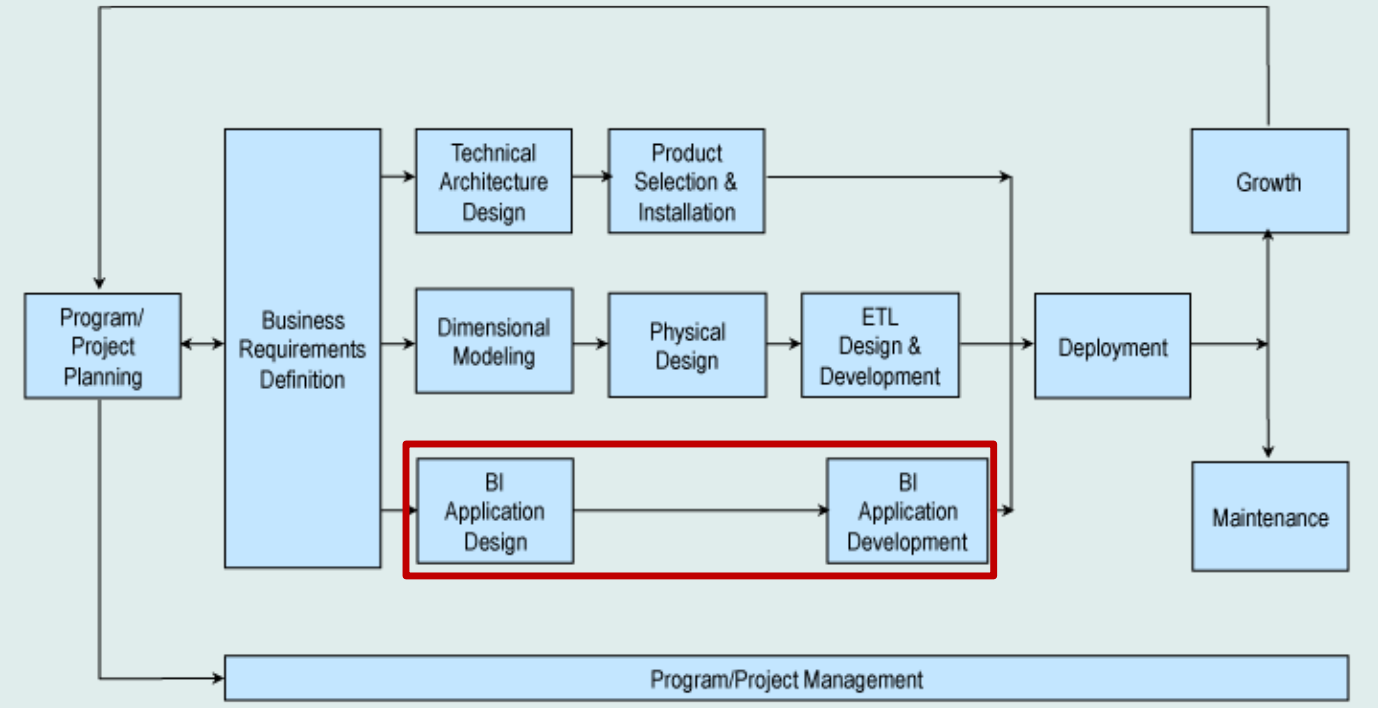
Introduction

School of Information Studies
Syracuse University

Agenda

- Define business intelligence
- Explain the six categories of BI
- Discuss how we measure BI
- Describe the process of developing BI
- Demonstrate how to build OLAP cubes in a multidimensional database

Kimball Lifecycle



BI Defined…

Inmon

"Systems that help make companies understand what makes the wheels of the corporation turn and to help predict the future impact of decisions."

Kimball

"A generic term to describe leveraging the organization's internal and external information assets to support improved business decision-making."

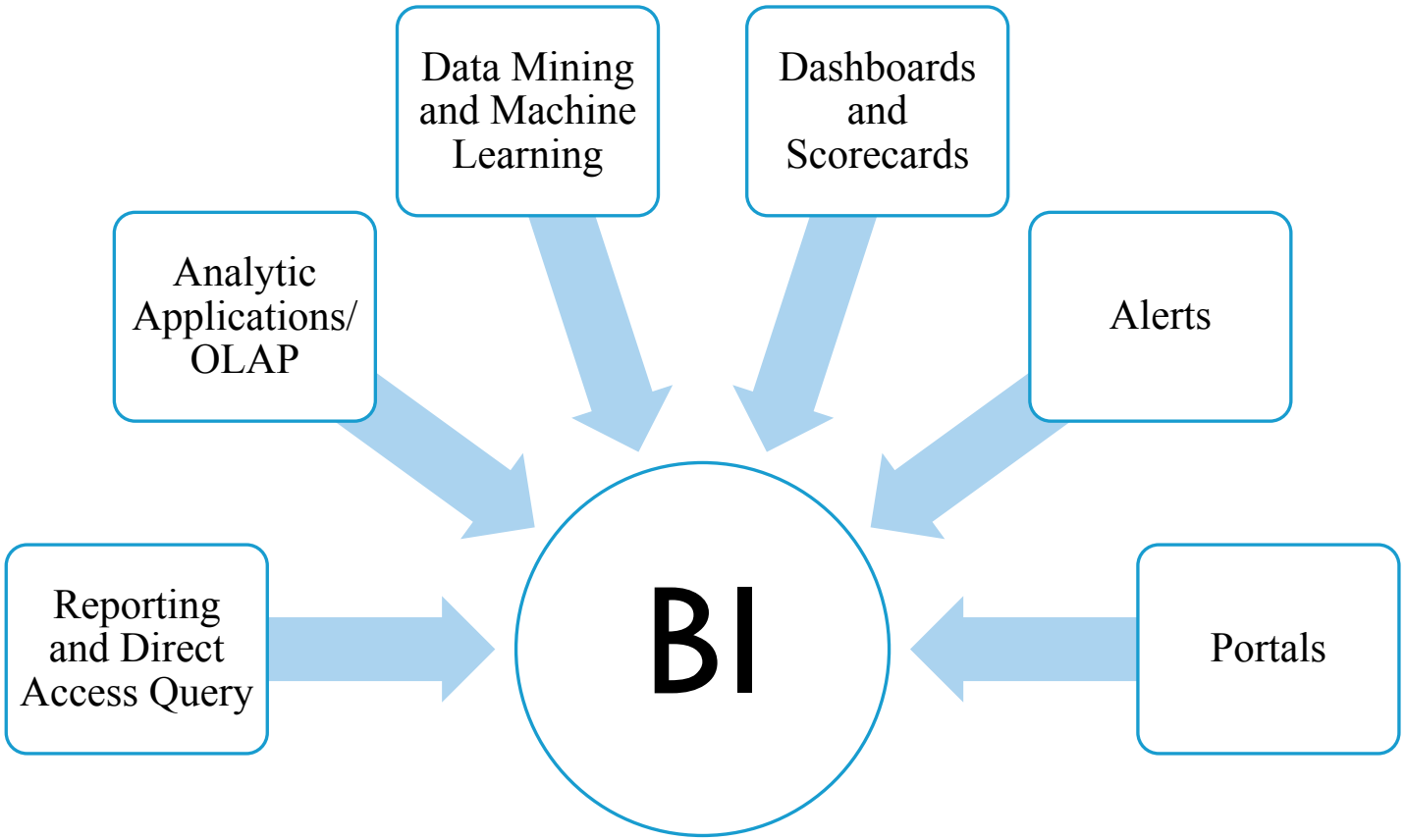
Aren't they saying the same thing?

DW Is the Foundation for BI

Introduction

School of Information Studies
Syracuse University

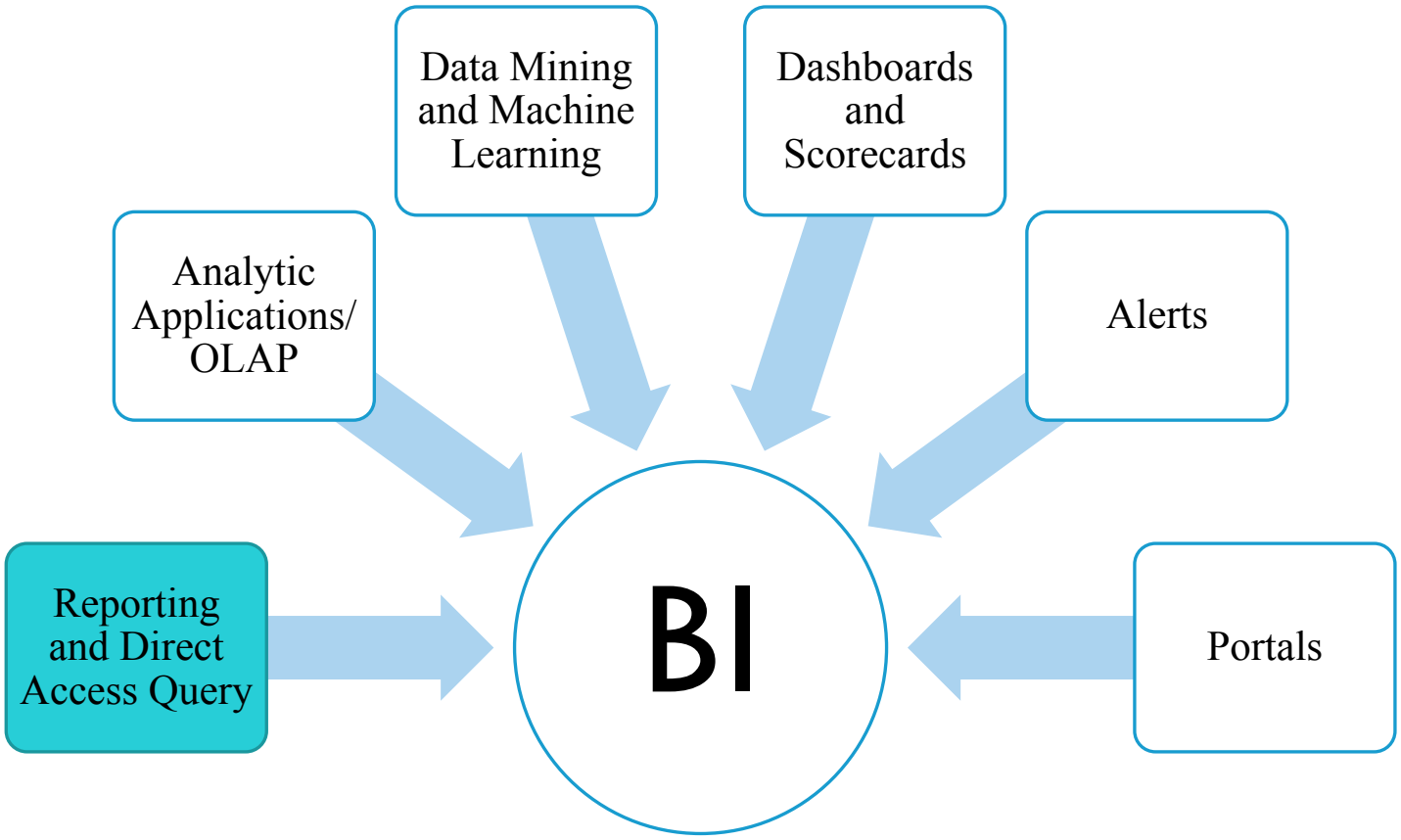
Six Categories of BI



Reporting

School of Information Studies
Syracuse University

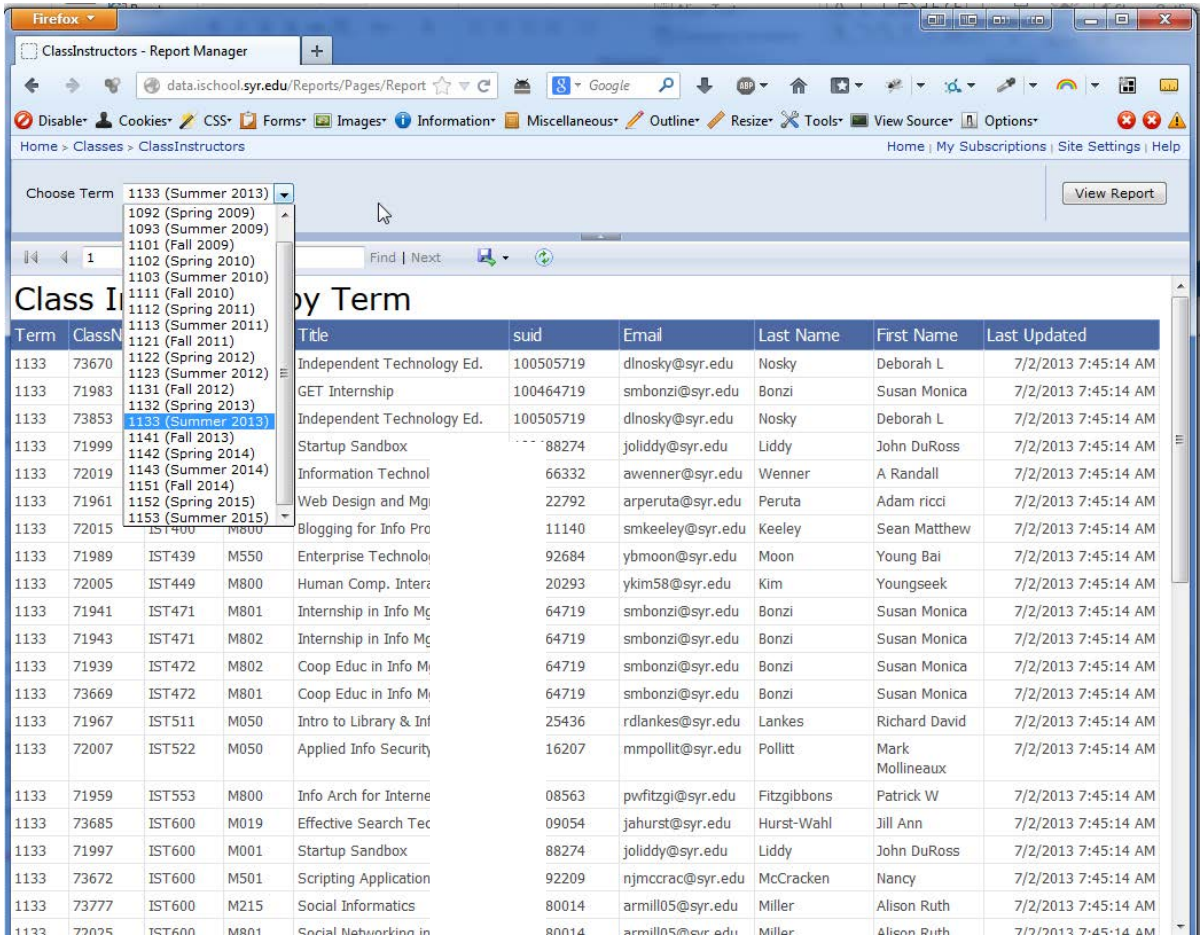
Six Categories of BI



Reports

- Information presented in a tabular, matrix format or as a chart.
- Parameters can be added to make them dynamic.
- Typically delivered to business user or available in the portal.
- Pros:
  - Simple to build and execute
  - Easy to understand without training
- Cons:
  - Not flexible
  - Meets only a specific informational need

Example: Reports



School of Information Studies
Syracuse University

Direct Access Query and Reporting Tools

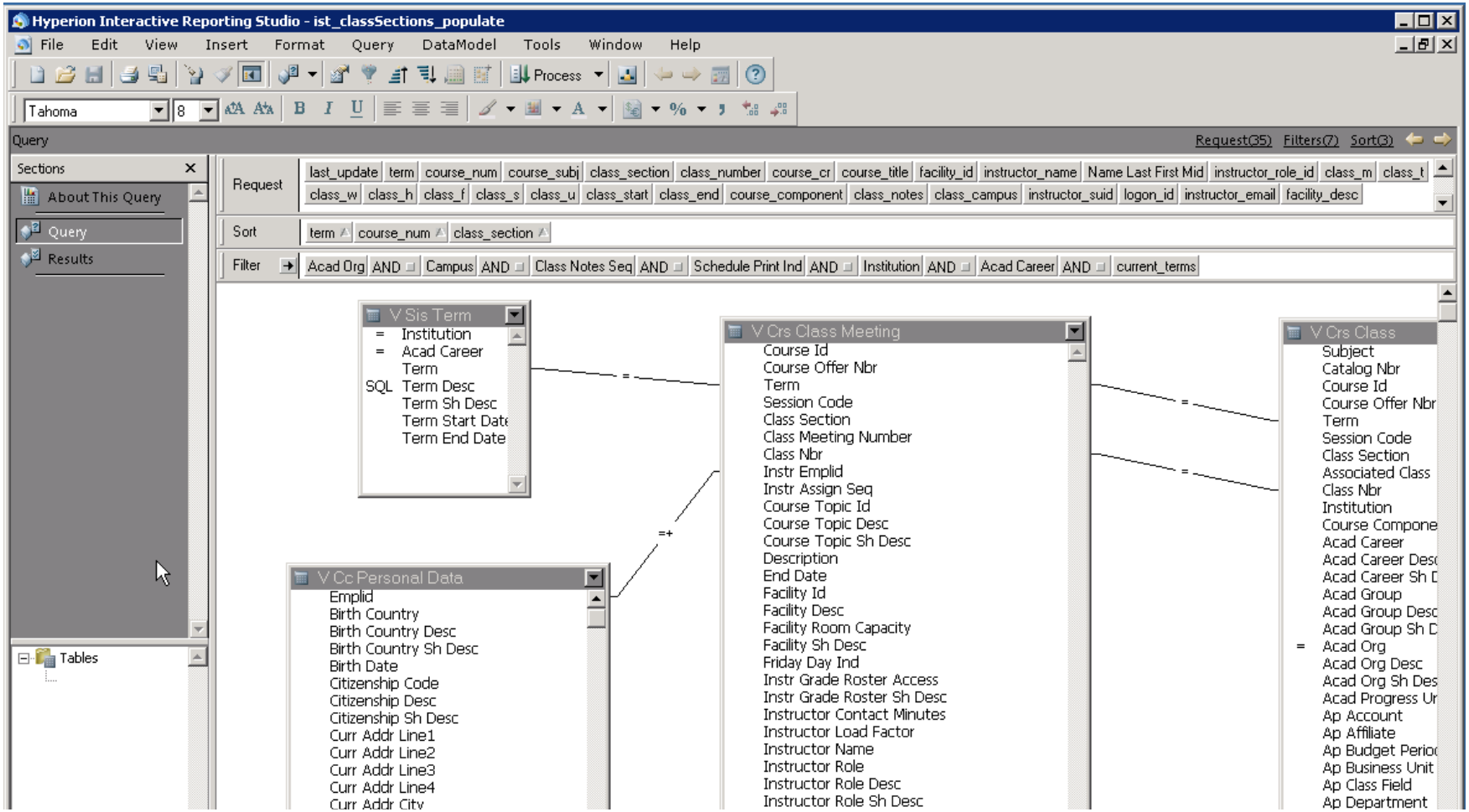
Power users have access to software and the dimensional model for writing their own queries.

Four key functions of these tools:

- **Query formulation:** assist with data queries
- **Analysis and presentation capabilities:** placing the data in "presentation quality" format
- **User experience:** metadata access, easy to use, prevent misuse of data
- **Technical features:** multitasking, scheduling, import/export

Examples: MS Excel, Hyperion

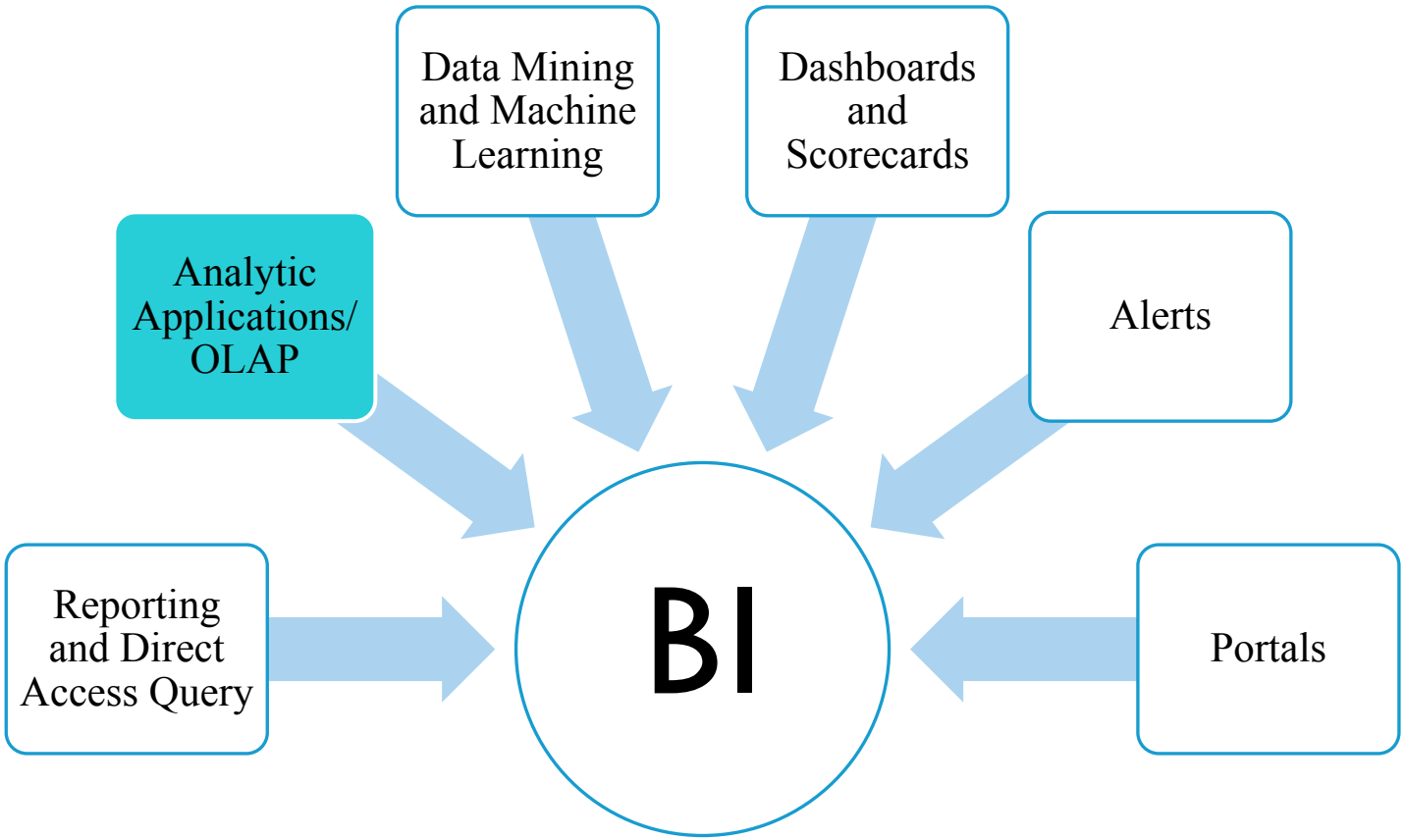
Example: Query Tool



Analytic Applications

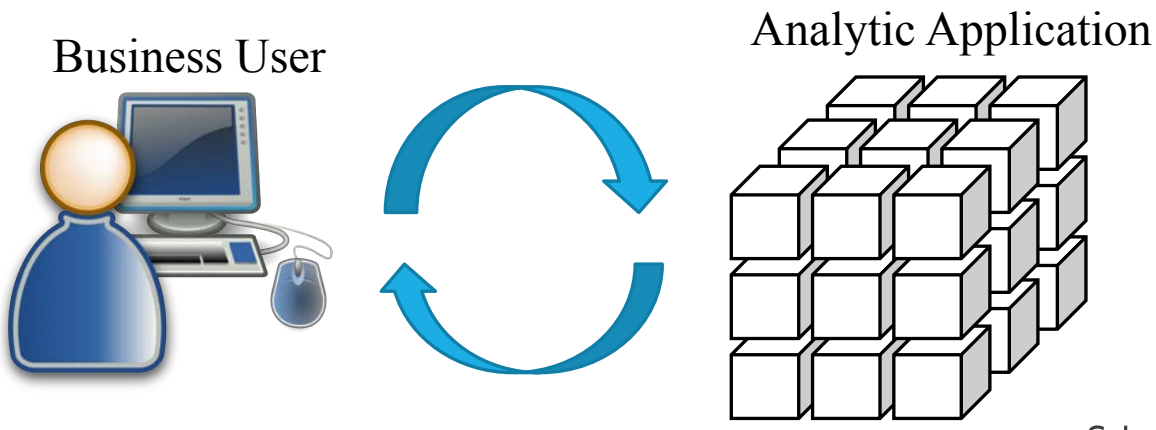
School of Information Studies
Syracuse University

Six Categories of BI



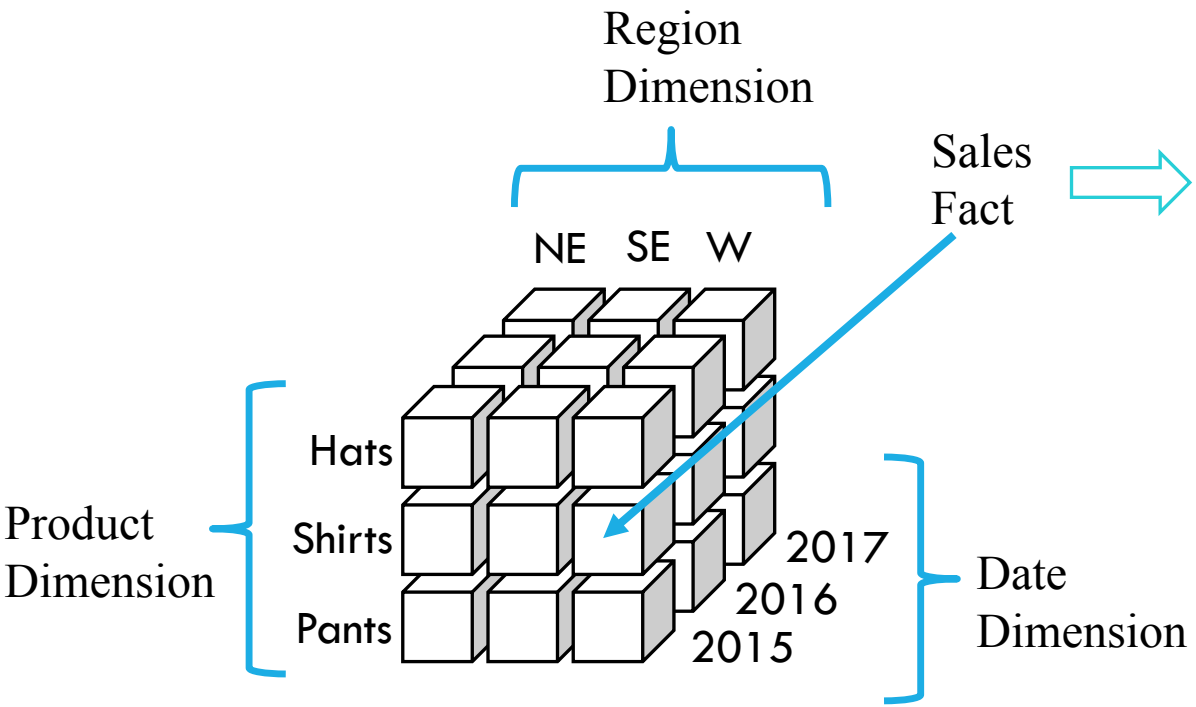
Analytic Applications

- Based on the DDS, use fact tables and dimensions.
- Users interact with the data and use them to answer their own questions.
- OLAP = online analytical processing. Instead of a report, presents an explorable dataset.
- Typically use a multidimensional database.



OLAP Cube

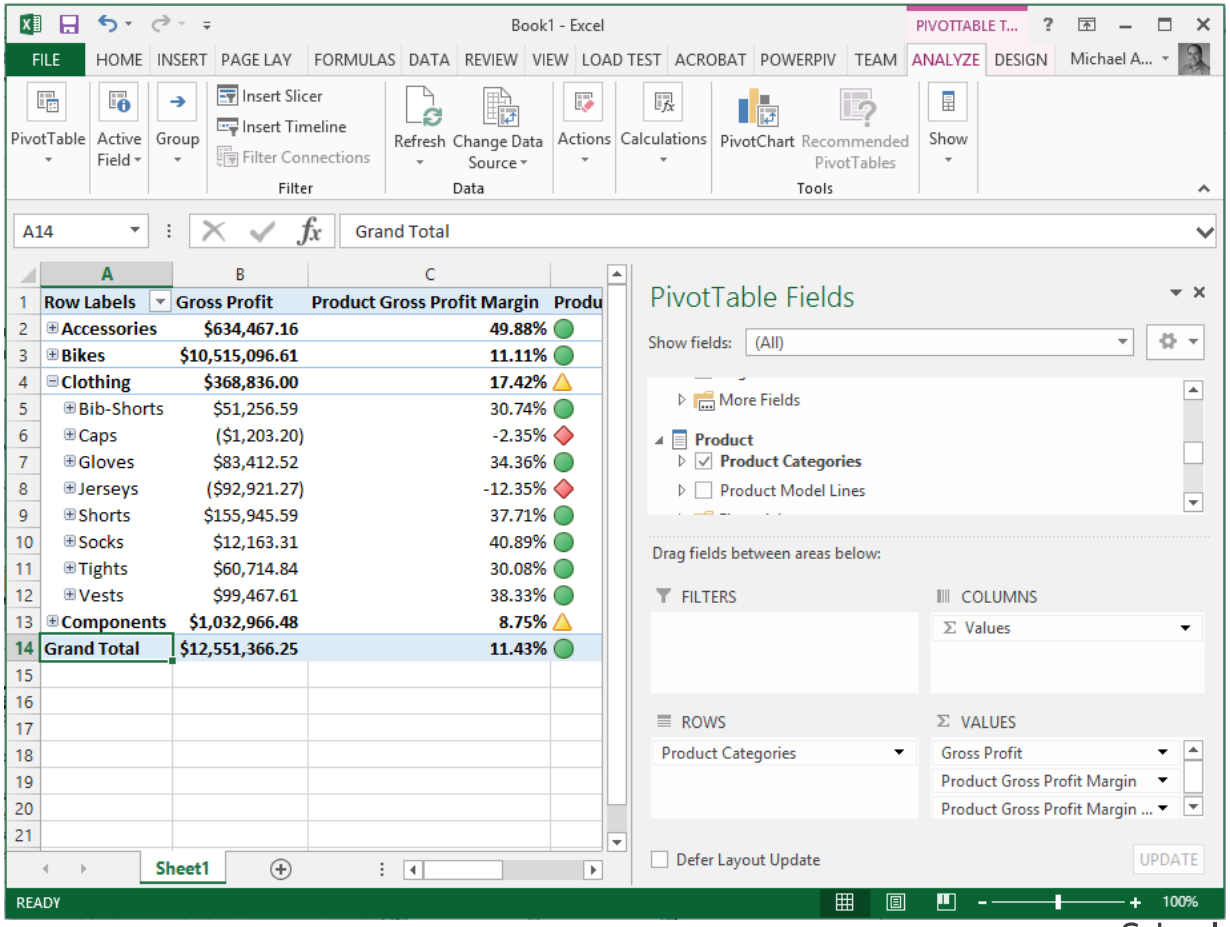
Data structure for representing analytic data.



Sold **150** **Shirts**
in the **W** region
in year **2015**.

Region: **W**
Product: **Shirts**
Year: **2015**
Sales Fact: **150**

Example: Analytic Application

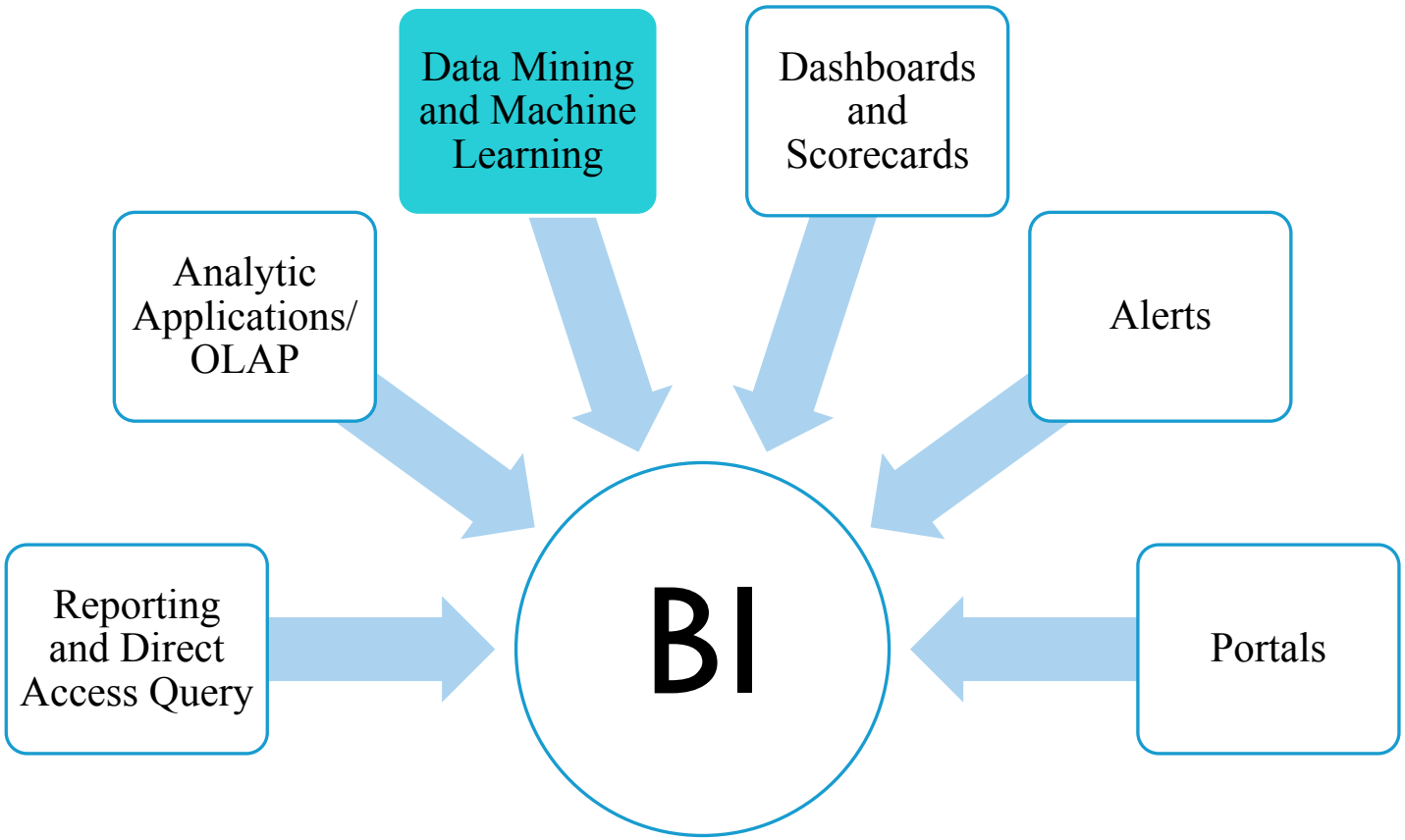


School of Information Studies
Syracuse University

Data Mining

School of Information Studies
Syracuse University

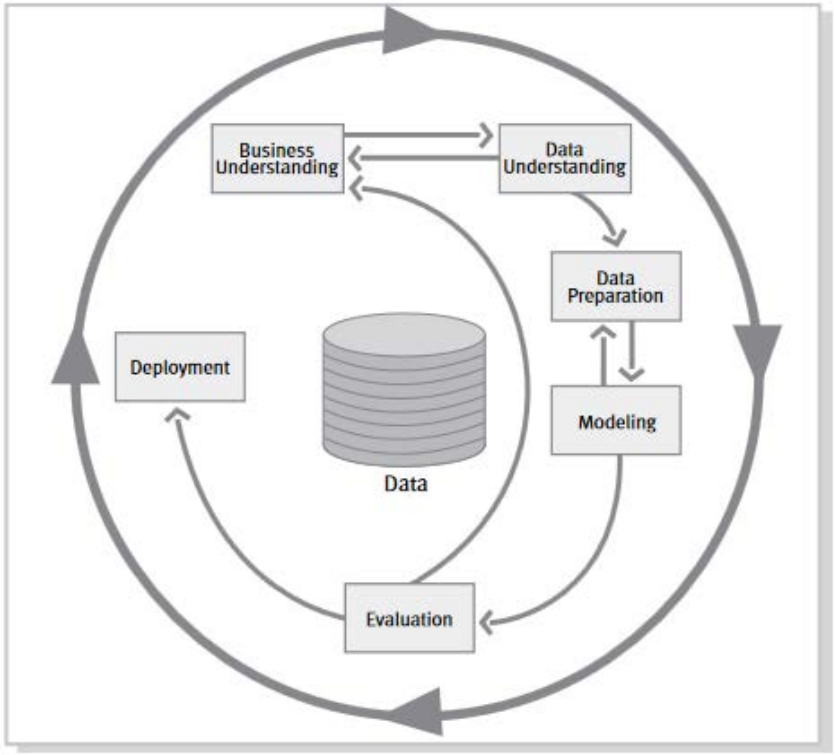
Six Categories of BI



Data Mining

- A process of **data exploration** with the intent to find **patterns, relationships, or insights** of organizational value.
- Data mining is part of **descriptive** and **predictive** analytics.
  - **Clustering and classifying data:** "Customer 1 is Type A, Customer 2 is Type B."
  - **Estimating and predicting:** "Customer Type A will spend $N this year."
  - **Affinity grouping:** "Customers who buy Product X are also likely to buy Product Y."
  - **Anomaly detection:** fraud detection, unusual patterns, outliers.

CRISP-DM: Cross-Industry Standard Process for Data Mining (2000)



CRISP-DM Step-By-Step Guide 1.0

https://www.the-modeling-agency.com/crisp-dm.pdf

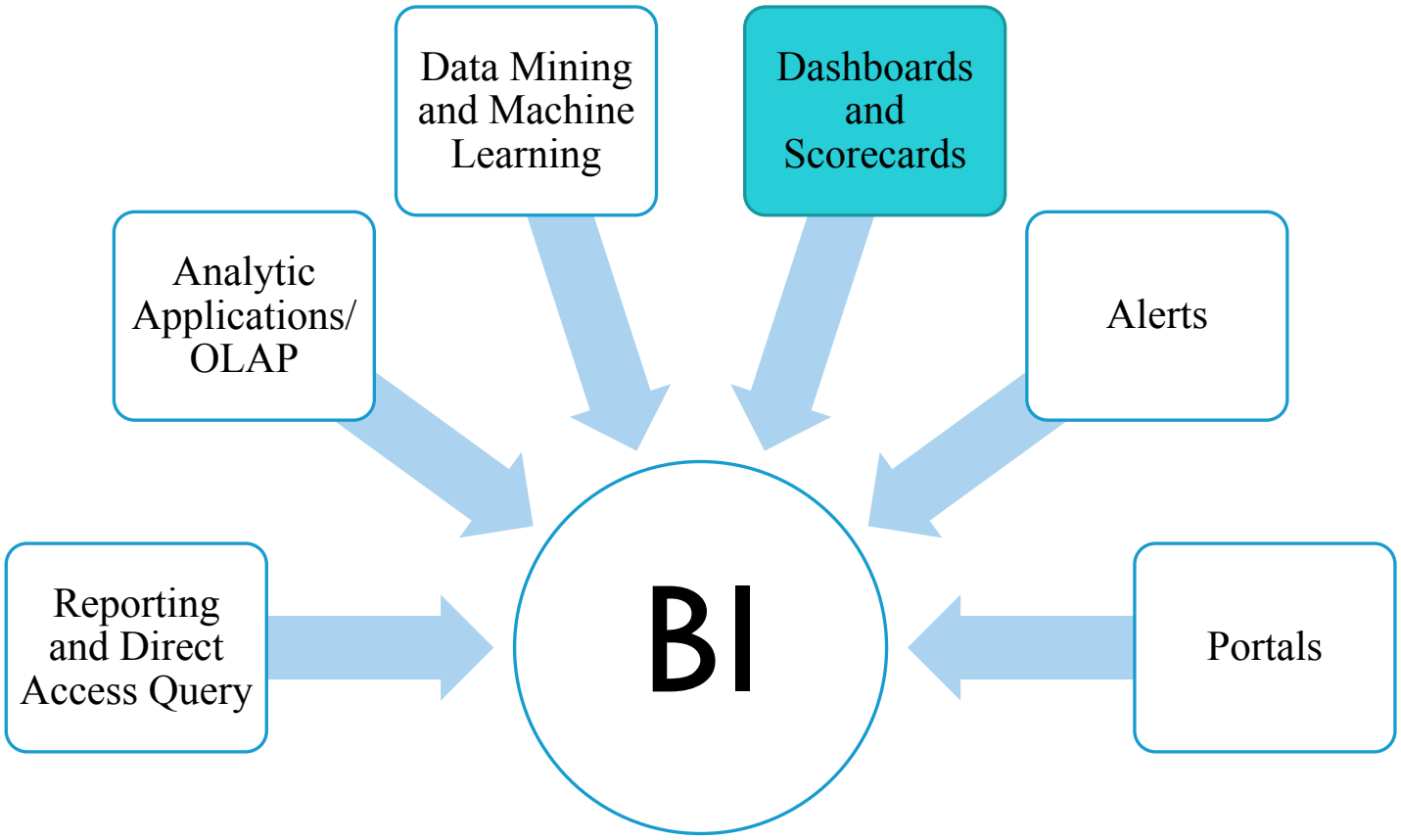
Data Mining and Machine Learning

- What is the relationship between these two?
- We use **machine learning** algorithms to perform **data mining**.
- Machine learning is divided into **supervised** and **unsupervised** algorithms.
  - **Supervised** algorithms are trained using a feature set and labeled targets to predict future unknown values.
  - **Unsupervised** algorithms discover patterns or relationships in data from a feature set.

Dashboards and Scorecards

School of Information Studies
Syracuse University

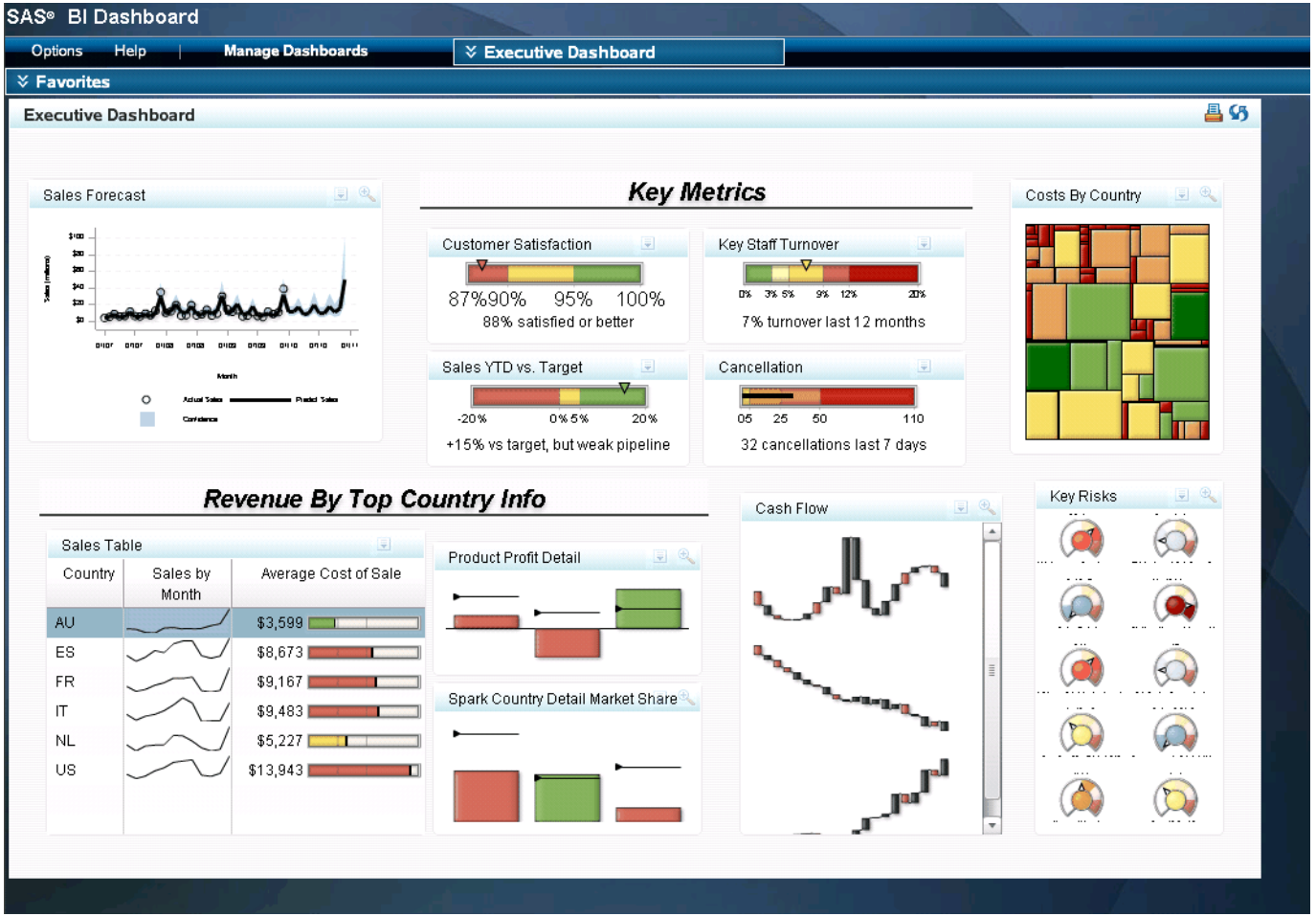
Six Categories of BI



Dashboard and Scorecards

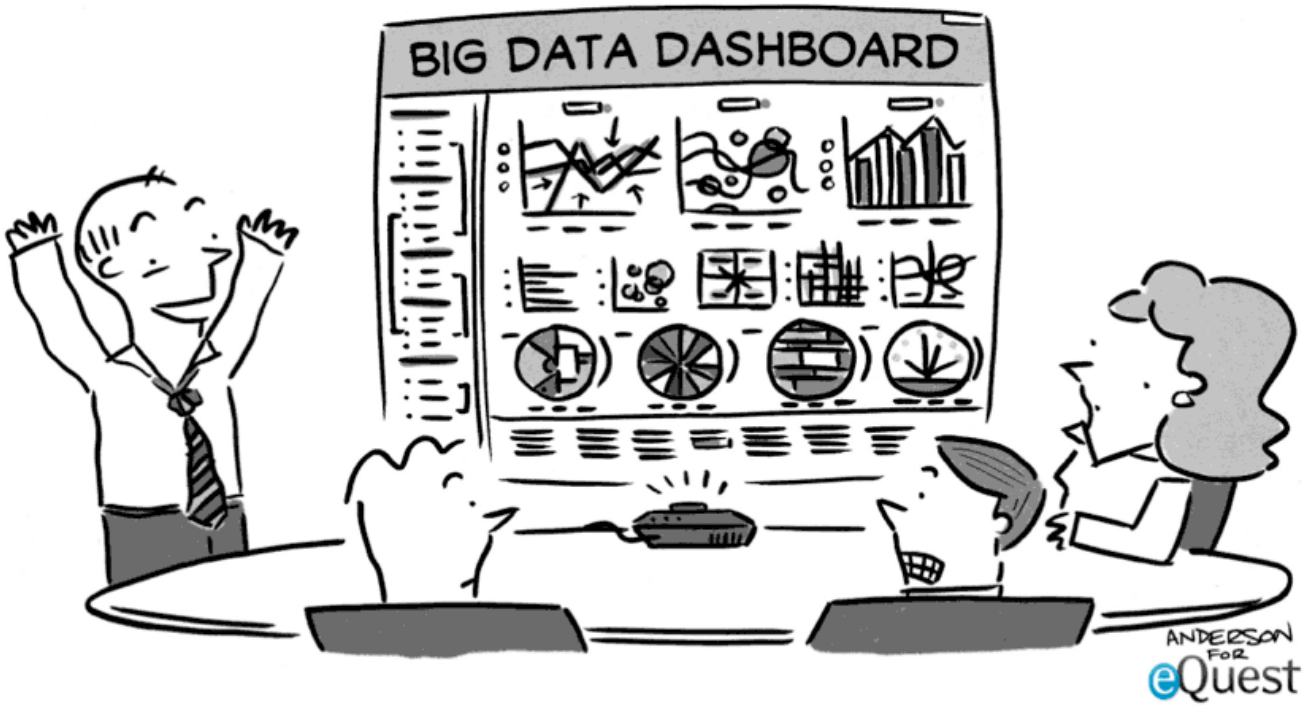
- Originally just executive interfaces, now organization-wide
- Provide a high-density and information-rich visual representation of data
- Usually web-based and highly interactive
- Consist of gauges, charts, and cards
- Contain KPIs (key performance indicators) for measuring goals
- As much of an organizational challenge as it is a technical one

Example: Dashboard



School of Information Studies
Syracuse University

Just Because You Can Doesn't Mean You Should

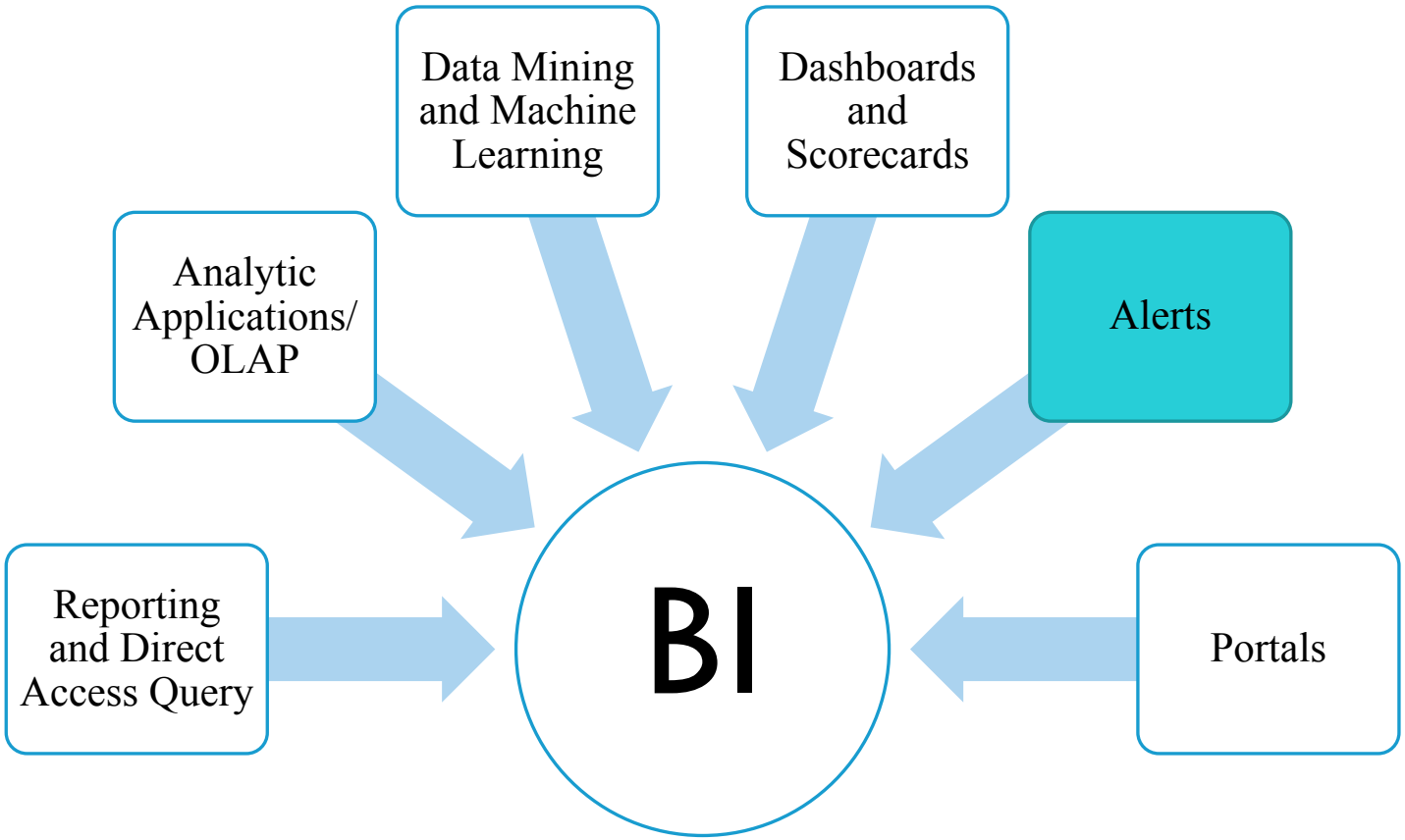


"After careful consideration of all 437 charts, graphs, and metrics, I've decided to throw up my hands, hit the liquor store, and get snockered. Who's with me?!"

Alerts

School of Information Studies
Syracuse University

Six Categories of BI



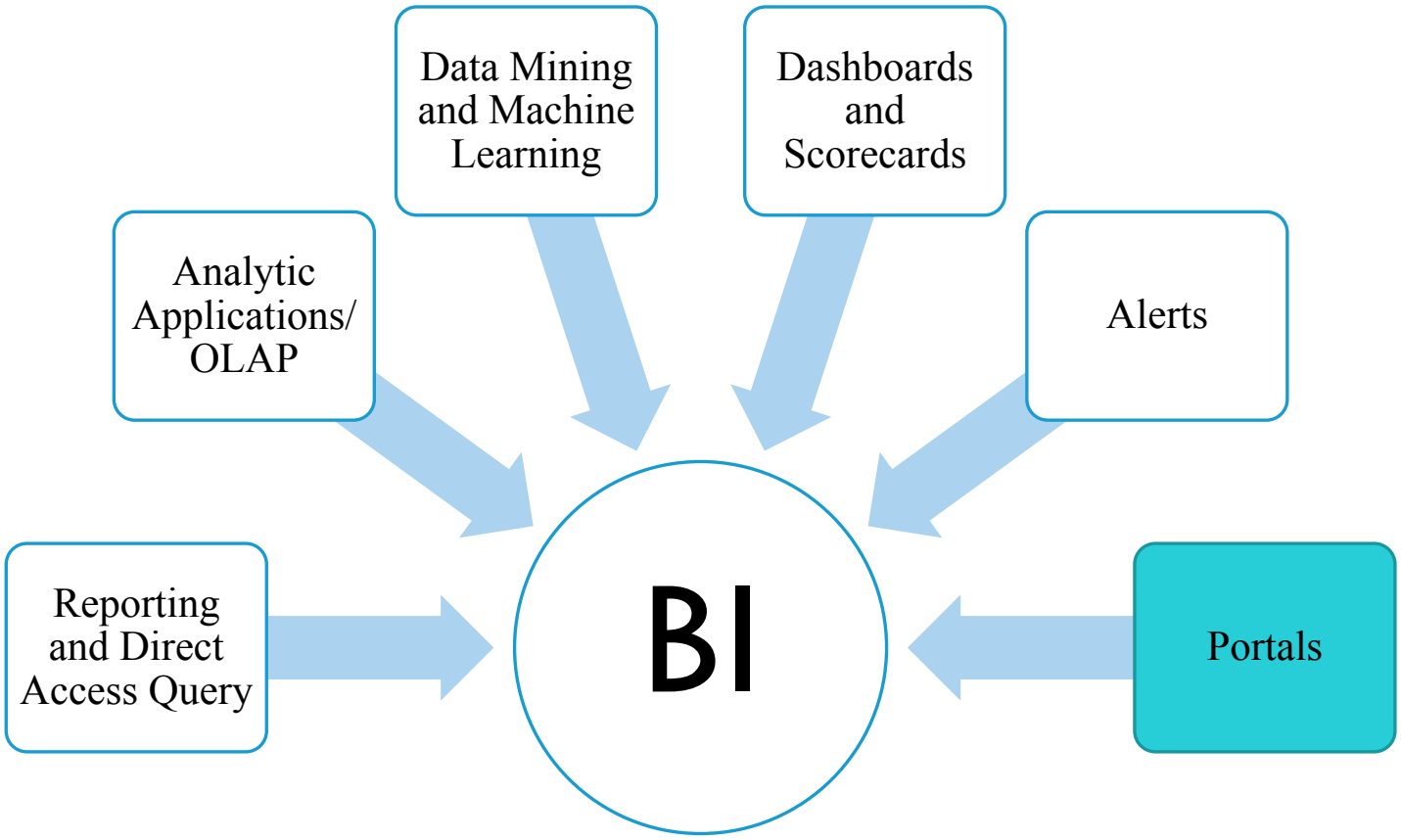
Alerts

- Notifications to business users that an event or condition has happened
- Allow business users to react or respond
- Examples:
  - An inventory manager receives notification when a stock level drops too low.
  - The customer account representative receives a notification when one of his assigned customers completes the feedback form on the company's website.
  - When the daily sales figure for a retail store drops under a certain percentage of the target.

Portals

School of Information Studies
Syracuse University

Six Categories of BI



The BI Portal

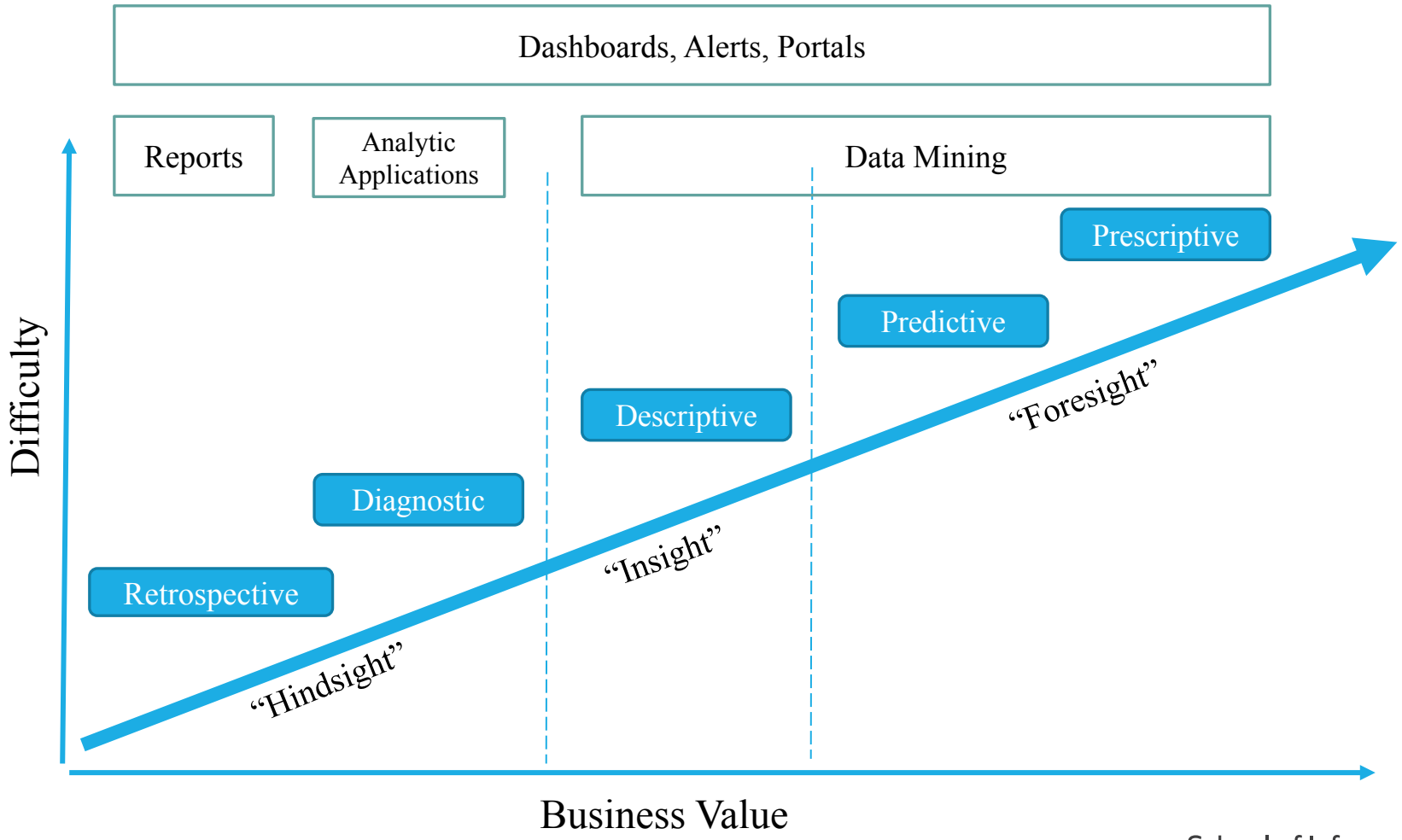
- Provides access to BI applications and data.
- Security can be set up to restrict access.
- Makes it easier to navigate a large BI implementation.
- Requirements of any BI portal:
  - **Useable: e**asy to find what you need
  - **Content rich: i**nformation shown in context
  - **Clean:** simple design not overwhelming
  - **Current:** new content often
  - **Interactive:** browse data, customization for relevance
  - **Value-oriented:** users need to see value in it

Example: Portal

Revisiting the Five Types of Analytics

School of Information Studies
Syracuse University

Six Categories and Analytics



Three Types of BI

School of Information Studies
Syracuse University

Three Types of BI

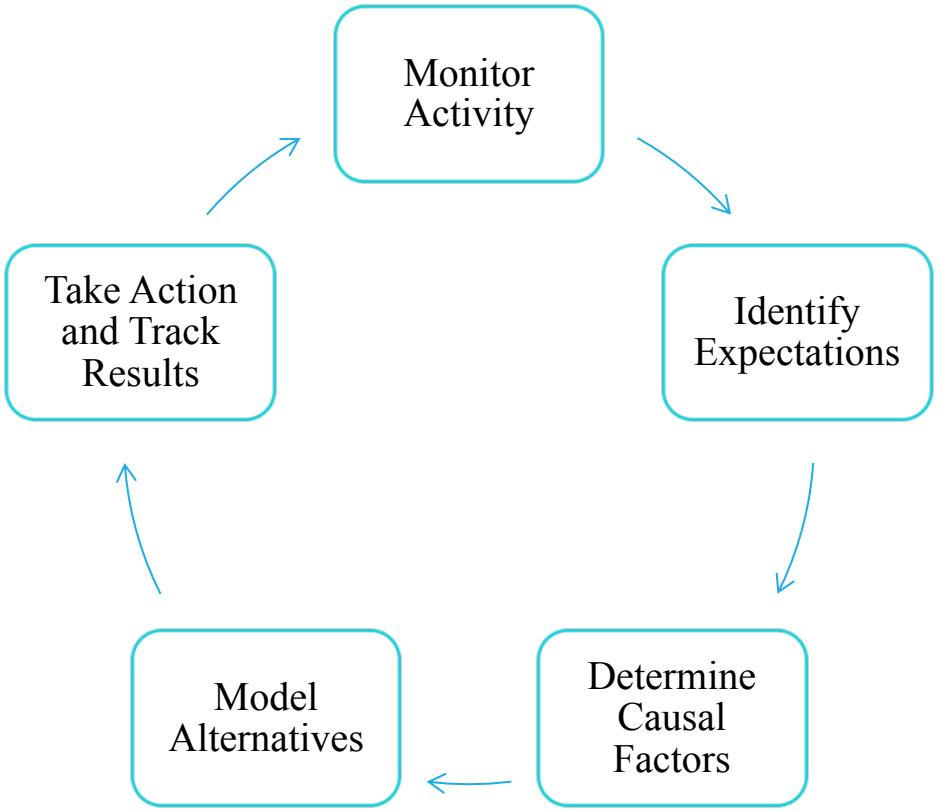
	Operational BI	**Tactical BI**	**Strategic BI**
Business Focus	Manage daily operations; integrate BI with operational systems	Conduct short-term analysis to achieve strategic goals	Achieve long-term organizational goals
Purpose	Make immediate decision	Short-term decision-making	Long-term planning
Primary Users	Analysis, operational users	Executives, managers	Executives, managers
Data	Real-time metrics	Historical metrics	Historical metrics

Analytic Cycle for BI

School of Information Studies
Syracuse University

Analytic Cycle for BI Analysis

- A model for BI development
- Improve BI through feedback of business users
- This helps us :
  - Understand how our users will use BI
  - Determine the tools we must provide to make their experience positive and productive



Multidimensional Database Management Systems

School of Information Studies
Syracuse University

What Is a Multidimensional Database?

- A multidimensional database is a storage mechanism for dimensional data. It consists of **dimensions** and **cubes**.
- Facts are stored in cube cells referenced by the atomic values of the dimension, similar to a multidimensional array.
- Aggregations are precalculated at higher levels in each dimension and stored in the cube cells, too.
- The underlying structure is very compact and fast compared to a relational star schema.
- The drawback being it takes longer to add/update data as the time to precalculate each aggregation can be time-intensive.

OLAP, ROLAP, HOLAP, MOLAP

- OLAP, or online analytical processing, is the act of exploring and interacting with dimensional data (facts and dimensions).
- OLAP can be done with relational data, multidimensional data, or both.

OLAP Type	DBMS	Data Stored in:
ROLAP	Relational	Star schema
MOLAP	Multidimensional	Cube
HOLAP	Hybrid (Both)	Data in star schema, aggregates in cube

Multidimensional Database Key Features

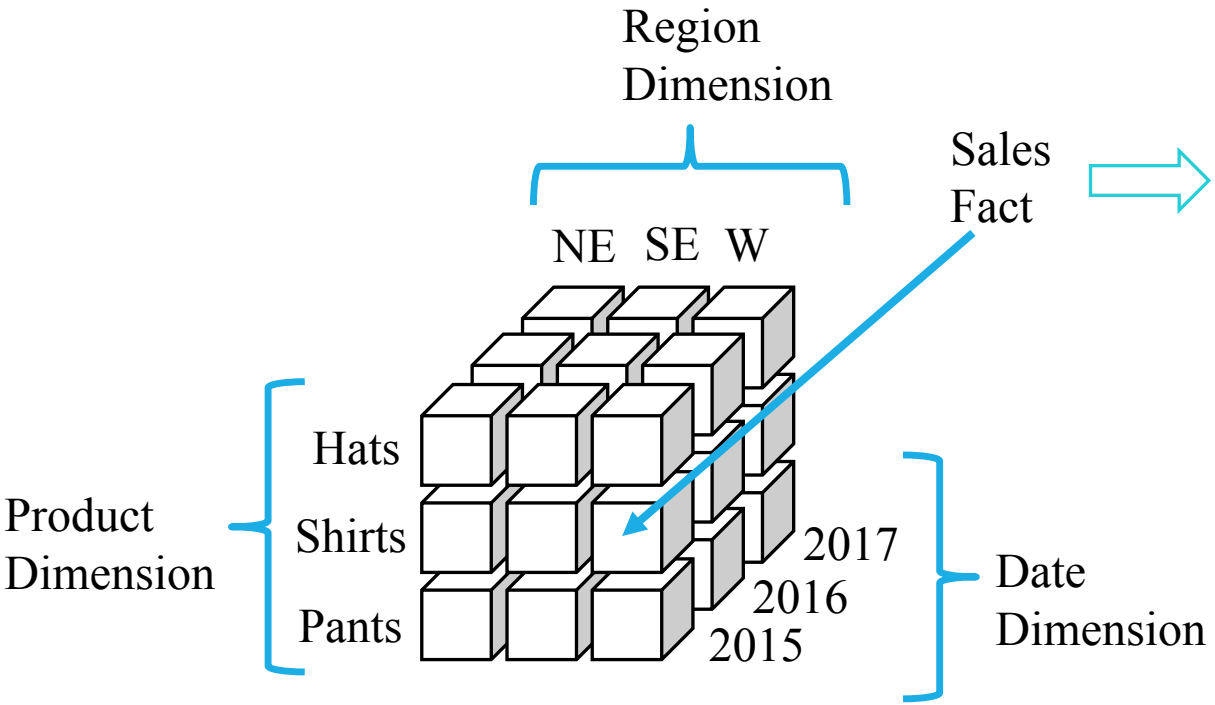
- Consolidate multiple data sources into a single cube
- Independent processing of facts and dimensions
- Semantic model:
  - Unknown members
  - Robust attribute properties
  - Define hierarchies
- Predefined aggregates (attribute relationships)
- Calculations/KPIs
- Perspectives

OLAP Terminology

School of Information Studies
Syracuse University

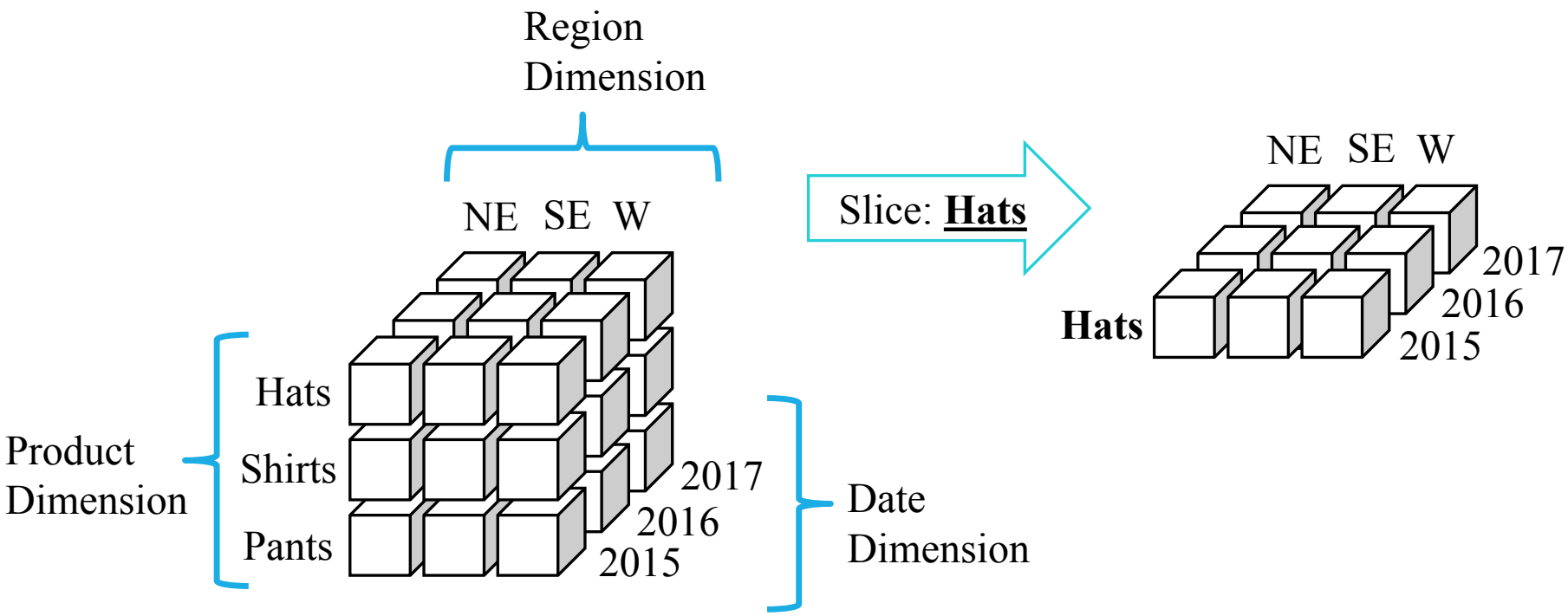
OLAP Cube

Data structure for representing analytic data.



OLAP Terminology: Slice

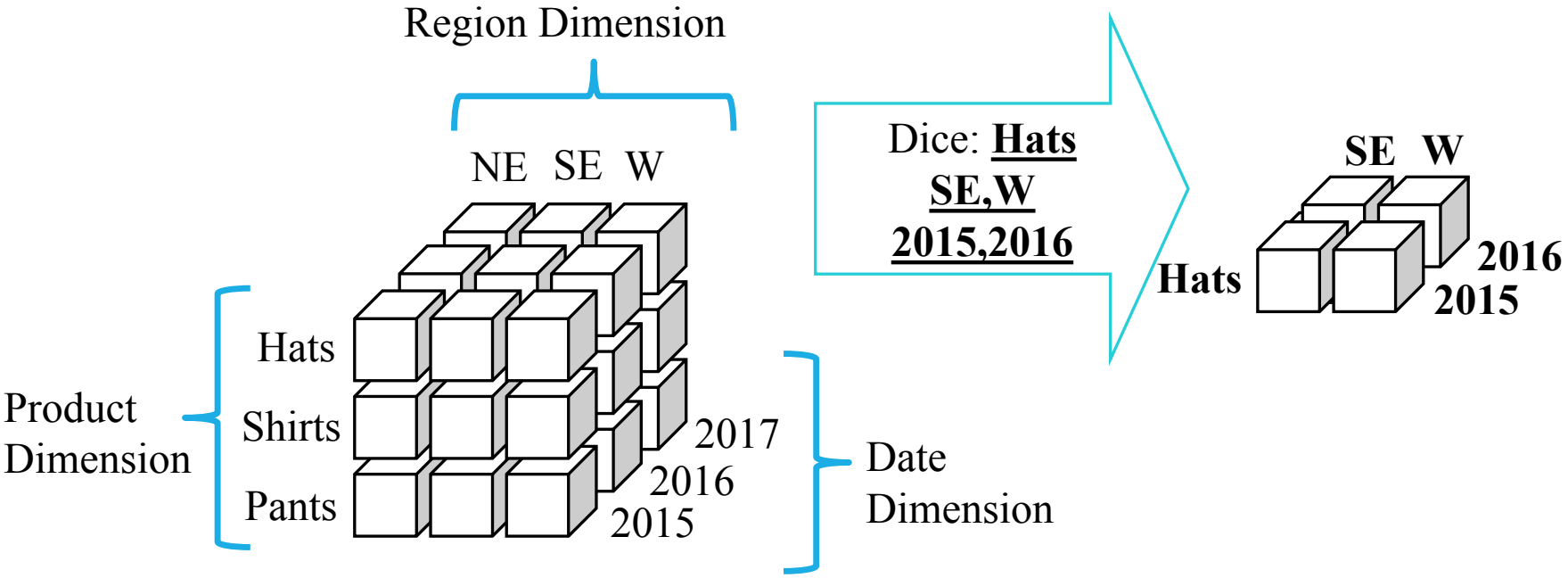
Slice: Filtering on one dimension



OLAP Terminology: Dice

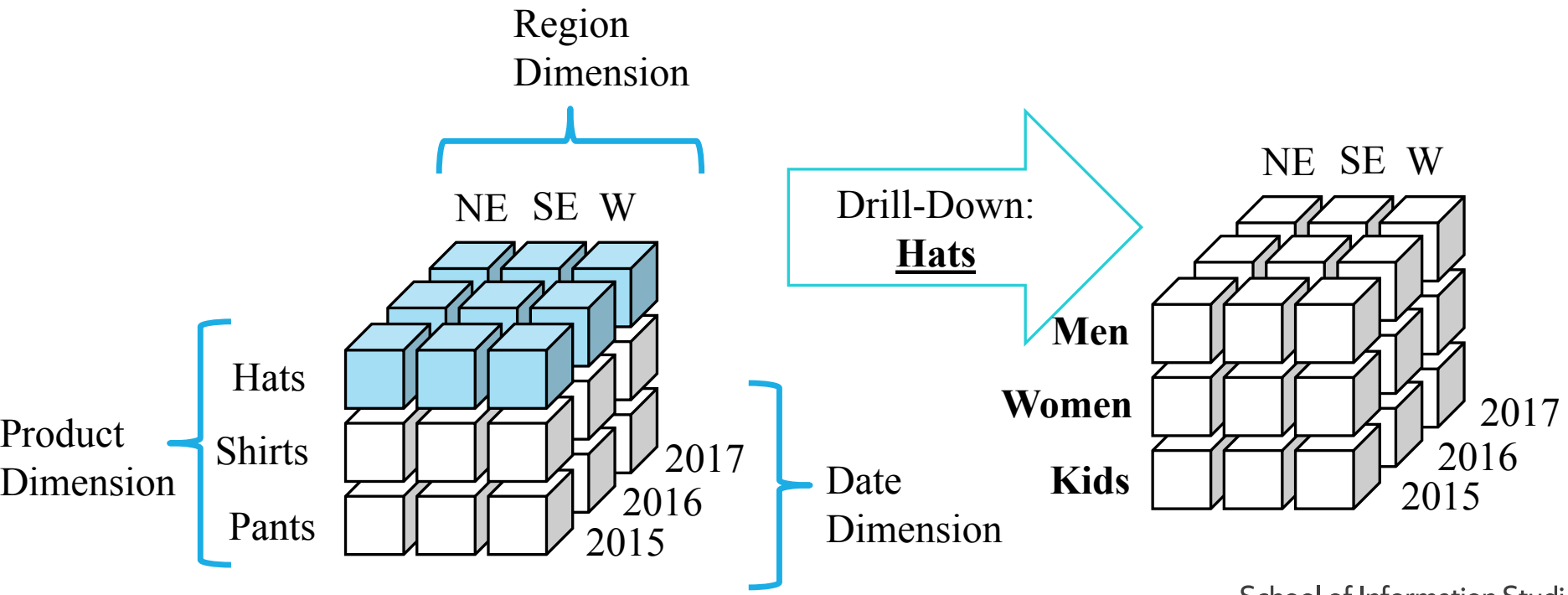
Dice: Filtering on more than one dimension to make a subset of the cube.

Perspective: A named dice of the cube



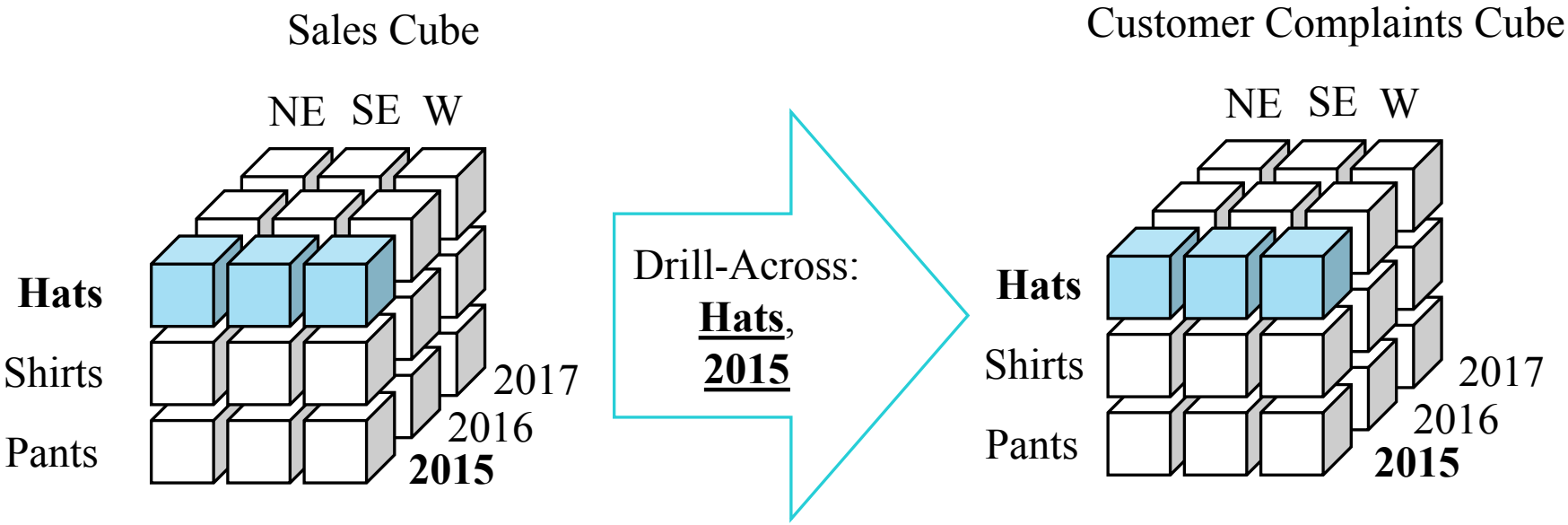
OLAP Terminology: Drill-Down/Up

- **Drill-down:** Moving down a level in a dimension hierarchy to view more detail
- **Drill-up:** Moving up a level in a dimension hierarchy



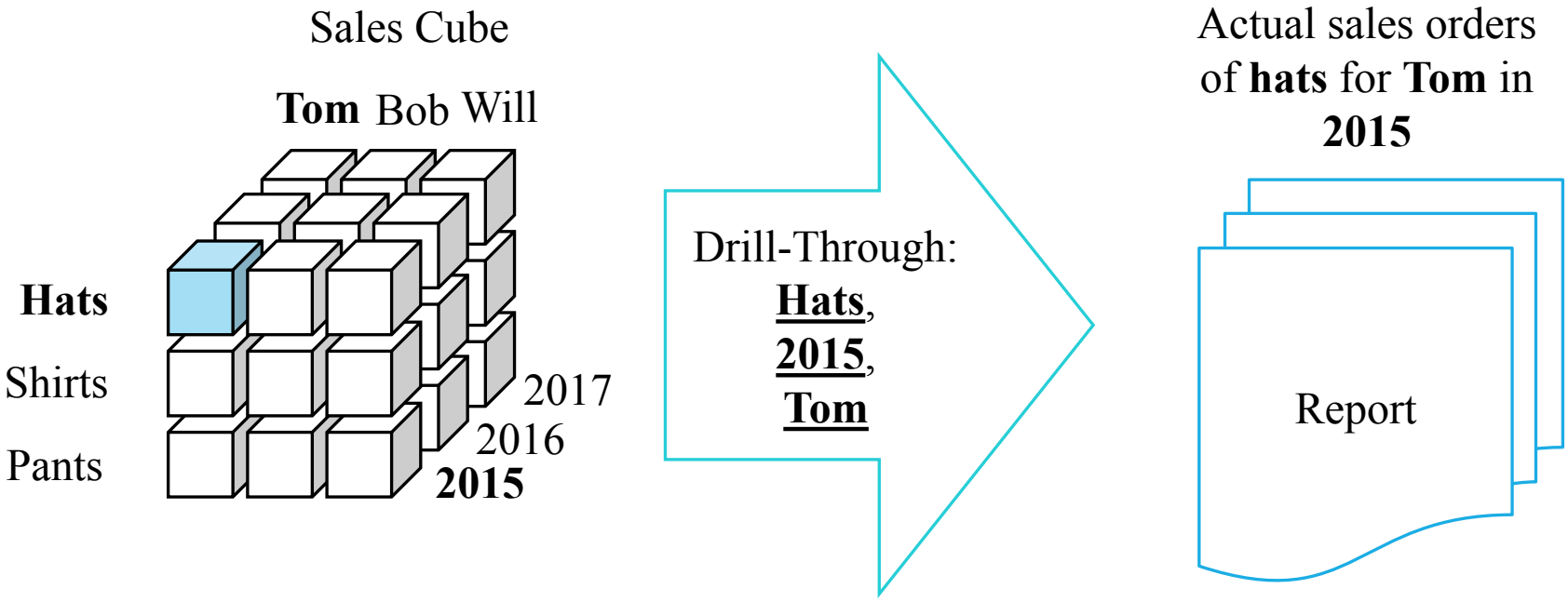
OLAP Terminology: Drill-Across

Drill-across: Use a sliced or diced dimension to view facts in a different cube.



OLAP Terminology: Drill-Through

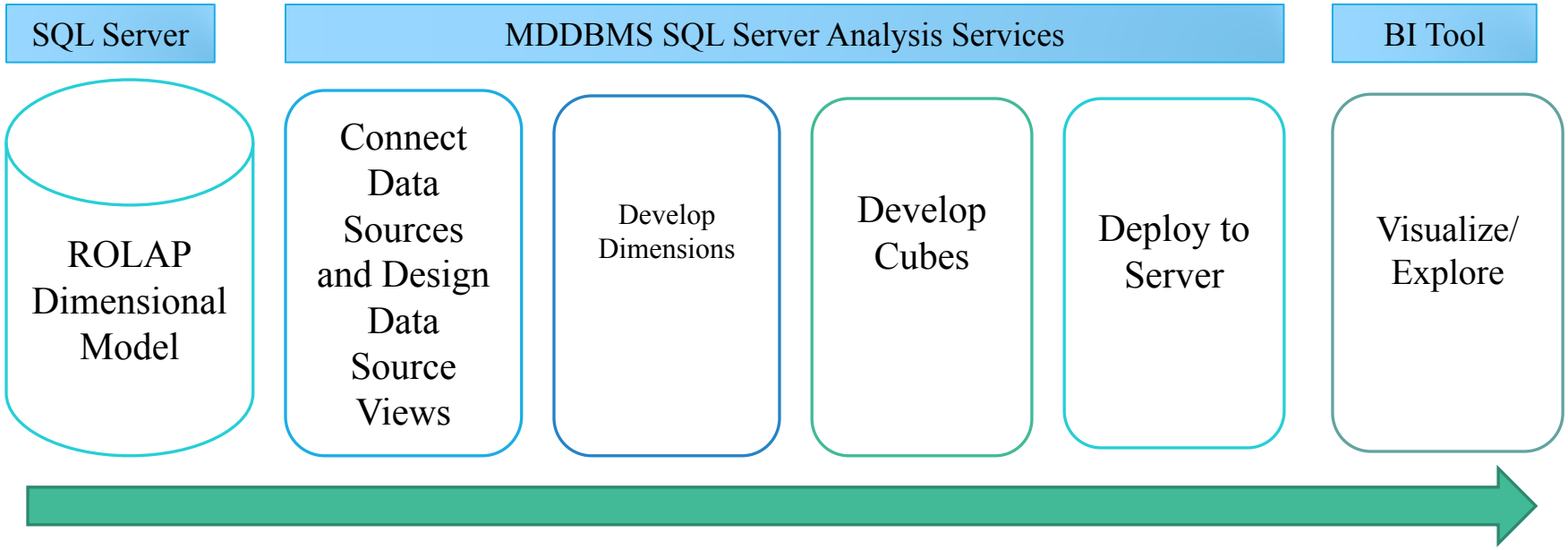
Drill-through: Retrieve the OLTP detail of the actual event in the cube.



An Overview of the Process

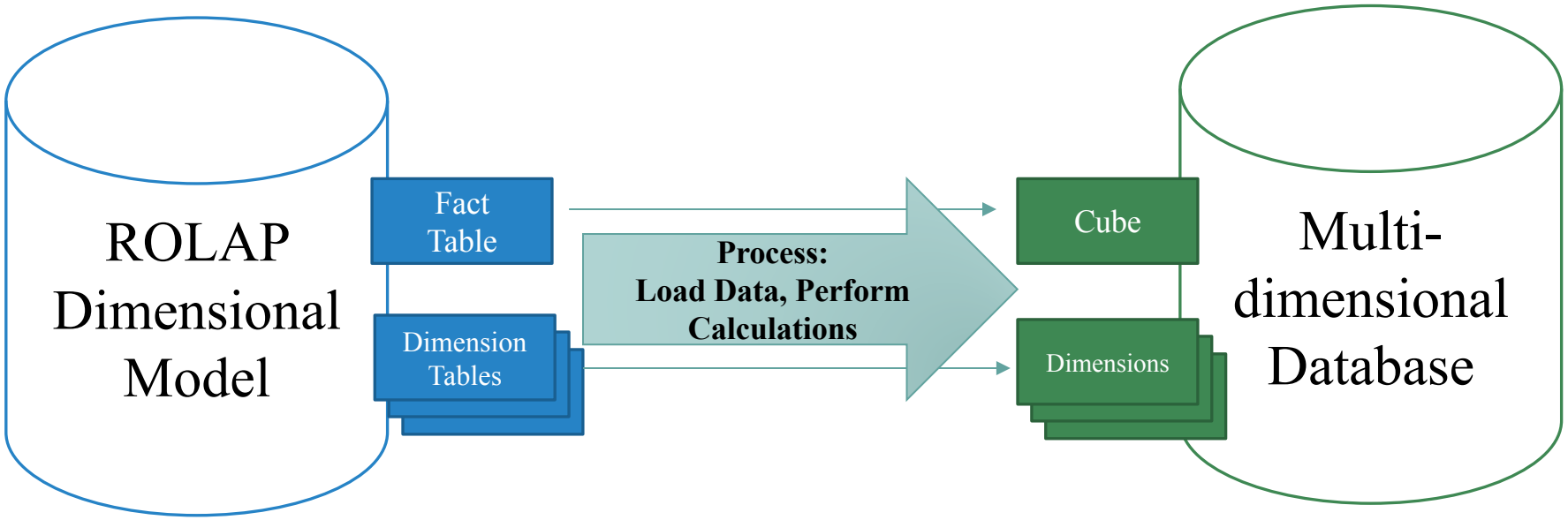
School of Information Studies
Syracuse University

The MOLAP Development Process



MOLAP: Processing

Before you can view changes in a MDDBMS, you must first **process** the data.

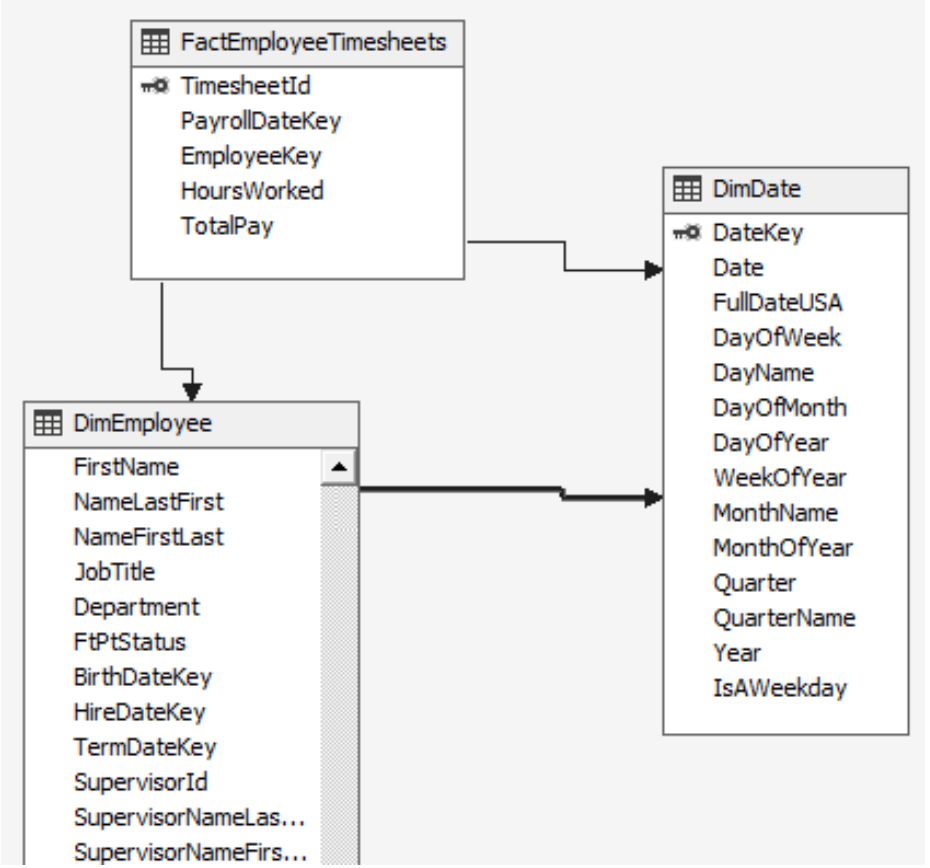


Connecting Data Sources and Creating Views in SSAS

School of Information Studies
Syracuse University

Connecting Data Sources and Views

- Connect to ROLAP sources.
- Create a data source view to combine tables from disparate sources.
- Derive calculated columns for values that do not exist.
- Establish relationships that are not in the ROLAP schema.



Building Dimensions

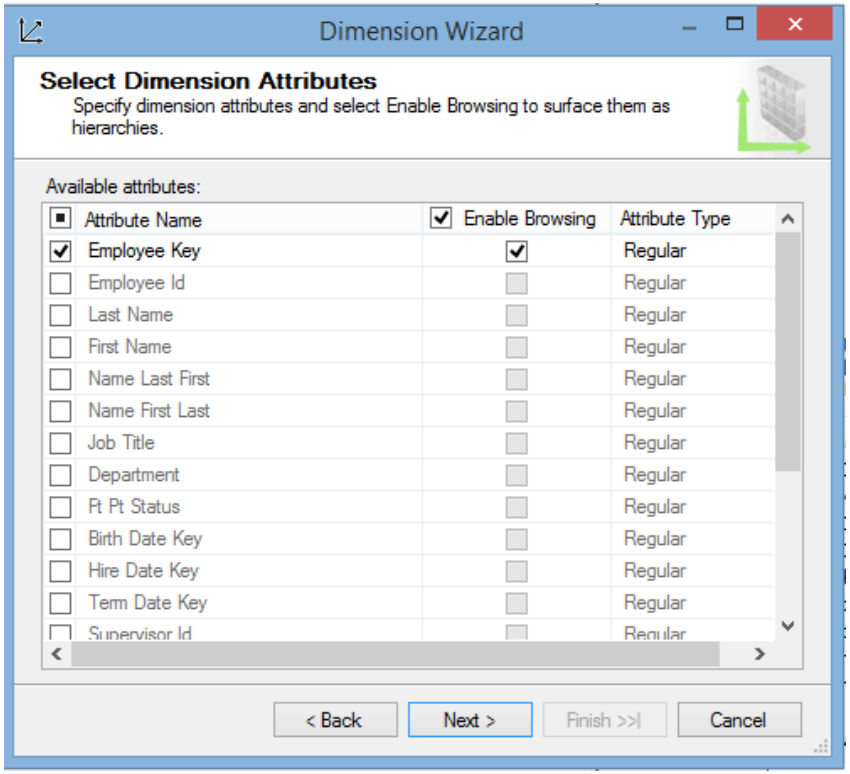
School of Information Studies
Syracuse University

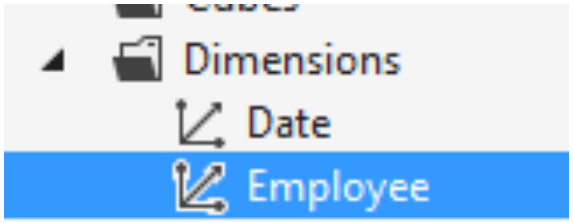
Developing Dimensions: The Steps

1. Create dimension from source.
2. Add and configure attributes.
3. Configure hierarchies.

1. Creating Dimensions From Source

- Choose the ROLAP dimension from the data source view.
- Start with only the key attribute.
- Add the other attributes manually.
- Best practice: Drop “Dim” from the name.





2. Configuring Attributes

- We configure how we want the attributes presented to the user.
- Important properties
  - Name
  - Key column
  - Name column
  - Value column
  - Order by (key or name?)

Example: Month

- Key: month of year
- Value: month name
- Order by: key

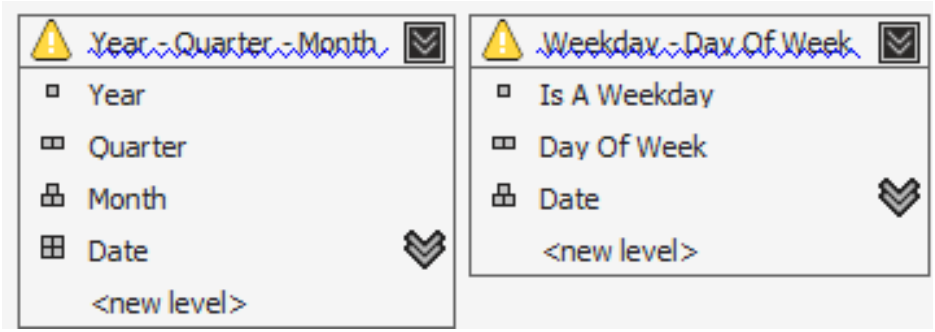
3. Configure Hierarchies

Allow us to drill-down through attributes.

Three types:

1. Natural: 1-M relationship among hierarchies
   - E.g., Year → Month → Day
2. Unnatural: no dependent relationship
   - E.g., Color → Size
3. Parent-child
   - E.g., Employee → Supervisor
   - Automatic when self-join exists in dimension.

Examples:



The Cube Development Process

School of Information Studies
Syracuse University

Develop Cubes: The Steps

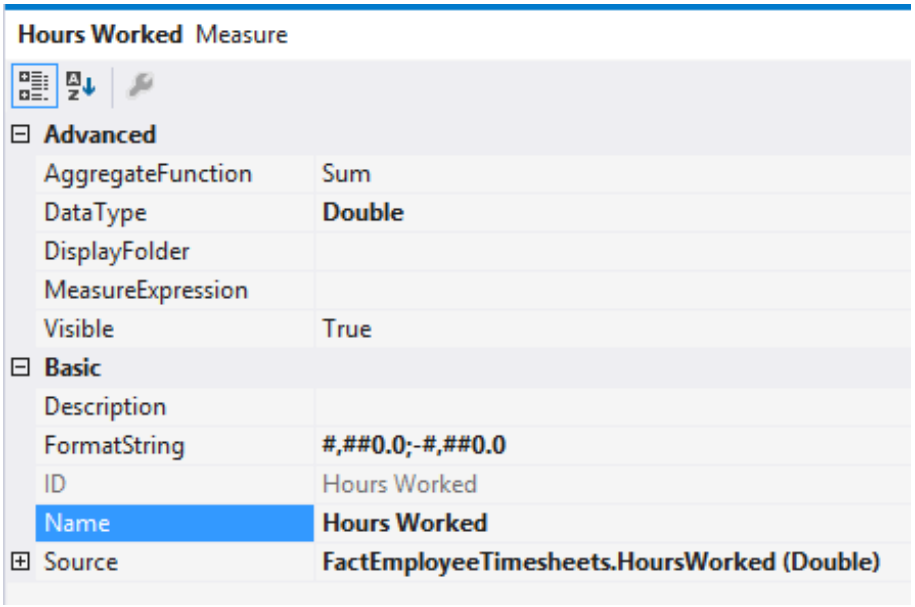
1. Select fact table: adds measures and determines dimensions.
2. Add and configure each measure (fact) properties.
3. Add calculations: business rules to the cube.
4. Add KPIs: key performance indicators.
5. Configure partitions: summary tables.
6. Configure perspectives: cube views.

