About this presentation

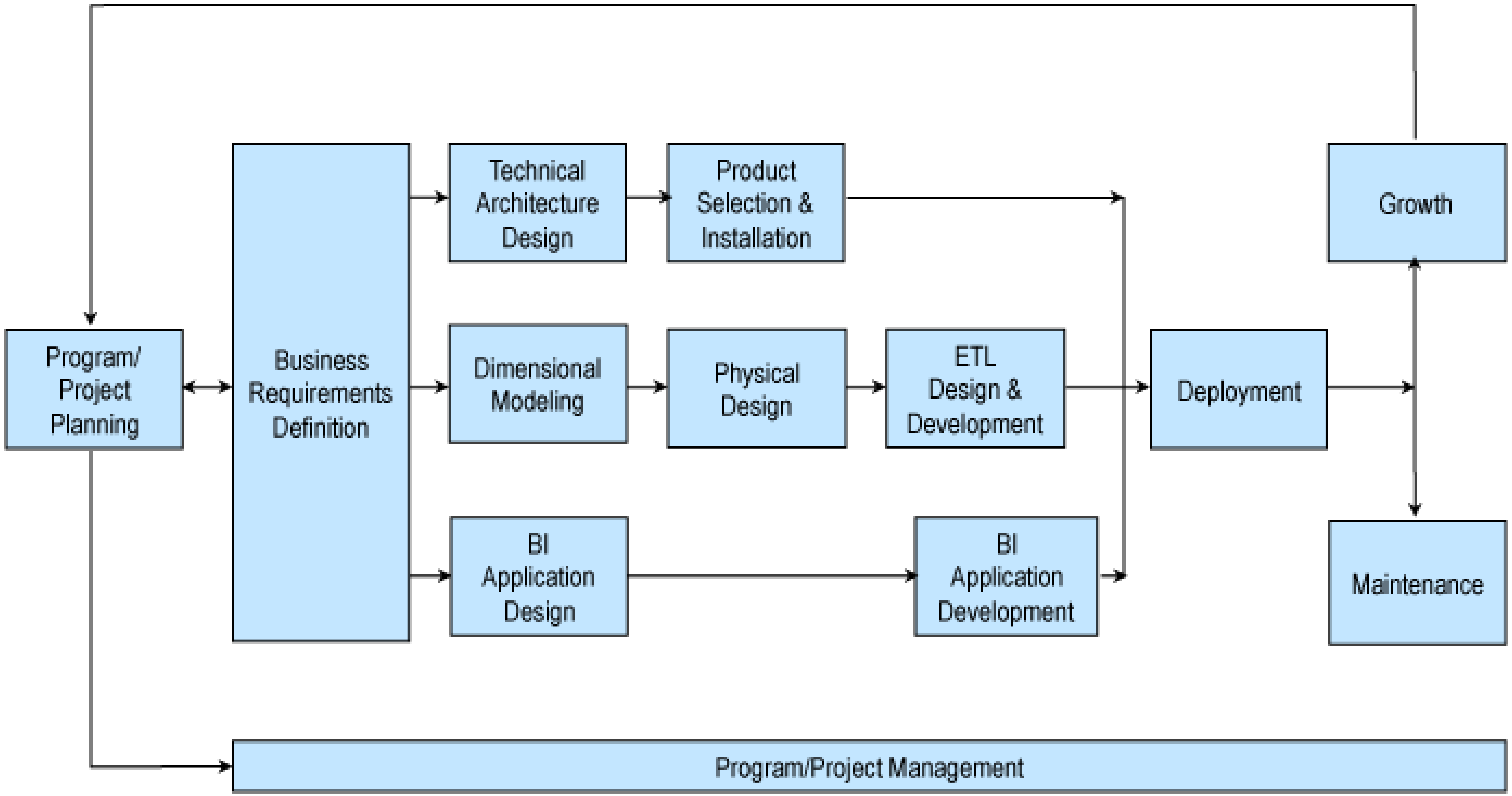
Now that we know how to draft our dimensional model, we can start to plan our low-level database design.

