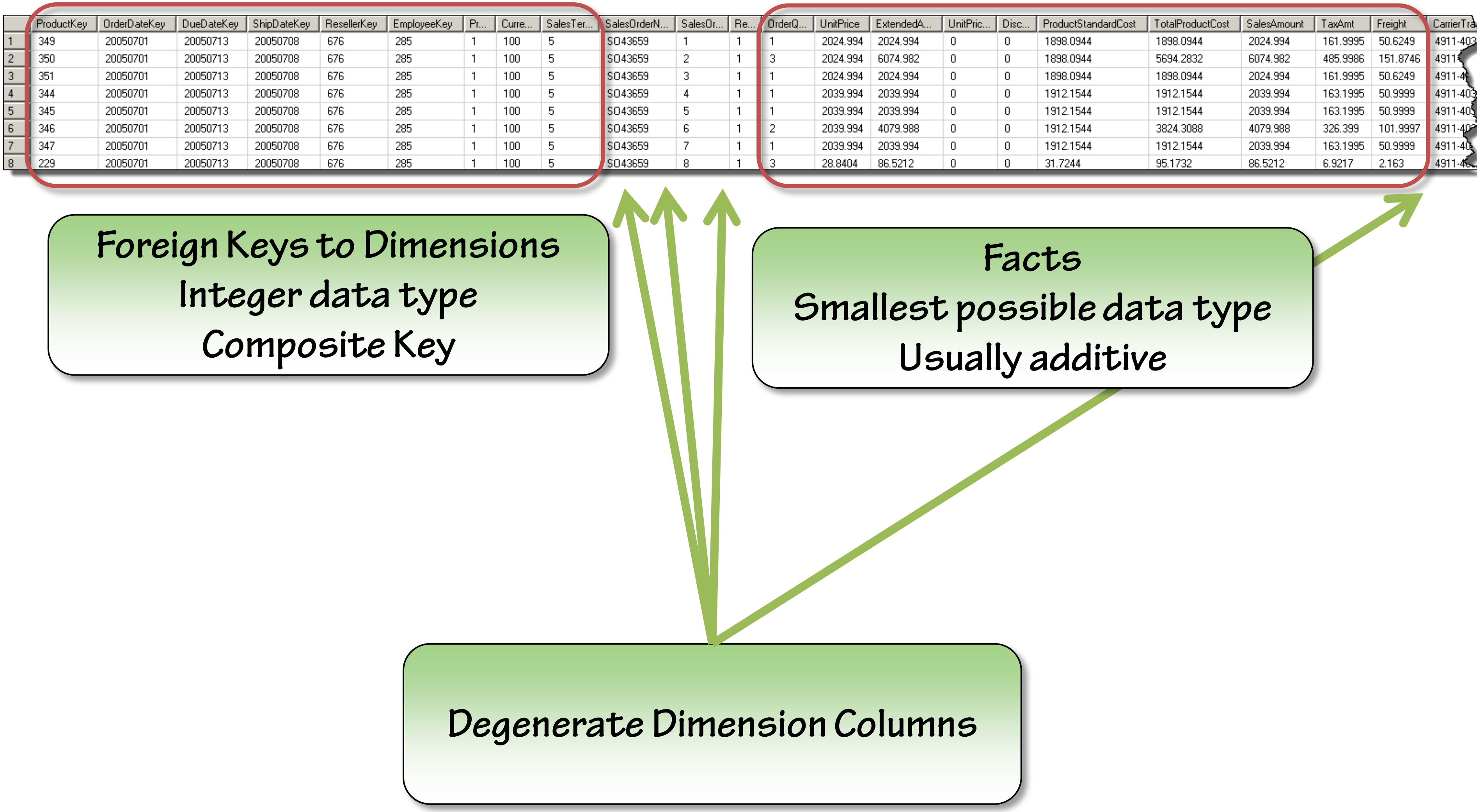
# Dimensional Modeling



Dimension Tables

# Dimensional Modeling

## Fact Table Characteristics



| | ProductKey | OrderDateKey | DueDateKey | ShipDateKey | ResellerKey | EmployeeKey | Pr... | Curre... | SalesTer... | SalesOrderN... | SalesOr... | Re... | OrderQ... | UnitPrice | ExtendedA... | UnitPric... | Disc... | ProductStandardCost | TotalProductCost | SalesAmount | TaxAmt | Freight | CarrierTra |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 349 | 20050701 | 20050713 | 20050708 | 676 | 285 | 1 | 100 | 5 | SO43659 | 1 | 1 | 1 | 2024.994 | 2024.994 | 0 | 0 | 1898.0944 | 1898.0944 | 2024.994 | 161.9995 | 50.6249 | 4911-403 |
| 2 | 350 | 20050701 | 20050713 | 20050708 | 676 | 285 | 1 | 100 | 5 | SO43659 | 2 | 1 | 3 | 2024.994 | 6074.982 | 0 | 0 | 1898.0944 | 5694.2832 | 6074.982 | 485.9986 | 151.8746 | 4911 |
| 3 | 351 | 20050701 | 20050713 | 20050708 | 676 | 285 | 1 | 100 | 5 | SO43659 | 3 | 1 | 1 | 2024.994 | 2024.994 | 0 | 0 | 1898.0944 | 1898.0944 | 2024.994 | 161.9995 | 50.6249 | 4911-4 |
| 4 | 344 | 20050701 | 20050713 | 20050708 | 676 | 285 | 1 | 100 | 5 | SO43659 | 4 | 1 | 1 | 2039.994 | 2039.994 | 0 | 0 | 1912.1544 | 1912.1544 | 2039.994 | 163.1995 | 50.9999 | 4911-403 |
| 5 | 345 | 20050701 | 20050713 | 20050708 | 676 | 285 | 1 | 100 | 5 | SO43659 | 5 | 1 | 1 | 2039.994 | 2039.994 | 0 | 0 | 1912.1544 | 1912.1544 | 2039.994 | 163.1995 | 50.9999 | 4911-40 |
| 6 | 346 | 20050701 | 20050713 | 20050708 | 676 | 285 | 1 | 100 | 5 | SO43659 | 6 | 1 | 2 | 2039.994 | 4079.988 | 0 | 0 | 1912.1544 | 3824.3088 | 4079.988 | 326.399 | 101.9997 | 4911-40 |
| 7 | 347 | 20050701 | 20050713 | 20050708 | 676 | 285 | 1 | 100 | 5 | SO43659 | 7 | 1 | 1 | 2039.994 | 2039.994 | 0 | 0 | 1912.1544 | 1912.1544 | 2039.994 | 163.1995 | 50.9999 | 4911-40 |
| 8 | 229 | 20050701 | 20050713 | 20050708 | 676 | 285 | 1 | 100 | 5 | SO43659 | 8 | 1 | 3 | 28.8404 | 86.5212 | 0 | 0 | 31.7244 | 95.1732 | 86.5212 | 6.9217 | 2.163 | 4911-4 |