1. Select Fact Table

- To build a cube you must include at least one fact table. This initiates the cube-building process.
- You can have more than one fact table in the cube.
- This allows you to consolidate logic into a single point of access for the user.
- Each fact grain will contain the same measure group.

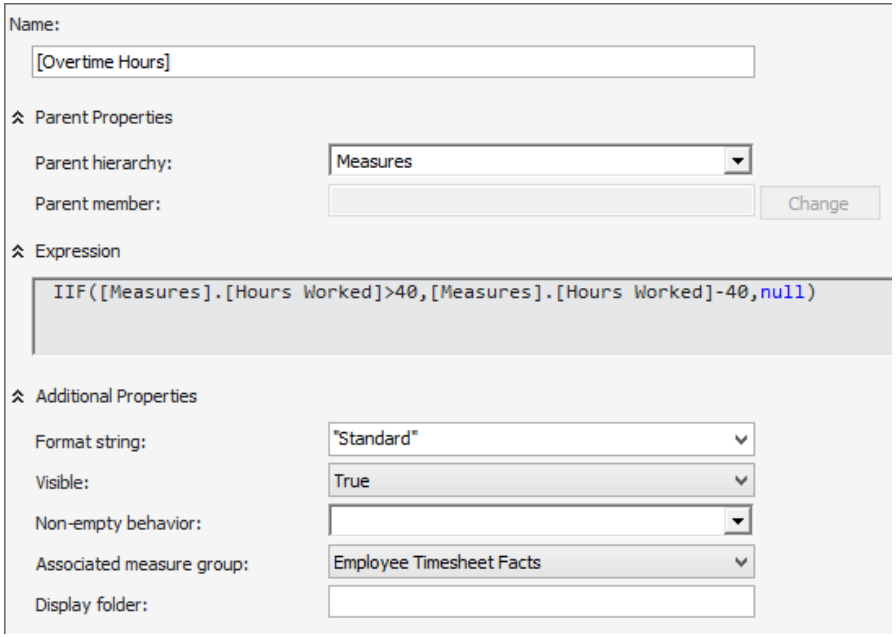
2. Configuring Measure Properties

- We configure how we want the measures presented to the user.
- Important properties:
  - Name
  - Format string
  - Aggregate function: sum, count, min/max, none
  - Visibility: show/hide
  - Measure group



3. Add Calculations

- Additional calculations based on current measures.
- Function builder to help you out.
- These are MDX expressions. MDX is the SQL of multidimensional databases.



4. KPIs—Key Performance Indicators

KPIs are metrics defined by the organization to measure performance.

You need:

1. Value: What you're measuring.
2. Goal: What you want to achieve.
3. Status: How close is the value to the goal? Range between −1 and 1.
4. Trend: Status over time periods. Range between −1 and 1.

Example:

1. Value: overtime hours
2. Goal: 0
3. Status:
   - When overtime = 0 then 1
   - When overtime >=10 then −1
   - Else 0
4. Trend:
   - When overtime < last time period overtime, then 1, etc.

5. Partitions

- Filter cube data rows based on an expression; to make smaller cubes
- Improves cube performance
- Example:
  - Partition by date so current FY is in a single partition
  - Partition by store so each branch has its data

6. Perspectives

- Reduce complexity of cubes by eliminating irrelevant facts and dimension attributes
- Similar to SQL views in relational databases
- Example:
  - Single physical cube with sales and projections can be simplified into two perspectives sales and projects

Visualizing Your Cube

School of Information Studies
Syracuse University

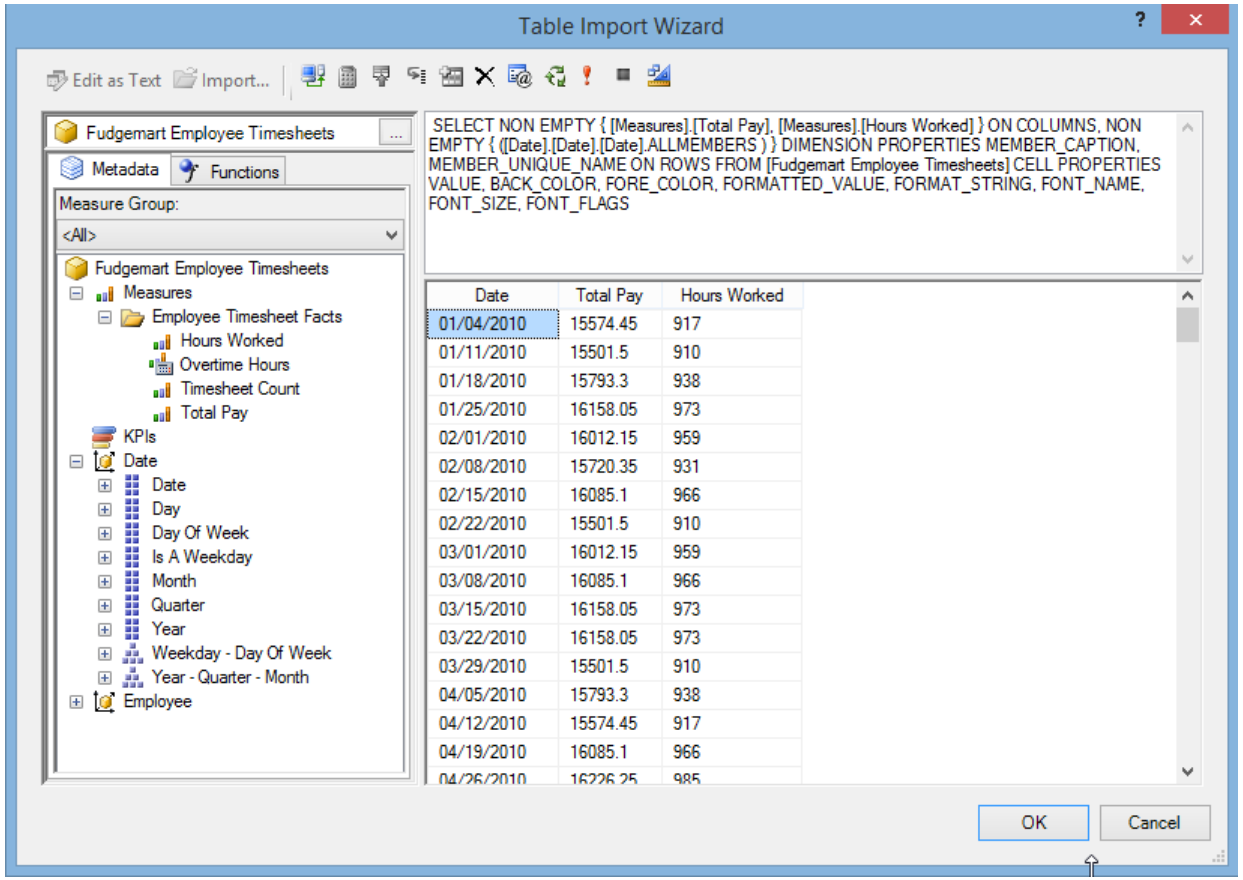
Visualizing Cubes

- SSAS Cube Browser
  - During cube development
- Excel Pivot Table
  - For power users and cube testing
- Power BI or other BI tool
  - For business users and BI testing

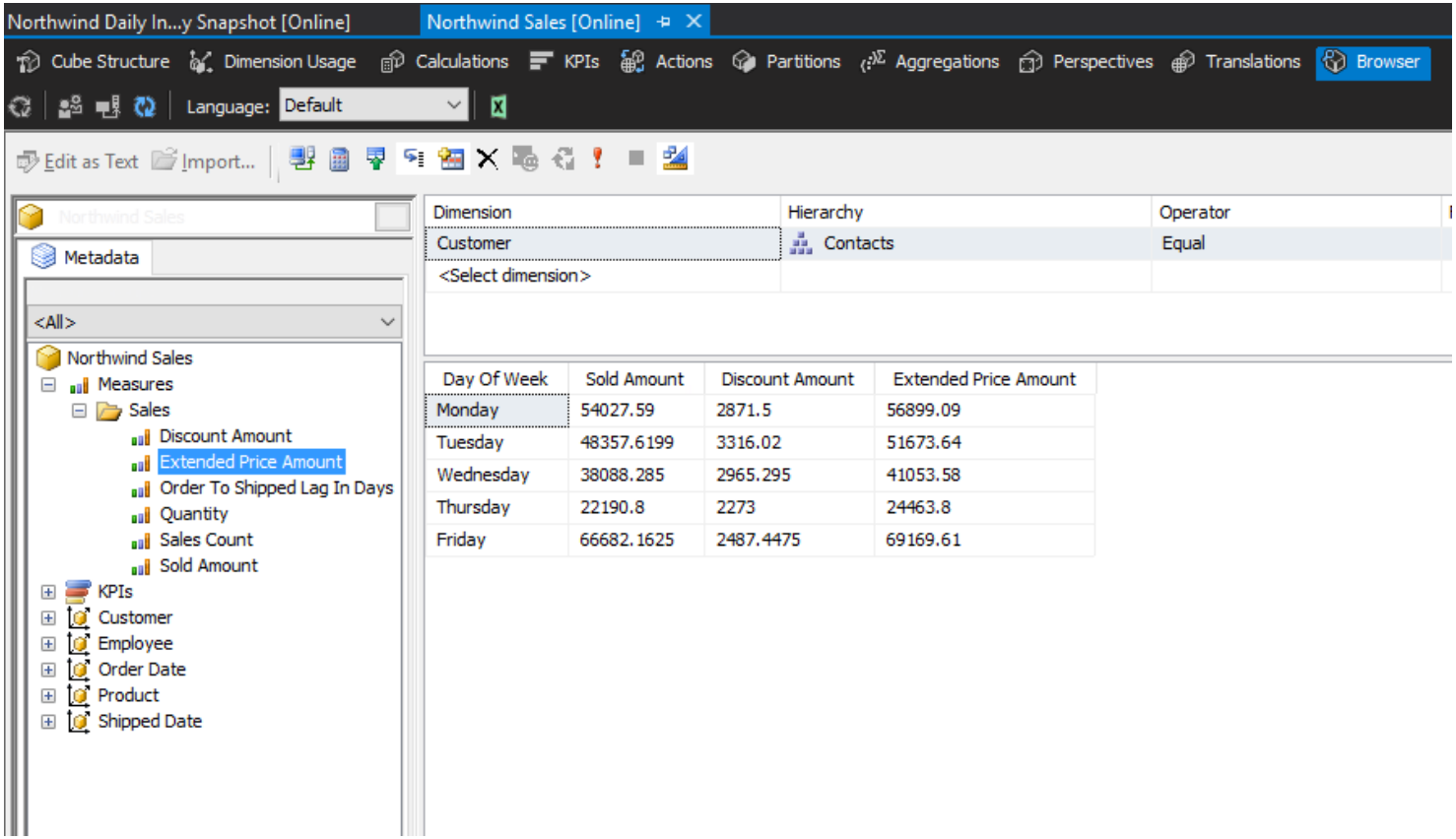
MDX Query

MDX is the SQL for MOLAP databases.

Allows you to query your cube and retrieve data in tabular form.



Example: Cube Browser

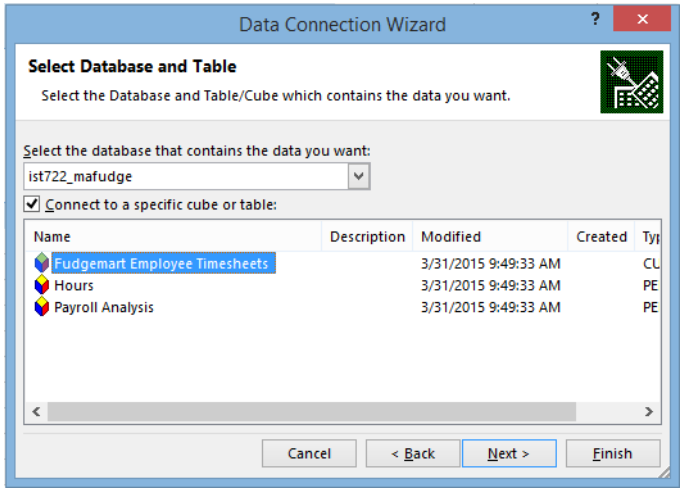


Pivot Tables

Use SSAS as a data source.

BI semantic model:

- Understands measures, data types, and hierarchies



Total Pay	Column Labels				
Row Labels	2010	2011	2012	2013	Grand Total
Clothing	$86,783.60	$94,996.43	$100,468.44	$99,744.19	$381,992.66
Customer Service	$302,224.00	$305,382.90	$320,597.21	$315,682.45	$1,243,886.56
Electronics	$127,192.00	$134,451.83	$148,067.41	$146,442.40	$556,153.64
Hardware	$95,295.05	$114,229.65	$119,920.56	$118,082.34	$447,527.59
Housewares	$113,744.00	$115,833.11	$122,052.11	$120,675.38	$472,304.60
Sporting Goods	$111,047.50	$109,174.16	$132,897.30	$132,777.07	$485,896.03
Grand Total	**$836,286.15**	**$874,068.06**	**$944,003.02**	**$933,403.84**	**$3,587,761.08**