Agenda

- Fact tables core concepts
- Types of fact tables
- Types of facts
- Common fact tables

Fall 2024

The data warehouse lifecycle


Program/ Project Planning
Business Requirements Definition
Technical Architecture Design
Product Selection & Installation
Dimensional Modeling
Physical Design
ETL Design & Development
BI Application Design
BI Application Development
Deployment
Growth
Maintenance
Program/Project Management

Reference: https://www.kimballgroup.com

Fact tables

Now that we have determined our dimensions, we need to identify our fact table attributes.

What attributes does a fact table contain?

- A unique identifier (e.g., primary key)
- Facts (e.g., sales amounts)
- Foreign keys from dimension tables (who, what, when, where, why, how)

Fact tables

Primary key considerations:

- Can be a surrogate key (e.g., autoincrement)

Surrogate Key (pk)	Invoice(nk)	SalePrice	Product	Date	CustID	EmpID	Location
1	12345|NA	$12.50	13421	10421	23	12	1039

- Can be a business/natural key (e.g., invoice number)

Invoice(PK)	SalePrice	Product	Date	CustID	EmpID	Location
12345|NA	$12.50	13421	10421	23	12	1039

- Can be a “composite key” (or compound key) of the dimensional foreign keys

Invoice(PK)	SalePrice	Product	Date	CustID	EmpID	Location
12345|NA	$12.50	13421	10421	23	12	1039

Reference: Kimball, Ralph, and Margy Ross. *The Data Warehouse Toolkit : The Definitive Guide to Dimensional Modeling*, John Wiley & Sons, Incorporated, 2013.

Types of fact tables

There are three primary types of fact tables:

- Transactional
- Periodic Snapshot
- Accumulating Snapshot

Reference: Kimball, Ralph, and Margy Ross. *The Data Warehouse Toolkit : The Definitive Guide to Dimensional Modeling*, John Wiley & Sons, Incorporated, 2013.

Transactional fact tables

Transactional fact tables “correspond to a measurement event at a point in space and time (pg. 43).”

Transactional fact tables are typically comprised of facts gathered at an atomic level of granularity.

Surrogate Key (pk)	Invoice(nk)	SalePrice	Product	Date	CustID	EmpID	Location
1	12345|NA	$12.50	13421	10421	23	12	1039
2	12346|NA	$8.50	13420	10422	24	12	1039
3	12347|NA	$12.50	13421	10422	23	13	1040

Reference: Kimball, Ralph, and Margy Ross. *The Data Warehouse Toolkit : The Definitive Guide to Dimensional Modeling*, John Wiley & Sons, Incorporated, 2013.

Periodic snapshot fact tables

Periodic snapshot fact tables contain aggregated (or rolled up) facts summarized over a specific time period (day, month, quarter, year).

Transactional fact tables are typically comprised of facts gathered at an atomic level of granularity.

Surrogate Key (pk)	Daily Sales	Date	Location
1	$12.50	10421	1039
2	$20.00	10422	1039

Reference: Kimball, Ralph, and Margy Ross. *The Data Warehouse Toolkit : The Definitive Guide to Dimensional Modeling*, John Wiley & Sons, Incorporated, 2013.

Accumulating snapshot fact tables

Accumulating snapshot fact tables are helpful for tracking a business process that has defined milestones throughout its lifecycle.

Surrogate Key (pk)	OrderID	OrderDate	ShipDate	DaysToShip	DeliverDate	DaysToDeliver
1	12345	10/4/21	10/7/21	3	10/10/21	3
2	12346	10/4/21	10/14/21	10	NULL	NULL

Reference: Kimball, Ralph, and Margy Ross. *The Data Warehouse Toolkit : The Definitive Guide to Dimensional Modeling*, John Wiley & Sons, Incorporated, 2013, https://www.nuwavesolutions.com/accumulating-snapshot-fact-tables/

Types of facts

Remember, facts are the outputs (or measurements) of a **business process**.

Facts are generally **numeric**, and tend to fall into one of three classifications:

- Additive (also known as “fully-additive”)
- Semi-additive
- Non-additive

Reference: Kimball, Ralph, and Margy Ross. *The Data Warehouse Toolkit : The Definitive Guide to Dimensional Modeling*, John Wiley & Sons, Incorporated, 2013.

Additive facts

Additive facts can be, well, added together regardless of the dimension context.