Foreign Keys to Dimensions
Integer data type
Composite Key

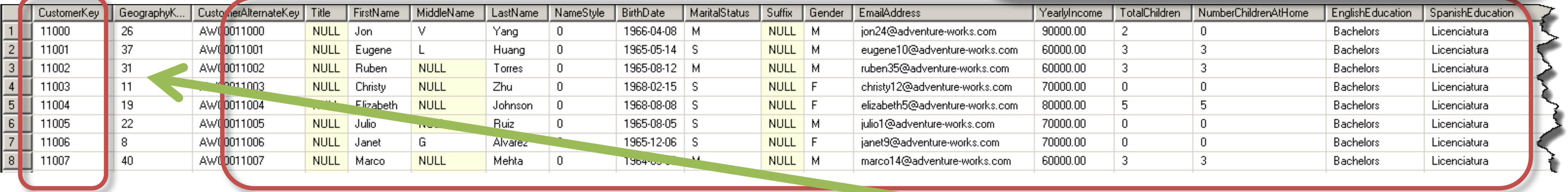
Facts
Smallest possible data type
Usually additive

Degenerate Dimension Columns

# Dimensional Modeling

## Dimension Table Characteristics

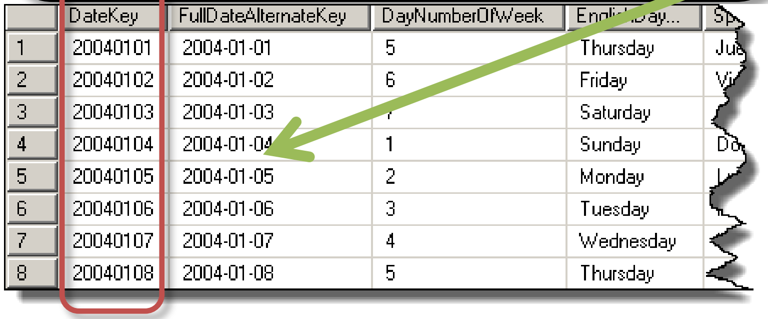
Dimension columns:
Labels, grouping, filtering, sorting, drilling
All data types allowed



| | CustomerKey | GeographyK... | CustomerAlternateKey | Title | FirstName | MiddleName | LastName | NameStyle | BirthDate | MaritalStatus | Suffix | Gender | EmailAddress | YearlyIncome | TotalChildren | NumberChildrenAtHome | EnglishEducation | SpanishEducation |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 11000 | 26 | AW00011000 | NULL | Jon | V | Yang | 0 | 1966-04-08 | M | NULL | M | jon24@adventure-works.com | 90000.00 | 2 | 0 | Bachelors | Licenciatura |
| 2 | 11001 | 37 | AW00011001 | NULL | Eugene | L | Huang | 0 | 1965-05-14 | S | NULL | M | eugene10@adventure-works.com | 60000.00 | 3 | 3 | Bachelors | Licenciatura |
| 3 | 11002 | 31 | AW00011002 | NULL | Ruben | NULL | Torres | 0 | 1965-08-12 | M | NULL | M | ruben35@adventure-works.com | 60000.00 | 3 | 3 | Bachelors | Licenciatura |
| 4 | 11003 | 11 | AW00011003 | NULL | Christy | NULL | Zhu | 0 | 1968-02-15 | S | NULL | F | christy12@adventure-works.com | 70000.00 | 0 | 0 | Bachelors | Licenciatura |
| 5 | 11004 | 19 | AW00011004 | NULL | Elizabeth | NULL | Johnson | 0 | 1968-08-08 | S | NULL | F | elizabeth5@adventure-works.com | 80000.00 | 5 | 5 | Bachelors | Licenciatura |
| 6 | 11005 | 22 | AW00011005 | NULL | Julio | NULL | Ruiz | 0 | 1965-08-05 | S | NULL | M | julio1@adventure-works.com | 70000.00 | 0 | 0 | Bachelors | Licenciatura |
| 7 | 11006 | 8 | AW00011006 | NULL | Janet | G | Alvarez | 0 | 1965-12-06 | S | NULL | F | janet9@adventure-works.com | 70000.00 | 0 | 0 | Bachelors | Licenciatura |
| 8 | 11007 | 40 | AW00011007 | NULL | Marco | NULL | Mehta | 0 | [illegible] | M | NULL | M | marco14@adventure-works.com | 60000.00 | 3 | 3 | Bachelors | Licenciatura |

