

Inform Align Engage – Another Smarter Planet Initiative



IBM Cognos Forum

Ignite knowledge, ideas, connections

Go Dimensional!

Venkat Srinivasan

Information Management

Cognos.
software

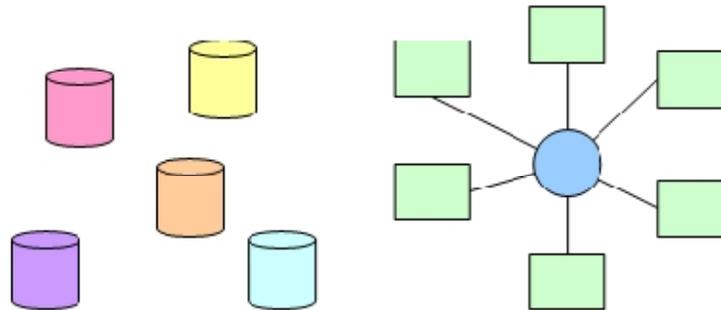
Agenda

- > **Why GO Dimensional?**
- > **Dimensional modelling process**
 - Grain
 - Facts
 - Dimensions
- > **Different ETL techniques**
- > **ETL tools**
 - Data Manager and Data Stage comparison
 - Data lineage

Why Data Warehouse?

- > **Different business requirements drive:**
 - 1) Single Source Truth and Consolidation of data across disparate systems. 2) Business View of the data
 - Access to History data and trend analysis
- > **Primary design goal is to support analytic queries**
 - Usable
 - Performance
- > **How does the De-normalise help**
 - De-normalise dimensions for usability
 - Normalise facts for performance
- > **Key Terms**
 - Facts = Business measures
 - Dimension = Business variables that participate in the business event

Why build a Data Warehouse?



Transaction Systems

Data Warehouse

Data Orientation:	Transactions	Reporting
Complexity:	High	Low
Accessibility:	Centralized	Distributed
Data Integrity:	Questionable	Validated
Focus:	Operating	Analysis
Sources:	Disparate	Integrated
Usefulness:	Data	Information

Building a Data Warehouse

- > Data Warehouse is **NOT** a IT solution

- > How does the business view their information?
 - Products, Stores, Customer, Customer Segment, Account, Demographics, Sales Rep, Branch

- > Data Warehouse is a **business solution** that models data for a business process and enables easy access to the data

- > Querying 3NF versus Dimensional data structure

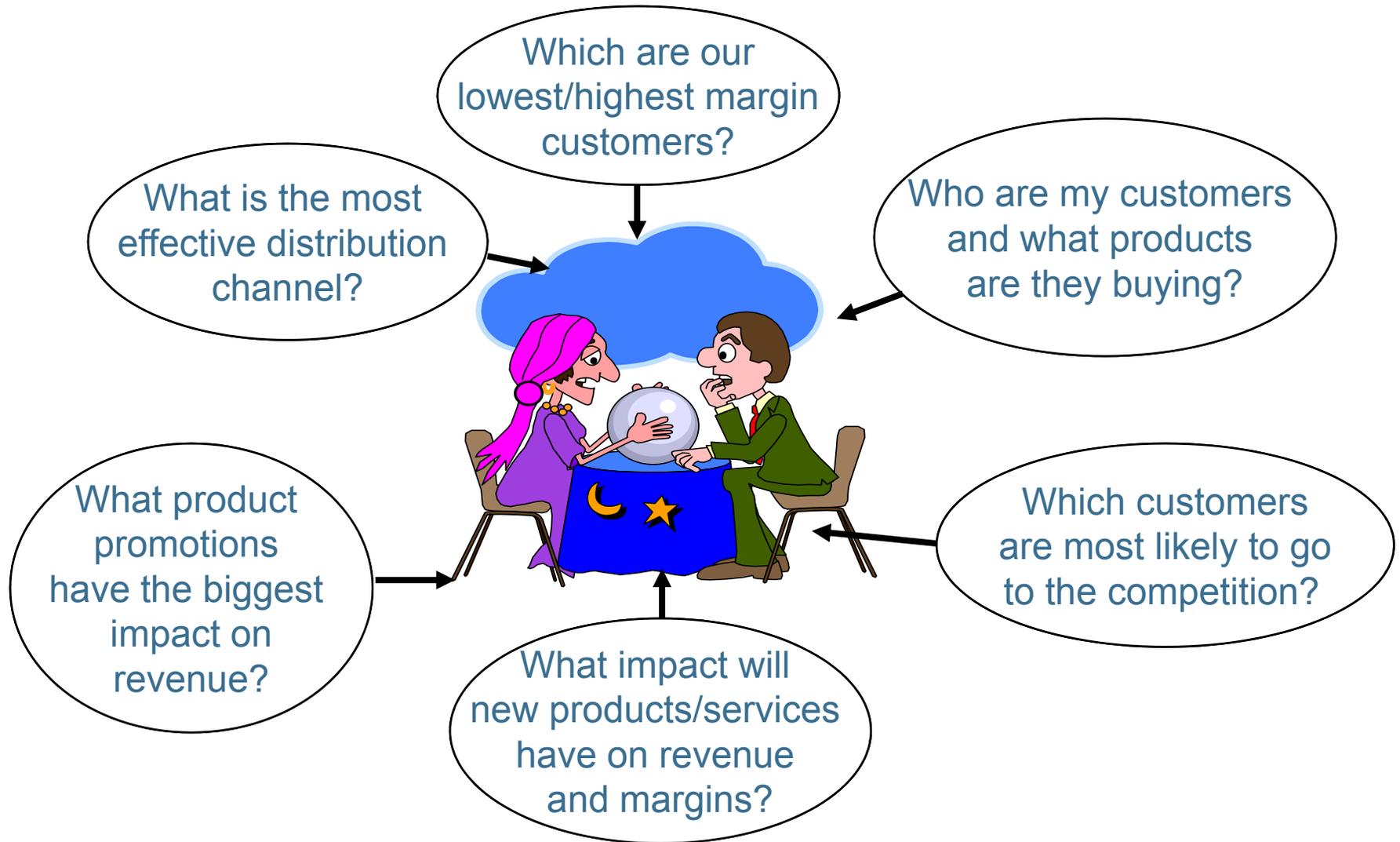
Facts and Fables

- > **Myth: Dimensional models support only summary data**
 - **Fact: The model depends on the grain of the data**
- > **Myth: Dimensional model support only departmental data marts**
 - **Fact: An Enterprise Dimensional Data Warehouse (EDDWH) is possible if designed properly**
- > **Myth: Bringing in a new data source breaks the existing dimensional model; requires new fact tables**
 - **Fact: If the new data source presents data at the same grain and dimensionality, it can be added gracefully**
- > **Myth: Design Principles used in building Operational Systems not valid for Data Warehouse**
 - **Fact: If no proper planning is in place then there will be several “web” like interfaces between systems. Simplifying the interfaces is valid for both Operational Systems and Data Warehouse.**