Surrogate Key (pk)	Invoice(nk)	SalePrice	Product	Date	CustID	EmpID	Location
1	12345|NA	$12.50	13421	10421	23	12	1039
2	12346|NA	$8.50	13420	10422	24	12	1039
3	12347|NA	$12.50	13421	10422	23	13	1040

SalePrice is our fact. We can add this fact across:

- Product (e.g., Total sales of 13421 = $25.00)
- Date (e.g., Total sales on 10422 = $20.00)
- Customer (e.g., Total sales to 23 = $25.00)
- Employee (e.g., Total sales by 12 = $20.00)
- Location (e.g., Total sales at 1039 = $20.00)

Reference: Kimball, Ralph, and Margy Ross. *The Data Warehouse Toolkit : The Definitive Guide to Dimensional Modeling*, John Wiley & Sons, Incorporated, 2013.

Semi-Additive facts

Semi-additive facts can be added across some dimensions, but not all.

Surrogate Key (pk)	Acct#(nk)	EoDBalance	Date	CustID
1	12345|NA	$10,000	10421	23
2	12344|NA	$8,000	10421	24
3	12345|NA	$11,000	10423	23

EoDBalance is our fact. We can add this fact across:

- CustID for a specific date (e.g., total balance of all customers on 10421 = $18,000)

But we **wouldn’t** add it across:

- Date (like year) for a specific customer (e.g., customer 23 for 10421 and 10423 does not add up to customer 23 having $21,000 in their account).

Reference: Kimball, Ralph, and Margy Ross. *The Data Warehouse Toolkit : The Definitive Guide to Dimensional Modeling*, John Wiley & Sons, Incorporated, 2013.

Semi-Additive facts

Let's see another example.

Surrogate Key (pk)	Location	Inventory Quantity	Product	Date
1	12345|NA	5,000	1	10421
2	12345|NA	2,000	1	10422
3	12345|NA	1,000	2	10421

We can't sum up the inventory for a specific product across the date dimension– we have 3,000 in inventory for product 1 on 10422, not 7,000.

However, we ***can*** add up inventory for a specific date across the product dimension– we can have 6,000 in total inventory (product 1 + product 2) on 10421

Reference: Kimball, Ralph, and Margy Ross. *The Data Warehouse Toolkit : The Definitive Guide to Dimensional Modeling*, John Wiley & Sons, Incorporated, 2013, https://network.informatica.com/thread/46827

Non-Additive facts

Non-additive facts cannot be summed across any dimension.

Surrogate Key (pk)	Invoice(nk)	SalePrice	Discount	Product	Date	CustID	EmpID	Location
1	12345|NA	$12.50	50%	13421	10421	23	12	1039
2	12346|NA	$8.50	40%	13420	10422	24	12	1039
3	12347|NA	$12.50	20%	13421	10422	23	13	1040

Discount is our fact. We ***cannot*** add this fact. Let's try anyway!

- Does customer 23's 50% and 20% discounts add up to a 70% discount? Nope.
- Do the total discounts across time and customers add up to a 110% discount? Nope.
- Has location 1039 (or employee 12) given 90% off of sold products? Nope.

Reference: Kimball, Ralph, and Margy Ross. *The Data Warehouse Toolkit : The Definitive Guide to Dimensional Modeling*, John Wiley & Sons, Incorporated, 2013.

Non-Additive facts

What should we do with non-additive facts?

We can compute and store the additive components of the non-additive fact:

Surrogate Key (pk)	Invoice(nk)	SalePrice	FullPrice	Discount	Product	Date	CustID	EmpID	Location
1	12345|NA	$12.50	$25.00	50%	13421	10421	23	12	1039
2	12346|NA	$8.50	$14.17	40%	13420	10422	24	12	1039
3	12347|NA	$12.50	$25.00	50%	13421	10422	23	13	1040

Reference: Kimball, Ralph, and Margy Ross. *The Data Warehouse Toolkit : The Definitive Guide to Dimensional Modeling*, John Wiley & Sons, Incorporated, 2013.

Handling NULL values (empty facts)

In a fact table, you can have a row with NULL facts. This will not impact your ability to conduct analytical operations.