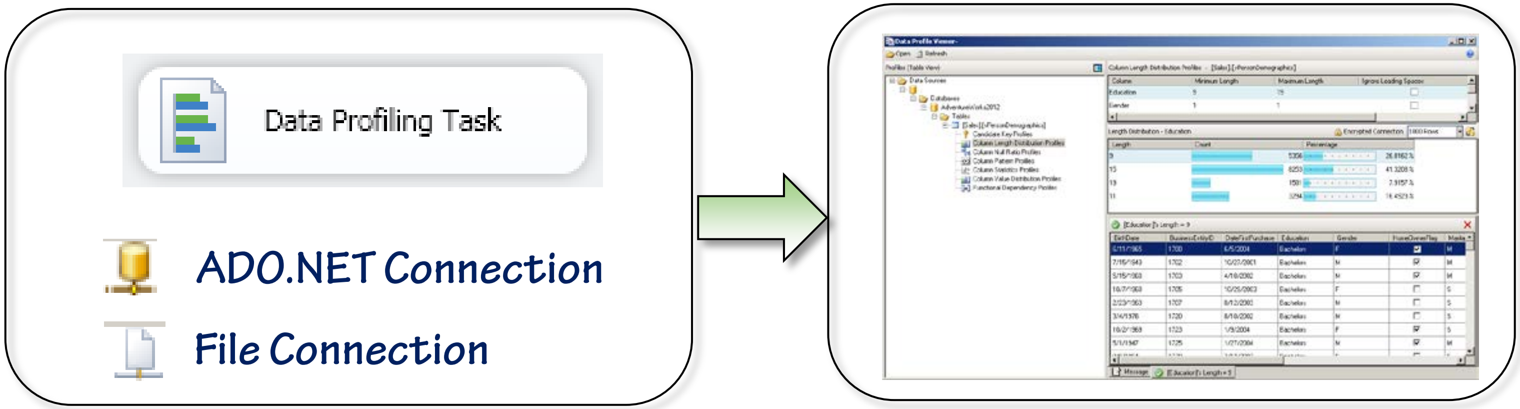
Primary key
Surrogate Key
or
YYYYMMDD for Date Dimension

Business or Natural Key



| | DateKey | FullDateAlternateKey | DayNumberOfWeek | EnglishDay... | Sp |
|---|---|---|---|---|---|
| 1 | 20040101 | 2004-01-01 | 5 | Thursday | Ju |
| 2 | 20040102 | 2004-01-02 | 6 | Friday | Vi |
| 3 | 20040103 | 2004-01-03 | 7 | Saturday | |
| 4 | 20040104 | 2004-01-04 | 1 | Sunday | Do |
| 5 | 20040105 | 2004-01-05 | 2 | Monday | L |
| 6 | 20040106 | 2004-01-06 | 3 | Tuesday | |
| 7 | 20040107 | 2004-01-07 | 4 | Wednesday | |
| 8 | 20040108 | 2004-01-08 | 5 | Thursday | |

# Data Profiling



Data Profiling Task

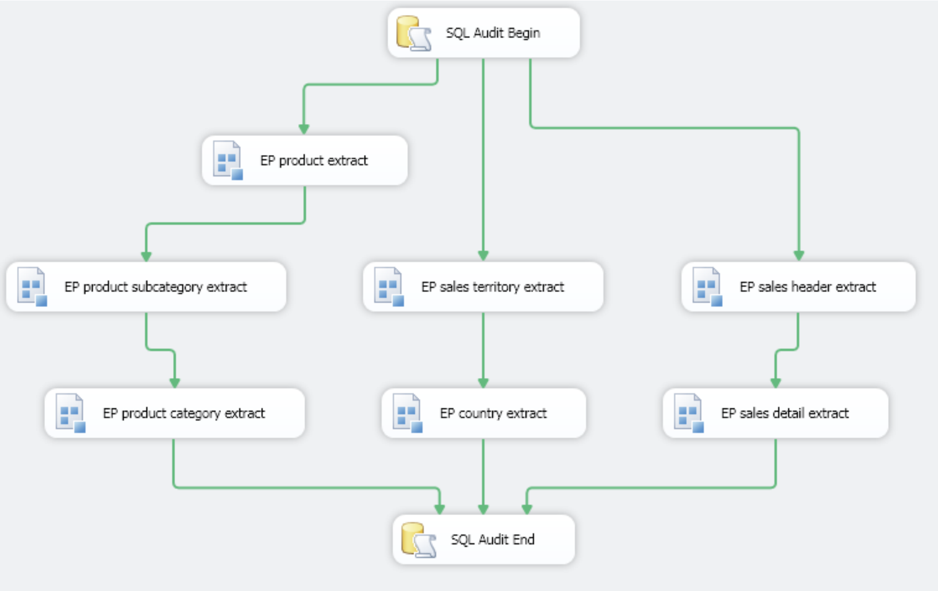
ADO.NET Connection

File Connection

| Profile  Type | Description |
|---|---|
| Candidate Key Profile | Uniqueness of values and violations |
| Column Length Distribution Profile | Distinct length of column values as percentage of total rows |
| Column Null Ratio Profile | Ratio of NULL records in column to overall |
| Column Pattern Profile | Regular expressions for found patterns |
| Column Statistics Profile | Min, max, mean, and std dev |
| Column Value Distribution Profile | Distinct values as count and percent |
| Functional Dependency Profile | Strength of dependency between columns |
| Value Inclusion Profile | Existence of value in lookup table |