Agenda

- > **Why GO Dimensional?**
- > **Dimensional modelling process**
 - Grain
 - Facts
 - Dimensions
- > **Different ETL techniques**
- > **ETL tools**
 - Data Manager and Data Stage comparison
 - Data lineage

Reporting and modeling



ER v Dimensional Modelling

Entity Relationship modelling

Transaction-orientated OLTP
Great number of tables and joins
Normalised
Transactional tables

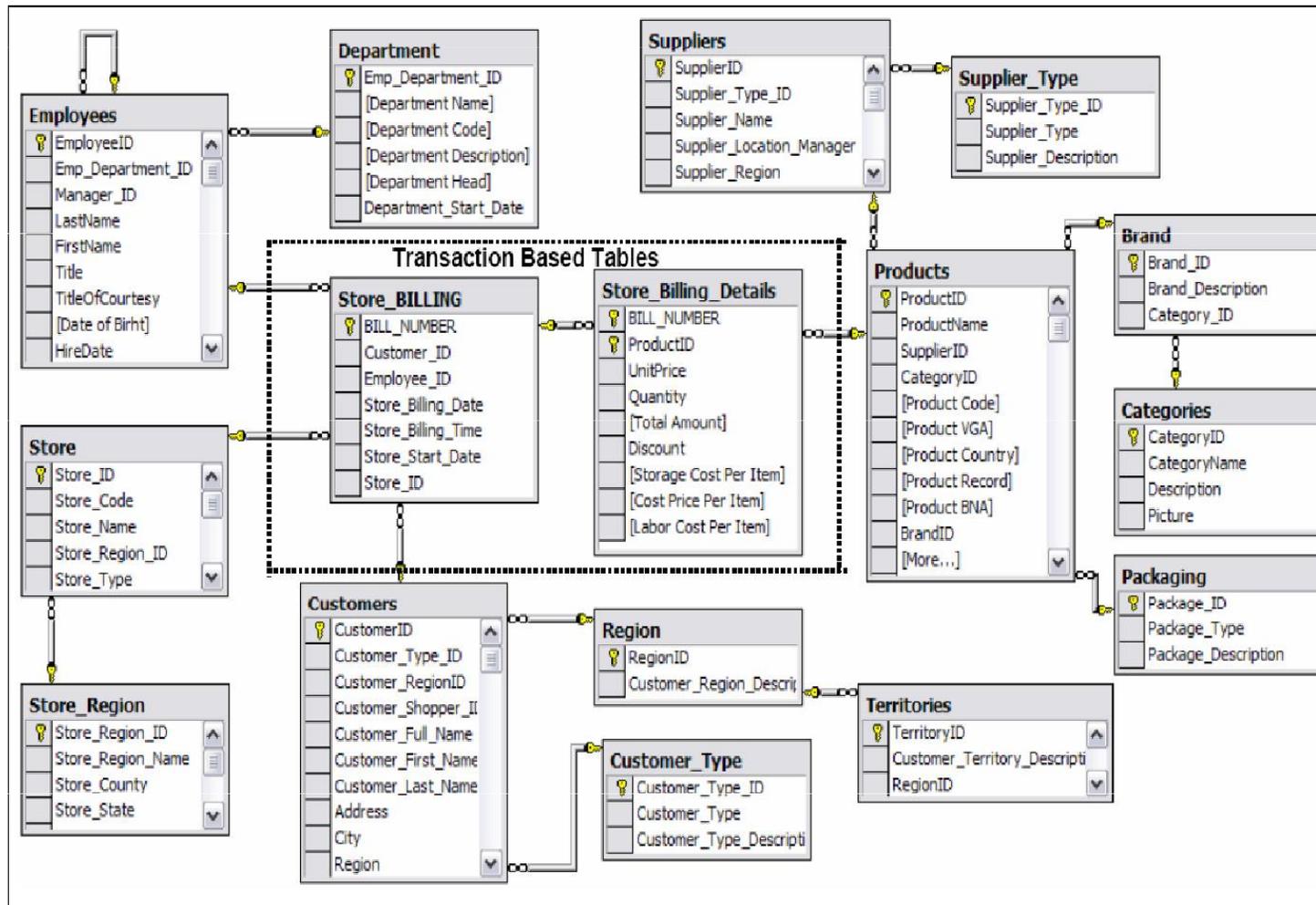
INSERT / UPDATE / DELETE

Dimensional modelling

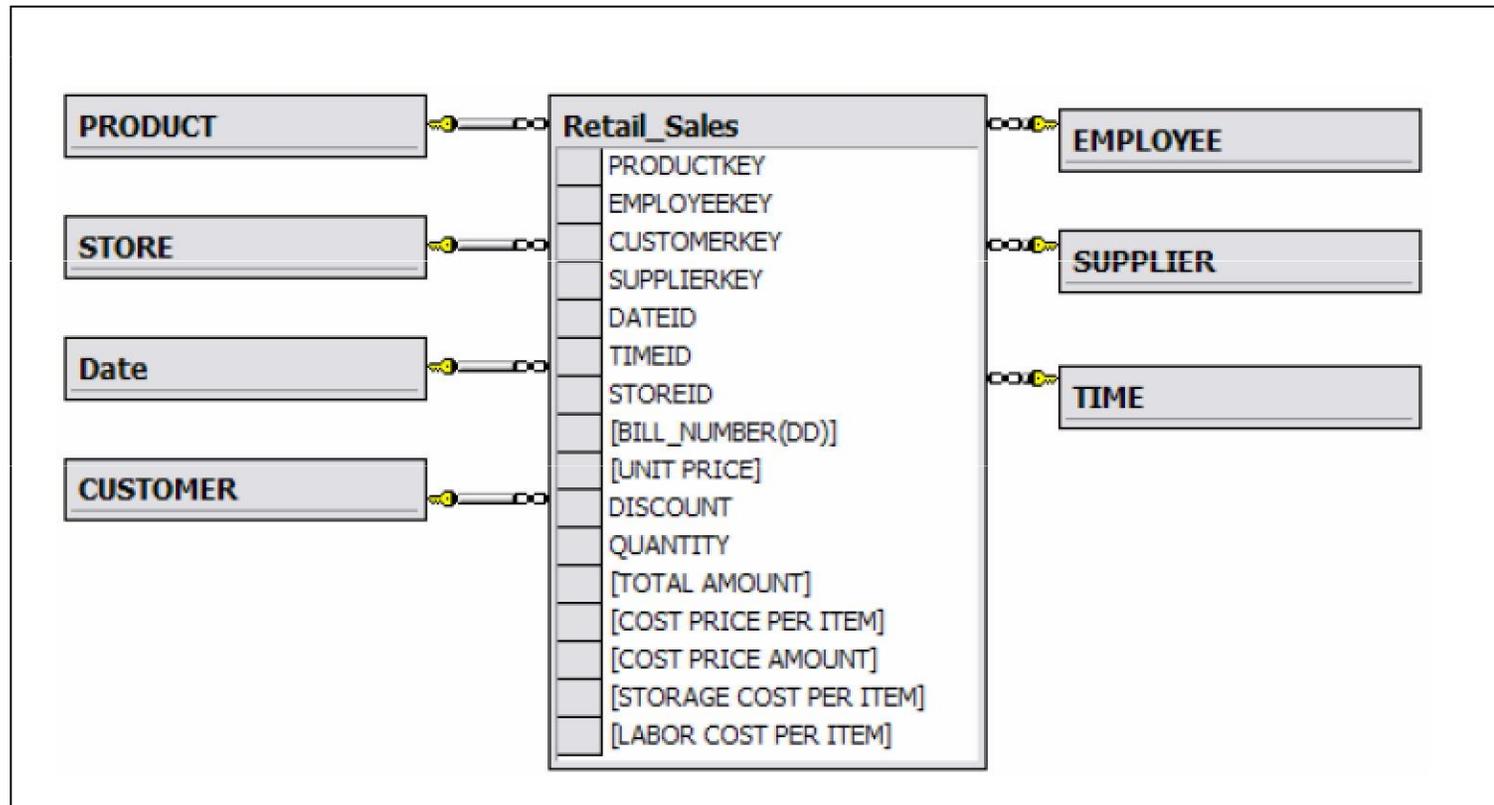
Data warehouse / Reporting
Few tables and joins
De-normalised
Fact / Dimensional tables
(Star Schema)