However, your dimension foreign keys cannot be null. They must reference a record in the corresponding dimension table. As we discussed in the dimensions lecture, consider creating a "wildcard" row in dimension tables to store references made from NULL facts.

DO NOT CONFUSE NULL WITH ZERO.

Surrogate Key (pk)	Invoice(nk)	SalePrice	FullPrice	Discount	Product	Date	CustID	EmpID	Location
1	12345|NA	$12.50	$25.00	50%	13421	10422	23	12	1039
2	12346|EU	$8.50	$0.00	8.50/0	13420	10422	24	12	1039
3	12347|EU	$12.50	NULL	NULL	13421	10422	23	13	1040

Did the customer really have a discount that divides by zero? Was the average FullPrice of goods $12.50? **No.**

Reference: Kimball, Ralph, and Margy Ross. *The Data Warehouse Toolkit : The Definitive Guide to Dimensional Modeling*, John Wiley & Sons, Incorporated, 2013.

Handling NULL values (empty facts)

There are some circumstances where a zero value would be acceptable instead of a NULL value. Consider a periodic snapshot fact table where there were no sales recorded for a specific day:

Surrogate Key (pk)	Daily Sales	Date	Location
1	$12.50	10421	1039
2	$0.00	10422	1039

Surrogate Key (pk)	Daily Sales	Date	Location
1	$12.50	10421	1039
2	NULL	10422	1039

Depending on the use case, you may wish to have a zero to reflect the sales as opposed to a null, so it is captured accurately in calculations.

Reference: Kimball, Ralph, and Margy Ross. *The Data Warehouse Toolkit : The Definitive Guide to Dimensional Modeling*, John Wiley & Sons, Incorporated, 2013.

Conformed Facts

Like a conformed dimension, a conformed fact is one that may exist across multiple fact tables. If these facts are computed the same way, and are expected to contain the same values, they are **conformed** facts.

If similar facts exist across tables but are not considered conformed facts (e.g., are in a different format or calculated in a different way), they should be named accordingly.

Reference: Kimball, Ralph, and Margy Ross. *The Data Warehouse Toolkit : The Definitive Guide to Dimensional Modeling*, John Wiley & Sons, Incorporated, 2013

“Factless” Facts

While facts are typically numeric, they don’t always have to be. A factless fact table is one that does not have a numeric element.

If you think you’ve encountered a factless fact table, try to make sure you haven’t missed any potential derivable numeric facts.

Example of a factless fact table (Attendance):

Surrogate Key (pk)	StudentID	Class	Attended	Teacher	Date	Location
1	12345	MGS657	Yes	14627	10421	1039
2	12346	MGS650	No	23019	10422	1038
3	12347	MGS616	Yes	10293	10422	1040

Reference: Kimball, Ralph, and Margy Ross. *The Data Warehouse Toolkit : The Definitive Guide to Dimensional Modeling*, John Wiley & Sons, Incorporated, 2013

Aggregate fact tables

An aggregate fact table is a numeric rollup of otherwise atomic data (e.g., sales). By aggregating fact tables, you can also create shrunken dimensions and optimize queries for specific analytical use cases.

