



Online Analytical Processing (OLAP)

IT 4713 BI

Jack G. Zheng

Spring 2020

<http://zheng.kennesaw.edu/teaching/it4713/>

Overview



- What is OLAP?
- How dimensional model and query are implemented differently in (dimensional) data warehouse and OLAP?
- Basic OLAP terms and operations
- OLAP types
- OLAP systems

OLTP (Online Transaction Processing)



- Data recorded in business systems and transactional databases
- Optimized for quick data entry and relational integrity
- Typically very complex schema design with many normalized tables that facilitate high volume throughput of transactions
- Not optimized for reporting and data analysis

OLAP



- OLAP is a category of technology or software that is optimized to answer queries that are multi-dimensional in nature.
 - Compared to transactional queries, multidimensional queries involves a lot of aggregates based on grouping of rows
- OLAP provides a multidimensional data model, storage, and query methods optimized for multidimensional analysis.
- Characteristics of dynamic multi-dimensional queries:
 - calculations and modeling applied across dimensions, through hierarchies and/or across members
 - trend analysis over sequential time periods
 - slicing subsets for on-screen viewing
 - drill-down to deeper levels of consolidation
 - reach-through to underlying detail data
 - rotation to new dimensional comparisons in the viewing area

Multidimensional Analysis



	SalesOrderID	SalesOrderDetailID	CarrierTrackingNumber	OrderQty	ProductID	SpecialOfferID	UnitPrice	UnitPriceDiscount	LineTotal
1	43659	1	4911-403C-98	1	776	1	2024.994	0.00	2024.994000
2	43659	2	4911-403C-98	3	777	1	2024.994	0.00	6074.982000
3	43659	3	4911-403C-98	1	778	1	2024.994	0.00	2024.994000
4	43659	4	4911-403C-98	1	771	1	2039.994	0.00	2039.994000
5	43659	5	4911-403C-98	1	772	1	2039.994	0.00	2039.994000
6	43659	6	4911-403C-98	2	773	1	2039.994	0.00	4079.988000
7	43659	7	4911-403C-98	1	774	1	2039.994	0.00	2039.994000
8	43659	8	4911-403C-98	3	714	1	28.8404	0.00	86.521200
9	43659	9	4911-403C-98	1	716	1	28.8404	0.00	28.840400
10	43659	10	4911-403C-98	6	709	1	5.70	0.00	34.200000
11	43659	11	4911-403C-98	2	712	1	5.1865	0.00	10.373000
12			03C-98	4	711	1	20.1865	0.00	80.746000
13			057-83	1	762	1	419.4589	0.00	419.458900

This is the transactional data report with line by line data.

A pivot table or crosstab is usually used for OLAP result view (aggregated data by dimensions).

Dimension

Dimension

3	Store Sales Net	Store Type					
4	Product Family	Deluxe Supermarket	Gourmet Supermarket	Mid-Size Grocery	Small Grocery	Supermarket	Grand Total
5	Drink	\$8,119.05	\$2,392.83	\$1,409.50	\$685.89	\$16,751.71	\$29,358.98
6	Food	\$70,276.11	\$20,026.18	\$10,392.19	\$6,109.72	\$138,960.67	\$245,764.87
7	Non-Consumable	\$18,884.24	\$5,064.79	\$2,813.73	\$1,534.90	\$36,189.40	\$64,487.05
8	Grand Total	\$97,279.40	\$27,483.80	\$14,615.42	\$8,330.51	\$191,901.77	\$339,610.90

Aggregated data by dimension members

Why OLAP?



- Both (dimensional) data warehouse/mart and OLAP use the (Kimball's) multidimensional data model to store data and support multidimensional queries.
 - What are the differences? Why OLAP is needed?
- Two important aspects
 - Dimensional data model definition and implementation
 - Dimensional data query methods and performance
- In short
 - The multidimensionality of data warehouse is implemented through the relational model, and queries uses SQL and are usually arithmetic operations (sum, average, etc.) on records grouped by multiple dimensions (attributes).
 - OLAP is a more specific implementation of the dimensional model to enhance multidimensional query writing and execution performance

Data Model Implementation



- Relational databases only provide generic implementations through tables and keys for the dimensional model.
 - For example, a data warehouse or data mart based on the dimensional model
- The OLAP cube defines many dimensional model elements specifically beyond tables and keys.
 - OLAP system is independent from a data warehouse/mart
- Examples:
 - Specific definition of measures and dimensions (and different types of them)
 - Specific hierarchy definition
 - Grouping of measures and dimensions
- More readings
 - <http://www.informationweek.com/software/information-management/dimensional-relational-vs-olap-the-final-deployment-conundrum/d/d-id/1054577>

Query



- Directly querying the multidimensional data warehouse is based on SQL
 - Query (structural) complexity
 - Lack of native support for OLAP operations like drill down
 - Low performance

```
SELECT SUM(dbo.SalesFact.SalesAmount) AS [Total Sales], DimDate.TimeYear,
DimDate.TimeQuarter, DimDate.TimeMonth, DimProduct.Category,
DimProduct.Brand, DimLocation.Region, DimLocation.State
FROM SalesFact INNER JOIN DimProduct ON SalesFact.ProductKey = DimProduct.ID
INNER JOIN DimLocation ON SalesFact.LocationKey = DimLocation.ID
INNER JOIN DimDate ON SalesFact.TimeKey = DimDate.ID
GROUP BY DimDate.TimeMonth, DimDate.TimeYear, DimDate.TimeQuarter,
DimProduct.Brand, DimProduct.Category, DimLocation.Region, DimLocation.State;
```

- An OLAP system provides its own query methods
 - Native query support to operations like drill up/down
 - Aggregated measurements are pre-calculated thus improves performance

Basic OLAP Concepts and Operations



- Cube
- Slice-and-dice
- Drill-up and drill-down
- Drill-through and drill-across
- Pivot/Rotate
- Aggregate (Roll-up or Consolidate)
- Server and client
- Implementation type and architecture

OLAP Cube



- OLAP cube is a presentation of the chosen *measure* with associated dimensions.
 - Measure is the data item (fact) of interest: sales, cost, etc.
 - Dimension is the characteristic of a measure: time, location, etc.
 - Cube is a representation of certain view points.
- Cell
 - A single data point that occurs at the intersection defined by selecting one member from each dimension in a multi-dimensional structure.
- Data cubes aren't restricted to just three dimensions. Most OLAP systems can build data cubes with many more dimensions allows up to 64 dimensions.
 - In practice, we often construct data cubes with many dimensions, but we tend to look at just three at a time.

Cube Visualizations



1D relational table view

Region	Category	Sales Amount
Southeast	Electronics	183952.01
Southeast	Electronics	147029.95
Southeast	Computer	169694.01
Southeast	Computer	170147.58
West	Electronics	172423.83
West	Electronics	207814.28
West	Computer	177584.21
West	Computer	181099.18

2D Pivot Table view

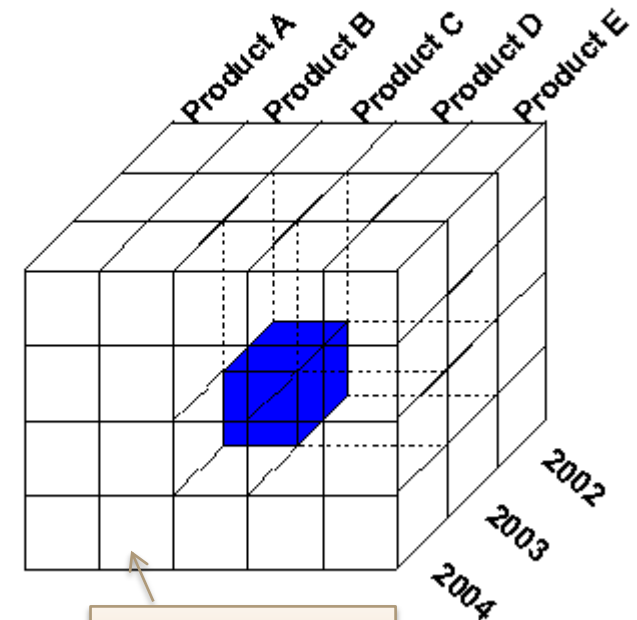
3	Store Sales Net	Store Type					
4	Product Family	Deluxe Supermarket	Gourmet Supermarket	Mid-Size Grocery	Small Grocery	Supermarket	Grand Total
5	Drink	\$8,119.05	\$2,392.83	\$1,409.50	\$685.89	\$16,751.71	\$29,358.98
6	Food	\$70,276.11	\$20,026.18	\$10,392.19	\$6,109.72	\$138,960.67	\$245,764.87
7	Non-Consumable	\$18,884.24	\$5,064.79	\$2,813.73	\$1,534.90	\$36,189.40	\$64,487.05
8	Grand Total	\$97,279.40	\$27,483.80	\$14,615.42	\$8,330.51	\$191,901.77	\$339,610.90

Chicago

Cincinnati

Dallas

Louisville



3D cube view

Slice and Dice

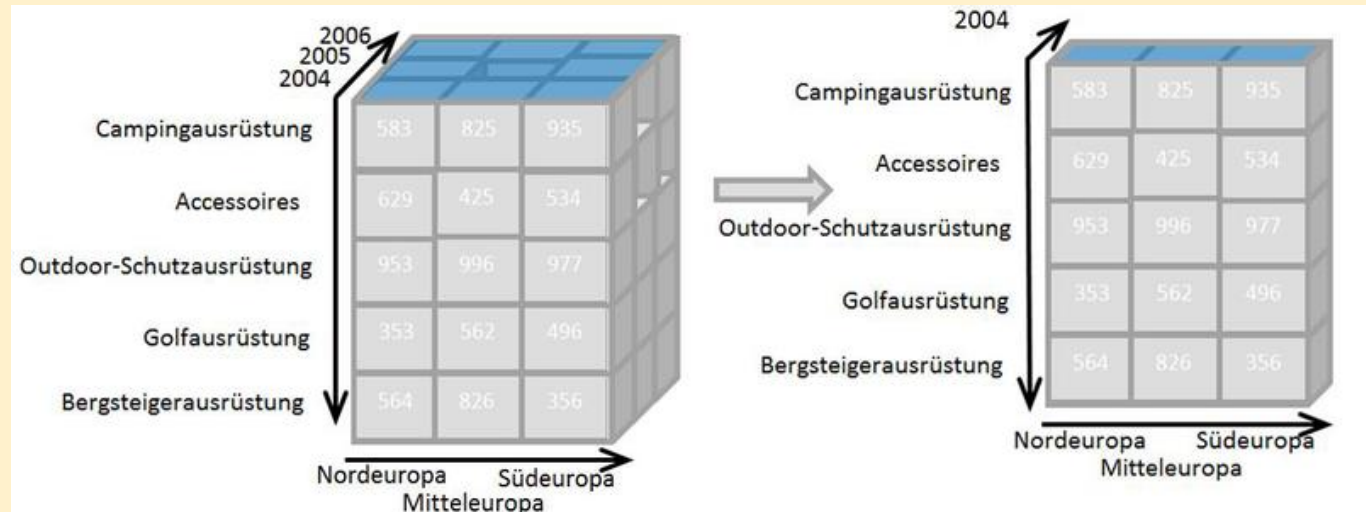


- Slice and Dice
 - Is a term for the user-initiated process of data browsing and analysis by interactive OLAP operations (slice, dice, pivot, and drill down/up).
- Slice
 - Slice is the act of picking subset of a cube by choosing a single value for one of its dimensions, creating a new cube with one fewer dimension
 - Like a filter on one dimension
- Dice
 - The dice operation produces a sub-cube by allowing the analyst to pick specific values of multiple dimensions
 - Like multiple filters on more dimensions

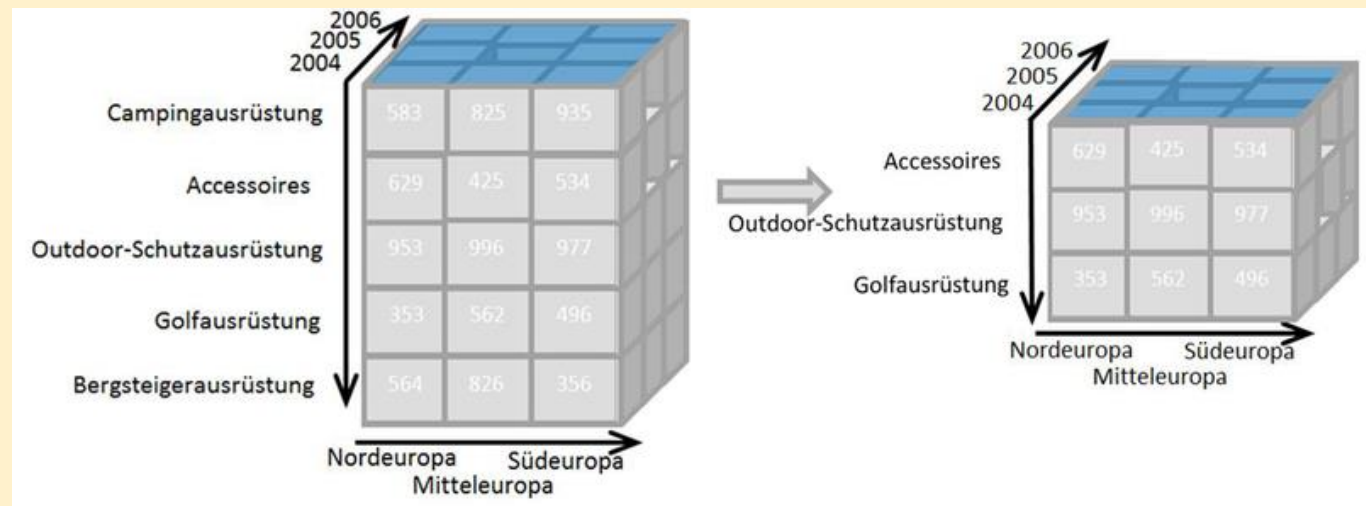
Slice and Dice



- Slice



- Dice

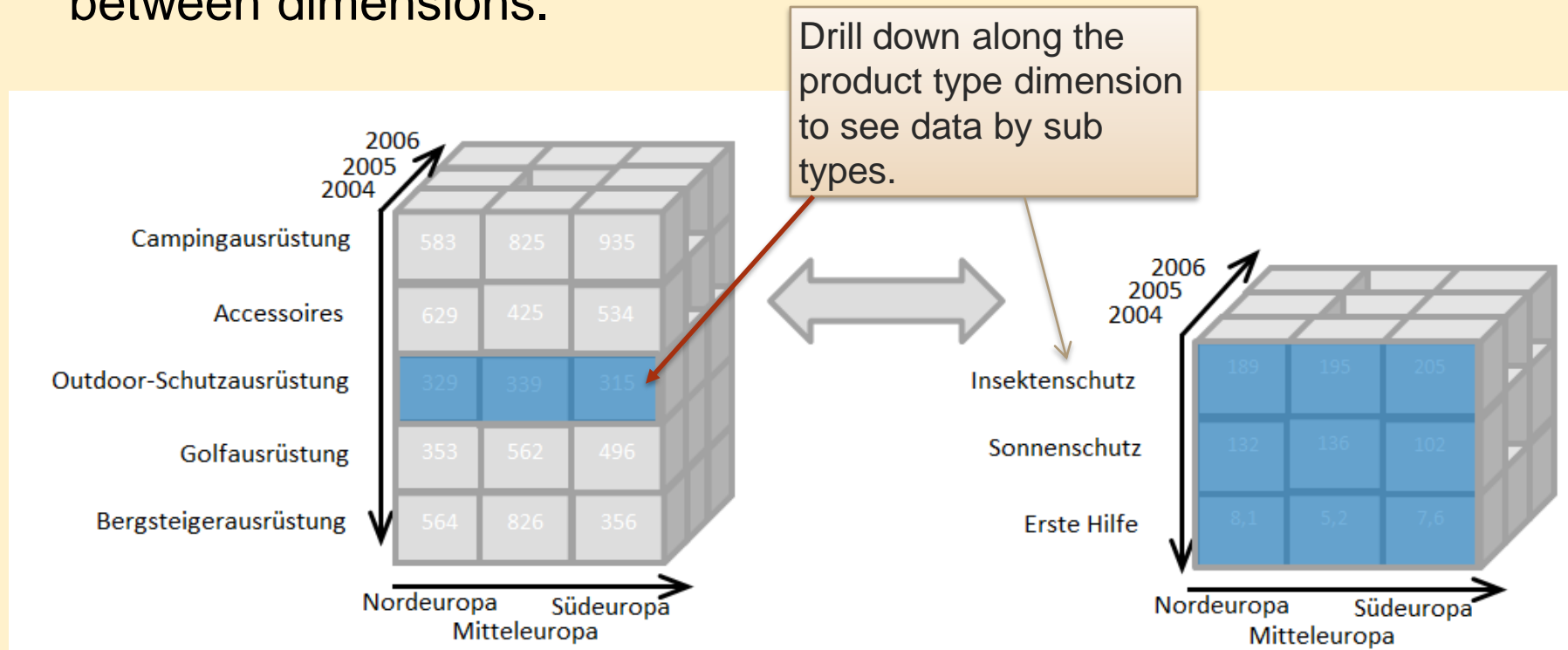


https://en.wikipedia.org/wiki/OLAP_cube

Drill-Up/Down



- Drilling down or up is a specific analytical technique whereby the user navigates among levels of data ranging from the most summarized (up) to the most detailed (down) level.
- The drilling paths may be defined by the hierarchies within dimensions or other relationships that may be dynamic within or between dimensions.



Drill Up/Down in a Pivot Table



Drill up				C	D		H	I	J
1									
2									
3	Store Sales Net				Store				
4	Store Country	Store Sta	Store City	Product Family	Delux		Small Grocery	Supermarket	Grand Total
5	USA	CA	Beverly Hills	Drink		\$2,392.83			\$2,392.83
6				Food		\$20,026.18			\$20,026.18
7				Non-Consumable		\$5,064.79			\$5,064.79
8			Beverly Hills Total			\$27,483.80			\$27,483.80
9			Los Angeles	Drink				\$2,870.33	\$2,870.33
10				Food				\$23,598.28	\$23,598.28
11				Non-Consumable				\$6,305.14	\$6,305.14
12			Los Angeles Total					\$32,773.74	\$32,773.74
13			San Diego	Drink				\$3,050.43	\$3,050.43
14				Food				\$23,627.83	\$23,627.83
15				Non-Consumable				\$6,039.34	\$6,039.34
16			San Diego Total					\$32,717.61	\$32,717.61
17			San Francisco	Drink			\$227.38		\$227.38
18				Food			\$1,960.53		\$1,960.53
19				Non-Consumable			\$474.35		\$474.35
20			San Francisco Total				\$2,662.26		\$2,662.26
21		CA Total				\$27,483.80	\$2,662.26	\$65,491.35	\$95,637.41
22		OR		Drink		\$4,438.49		\$2,862.45	\$7,300.94
23				Food		\$37,778.35		\$23,818.87	\$61,597.22
24				Non-Consumable		\$10,177.89		\$6,428.53	\$16,606.41
25		OR Total				\$52,394.72		\$33,109.85	\$85,504.57
26		WA		Drink		\$3,680.56	\$1,409.50	\$458.51	\$7,968.50
27				Food		\$32,497.76	\$10,392.19	\$4,149.19	\$67,915.69
28				Non-Consumable		\$8,706.36	\$2,813.73	\$1,060.54	\$17,416.38
29		WA Total				\$44,884.68	\$14,615.42	\$5,668.24	\$93,300.57
30	USA Total					\$97,279.40	\$27,483.80	\$14,615.42	\$8,330.51
31	Grand Total					\$97,279.40	\$27,483.80	\$14,615.42	\$8,330.51

Drill down along the hierarchy of the same dimension (region: country, state, city)

Drill down between dimensions (region and product)

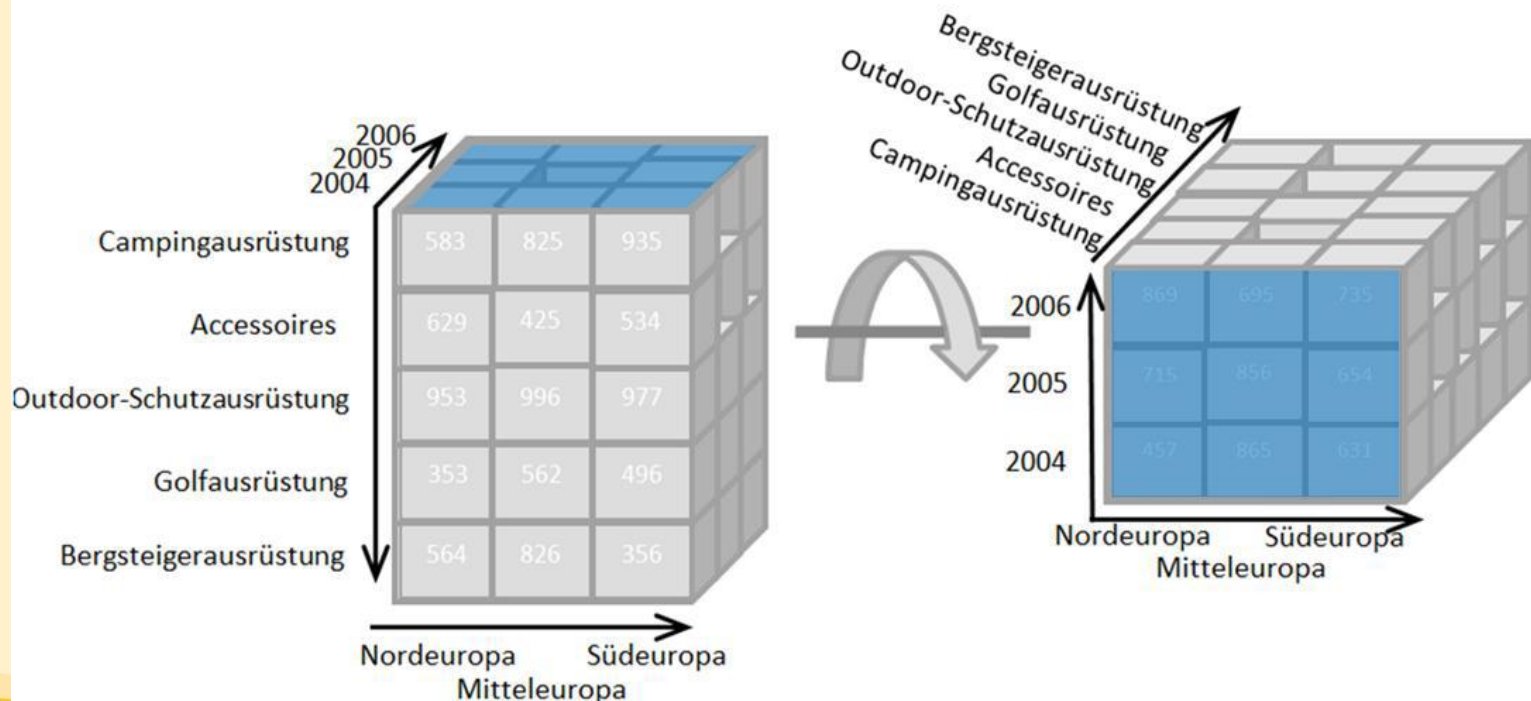
Drill up

Drill down

Pivot/Rotate



- To change the dimensional orientation of a report or page display.
- Examples
 - swapping the rows and columns
 - moving one of the row dimensions into the column dimension
 - swapping an off-spreadsheet dimension with one of the dimensions in the page display (either to become one of the new rows or columns), etc.



Other Operations



- Drill-across
 - Accesses more than one fact table that is linked by common dimensions.
 - Combines cubes that share one or more dimensions.
- Drill-through
 - Drill down to the bottom level of a data cube down to its back-end relational tables.

Aggregate



- Involves computing all of child level data for one or more dimensions, for example, adding up all Departments to get Total Division data.
- While such relationships are normally summations, any type of computational relationship or formula might be defined.
- Aggregate function/formula
 - Sum, count, average, max, min, etc.

OLAP Server and Client



- Traditional OLAP systems are of a client/server architecture.
- Server
 - An OLAP server (e.g. SSAS Multidimensional Model) is a high-capacity, multi-user data management engine specifically designed to support and operate on multi-dimensional data structures. It defines and prepares OLAP cubes, with measures, dimensions, and other data model elements.
 - The OLAP Server may either physically stage the processed (pre-calculated) multi-dimensional information, or it may populate its data structures in real-time from relational or other databases, or offer a choice of both.
- Client
 - Client programs send queries to the server and present results to users in various format (pivot tables, charts, etc.)
 - They also support user interactions, such as data exploration, sorting, custom calculation, user defined data, formatting, etc.
 - OLAP clients may be as simple as a spreadsheet program or as high-functioned as a financial modeling or sales analysis application.
 - OLAP client query is covered in milestone #4 (module 9 and 10).

OLAP Server Types



- MOLAP (Multidimensional OLAP)
 - The 'classic' and popular form of OLAP
 - Stores both cube structure and data (leaf level data and preprocessed aggregates) in an optimized multi-dimensional storage, rather than in a relational database.
 - Requires the pre-computation and storage of information in the cube - the operation known as processing.
 - Faster in OLAP operations but takes extra time to copy measures – latency
 - It is difficult to include a large amount of data in the cube itself
- ROLAP (Relational OLAP)
 - Stores cube structure in multidimensional databases
 - Stores data (leaf level data and preprocessed aggregates) in relational databases
 - In essence, each action of slicing and dicing is equivalent to adding a "WHERE" clause in the SQL statement.
 - Can handle large amounts of data and leverage functionalities inherent in the relational database.
 - Slow query performance and limited by SQL functionalities (like complex calculation)
- HOLAP (Hybrid OLAP)
 - Stores cube structure and high level aggregate data in multidimensional databases
 - Stores more detailed-level data in relational databases
 - Combines the advantages of MOLAP and ROLAP
- Types of OLAP Systems: <http://olap.com/types-of-olap-systems/>

OLAP Server Type Diagram

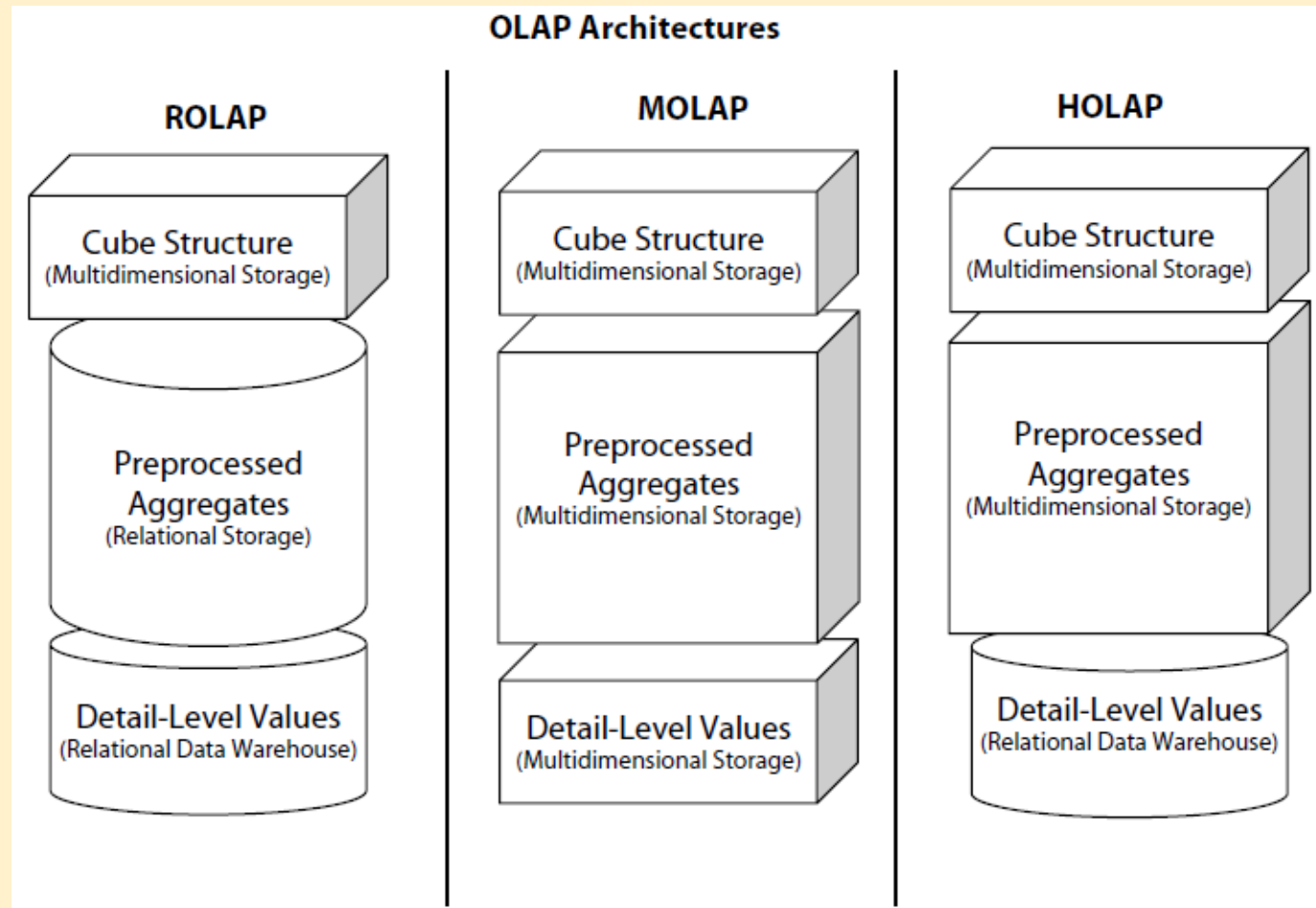


Figure from book: Brian Larson, Delivering Business Intelligence with Microsoft SQL Server

OLAP Storage Setting in SSAS Cube Design



Cube Struct... Dimension Usage Calculations KPIs Actions **Partitions** Aggregations Perspectives Translations Browser

Fact Internet Sales (1 Partition)

Partition Name	Source	Storage Mode	Aggregation Design
1 Fact Internet Sales	FactInternetSales	MOLAP	

[New Partition...](#)

1. Partitions

[Storage Settings...](#)

2. Storage settings ...

3. Use the standard setting or custom setting

Partition Storage Settings - Fact Internet Sales

☒ Standard setting

Real-time HOLAP Medium-latency MOLAP Scheduled MOLAP

Real-time ROLAP Low-latency MOLAP Automatic MOLAP **MOLAP**

- Measure group data and aggregations are stored in a multidimensional format.
- Notifications are not received when data changes.
- Processing must be either scheduled or performed manually.

☐ Custom setting

To view or modify settings, click Options.

[Options...](#)

OK Cancel Help

OLAP Systems and Products



- Traditional OLAP is at the enterprise level
- Major products
 - Microsoft SSAS (SSAS is more than OLAP): Analysis Service is Microsoft's Online Analytical Processing (OLAP) Engine that resides on an Enterprise Infrastructure
 - IBM Cognos
<http://www.ibm.com/analytics/us/en/technology/cognos-software/>
 - https://en.wikipedia.org/wiki/Comparison_of_OLAP_Servers
- Desktop OLAP or in-memory OLAP is emerging as users demands the control and flexibility of data analysis (self-service BI, agile BI, etc.)
 - <http://www.bi-dw.info/in-memory-olap.htm>

SSAS OLAP and MSBI Stack

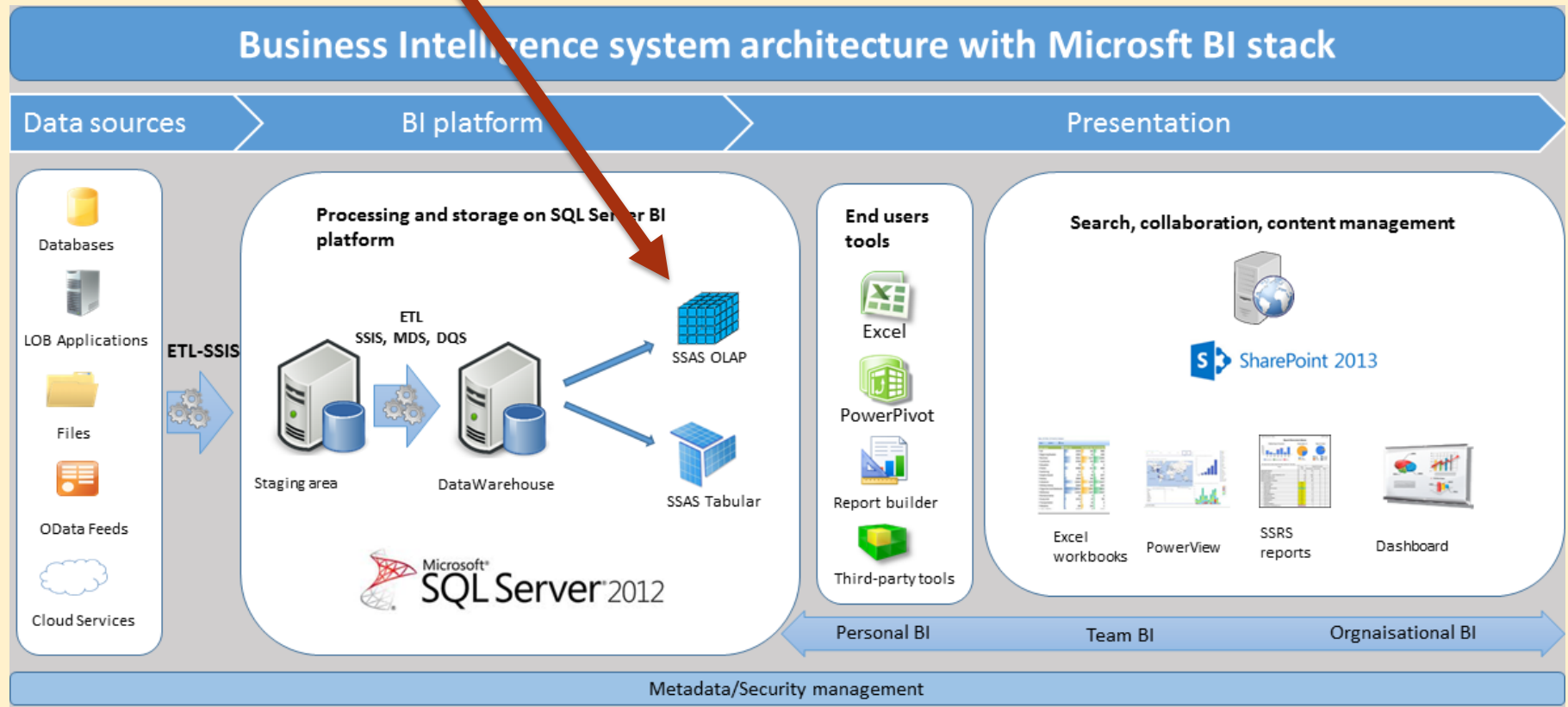


Image from <https://bipointblog.wordpress.com/2014/05/28/implementation-of-a-bi-system-using-microsoft-bi-stack-introduction/>

SSAS OLAP Client Access



- The following Microsoft applications support native connections to SSAS multidimensional data
 - SSRS
 - SSDT
 - SSMS
 - Excel
 - Power BI Desktop
 - <https://docs.microsoft.com/en-us/sql/analysis-services/multidimensional-models/mdx/multidimensional-model-data-access-analysis-services-multidimensional-data?view=sql-server-2017>
- OLAP clients and reporting will be covered in milestone #4
- Client libraries like ADOMD.Net can also query SSAS cubes programmatically

Challenges of OLAP



- One challenge with all OLAP systems was that a separate data storage layer was required.
- This meant that in addition to a corporate data warehouse - typically based on a relational database - IT departments had to manage a separate “OLAP database”.
- The only exception here are ROLAP (relational OLAP) systems, these though, were typically suboptimal from a performance perspective and mostly didn't support write back.
- This typically required complex and time-consuming extraction transformation and loading (ETL) processes to e.g. get the planning data in the data warehouse and vice versa the actuals in the planning system.

Good Resources



- Core learning
 - Video introduction
 - What is OLAP from Intricity? <http://www.youtube.com/watch?v=2ryG3Jy6eIY>
 - What is Business Intelligence and an OLAP Cube from ExcelCentral.com: <https://www.youtube.com/watch?v=yoE6bgJv08E>
 - Comparing Relational Databases to Multidimensional Databases: https://www.youtube.com/watch?v=q8p_2j5yTHI
 - OLAP tutorial from Tutorialspoint: http://www.tutorialspoint.com/dwh/dwh_olap.htm and also some good visuals from http://en.wikipedia.org/wiki/OLAP_cube
 - Tutorial: Database Queries, Data Mining, and OLAP, by Lutz Hamel, in The Encyclopedia of Data Warehousing and Mining, <http://homepage.cs.uri.edu/faculty/hamel/pubs/hamel-81-921.pdf>
 - OLAP: Past, Present ... and Future? <https://www.senturus.com/blog/olap-past-present-future/>
- More
 - The origins of today's OLAP products: <http://dssresources.com/papers/features/pendse10062002.html>
 - https://en.wikipedia.org/wiki/Comparison_of_OLAP_Servers
 - <http://olap.com>
- OLAP definitions
 - <http://www.olapcouncil.org/research/glossaryly.htm> or
 - <http://altaplana.com/olap/glossary.html>
- SSAS resources
 - <http://www.ssas-info.com>
 - <https://msdn.microsoft.com/en-us/library/hh231701>
 - <https://blog.crossjoin.co.uk/>