READ

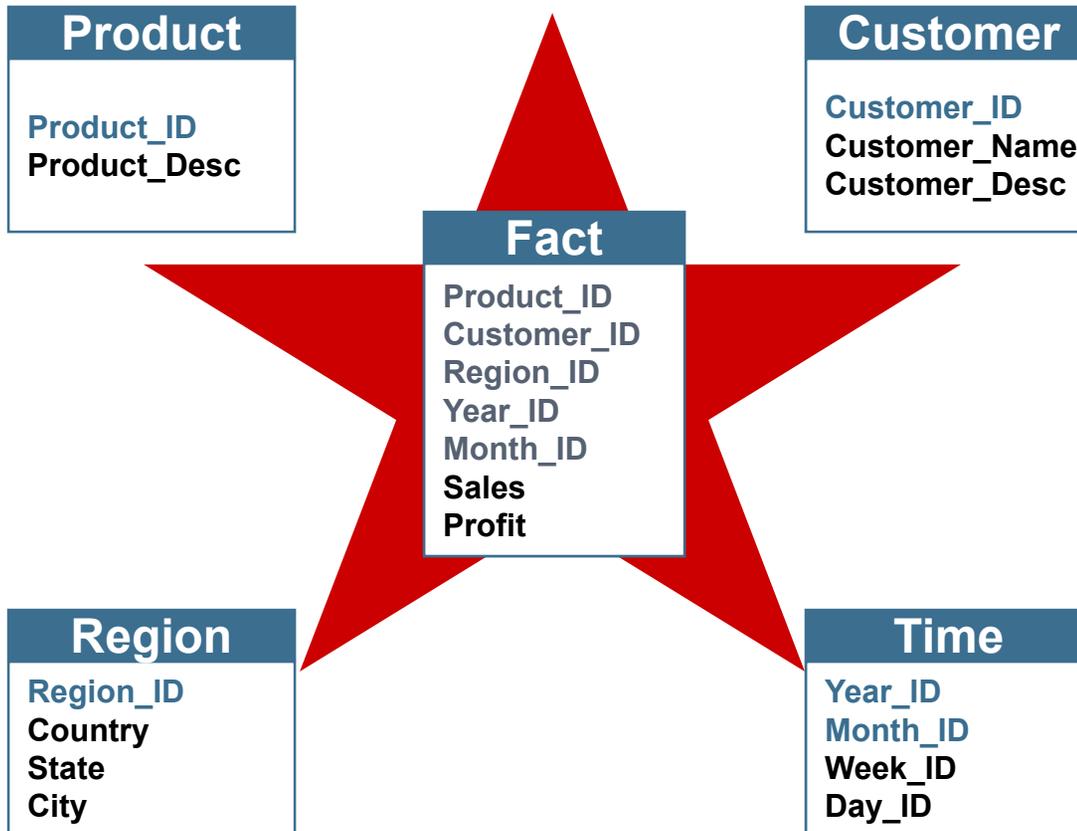
E/R Model



Dimensional Modelling



Dimensional Model Structure



> Dimensional Model consists of:

- Dimensions
- Facts

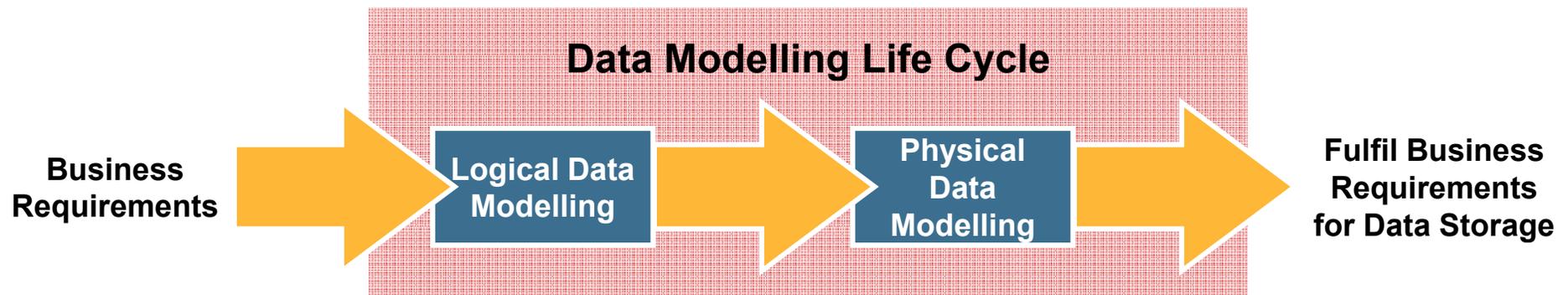
> Advantage

- One fact table per business event
- Easier to use
- Better query performance
- Extensible to handle change

Benefits of Dimensional Modeling

- > Focus on business processes and business measures**
- > Reporting performance better on dimensional model**
 - **Highly de-normalised and faster throughput (less joins)**
- > Supports tracking and changes to business variables**
- > Reporting on different business areas using common variables (conformed dimensions)**
- > Empower the business users with interactive reporting and query functionality (Flexible and easy using conformed dimensions)**
- > Provide self service because of easy data navigation:**
 - **IBM Cognos 8 Query Studio**
 - **IBM Cognos 8 Analysis Studio**