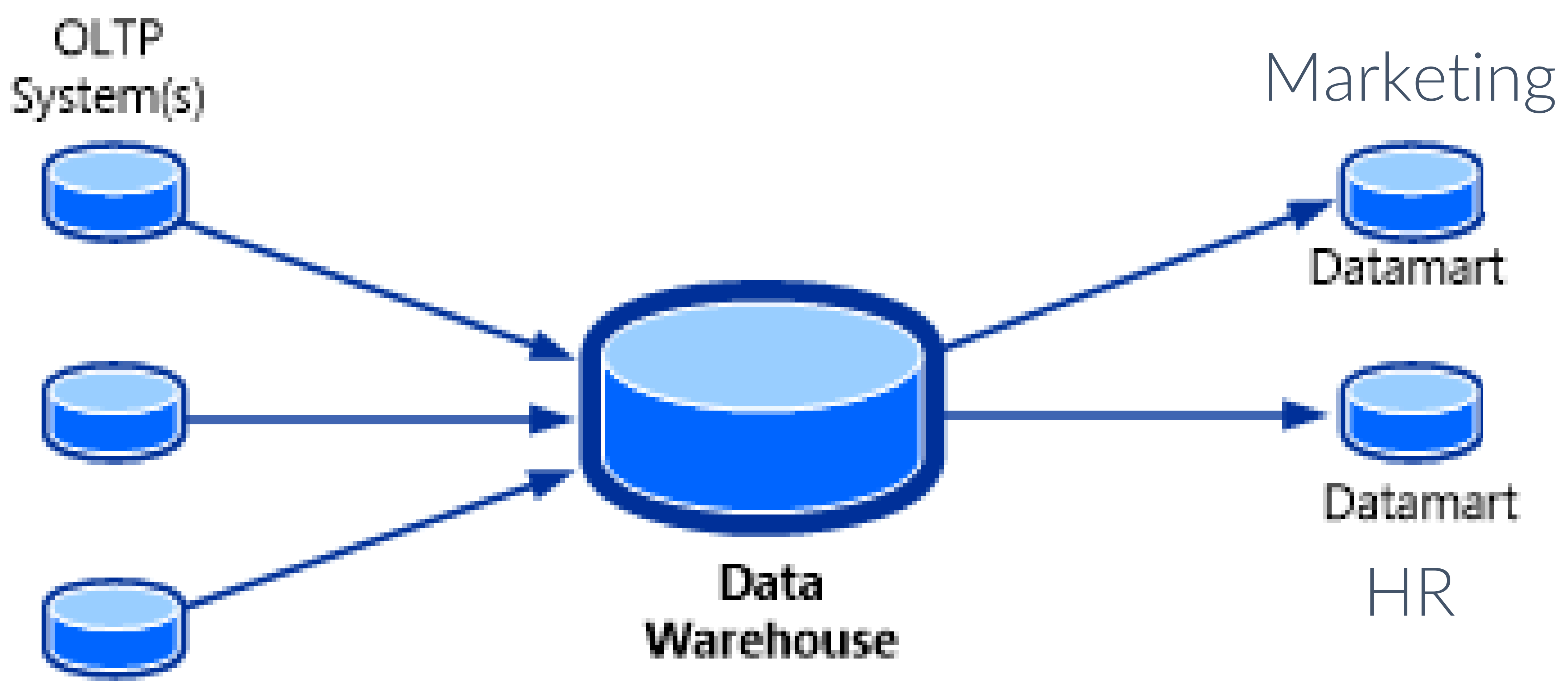
Surrogate Key (pk)	Invoice(nk)	SalePrice	Product	Date	CustID	EmpID	Location
1	12345|NA	$12.50	13421	10422	23	12	1039
2	12346|NA	$8.50	13420	10422	24	12	1039
3	12347|NA	$12.50	13421	10422	23	13	1039

Surrogate Key (pk)	AggregateSalePrice	Date	Location
1	$12.50	10421	1039
2	$20.00	10422	1039

Reference: Kimball, Ralph, and Margy Ross. *The Data Warehouse Toolkit : The Definitive Guide to Dimensional Modeling*, John Wiley & Sons, Incorporated, 2013

Data Marts

“A data mart is a subset of a data warehouse focused on a particular line of business, department, or subject area.” -IBM