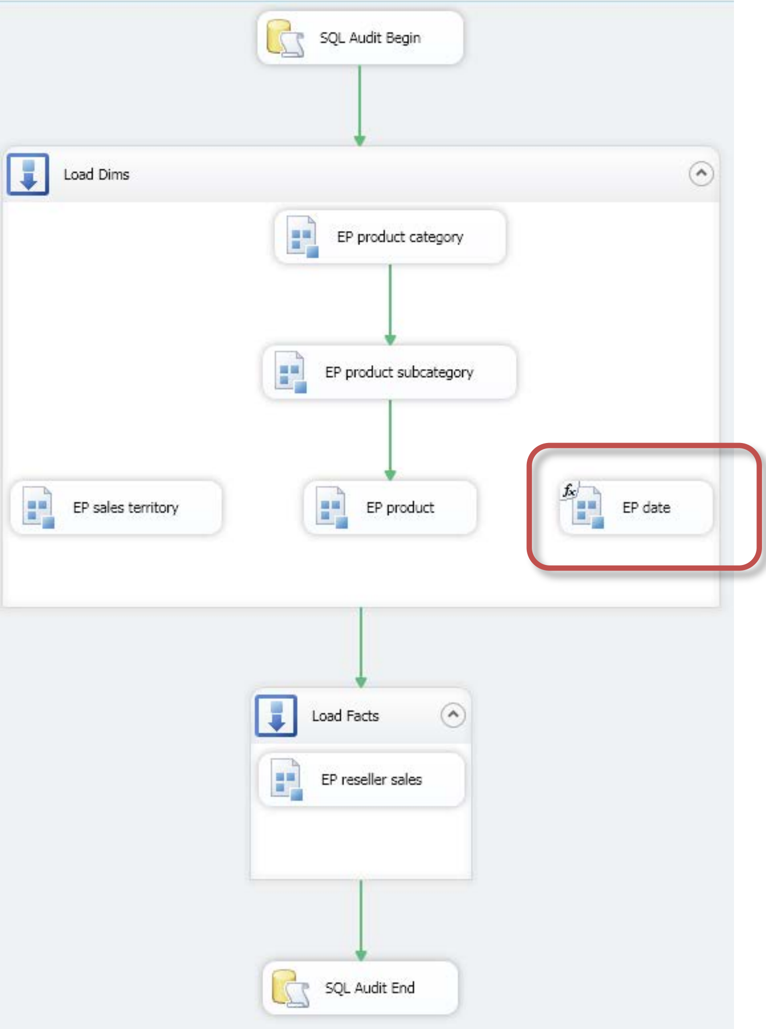
# ETL Design Patterns

## Master Extract Package

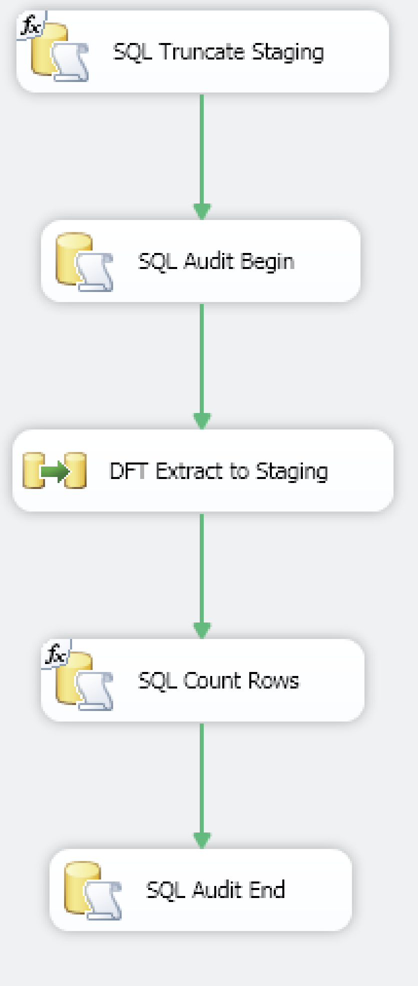


Create snapshot of database to simplify error recovery

## Master Transform-Load Package



# Extract Package



Truncate staging before extraction

Add record to audit table

Extract data from source and load into staging table

Count records in staging table

Update audit record

# Extract Data Flow

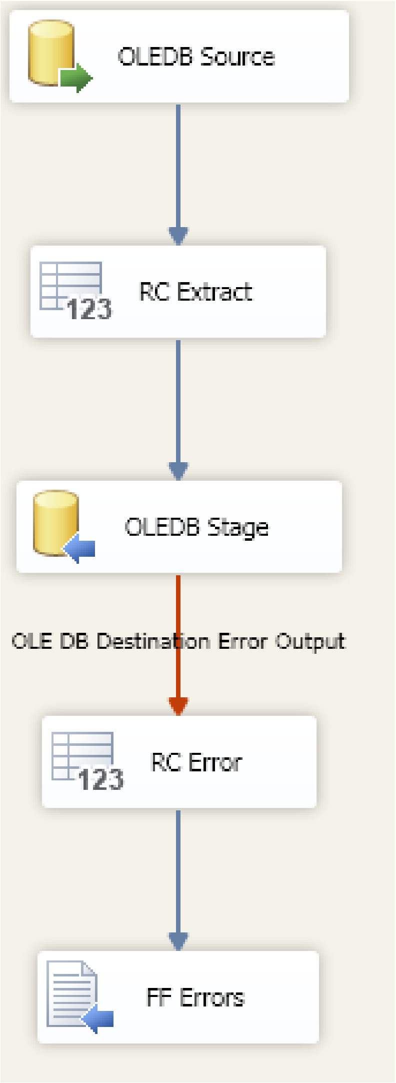
Connect to source

Store extracted row count in variable

Load extracted records into staging table

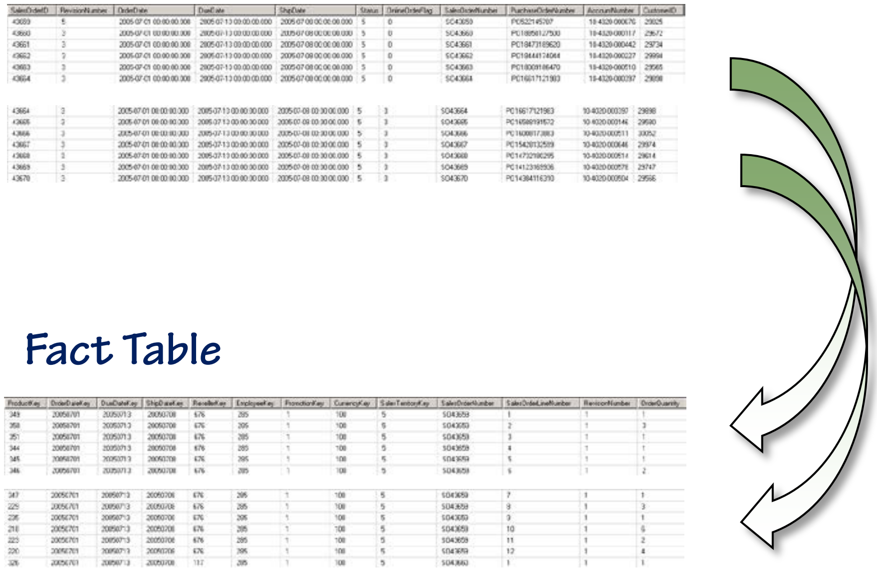
Store error row count in variable

Save error records



# Load Patterns

## Fact Loads



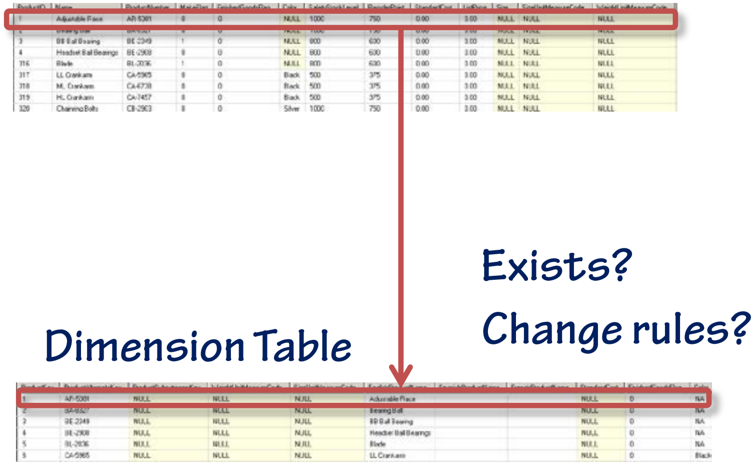
Source

| SalesOrderID | RevisionNumber | OrderDate | DueDate | ShipDate | Status | OnlineOrderFlag | SalesOrderNumber | PurchaseOrderNumber | AccountNumber | CustomerID |
|---|---|---|---|---|---|---|---|---|---|---|
| 43659 | 5 | 2005-07-01 00:00:00.000 | 2005-07-13 00:00:00.000 | 2005-07-08 00:00:00.000 | 5 | 0 | SO43659 | PO522145787 | 10-4020-000676 | 29825 |
| 43660 | 3 | 2005-07-01 00:00:00.000 | 2005-07-13 00:00:00.000 | 2005-07-08 00:00:00.000 | 5 | 0 | SO43660 | PO18850127500 | 10-4020-000117 | 29672 |
| 43661 | 3 | 2005-07-01 00:00:00.000 | 2005-07-13 00:00:00.000 | 2005-07-08 00:00:00.000 | 5 | 0 | SO43661 | PO18473189620 | 10-4020-000442 | 29734 |
| 43662 | 3 | 2005-07-01 00:00:00.000 | 2005-07-13 00:00:00.000 | 2005-07-08 00:00:00.000 | 5 | 0 | SO43662 | PO18444174044 | 10-4020-000227 | 29994 |
| 43663 | 3 | 2005-07-01 00:00:00.000 | 2005-07-13 00:00:00.000 | 2005-07-08 00:00:00.000 | 5 | 0 | SO43663 | PO18009186470 | 10-4020-000510 | 29565 |
| 43664 | 3 | 2005-07-01 00:00:00.000 | 2005-07-13 00:00:00.000 | 2005-07-08 00:00:00.000 | 5 | 0 | SO43664 | PO16617121983 | 10-4020-000397 | 29898 |
| 43664 | 3 | 2005-07-01 00:00:00.000 | 2005-07-13 00:00:00.000 | 2005-07-08 00:00:00.000 | 5 | 0 | SO43664 | PO16617121983 | 10-4020-000397 | 29898 |
| 43665 | 3 | 2005-07-01 00:00:00.000 | 2005-07-13 00:00:00.000 | 2005-07-08 00:00:00.000 | 5 | 0 | SO43665 | PO16588191572 | 10-4020-000146 | 29580 |