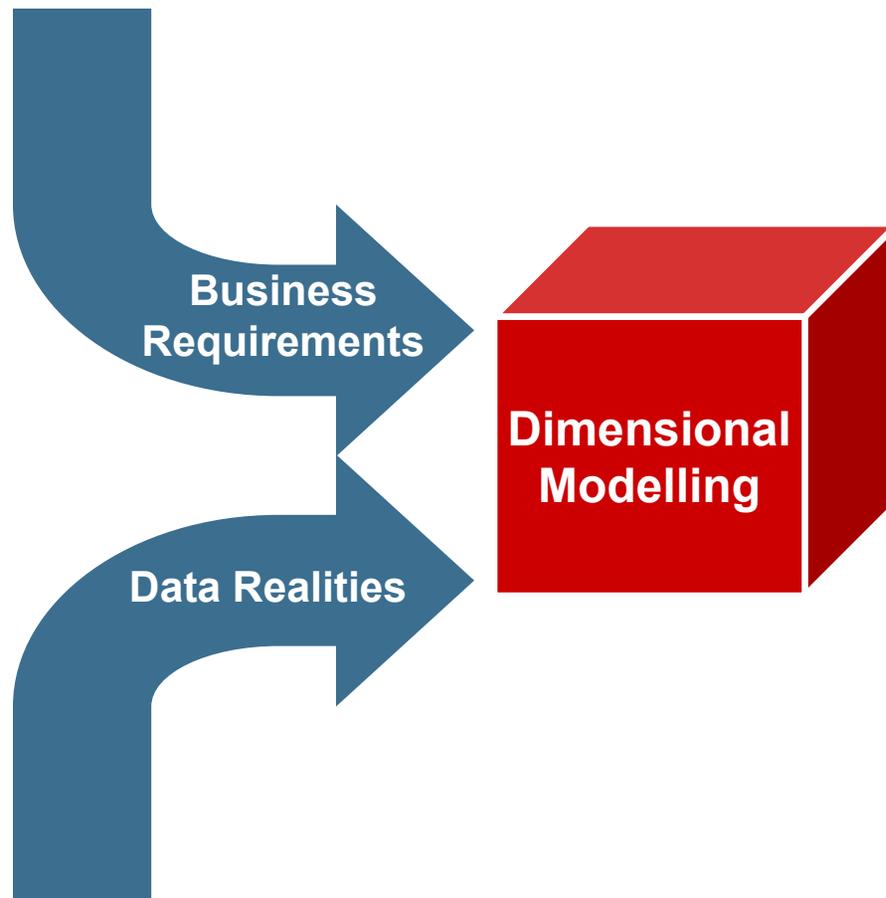
Steps to Build Dimensional Model



> 1. Identify the Business Process

- *From the Business Process you should be able to:*
 - 2. Identify the Grain
 - 3. Identify the Dimensions
 - 4. Identify the Facts

Key Input to Dimensional Modelling



- > Top down approach driven by **business reporting requirements** is best practice
- > These are crucial to a successful design
- > Need to avoid the usual approach of looking at what currently exists in a data source and modelling it
- > Kick off with a workshop with business users, stakeholders, report authors, DBAs

Identify the Business Process

- > What is a Business Process?**
 - Sales, Inventory, Marketing, Order Management

- > Business Process begins with Customer Need and ends with a delivery of Goods or Service to the Customer**

- > One or more Fact tables will exist in a data mart to support each key business process**

- > To help identify a business process look at the main activities in the business that collect data**

Example Questions

- > **Who** are the people, groups, and organisations of interest?
- > **What** functions need to be analysed?
- > **Why** is the data required?
- > **When** does the data need to be recorded?
- > **Where**, geographically and organisationally, do relevant processes occur?
- > **How** do we measure the performance of the functions being analysed?
- > **How** is performance of the business process measured? What factors determine the success or failure?
- > **What** is the method of information distribution? Is it a data report, paper, pager, or e-mail (examples)?

Identify the Dimensions

- > Dimensions are those business attributes that qualify facts
- > Typically the “by” words used to describe analytical requirements
- > Dimension provide reference structure to facts
- > **GRAIN**: 1 row per branch, product, customer per transaction date is the grain which shows dimensions

Identify the Facts

- > What metrics result from the business process?
- > Facts are measured, continually valued, rapidly changing information
- > For example, unit quantity, dollar sales, price
- > Can be calculated or derived
- > Can be additive, non additive or semi-additive
- > The data structure should support a business process.

- > Watch out!!
 - Don't design a fact table for each report

Bus Matrix – Powerful Tool

Business Process	Business Priority	Conformed Dimensions										
		Date (Order, Start, Ship)	Product	Promotion	Customer	Employee	Page	Internet Registered User	Part	Vendor	Shipper	Problem
Orders Forecasting	2	X	X	X	X	X						
Orders	1	X	X	X	X	X						
Purchasing		X	X		X	X			X	X	X	
Parts Inventory		X	X	X					X	X		
Manufacturing	6	X	X						X			
Finished Goods Inventory		X	X	X								
Shipping		X	X	X	X	X					X	
Returns	5	X	X		X	X					X	
Registration Cards		X	X		X							
Customer Calls	4	X	X	X	X	X			X			X
Web Support		X	X		X	X						X
Financial Forecasting		X	X	X	X	X	x	x		x		
Exchange Rate Management	3	X										

- Useful tool for highlighting conformed dimensions
- Useful for prioritising development
- Conformed dimensions are critical in an dimensionally modelled data marts

Workshop Case Study

> Background

- *Retail chain consisting of over 100 stores in five states*
- *Stores average 60,000 stock keeping units (SKU's) in departments such as frozen foods, dairy etc*
- *Bar codes are scanned directly into the cash registers' POS system*
- *Products are promoted via coupons, temporary price reductions, ads and in-store promotions*

> Analytical Requirements

- *Need to know what is selling in the stores each day in order to evaluate product movement, as well as to see how sales are impacted by promotions*
- *Need to understand the product mix in a consumers market basket*

Step 1: Identify Business Process

> Point of Sale (POS) Process

Analytical Requirements

Need to know what is selling in the stores each day in order to evaluate product movement, as well as to see how sales are impacted by promotions

Need to understand the product mix in a consumers market basket

- *Bar codes are scanned directly into the cash registers' POS system*

Step 2: Identify Grain

> Point of Sale (POS) Process

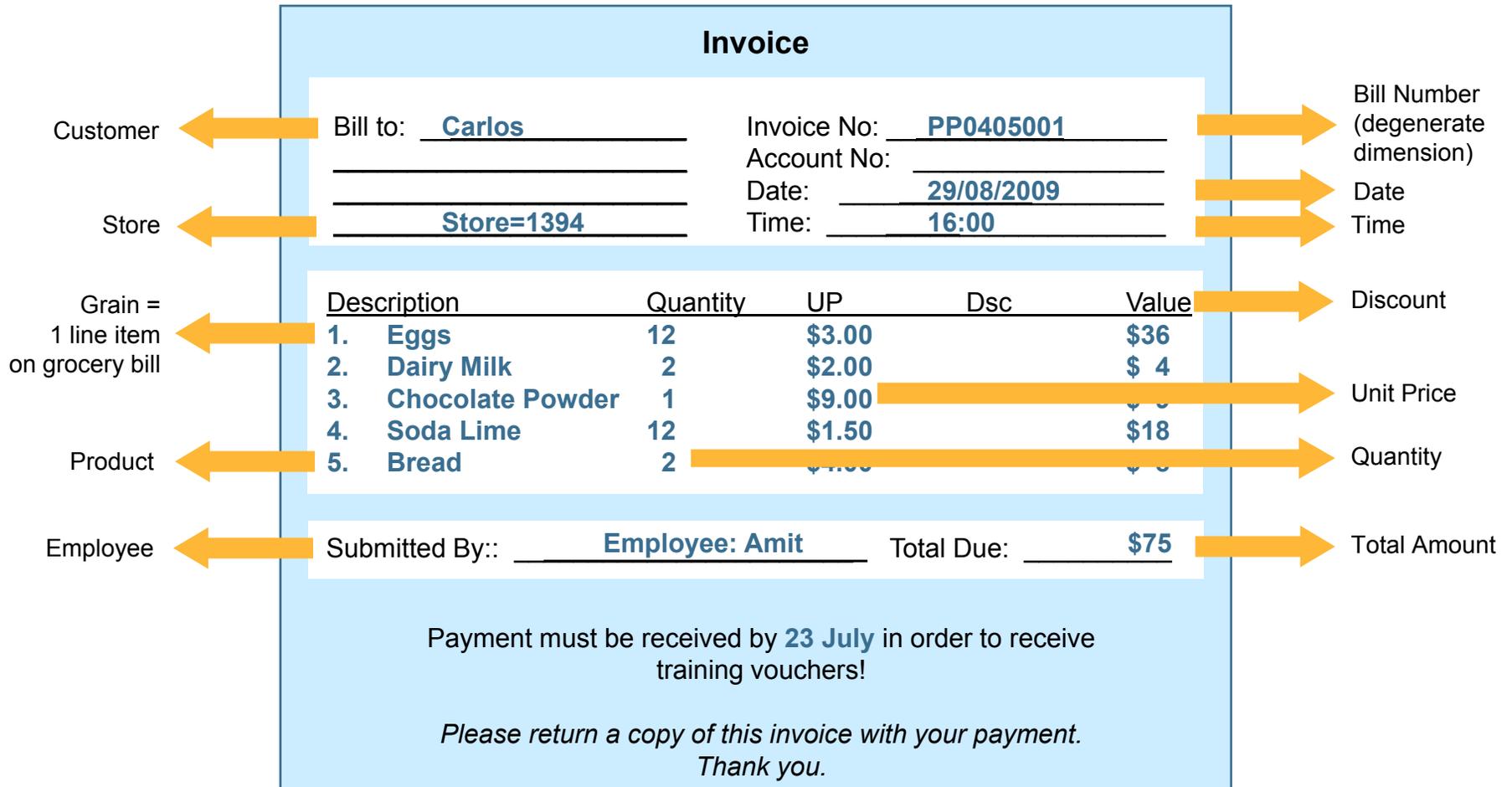
Analytical Requirements

Need to know what is selling in the stores each day in order to evaluate product movement, as well as to see how sales are impacted by promotions

Need to understand the product mix in a consumers market basket

- *Bar codes are scanned directly into the cash registers' POS system*

Inputs to Dimensional Model



Step 3: Identify Dimensions and Facts

Background

Retail chain consisting of over 100 stores in five states

Stores average 60,000 stock keeping units (SKU's) in departments such as frozen foods, dairy etc

Bar codes are scanned directly into the cash registers' POS system

Products are promoted via coupons, temporary price reductions, ads and in-store promotions

Analytical Requirements

Need to know what is selling in the stores each day in order to evaluate product movement, as well as to see how sales are impacted by promotions

Need to understand the product mix in a consumers market basket

Step 4: Star Schema

Agenda

- > **Why GO Dimensional?**
- > **Dimensional modelling process**
 - Grain
 - Facts
 - Dimensions
- > **Different ETL techniques**
- > **ETL tools**
 - Data Manager and Data Stage comparison
 - Data lineage

Information Management Issues

> Current Situation

- ETL plays an important role in any Data Warehouse. This is the backbone of the Data Warehouse.
- ETL development and maintenance takes too long.
- Uncontrolled ETL development is resulting in “web” like interfaces where each system exchange information with each other. System interfaces are too many to manage!!
- Users can not answer WHAT, HOW and WHY easily.
- Single Source of Truth is a ideal state
- **OPTIMISE** the system interfaces is the way to go.

Information Management Issues

> Information needs to reach more people



> Each person has very diverse needs



> Infrastructure is increasingly complex



> Data is siloed in multiple disparate systems



Different Data Transformation Techniques

- > **Option 1 – Using a metadata Modelling tool (such as Cognos FM)**
 - Use Framework manager and present the business view

- > **Option 2 – DTL (Push method)**
 - Source System providing extracts which are dimensionally aware

- > **Option 3 – ETL (Pull method)**
 - Create a separate ETL sub-system

OPTION 1 -Use IBM Cognos FM

> How

- Use IBM Cognos Framework Manager to create the metadata and create Dimensionally Modelled Relational structure for reporting
- Use FM and convert the operational model to a reporting model
- No ETL is used in this approach

> Pros

- No ETL complexities. No ETL skills required to maintain. Require data modelling skills

> Cons

- The underlying data is still highly normalised. The transformations are done at the report runtime for every report
- This method does not support complex transformation and large volumes
- Running reports can impact operational system. Potential performance issues in this approach

OPTION 2 – DTL (Push method)

> How does it work?

- New breed Source Systems are SQL aware and they can do initial transformations at source
- Provide Daily, Weekly or Monthly Strip files for other systems
- Single place where transformations are controlled
- In this method, additional transformations done by the recipients systems in the areas such as Profitability, Ranking, Predictive modelling
- Extracts and loads are simple. Transformations take 80% of the effort
- Interfaces Streamlined and Optimised
- Best of both worlds: Legacy System Design principles with latest technology
- **OUTCOME: Reduce or Optimise the number of System Interfaces.**

OPTION 2 – DTL (Push method)

> How

- Using the “push” method, the source system acts as an enhanced ETL feeder system. The feeds are made up of customised SQLs maintained by the Source System
- As part of the data extracts, the required transformations and filters are applied by the source systems and presented to the recipient system.
- Only transformations that could not be done at the source will be done by the recipient system

> Pros

- The business logic and transformation is held within the source system. Single point of control
- The skill set to maintain the ETL is controlled in the source system
- Dependent on the flexibility of the source system, this approach may take less effort and maintenance
- Additional transformation (such as CCD, profitability, at risk etc) are done by the recipient system

> Cons

- Require more planning

OPTION 3 – ETL (Pull method)

> How

- This is the traditional ETL development method. The steps involve creating staging tables, transformation and presentation areas
- Require definition of source to target ETL mappings and related transformation rules
- The staging tables reflect the source system structure and transformation is carried out by the ETL layer
- Require a separate ETL tool that performs these functions

> Pros

- This approach has benefits similar to Option 2. Changes to the Source system data structures will have an impact on the ETL modules as well.