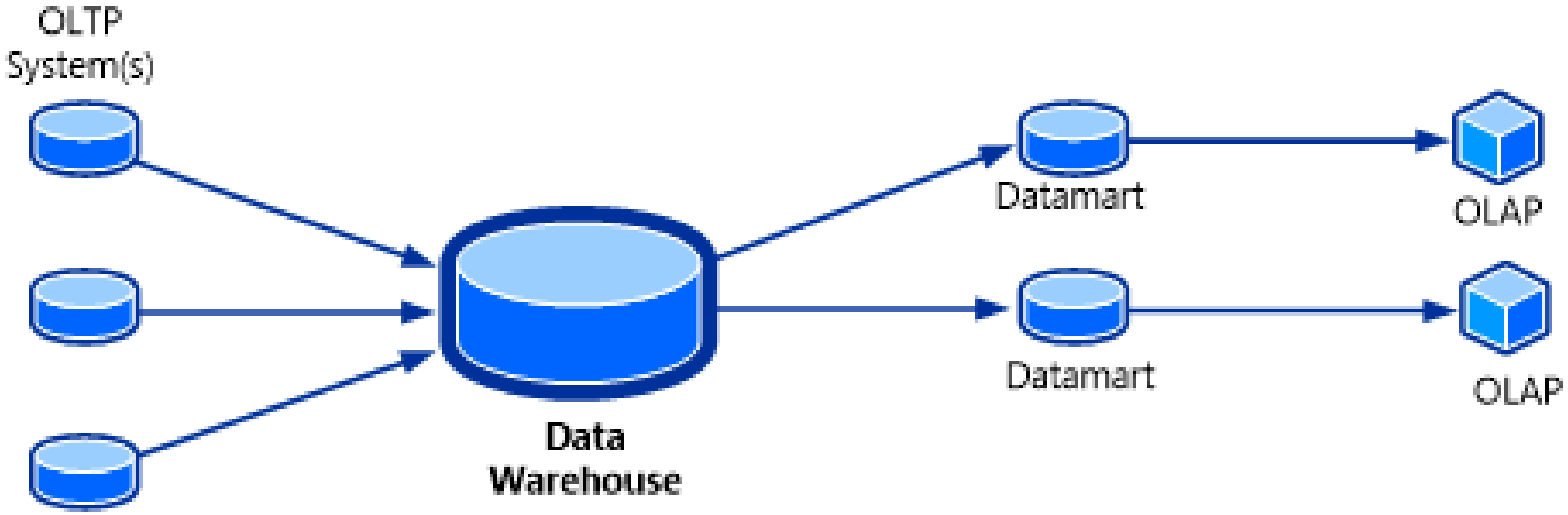


Reference: Kimball, Ralph, and Margy Ross. *The Data Warehouse Toolkit : The Definitive Guide to Dimensional Modeling*, John Wiley & Sons, Incorporated, 2013, https://docs.microsoft.com/en-us/system-center/scsm/olap-cubes-overview?view=sc-sm-2019, https://www.ibm.com/cloud/learn/data-mart#:~:text=A%20data%20mart%20is%20a,through%20an%20entire%20data%20warehouse.

OLAP Cubes

OLAP cubes are **multidimensional databases** that store data in a format that is designed for rapid analysis instead of in a relational format.

