



Introduction to Business Intelligence for SAS[™] Users

Tom Miron

Systems Seminar Consultants

Madison, WI



**Business intelligence is
the end-user or analyst's view
of your business data.**

What is Business Intelligence?

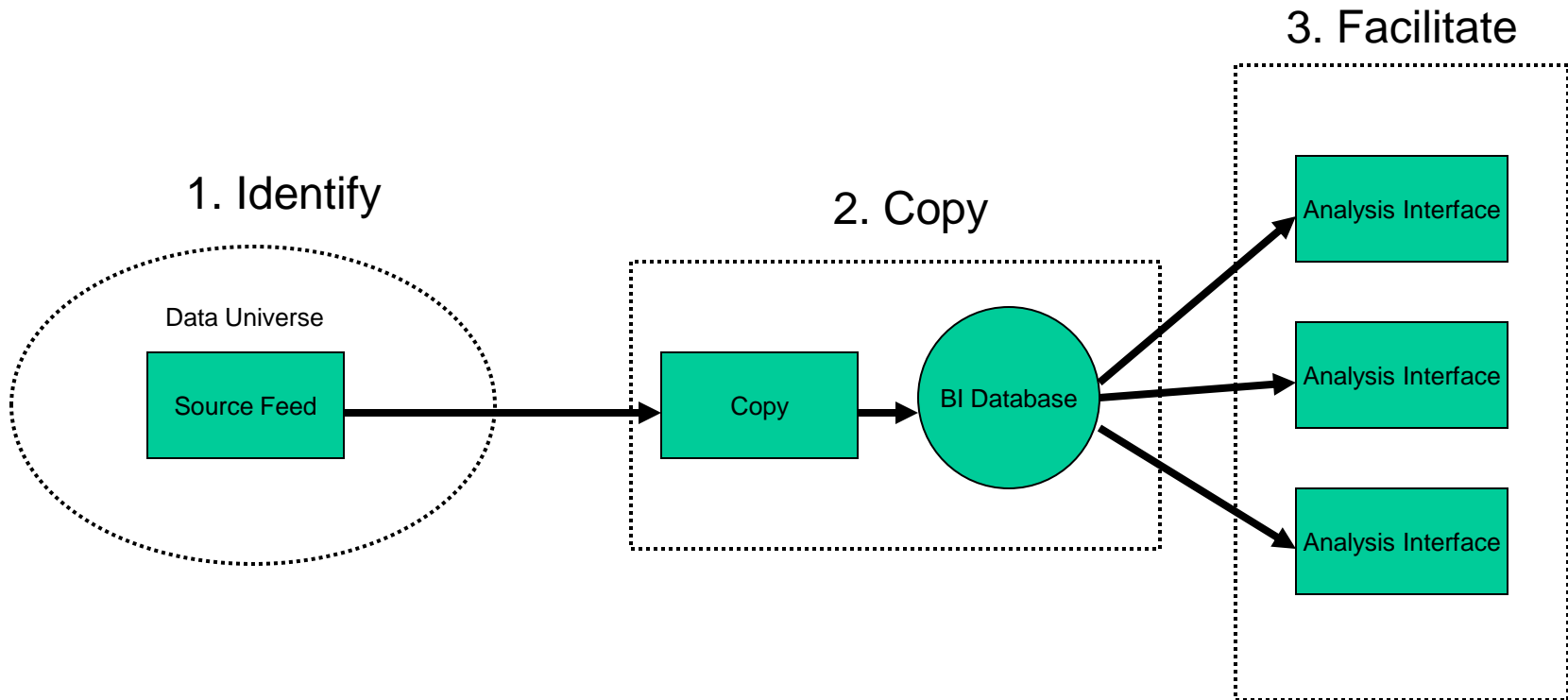


Three elements of business intelligence:

1. Identify and understand source data (operational data)
2. Create a processed copy of source data in a user accessible BI database
3. Develop or facilitate a user interface that allows analysis and presentation of the BI data

We will concentrate on 1. (Identify) and 2. (Create)

What is Business Intelligence?