> Cons

- Usually the difficult part of any BI project, and it is prone to lot of problems and maintenance unless it is designed properly
- If the SME for the source systems are not readily available, the ETL is an impossible task

Workshop Example

- > Scenario: Interface Oracle Financials ERP System**

- > Option 1 – IBM Cognos FM**

- > Option 2 – DTL**

- > Option 3 – ETL**

Agenda

- > **Why GO Dimensional?**
- > **Dimensional modelling process**
 - Grain
 - Facts
 - Dimensions
- > **Different ETL techniques**
- > **ETL tools**
 - **Data Manager and Data Stage comparison**
 - **Data lineage**

ETL Product Comparison

- > The skill set of the user and their experience is more important than the tool itself**

- > Several ETL tools are in the market place. In this example we are covering:**
 - IBM Cognos Data Manager**
 - IBM Information Manager**
 - Integration between Information Manager and Data Stage**

Product Comparison Summary

- > **Cognos Data Manager is a BI integrated ETL tool**
 - It does not provide generic ETL capabilities
 - but it provides elegant interface to generate a dimensional model
 - It is almost a “code generator” to build data marts using Kimball methodology
- > **In terms of ETL functionality IBM Information Server has nice features such as:**
 - Version control
 - Impact analysis
 - Metadata management
 - Security
 - Data lineage
- > **IBM Quality Stage is a leader in the Data Quality. This can do matching, de-dupes, survivorship, Name and Address cleansing**

ETL Functionality Comparison

Functionality	Data Manager	Information Server
Dimensional Modelling	Very Good	Good
Integration with third party	Limited Only Cognos BI	Very Good
Data Quality	None	Very Good
Meta data Mgmt	Good	Very Good
Transformations	Limited	Very Good

Data Lineage

- > **Features:**
- > **Lineage info traces the metadata of an item back to the package.**
- > **Lineage also displays any filter that were referenced by the report author**
- > **Lineage will show model caculations used in the Framework Manager**
- > **Provide Business View**
 - **High level textual information providing the lineage information. This comes from IBM Cognos Connection and Framework Manager**
- > **Provide Technical View**
 - **Give graphical representation of lineage of a data item**
- > **Lineage facility is available only on saved HTML reports**
- > **Watch This space – This is only a start**
 - **Empowering the users with more information**
 - **Improving the WHAT, HOW and WHY facility. Expose users with the underlying transformations**

Lineage within Cognos 8

Business View
Technical View

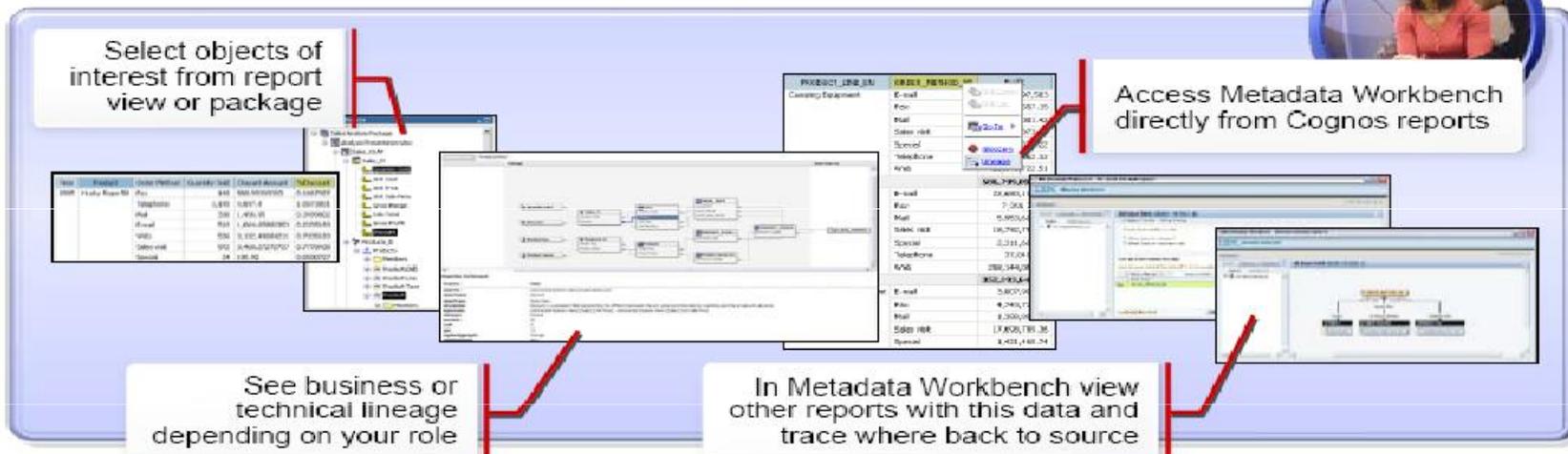
Package
Data Sources

Properties for Order value

Property	Value
objectId	[Filters and calculations].[Order value]
objectName	Order value
objectType	Calculation
expression	total ([gosales].[Sales].[Quantity] * [gosales].[Sales].[Unit sale price] for [gosales].[Sales].[Order number])
datatype	Decimal
precision	77
scale	2
size	39
regularAggregate	Automatic
semiAggregate	Automatic

Lineage within Cognos 8 and IM

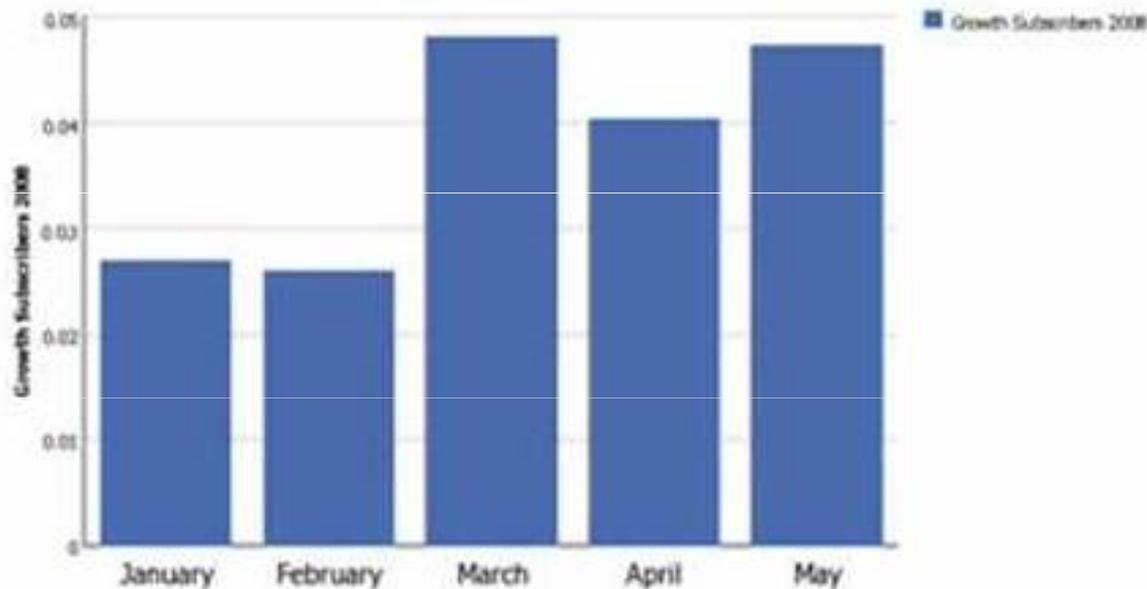
Data Lineage



- Business users easily see where data originated to gain confidence in the data they use to drive decisions
- Data modelers and report authors see lineage from report to source, saving time with report development and troubleshooting

IM with Cognos BI (Example)

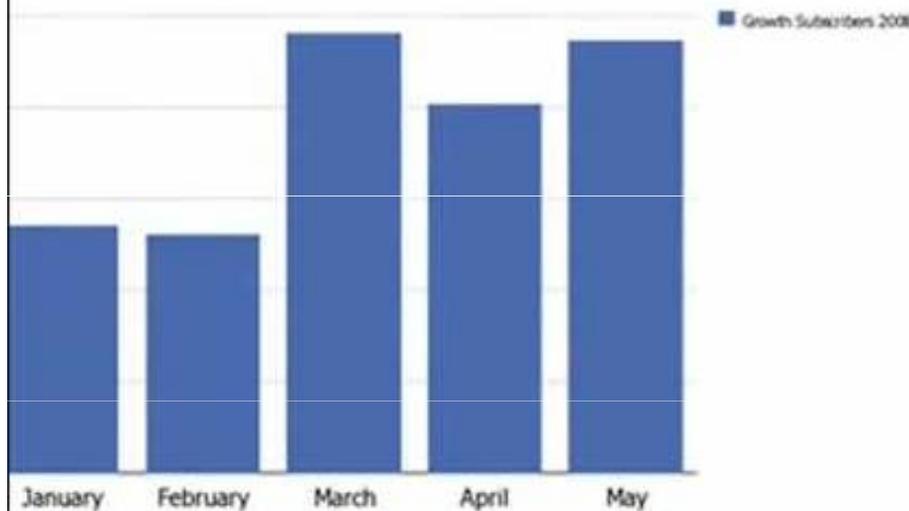
Impact of Marketing Campaign in February 2008



Month	Total Number of Subscribers	New Subscribers	Churned Subscribers 2008	Growth Subscribers 2008	Plan New Subscribers 2009
January	5,343,428	363,454	219,505	2.69%	407,068
February	5,090,953	343,995	212,058	2.59%	385,274
March	5,363,855	477,367	219,505	4.81%	534,651
April	5,297,539	429,984	216,366	4.03%	481,582
May	5,451,277	477,367	219,505	4.73%	0

Lineage from the report

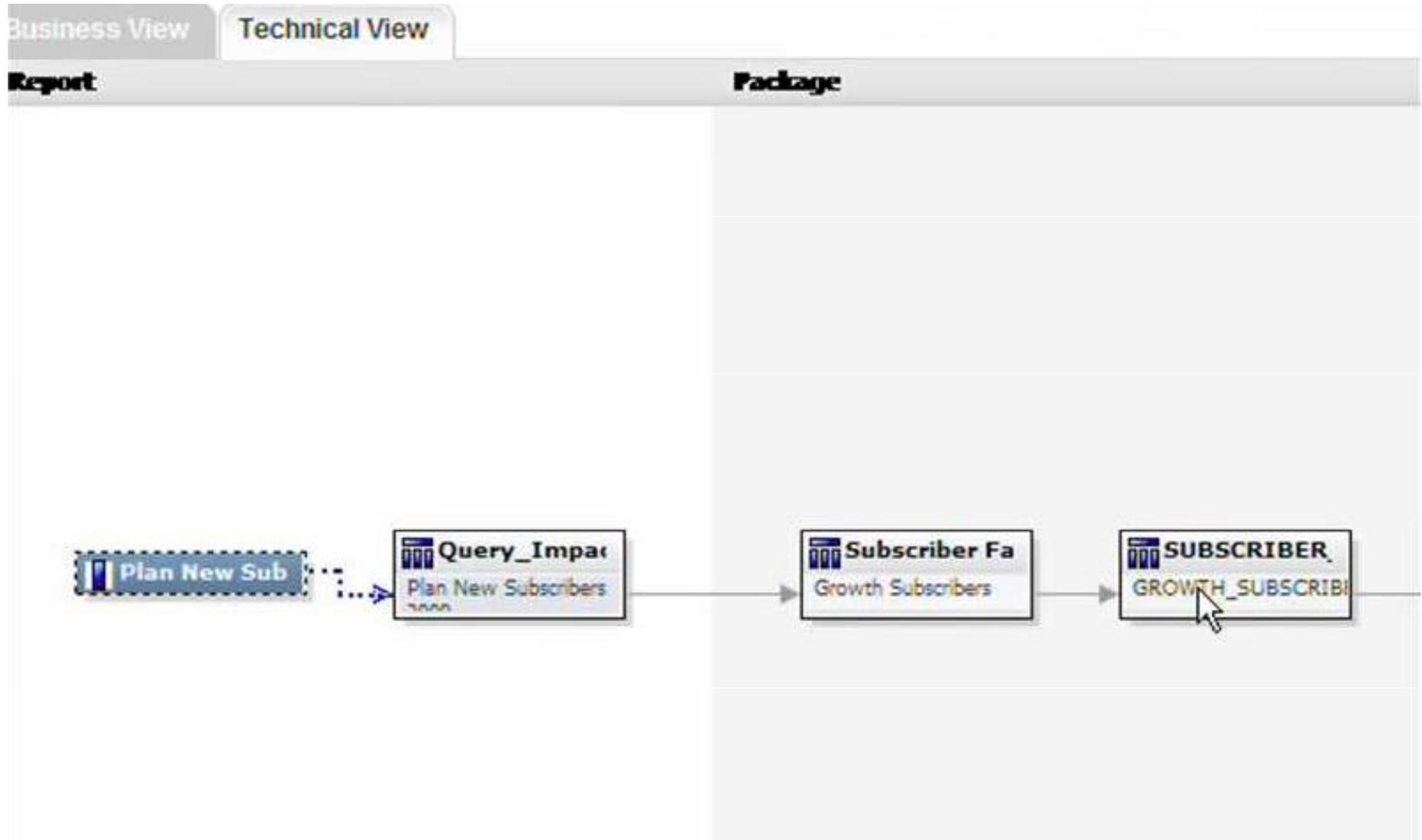
Impact of Marketing Campaign in February 2008



Total Number of Subscribers	New Subscribers	Churned Subscribers 2008	Growth Subscribers 2008	Plan New Subscribers 2009
5,343,428	363,454	219,505	2.69%	407,068
5,090,953	343,995	212,058	2.59%	385,274
5,363,855	477,367	219,505	4.81%	534,651
5,297,539	429,984	216,366	4.03%	481,582
5,451,277	477,367	219,505	4.73%	0

- Home
- Go To
- Lineage

Lineage - Technical View



Lineage - Technical View

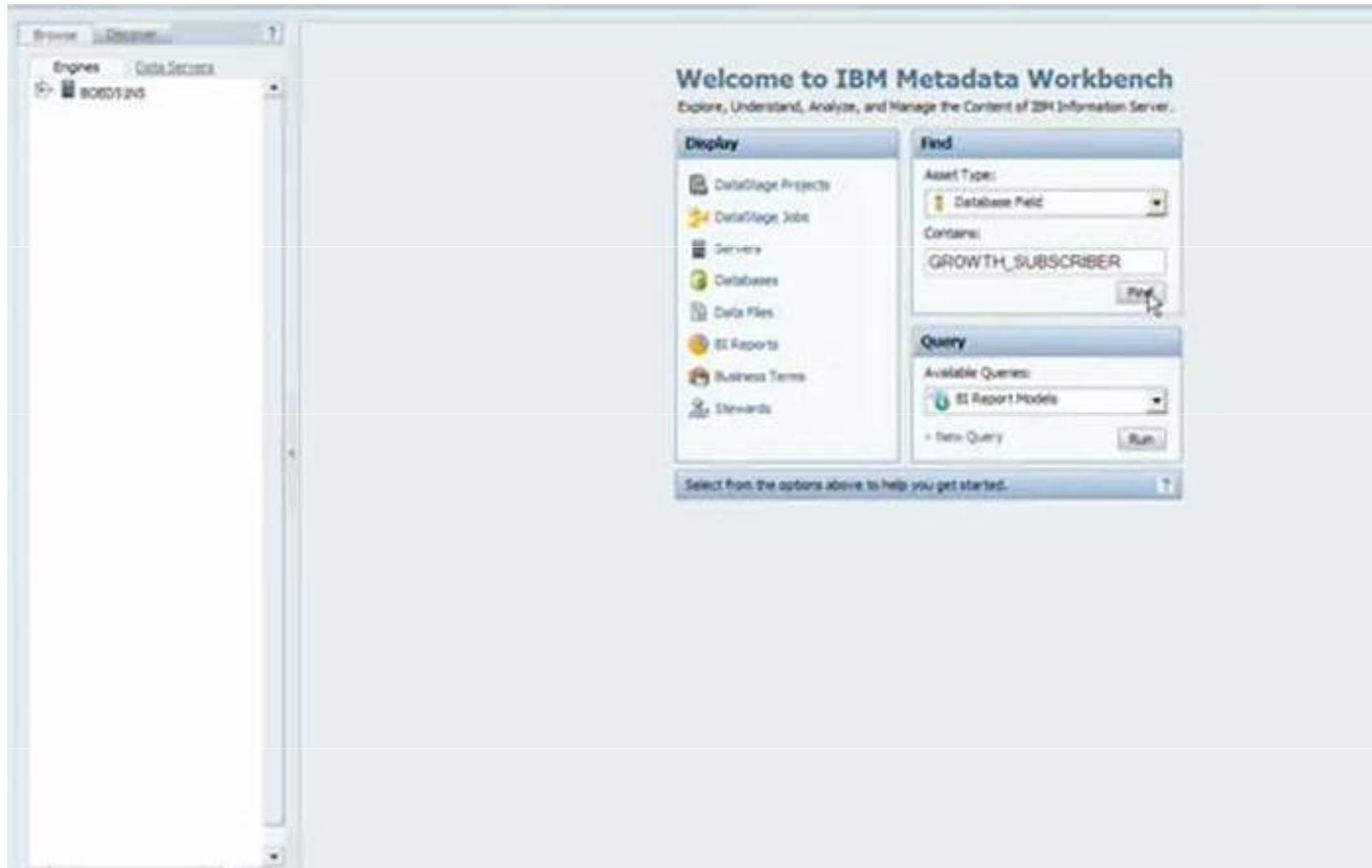
The diagram illustrates a data lineage path. It starts with a source object 'Plan New Sub' (dashed border), which flows into 'Plan New Subscribers' (solid border). This then flows into 'Growth Subscribers' (solid border), which flows into 'GROWTH_SUBSCRIBER' (solid border). Finally, 'GROWTH_SUBSCRIBER' flows into the target object 'IOD2008' (solid border). The 'GROWTH_SUBSCRIBER' object is highlighted with a blue border.

Properties for GROWTH_SUBSCRIBER

Property	Value
ID	[Data Source View].[SUBSCRIBER_FACT].[GROWTH SUBSCRIBER]
Name	GROWTH_SUBSCRIBER
Type	Query Item
External Name	GROWTH_SUBSCRIBER
Datatype	Decimal
Precision	11
Scale	6
Size	6
Regular Aggregate	Sum

Done

Trace in Metadata Workbench



Trace the Expression Error

Column Name	GROWTH_SUBSCRIBER
Expression	If SOR_TO_CALC_Link.MEASUREMENT_PERIOD_ID > 121 AND SOR_TO_CALC_Link.MEASUREMENT_PERIOD_ID < 153 Then SOR_TO_CALC_Link.NEW_SUBSCRIBERS * 0 Else SOR_TO_CALC_Link.NEW_SUBSCRIBERS * 1.12
DataStage Link	FACT Load Link
DataStage Job	LOAD FACT FROM SOR TO MART
Links from Stage	CALC_GROWTH_SUBS
Links to Stage	SUBSCRIBER_FACT_Load
Description	Generated by DataStage Job
Business Term	None
Column Type	DECIMAL
Column Length	11
Column Scale	0
Column Nullable	false
Column Visible	true
DataStage Stage Column Design Information	
DataStage Stage Column Operational Information	
DataStage Stage Column User Defined Information	
Modification Details	

Evaluation Forms

SHARE YOUR FEEDBACK AND WIN

Visit IBM Cognos Central to fill out your session evaluations online.

- > **Each completed evaluation** qualifies you to win one of five \$100 daily prizes.
- > **Complete evaluations for every session you attend** and qualify to win an additional \$500! An overall conference survey will be available at IBM Cognos Central on Friday morning, and will also be emailed to you.