| 43666 | 3 | 2005-07-01 00:00:00.000 | 2005-07-13 00:00:00.000 | 2005-07-08 00:00:00.000 | 5 | 0 | SO43666 | PO16008173883 | 10-4020-000511 | 30052 |
| 43667 | 3 | 2005-07-01 00:00:00.000 | 2005-07-13 00:00:00.000 | 2005-07-08 00:00:00.000 | 5 | 0 | SO43667 | PO15428132599 | 10-4020-000646 | 29974 |
| 43668 | 3 | 2005-07-01 00:00:00.000 | 2005-07-13 00:00:00.000 | 2005-07-08 00:00:00.000 | 5 | 0 | SO43668 | PO14732180295 | 10-4020-000514 | 29614 |
| 43669 | 3 | 2005-07-01 00:00:00.000 | 2005-07-13 00:00:00.000 | 2005-07-08 00:00:00.000 | 5 | 0 | SO43669 | PO14123169936 | 10-4020-000578 | 29747 |
| 43670 | 3 | 2005-07-01 00:00:00.000 | 2005-07-13 00:00:00.000 | 2005-07-08 00:00:00.000 | 5 | 0 | SO43670 | PO14384116310 | 10-4020-000504 | 29566 |

Fact Table

| ProductKey | OrderDateKey | DueDateKey | ShipDateKey | ResellerKey | EmployeeKey | PromotionKey | CurrencyKey | SalesTerritoryKey | SalesOrderNumber | SalesOrderLineNumber | RevisionNumber | OrderQuantity |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 349 | 20050701 | 20050713 | 20050708 | 676 | 285 | 1 | 100 | 5 | SO43659 | 1 | 1 | 1 |
| 350 | 20050701 | 20050713 | 20050708 | 676 | 285 | 1 | 100 | 5 | SO43659 | 2 | 1 | 3 |
| 351 | 20050701 | 20050713 | 20050708 | 676 | 285 | 1 | 100 | 5 | SO43659 | 3 | 1 | 1 |
| 344 | 20050701 | 20050713 | 20050708 | 676 | 285 | 1 | 100 | 5 | SO43659 | 4 | 1 | 1 |
| 345 | 20050701 | 20050713 | 20050708 | 676 | 285 | 1 | 100 | 5 | SO43659 | 5 | 1 | 1 |
| 346 | 20050701 | 20050713 | 20050708 | 676 | 285 | 1 | 100 | 5 | SO43659 | 6 | 1 | 2 |
| 347 | 20050701 | 20050713 | 20050708 | 676 | 285 | 1 | 100 | 5 | SO43659 | 7 | 1 | 1 |
| 229 | 20050701 | 20050713 | 20050708 | 676 | 285 | 1 | 100 | 5 | SO43659 | 8 | 1 | 3 |
| 235 | 20050701 | 20050713 | 20050708 | 676 | 285 | 1 | 100 | 5 | SO43659 | 9 | 1 | 1 |
| 218 | 20050701 | 20050713 | 20050708 | 676 | 285 | 1 | 100 | 5 | SO43659 | 10 | 1 | 6 |
| 223 | 20050701 | 20050713 | 20050708 | 676 | 285 | 1 | 100 | 5 | SO43659 | 11 | 1 | 2 |
| 220 | 20050701 | 20050713 | 20050708 | 676 | 285 | 1 | 100 | 5 | SO43659 | 12 | 1 | 4 |
| 326 | 20050701 | 20050713 | 20050708 | 117 | 285 | 1 | 100 | 5 | SO43660 | 1 | 1 | 1 |