If you're creating reports, graphics, predictive models, or other data presentations that drive business decisions then **you are involved with business intelligence.**

What We'll Cover



This presentation has three parts:

1. Understanding Operational vs. BI Data
2. ETL or something like it
3. Dimensional Modeling

Understanding Operational vs. BI Data



Operational data

- Operational data (op-data) originates with business events.
- An operational data item is the record of a business event.

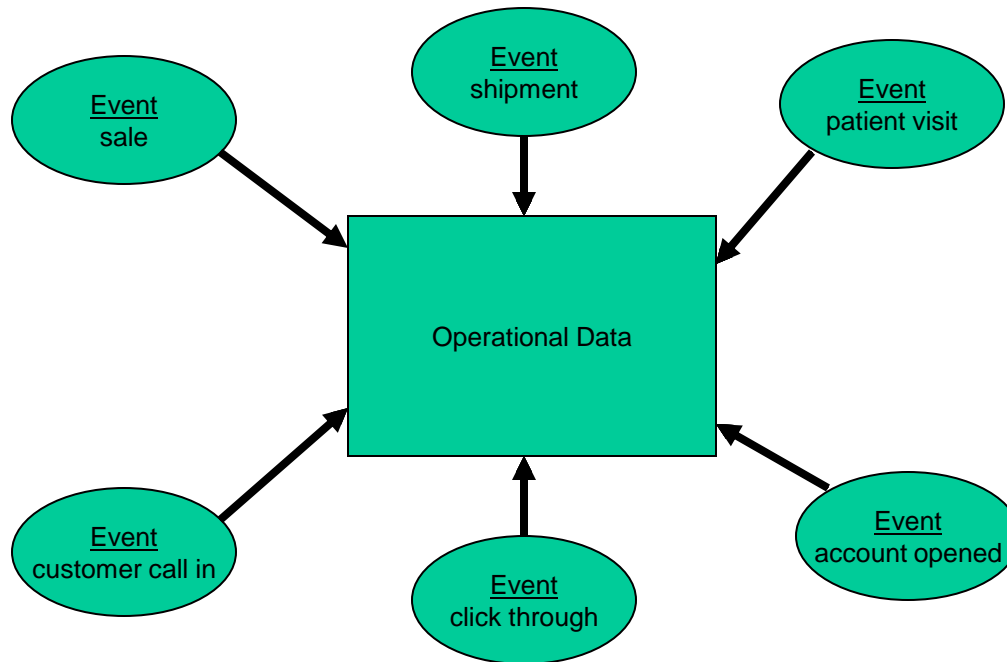
Examples:

- customer order
- patient visit
- banner ad click through
- account opened

Operational Data



Any business event worth recording is *operational data*



Operational Data – As It Happens



Op-data is typically recorded as it happens.

Examples:

- A server log record is cut when the page under a banner ad is served.
- An order entry record is added when a customer makes an order.
- A ship log record is entered when an item goes out the door.

System of record



Operational data systems are often systems of record.

An event doesn't really exist until it's in the system of record.

- Orders
- Payments
- Shipments
- Work/process completed



- Insert efficiency
- Up time
- Integrity

Operational data systems are not concerned with...

- Query efficiency
- A database optimized for reports and data analysis

...these are the concerns of Business Intelligence



For any data where you do not control...

- timing
- quality
- change management

That data can be considered as operational data...

...even if not truly operational data from the business viewpoint.

“Virtual op-data”



1. *Clearly define and understand the content, timing, and quality of your (virtual) op-data.*

BI = not op-data



- Business Intelligence is not concerned with capturing data.
- Business Intelligence is the system we use to analyze op-data.