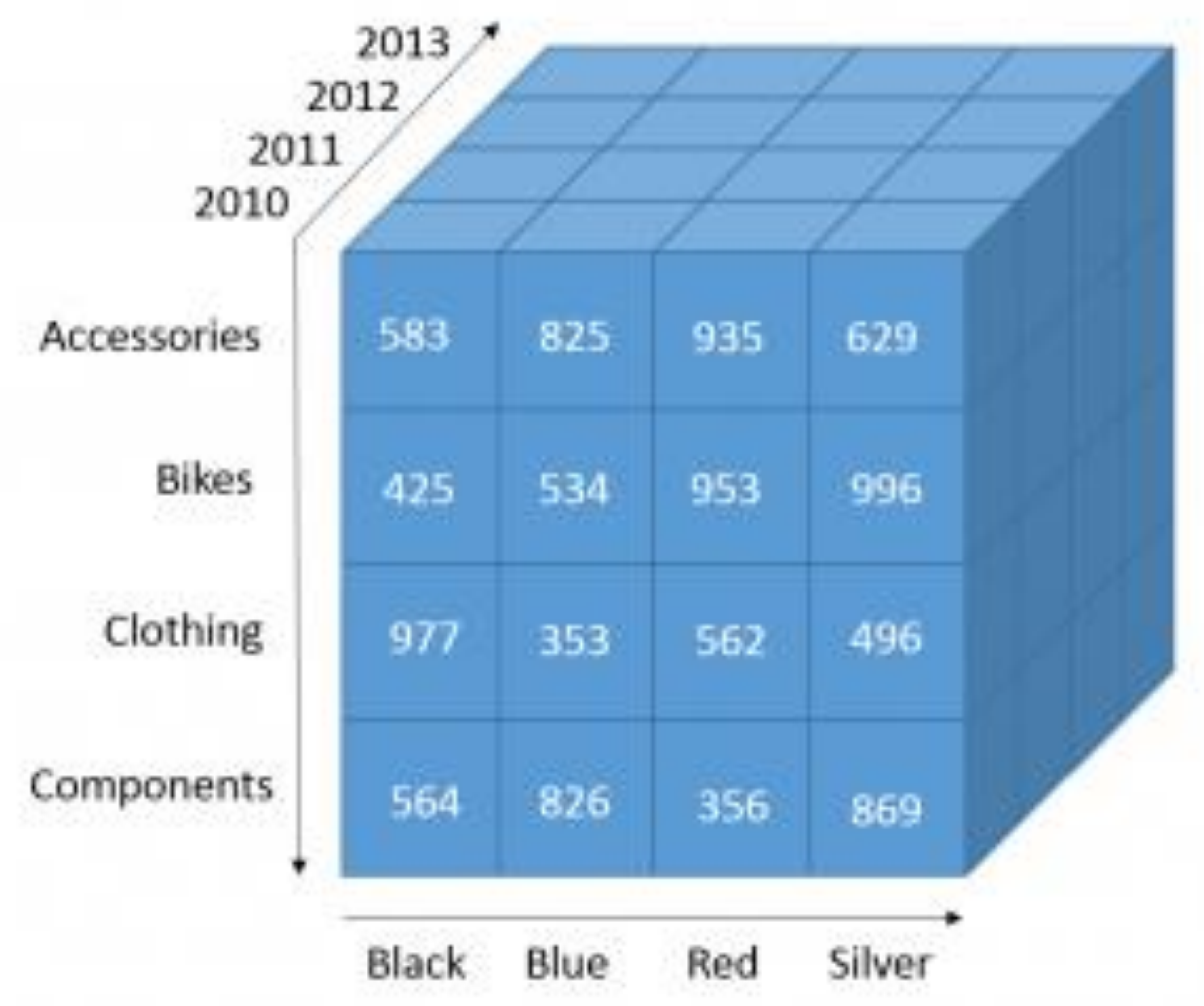
OLAP cubes are often highly specified.



Reference: Kimball, Ralph, and Margy Ross. *The Data Warehouse Toolkit : The Definitive Guide to Dimensional Modeling*, John Wiley & Sons, Incorporated, 2013, https://docs.microsoft.com/en-us/system-center/scsm/olap-cubes-overview?view=sc-sm-2019

OLAP Cubes

Minimize real-time processing, by pre-computing combinations of values for instant reporting.



	Black	Blue	Red	Silver
Accessories	583	825	935	629
Bikes	425	534	953	996
Clothing	977	353	562	496
Components	564	826	356	869

Reference: Kimball, Ralph, and Margy Ross. *The Data Warehouse Toolkit : The Definitive Guide to Dimensional Modeling*, John Wiley & Sons, Incorporated, 2013, https://www.grapecity.com/blogs/working-with-olap-cubes

Consolidated fact tables

Sometimes, it may make sense to combine the facts from two business processes into one table. This can make loading data more complex, but drastically simplifies business intelligence and analytics use cases.

Surrogate Key (pk)	Daily Sales	Sales Forecast	Date	Location
1	$12.50	$15.00	10421	1039
2	$0.00	$10.00	10422	1039

Reference: Kimball, Ralph, and Margy Ross. *The Data Warehouse Toolkit : The Definitive Guide to Dimensional Modeling*, John Wiley & Sons, Incorporated, 2013

Common fact tables

There are a number of common facts you'll repeatedly see across warehouse design initiatives:

- Sales
- Sales forecasts
- Accounts receivable
- Accounts payable
- Shipping
- Customer support
- Health insurance claims

Reference: Kimball, Ralph, and Margy Ross. *The Data Warehouse Toolkit : The Definitive Guide to Dimensional Modeling*, John Wiley & Sons, Incorporated, 2013, https://www.nuwavesolutions.com/slowly-changing-dimensions/

Demo

Let’s take a look at some common facts…

Reference: Kimball, Ralph, and Margy Ross. *The Data Warehouse Toolkit : The Definitive Guide to Dimensional Modeling*, John Wiley & Sons, Incorporated, 2013, https://www.nuwavesolutions.com/slowly-changing-dimensions/