## Dimension Loads



Source

| ProductID | Name | ProductNumber | MakeFlag | FinishedGoodsFlag | Color | SafetyStockLevel | ReorderPoint | StandardCost | ListPrice | Size | SizeUnitMeasureCode | WeightUnitMeasureCode |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Adjustable Race | AR-5381 | 0 | 0 | NULL | 1000 | 750 | 0.00 | 0.00 | NULL | NULL | NULL |
| 2 | Bearing Ball | BA-8327 | 0 | 0 | NULL | 1000 | 750 | 0.00 | 0.00 | NULL | NULL | NULL |
| 3 | BB Ball Bearing | BE-2349 | 1 | 0 | NULL | 800 | 600 | 0.00 | 0.00 | NULL | NULL | NULL |
| 4 | Headset Ball Bearings | BE-2908 | 0 | 0 | NULL | 800 | 600 | 0.00 | 0.00 | NULL | NULL | NULL |
| 316 | Blade | BL-2036 | 1 | 0 | NULL | 800 | 600 | 0.00 | 0.00 | NULL | NULL | NULL |
| 317 | LL Crankarm | CA-5965 | 0 | 0 | Black | 500 | 375 | 0.00 | 0.00 | NULL | NULL | NULL |
| 318 | ML Crankarm | CA-6738 | 0 | 0 | Black | 500 | 375 | 0.00 | 0.00 | NULL | NULL | NULL |
| 319 | HL Crankarm | CA-7457 | 0 | 0 | Black | 500 | 375 | 0.00 | 0.00 | NULL | NULL | NULL |
| 320 | Chainring Bolts | CB-2903 | 0 | 0 | Silver | 1000 | 750 | 0.00 | 0.00 | NULL | NULL | NULL |

Exists?

Change rules?

Dimension Table

| ProductKey | ProductAlternateKey | ProductSubcategoryKey | WeightUnitMeasureCode | SizeUnitMeasureCode | EnglishProductName | SpanishProductName | FrenchProductName | StandardCost | FinishedGoodsFlag | Color |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | AR-5381 | NULL | NULL | NULL | Adjustable Race | | | NULL | 0 | NA |
| 2 | BA-8327 | NULL | NULL | NULL | Bearing Ball | | | NULL | 0 | NA |
| 3 | BE-2349 | NULL | NULL | NULL | BB Ball Bearing | | | NULL | 0 | NA |
| 4 | BE-2908 | NULL | NULL | NULL | Headset Ball Bearings | | | NULL | 0 | NA |
| 5 | BL-2036 | NULL | NULL | NULL | Blade | | | NULL | 0 | NA |
| 6 | CA-5965 | NULL | NULL | NULL | LL Crankarm | | | NULL | 0 | Black |

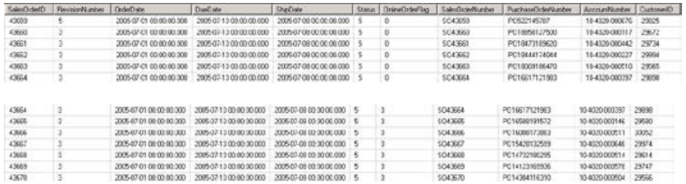
Extract only new records
and load into
transactional fact table

Extract all records
Exists? Update if changes (maybe)

# Load Patterns

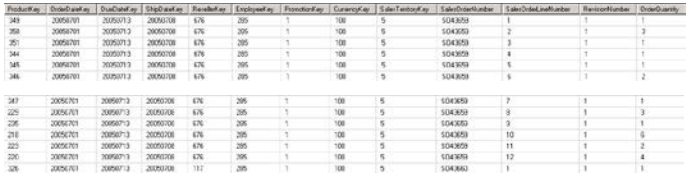
## Fact Loads

### Source



| SalesOrderID | RevisionNumber | OrderDate | DueDate | ShipDate | Status | OnlineOrderFlag | SalesOrderNumber | PurchaseOrderNumber | AccountNumber | CustomerID |
|---|---|---|---|---|---|---|---|---|---|---|
| 43659 | 5 | 2005-07-01 00:00:00.000 | 2005-07-13 00:00:00.000 | 2005-07-08 00:00:00.000 | 5 | 0 | SO43659 | PO522145787 | 10-4020-000676 | 29825 |
| 43660 | 3 | 2005-07-01 00:00:00.000 | 2005-07-13 00:00:00.000 | 2005-07-08 00:00:00.000 | 5 | 0 | SO43660 | PO18850127500 | 10-4020-000117 | 29672 |
| 43661 | 3 | 2005-07-01 00:00:00.000 | 2005-07-13 00:00:00.000 | 2005-07-08 00:00:00.000 | 5 | 0 | SO43661 | PO18473189620 | 10-4020-000442 | 29734 |
| 43662 | 3 | 2005-07-01 00:00:00.000 | 2005-07-13 00:00:00.000 | 2005-07-08 00:00:00.000 | 5 | 0 | SO43662 | PO18444174044 | 10-4020-000227 | 29994 |
| 43663 | 3 | 2005-07-01 00:00:00.000 | 2005-07-13 00:00:00.000 | 2005-07-08 00:00:00.000 | 5 | 0 | SO43663 | PO18009186470 | 10-4020-000510 | 29565 |
| 43664 | 3 | 2005-07-01 00:00:00.000 | 2005-07-13 00:00:00.000 | 2005-07-08 00:00:00.000 | 5 | 0 | SO43664 | PO16617121983 | 10-4020-000397 | 29898 |
| | | | | | | | | | | |
| 43664 | 3 | 2005-07-01 00:00:00.000 | 2005-07-13 00:00:00.000 | 2005-07-08 00:00:00.000 | 5 | 0 | SO43664 | PO16617121983 | 10-4020-000397 | 29898 |
| 43665 | 3 | 2005-07-01 00:00:00.000 | 2005-07-13 00:00:00.000 | 2005-07-08 00:00:00.000 | 5 | 0 | SO43665 | PO16588191572 | 10-4020-000146 | 29580 |
| 43666 | 3 | 2005-07-01 00:00:00.000 | 2005-07-13 00:00:00.000 | 2005-07-08 00:00:00.000 | 5 | 0 | SO43666 | PO16008173883 | 10-4020-000511 | 30052 |
| 43667 | 3 | 2005-07-01 00:00:00.000 | 2005-07-13 00:00:00.000 | 2005-07-08 00:00:00.000 | 5 | 0 | SO43667 | PO15428132599 | 10-4020-000646 | 29974 |
| 43668 | 3 | 2005-07-01 00:00:00.000 | 2005-07-13 00:00:00.000 | 2005-07-08 00:00:00.000 | 5 | 0 | SO43668 | PO14732180295 | 10-4020-000514 | 29614 |
| 43669 | 3 | 2005-07-01 00:00:00.000 | 2005-07-13 00:00:00.000 | 2005-07-08 00:00:00.000 | 5 | 0 | SO43669 | PO14123169936 | 10-4020-000578 | 29747 |
| 43670 | 3 | 2005-07-01 00:00:00.000 | 2005-07-13 00:00:00.000 | 2005-07-08 00:00:00.000 | 5 | 0 | SO43670 | PO14384116310 | 10-4020-000504 | 29566 |