- BI is after the fact.
- BI does not show us what's happening right now.
- Time and timing are critical.
- BI data represent a discrete slice of time...
 - last year
 - last month
 - yesterday

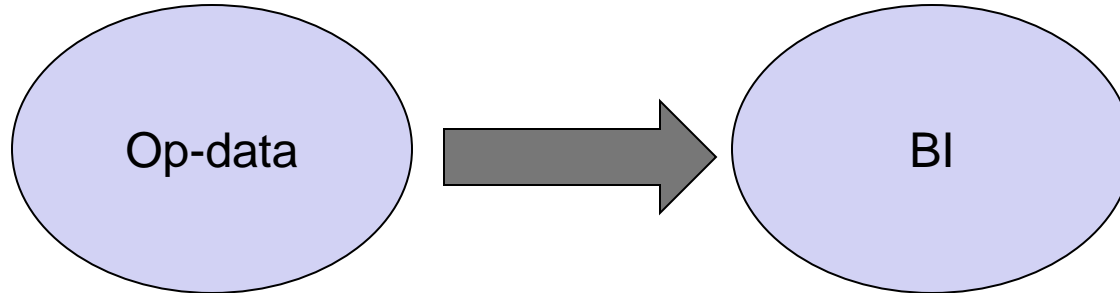


- 2. Make sure your input data's time attribute is at a finer level of detail (grain) than required for your BI system...
...or ensure that its time aligns with your presentation requirements.*

One way data flow



In a BI system data flows one way...
...from the operational system to the BI system.



Op-data is part of the system of record.
BI systems do (should) not change business events after the fact.



- 3. Your BI system should have explicit or implicit access controls that prevent update of op-data.*

When dealing with true operational data this principle is critical for internal control, compliance audits, and system integrity.



A Business Intelligence system is any system that...

1. Uses, but does not create, operational or virtual operations data to...
2. Create a database accessible by analyst's and end-users as it...
3. Isolates and protects operational data from the impact of those end-user access and...
4. Reformats the operational data to optimize end user tasks.



ETL is...

- Extract - get a copy of op-data
- Transform - validate, summarize, etc.
- Load - add to user-accessible database

What ETL Does



Basic functions:

- Copy operational data to the BI side in a timely and innocuous (to the operational system) manner.
- Create a fixed-in-time snapshot of the op-data.

Optional functions:

Validate

Backup

Check-point

Summarize

Version stamp

Create meta-data

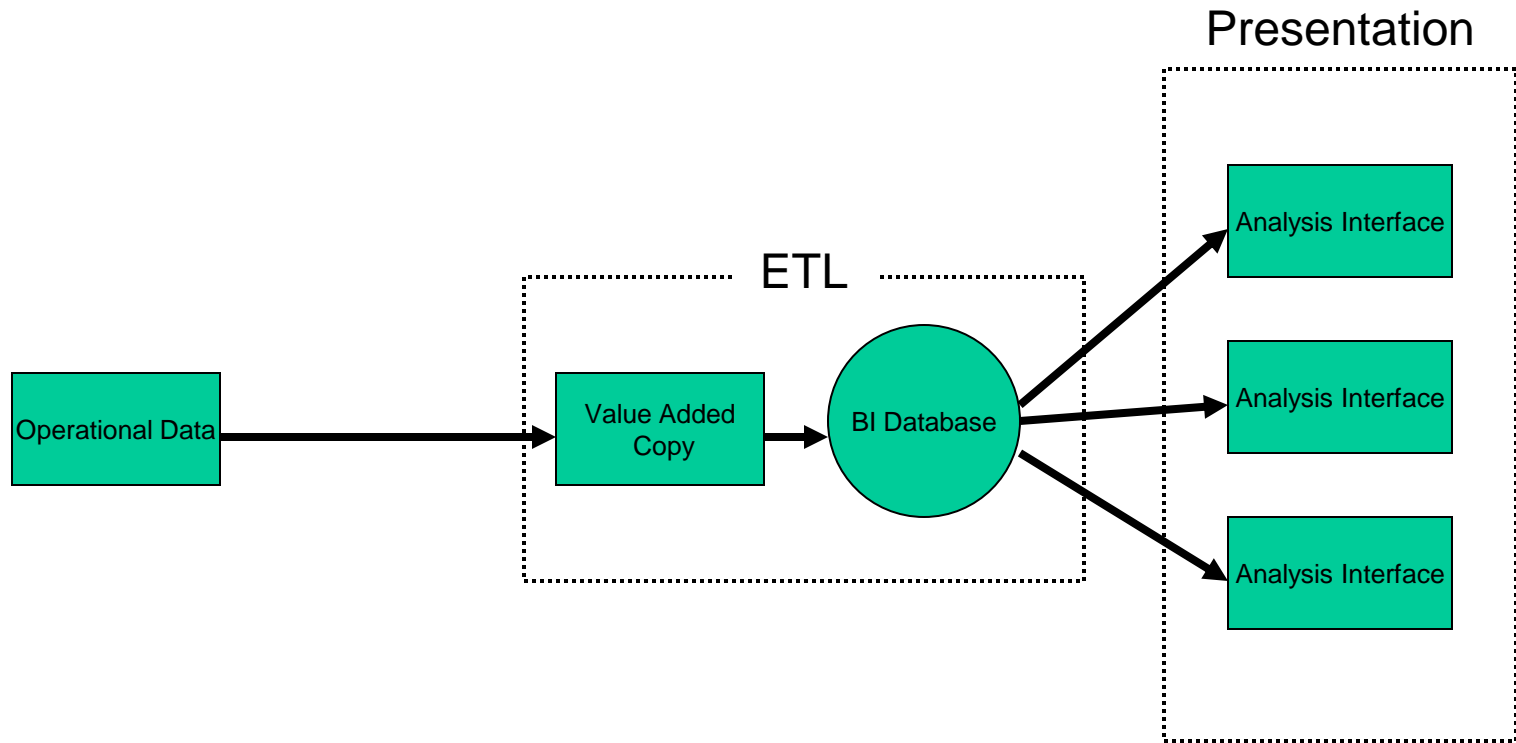
Apply access rules

...and on and on...

The Place of ETL



ETL isolates op-data from the unpredictable needs (and misdeeds) of users.



ETL loads the presentation database.



Copy

Copy data from op-data systems because BI does not work with actual op-data.

Validate

Apply rules for data validation and quality.

Transform

Decode, summarize, filter, etc.

Load

Insert into the user-accessible database.
(Data warehouse/mart, reporting tables, etc.)



In most BI systems, ETL requires the most effort and maintenance.

ETL must accommodate:

- op-data feeds from multiple sources
- data feeds with differing or perhaps random time alignments
- data feeds of dubious quality
- data feeds apt to change format without warning



Separate ETL from other processes, meaning...

ETL should not be integrated with reporting or operational data systems.

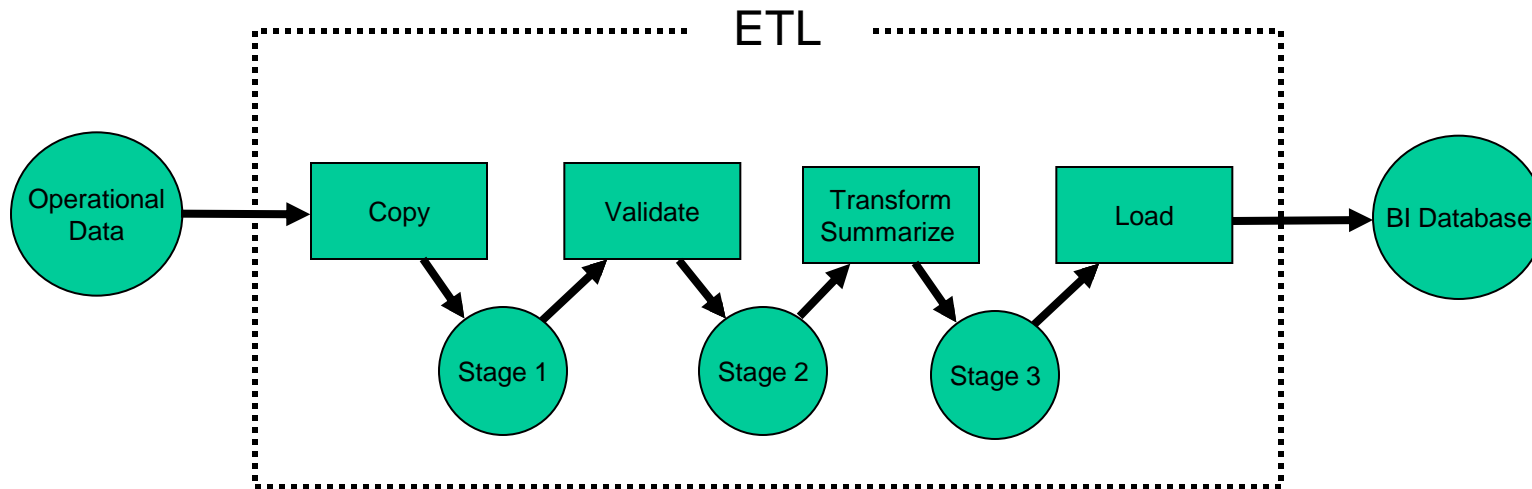
Machine efficiency is (usually) not the concern of ETL.