### Fact Table

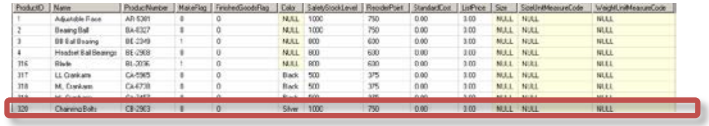


| ProductKey | OrderDateKey | DueDateKey | ShipDateKey | ResellerKey | EmployeeKey | PromotionKey | CurrencyKey | SalesTerritoryKey | SalesOrderNumber | SalesOrderLineNumber | RevisionNumber | OrderQuantity |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 349 | 20050701 | 20050713 | 20050708 | 676 | 285 | 1 | 100 | 5 | SO43659 | 1 | 1 | 1 |
| 350 | 20050701 | 20050713 | 20050708 | 676 | 285 | 1 | 100 | 5 | SO43659 | 2 | 1 | 3 |
| 351 | 20050701 | 20050713 | 20050708 | 676 | 285 | 1 | 100 | 5 | SO43659 | 3 | 1 | 1 |
| 344 | 20050701 | 20050713 | 20050708 | 676 | 285 | 1 | 100 | 5 | SO43659 | 4 | 1 | 1 |
| 345 | 20050701 | 20050713 | 20050708 | 676 | 285 | 1 | 100 | 5 | SO43659 | 5 | 1 | 1 |
| 346 | 20050701 | 20050713 | 20050708 | 676 | 285 | 1 | 100 | 5 | SO43659 | 6 | 1 | 2 |
| | | | | | | | | | | | | |
| 347 | 20050701 | 20050713 | 20050708 | 676 | 285 | 1 | 100 | 5 | SO43659 | 7 | 1 | 1 |
| 229 | 20050701 | 20050713 | 20050708 | 676 | 285 | 1 | 100 | 5 | SO43659 | 8 | 1 | 3 |
| 235 | 20050701 | 20050713 | 20050708 | 676 | 285 | 1 | 100 | 5 | SO43659 | 9 | 1 | 1 |
| 218 | 20050701 | 20050713 | 20050708 | 676 | 285 | 1 | 100 | 5 | SO43659 | 10 | 1 | 6 |
| 223 | 20050701 | 20050713 | 20050708 | 676 | 285 | 1 | 100 | 5 | SO43659 | 11 | 1 | 2 |
| 220 | 20050701 | 20050713 | 20050708 | 676 | 285 | 1 | 100 | 5 | SO43659 | 12 | 1 | 4 |
| 326 | 20050701 | 20050713 | 20050708 | 117 | 285 | 1 | 100 | 5 | SO43660 | 1 | 1 | 1 |

Extract only new records
and load into
transactional fact table

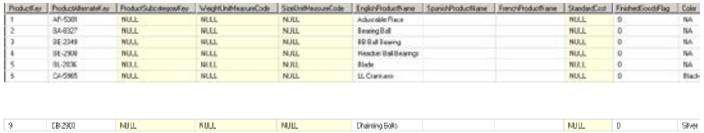
## Dimension Loads

### Source



| ProductID | Name | ProductNumber | MakeFlag | FinishedGoodsFlag | Color | SafetyStockLevel | ReorderPoint | StandardCost | ListPrice | Size | SizeUnitMeasureCode | WeightUnitMeasureCode |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Adjustable Race | AR-5381 | 0 | 0 | NULL | 1000 | 750 | 0.00 | 0.00 | NULL | NULL | NULL |
| 2 | Bearing Ball | BA-8327 | 0 | 0 | NULL | 1000 | 750 | 0.00 | 0.00 | NULL | NULL | NULL |
| 3 | BB Ball Bearing | BE-2349 | 1 | 0 | NULL | 800 | 600 | 0.00 | 0.00 | NULL | NULL | NULL |
| 4 | Headset Ball Bearings | BE-2908 | 0 | 0 | NULL | 800 | 600 | 0.00 | 0.00 | NULL | NULL | NULL |
| 316 | Blade | BL-2036 | 1 | 0 | NULL | 800 | 600 | 0.00 | 0.00 | NULL | NULL | NULL |
| 317 | LL Crankarm | CA-5965 | 0 | 0 | Black | 500 | 375 | 0.00 | 0.00 | NULL | NULL | NULL |
| 318 | ML Crankarm | CA-6738 | 0 | 0 | Black | 500 | 375 | 0.00 | 0.00 | NULL | NULL | NULL |
| 319 | HL Crankarm | CA-7457 | 0 | 0 | Black | 500 | 375 | 0.00 | 0.00 | NULL | NULL | NULL |
| 320 | Chainring Bolts | CB-2903 | 0 | 0 | Silver | 1000 | 750 | 0.00 | 0.00 | NULL | NULL | NULL |

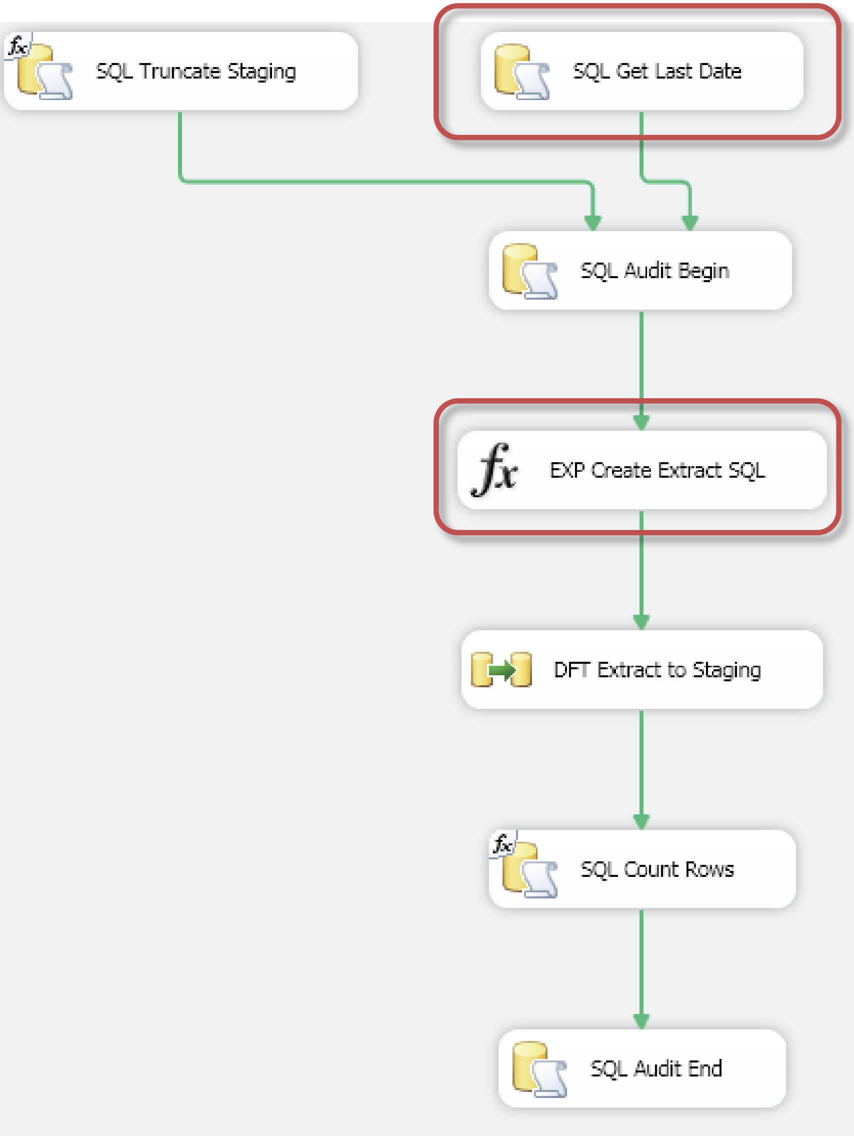
### Dimension Table



| ProductKey | ProductAlternateKey | ProductSubcategoryKey | WeightUnitMeasureCode | SizeUnitMeasureCode | EnglishProductName | SpanishProductName | FrenchProductName | StandardCost | FinishedGoodsFlag | Color |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | AR-5381 | NULL | NULL | NULL | Adjustable Race | | | NULL | 0 | NA |
| 2 | BA-8327 | NULL | NULL | NULL | Bearing Ball | | | NULL | 0 | NA |
| 3 | BE-2349 | NULL | NULL | NULL | BB Ball Bearing | | | NULL | 0 | NA |
| 4 | BE-2908 | NULL | NULL | NULL | Headset Ball Bearings | | | NULL | 0 | NA |
| 5 | BL-2036 | NULL | NULL | NULL | Blade | | | NULL | 0 | NA |
| 6 | CA-5965 | NULL | NULL | NULL | LL Crankarm | | | NULL | 0 | Black |
| | | | | | | | | | | |
| 9 | CB-2903 | NULL | NULL | NULL | Chainring Bolts | | | NULL | 0 | Silver |

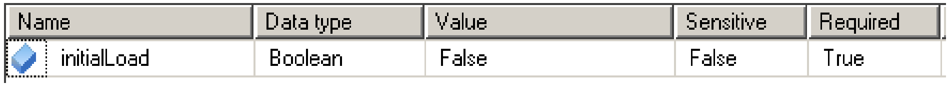
Extract all records
Exists? Update if changes (maybe)
Load new records

# Fact Extract for Ongoing Load



Look up max date for related fact table, store in MaxDateTime variable

Project Parameter



| Name | Data type | Value | Sensitive | Required |
|---|---|---|---|---|
| initialLoad | Boolean | False | False | True |

Create SQL statement for extract based on initialLoad value

```
@[User::ExtractSQL]
=@[$Project::initialLoad] ?
@[User::ExtractSQL] :
@[User::ExtractSQL] + " where OrderDate
> '" +  (DT_WSTR, 50) (DT_DBTIMESTAMP)
@[User::MaxDateTime] + "' "
```

# Summary

- **Introduction to Data Warehousing**
  - Resources, history,  business rules, data structures
- **Dimensional Modeling**
  - Star schema, act tables, dimension tables
- **Data Profiling**
  - ADO.NET connection, File connection, Data Profile Viewer
- **ETL Design Patterns**
  - Master packages, extract package template, fact versus dimension loads

# Resources

- **Microsoft Data Warehouse Toolkit, Second Edition**
    - http://tinyurl.com/qf84vl7
- **Kimball Group**
    - http://kimballgroup.com
- **Using Star Join and Few-Outer-Row Optimizations to Improve Data Warehousing Queries**
    - http://tinyurl.com/nyvww6j
- **Integration Services Error and Message Reference**
    - http://tinyurl.com/mo796jp