- Ease of maintenance vs. absolute machine efficiency
- Standardize (ISO, business codes, name standards)
- Discrete, modular, object-like vs. integrated

More ETL Stuff



- Life-cycle efficiency is the goal.
- Break your process into stages, modules, objects.





4. *Your BI system should have a discreet ETL (or ETL-like) process component that begins with a data copy or extract step.*

Dimensional Modeling: What and Why



- Dimensional modeling is a technical specification for database design.
- Dimensional modeling requires a relational database system.
- Dimensional modeling relies on tables with indexes and keys.
- SAS is a great environment for dimensional modeling.



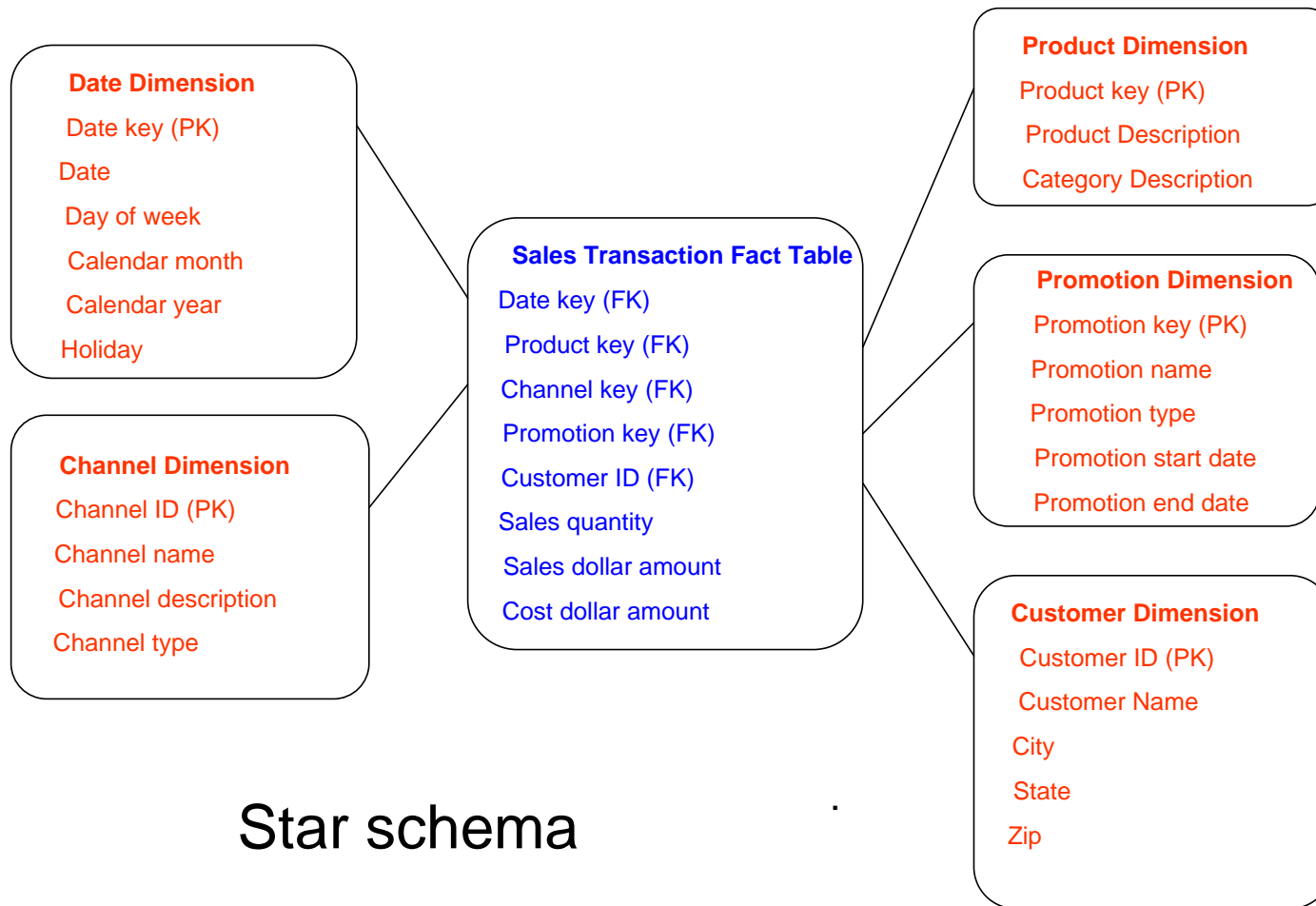
A dimensional data model is based on two types of tables:

1. Fact tables (*aka measures*)
2. Dimensional tables (*aka attributes*)

These tables are linked by keys.



Fact and dimension tables



Star schema

What's the point?



The dimensional model should be considered when...

1. *The potential number of fact rows is large compared to the number of unique dimension values.*
2. *Your dimensions are shared and/or must conform to a business standard.*
3. *You want to standardize queries or talk to a user interface that expects a dimensional model.*
4. *You are creating, or may evolve to, a data warehouse system.*

More DM Good Stuff



1. DM allows straightforward summarization, disaggregation, and joining of business events.

That is, drill up, drill down, and drill across.

2. DM facilitates control of dimension attribute changes, e.g., what happens when a customer changes name?
3. DM is relatively simple vs. a fully normalized table structure.
4. DM facilitates conformed data. Enforcing a single version of key business information such as customer and product attributes.



5. The dimensional model allows you to segregate sensitive personal (PII) or proprietary information.
6. Standalone DM systems can be integrated into larger warehouse (or warehouse-like) systems.
7. DM is time-tested, well understood, and well documented.

You don't have to reinvent the wheel!



The base SAS product provides all the tools you need to set up a dimensional database:

- Tables (SAS datasets)
- Indexing
- Integrity constraints (primary key, foreign key)
- SQL

Plus all the power of the DATA step language and PROCs.

Additional SAS products provide industry-specific processes, management facilities, user interface schemes.



Business Intelligence is (loosely) defined as:

Any system that deals with the end-user view of your data as opposed to operational data.



Four BI Principles

1. *Clearly define and understand the content, timing, and quality of your (virtual) op-data.*
2. *Make sure your input data's time attribute is at a finer level of detail (grain) than required for your BI system or ensure that it time aligns with your presentation requirements.*
3. *Your BI system should have explicit or implicit access controls that prevent update of op-data by the BI system.*
4. *Your BI system should have a discreet ETL component that begins with a data copy step.*



Consider the dimensional data model as the target database for your BI system.

DM benefits:

1. *Efficient for large amounts of data*
2. *Standard form, fast queries*
3. *Established methods for dealing with attribute changes.*
4. *Well documented, well tested in many business environments.*



Contact the author at:

Tom Miron

Systems Seminar Consultants

2997 Yarmouth Greenway Dr.

Madison, WI 53711

608 625-4541

608 278-9964

tmiron@sys-seminar.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies.