About this presentation

This presentation is designed to provide an overview of the data warehouse planning process.

Agenda

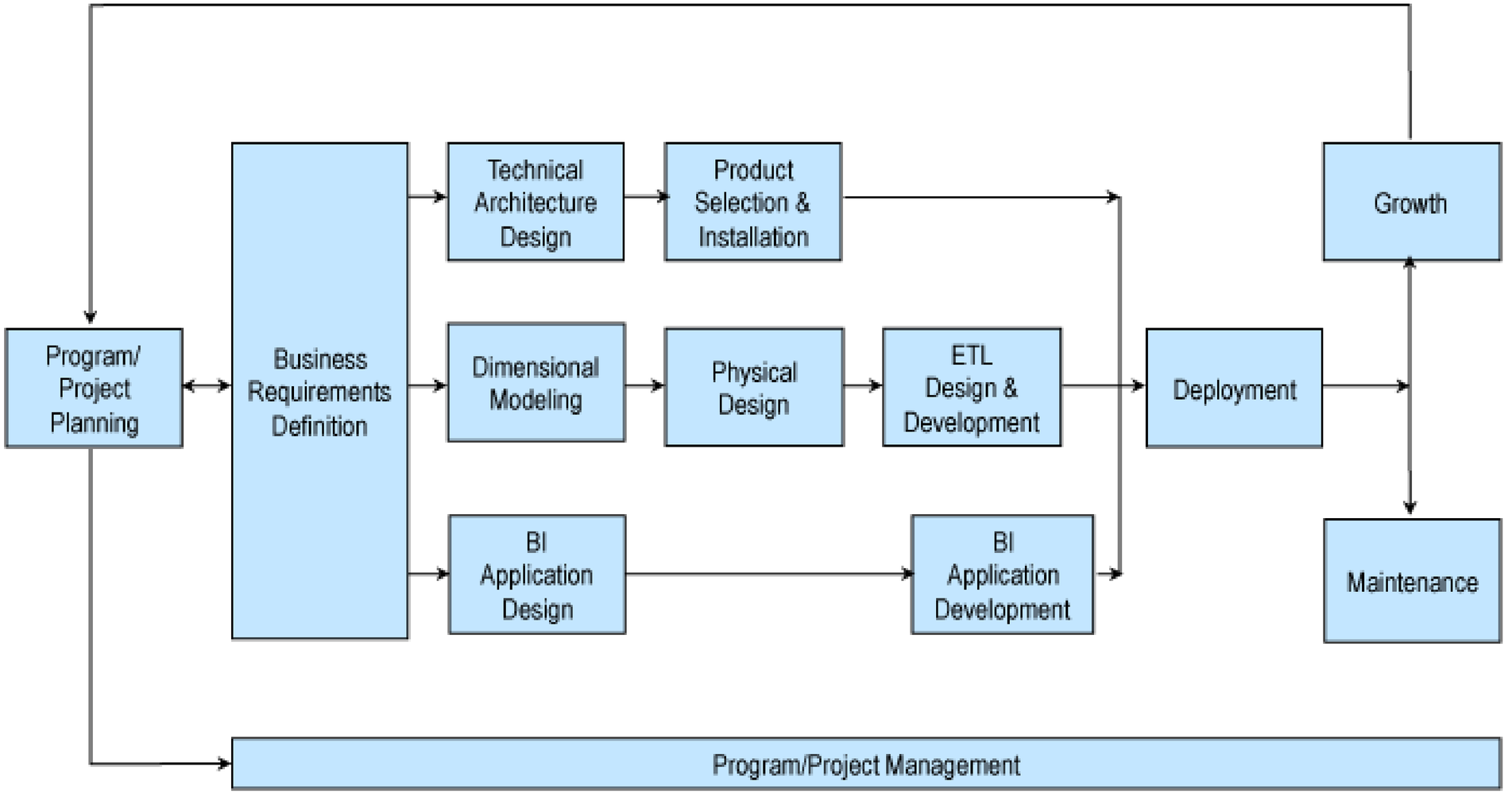
Dimensional Data Warehouse Lifecycle

Dimensional Design

Examples

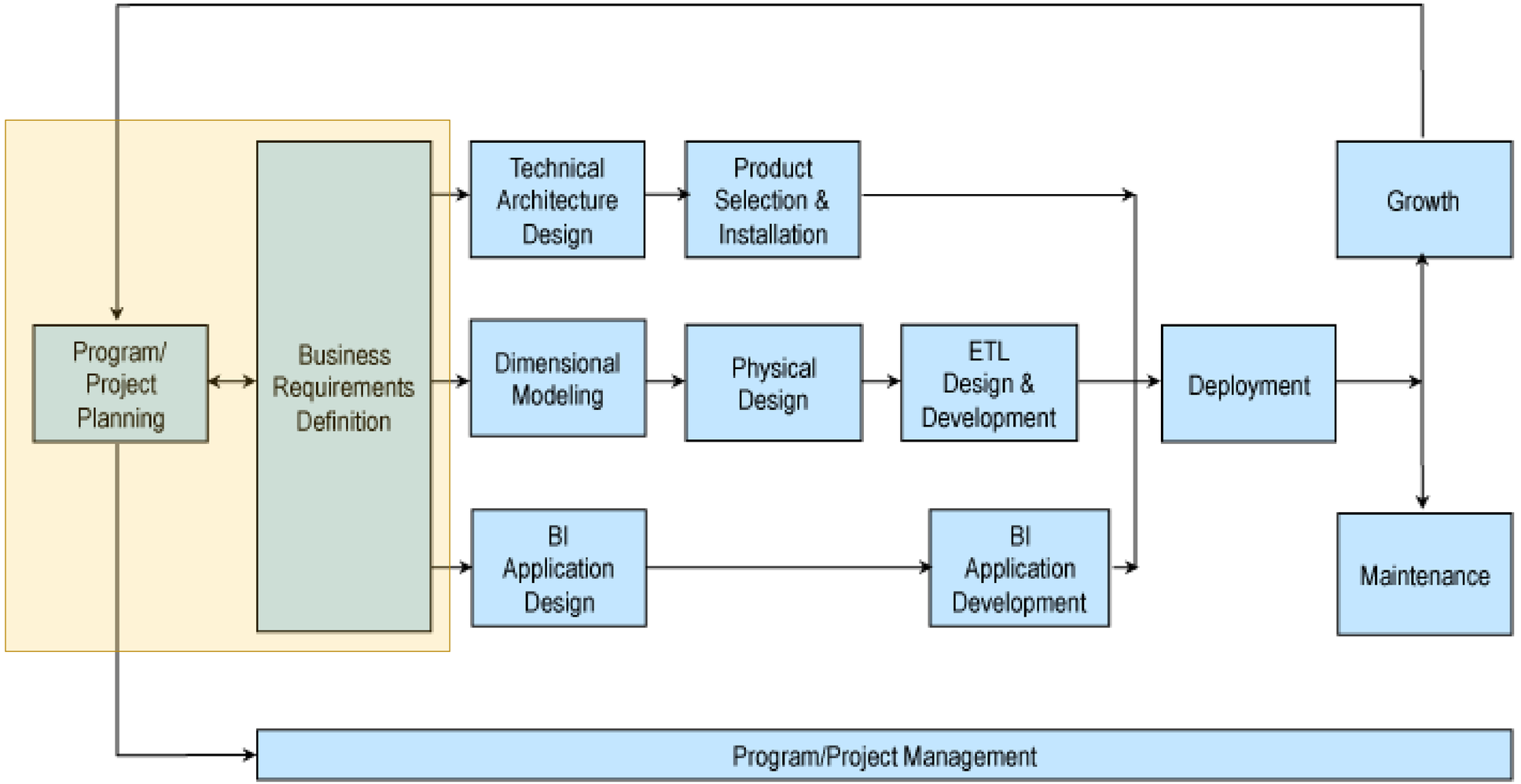
Fall 2024

The data warehouse lifecycle



Reference: https://www.kimballgroup.com

The data warehouse lifecycle



Reference: https://www.kimballgroup.com

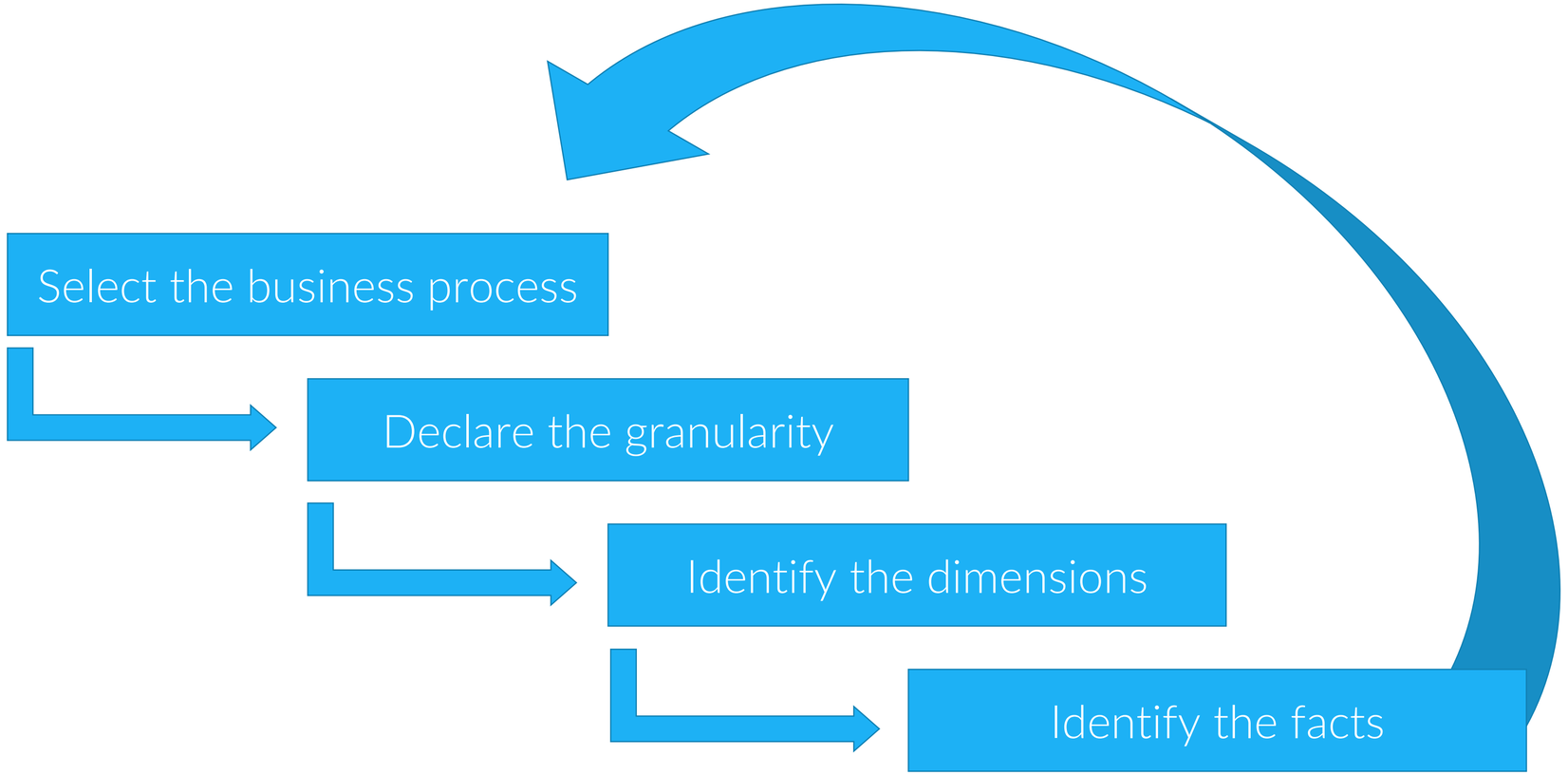
Before you begin modeling...

Gather business requirements.

Remember, you're building a data warehouse to support business analytics. You ***need*** to start your project by gathering business requirements.

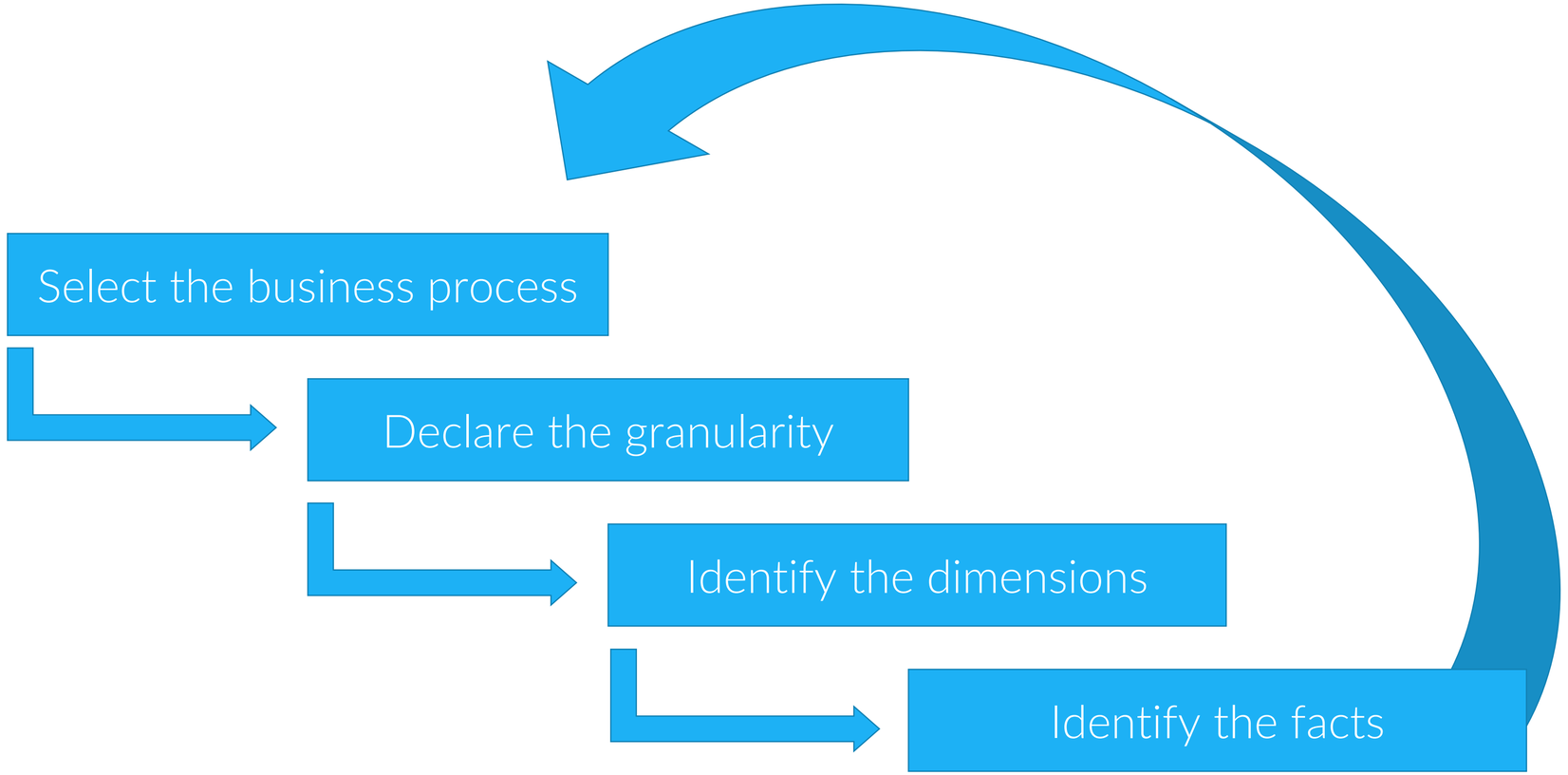
Reference: Kimball, Ralph, and Margy Ross. *The Data Warehouse Toolkit : The Definitive Guide to Dimensional Modeling*, John Wiley & Sons, Incorporated, 2013.

Dimensional Design Process



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

Dimensional Design Process



This process ***should not*** happen in isolation. Conduct collaborative workshops with business representatives and technical resources.

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

Business Processes

A business process represents the operational activities conducted by an organization and the steps performed by people/technology to achieve a goal– like making a sale.

Examples of business processes:

- Customer sales, invoicing, and billing processes
- Customer onboarding
- Purchasing

Reference: Kimball, Ralph, and Margy Ross. *The Data Warehouse Toolkit : The Definitive Guide to Dimensional Modeling*, John Wiley & Sons, Incorporated, 2013, https://www.comindware.com/blog-6-business-process-examples-automation-ideas/

Business Processes

But, how do we identify and select a business process? **We talk to the business!** This is typically conducted through collaborative interviews with stakeholders.

What should we be asking?

- How do you measure business process outcomes (e.g., sales)?
- What are your (or your functional group's) key performance indicators and metrics?
- What are your (or your functional group's) strategic goals and initiatives?
- What is a day-in-the life for your team in support of these goals and initiatives?
- What technologies are used to support your activities?
- Where do you feel you are lacking visibility today?

What *shouldn't* we be asking?

- What business processes are you interested in? (non-specific questions)
- Would you like us to build XYZ? (leading questions)

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

Business Processes

Business processes are often related to one another (e.g., the outputs of one process may be the inputs of another). It’s important to segment these processes out into their smallest “input/output” components.

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

Documenting Business Processes

DATE CREATED	PROCESS NAME
VERSION NO.	CREATED BY
PROCEDURE NO.	PROCESS OWNER
DATE OF LAST UPDATE	LAST UPDATED BY

I. INTRODUCTION

PURPOSE	
SCOPE	
DOCUMENT MANAGEMENT	
ROLES AND RESPONSIBILITIES	

ROLE	RESPONSIBILITY

MATERIALS	

MATERIAL TYPE	NAME	LOCATION / LINK

II. PROCESS

OVERVIEW	

STEP	ACTIVITY

III. MEASUREMENTS

MEASUREMENT CONVENTIONS	

IV. VERIFICATION

VERIFICATION, VALIDATION, AND TESTING PROCESS	

V. REFERENCES

MATERIAL TYPE	NAME	LOCATION / LINK

Reference: https://www.smartsheet.com/free-process-document-templates

Granularity

Once we have our business processes, we need to decide on the level of detail we want to capture. The granularity will determine the detail contained in each row of your eventual fact table.

Atomic grain is the lowest level at which data is captured by a particular process.

Rolled-up summary grains are aggregations of business process outputs.

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

Granularity

Let’s take a customer sales process as an example.

1. Customer enters retail store
2. Customer finds one or more items
3. Customer checks out

The information you want to capture is sales, perhaps dollar amounts. Later, this will become your fact.

In this case, the “atomic grain” may be each individual product sold, by its sales price, while the “rolled up summary grain” may be the entire purchase order by the customer in its totality.

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

Dimensions

Dimensions provide context to the outputs of a business process. For example, they can **describe** the who, what, when, where, why, and how of a specific business process event.

In the case of a customer sales process, the dimensions may include:

- Customer (who)
- Salesperson (who)
- Products (what)
- Store location (where)
- Date/time (when)
- Promotional details (why)
- Payment type (how)

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

Enterprise Bus Matrix

The **Enterprise Data Warehouse Bus Matrix** is a dimensional modeling technique designed to assist in the mapping of processes to potential common dimensions. These will form the foundations for your Star schema relationships when conducting the logical design.

	COMMON DIMENSIONS						
BUSINESS PROCESSES	Date	Product	Warehouse	Store	Promotion	Customer	Employee
Issue Purchase Orders	X	X	X				
Receive Warehouse Deliveries	X	X	X				X
Warehouse Inventory	X	X	X				
Receive Store Deliveries	X	X	X	X			X
Store Inventory	X	X		X			
Retail Sales	X	X		X	X	X	X
Retail Sales Forecast	X	X		X			
Retail Promotion Tracking	X	X		X	X		
Customer Returns	X	X		X	X	X	X
Returns to Vendor	X	X		X			X
Frequent Shopper Sign-Ups	X			X		X	X

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

Enterprise Bus Matrix

Things to keep in mind:

- The bus matrix columns (dimensions) and rows (facts) will directly depend on the processes and descriptors relevant for the industry and specific business.
- The bus matrix mapping of dimensions to processes should be derived from (and confirmed through) collaborative conversations with business stakeholders.
- The enterprise bus matrix (and this entire process) is technology agnostic. We are not considering the technology and infrastructure requirements at this point.

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

Facts

Facts are measurements of a business process. These are typically numeric (e.g., sales).

In a dimensional model, facts are stored in ***fact tables***. Each fact (or, measurement) is an individual record in the fact table that is captured at the specified granularity.

For example:

- A “Sales” fact table may exist as an outcome of a sales business process
- The grain may be declared at the individual product/service level, or the total invoice level, or the store level, etc.
- Each record of a sale in the fact table corresponds with the output from the business process

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

“What about the source data?”

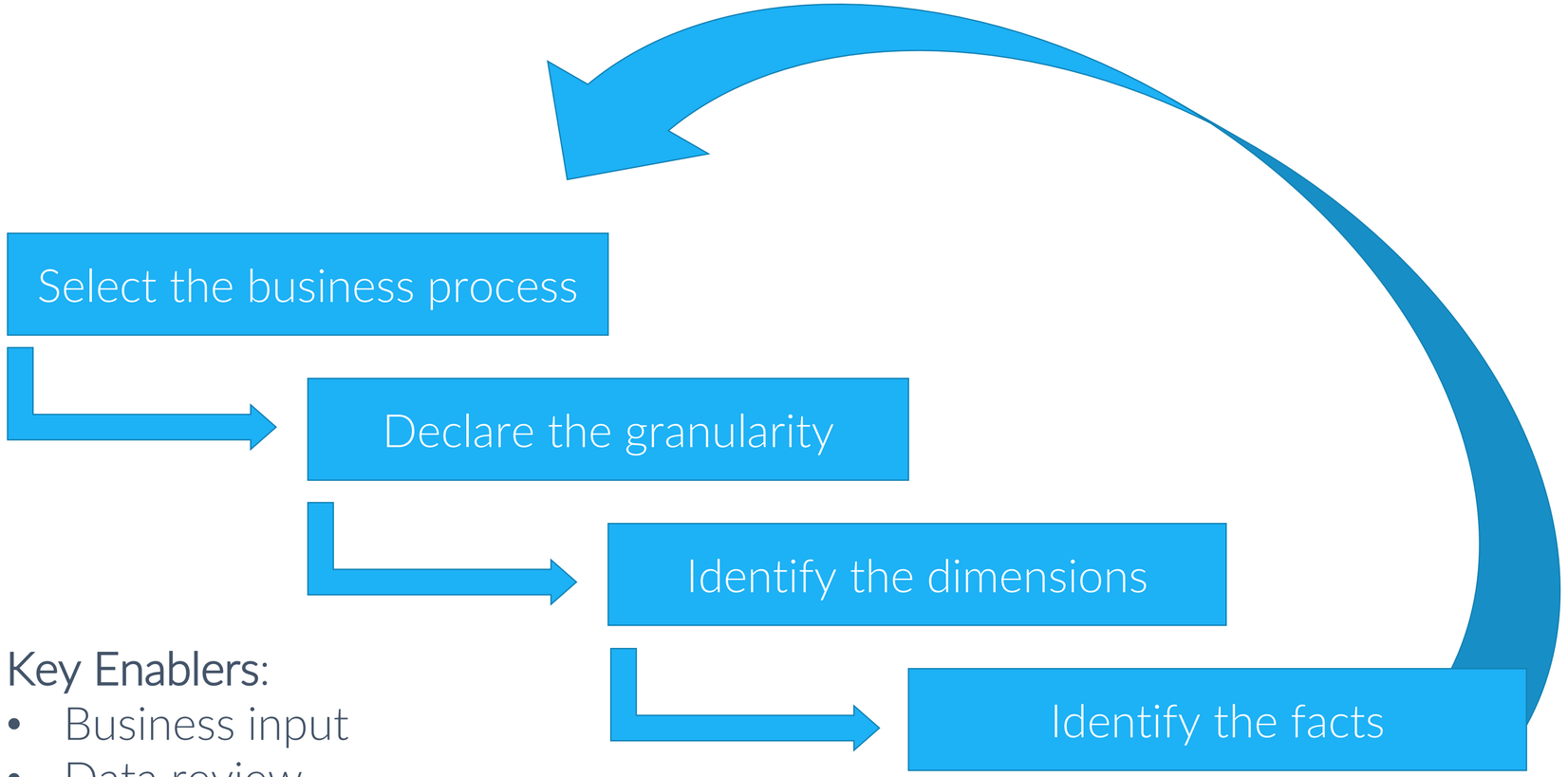
A common question is “can’t we cut out the whole modeling process and just work off of the source data by querying the transactional databases?”

No. Glad we cleared that up.

Jokes aside, the answer is “yes, but we shouldn’t.” The whole point of a data warehouse is to combine multiple data sources to drive meaningful insights.

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

Dimensional Design Process



Key Enablers:

- Business input
- Data review

This process ***should not*** happen in isolation. Conduct collaborative workshops with business representatives and technical resources.

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

Understanding source data

Understanding your source data is a key component of building a strong data warehouse.

Understanding the **available data** helps to:

- Provide context to business process
- Balance business requirements with reality
- Prevent you from overcommitting (**you can’t build a warehouse with data that doesn’t exist**)

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

Conducting source data reviews

Some key tasks:

- Talk to business users to identify systems they work with
- Talk to IT to understand any back-office systems that may factor into business processes
- Review existing system outputs (e.g., reports)
- Profile data (e.g., conduct an analysis of information for its quality, structure, relationships, and content)

Like with business processes, make sure you document all system information alongside business processes.

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

Conducting source data reviews

Some key questions:

- Is this data complete and reliable?
- Is the data in a consistent format?
- How easy is it to access the data?
- Are there legal/regulatory restrictions to using the data?

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

Conducting source data reviews

	Type	Field Name	Original Field Name	Changes	Preview
☐	#	Row ID	Row ID		9,825, 1,973, 436
☐	Abc	Order Date	Order Date		August 15, 2014, December 14, 2014, Dec
☑	Abc	Order ID	Order ID		US-2015-164406, CA-2015-148950, US-2C
☑	Abc	Product ID	Product ID		OFF-BI-10002309, OFF-BI-10001249, TEC-
☐	Abc	Sub-Category	Sub-Category		Binders, Accessories
☐	Abc	Manufacturer	Manufacturer		Avery, Belkin
☐	Abc	Product Name	Product Name		Avery Heavy-Duty EZD Binder With Lockin
☑	Abc	Return Reason	Return Reason		Defective
☑	Abc	Notes	Notes		One ring won't close - E Williams, not all r

Abc Approver 20	Abc Return Notes 135	Abc Order ID 253	Abc Product ID 237	Abc Return Reason 6
null	*null*	CA-2015-100762	FUR-BO-10000362	Defective
C Arnold	Appeared to be corrupt	CA-2015-100867	FUR-BO-10002268	General
C Arnold	armrests would not ad...	CA-2015-102652	FUR-BO-10003159	Incorrect Product Ordered
c. arnold	belts didn't fit custom...	CA-2015-103373	FUR-BO-10004218	Incorrect Product Shipped
C.. Arnold	bindings were not com...	CA-2015-103744	FUR-CH-10000847	No Longer Needed
C/ Arnold	Black chair was shipped	CA-2015-103940	FUR-CH-10000863	Product Listed Incorrectly
E Williams	corner was bent, so it ...	CA-2015-105270	FUR-CH-10000988	
E Wi lliams	Csustomer ordered too...	CA-2015-109918	FUR-CH-10001146	
E Williams	cust didn't realize item...	CA-2015-110786	FUR-CH-10002647	
EWilliams	cust. thought that this ...	CA-2015-111871	FUR-CH-10002774	
F Azad	custmer didn't know it...	CA-2015-116785	FUR-CH-10002961	
	custmer ordered too m...	CA-2015-123225	FUR-CH-10003396	

Let’s put it all together

A member of the business approaches you and asks you if there’s a way to get more meaningful, historical, information about their global sales.

What is your first step?

Identify and document the business processes:

- What does the business mean by “global sales”? How are sales handled globally?
- Are there different types of sales processes? Do they share commonalities (e.g., we have e-commerce and brick and mortar, but they both are the same type of sales)?
- Who is involved in the sales process?
- How do we measure sales?
- What are our KPI’s?
- What systems support the sales process?

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

Let’s put it all together

What is your second step?

Declare the grain:

- At what level does the business want to collect the data?
- What are the business reporting wants/needs?
- What data do we ***actually have***? Does it support the level of granularity requested?

Remember: Grain is a foundational element of your dimensional model. Making changes to grain later takes time and costs money. This is why I recommend **atomic grain** barring other constraints (data source, space).

For global sales… let’s select “individual product sale” as our level of granularity.

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

Let’s put it all together

What is your third step?

Identify your dimensions:

- Who is involved in the sales process? **Customer, salesperson**
- What is being sold? **Products**
- When is it being sold? **Time of sale**
- Where is it being sold? **Location**
- How is it being sold? **Payment method**
- Why is it being sold? **Promotion (e.g., coupon)**

In this case, our dimensions may be: Customer, Salesperson, Product, Time, Location, Payment Type, Promotion

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

Let’s put it all together

What is your third step?

Document your Bus Matrix:

	Customer	Salesperson	Product	Time	Location	Promotion	Payment
Retail Sales	X	X	X	X	X	X	X
E-Commerce Sale	X		X	X		X	X

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

Let’s put it all together

What is your final step?

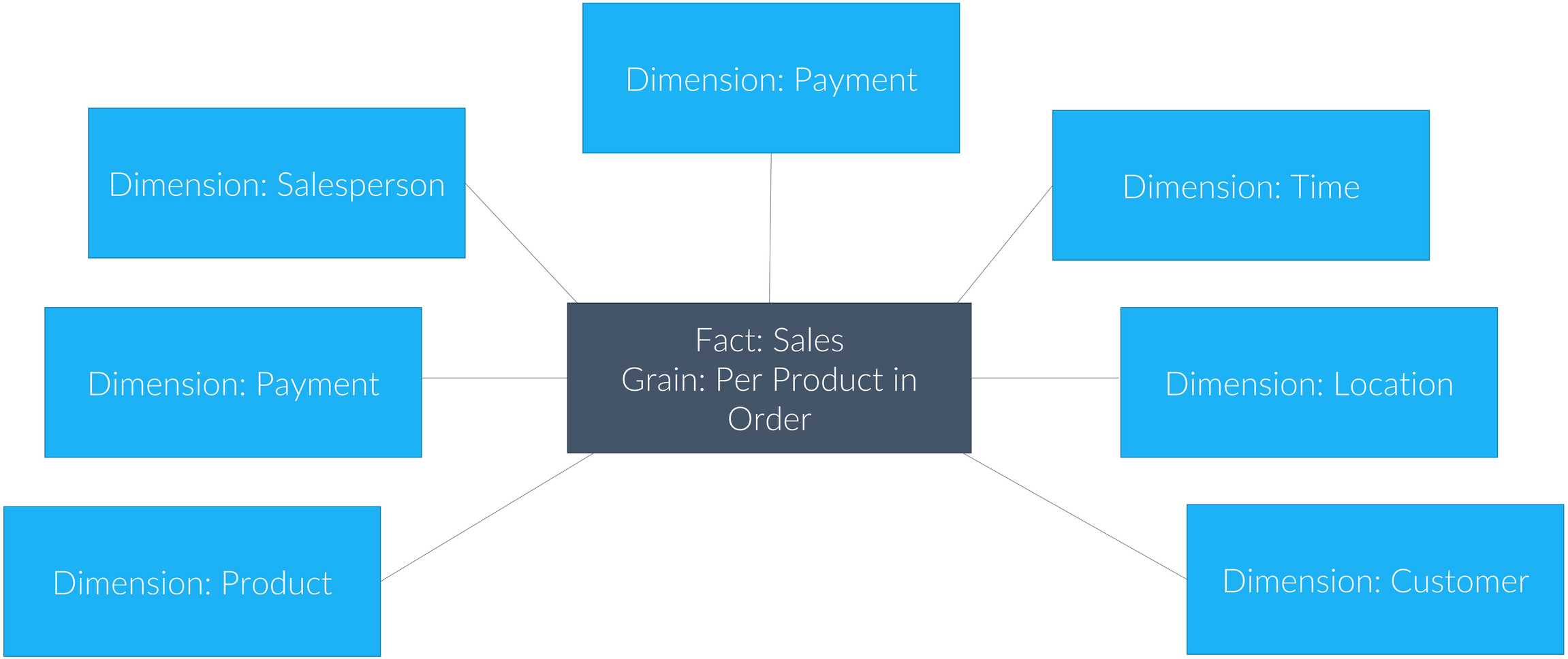
Identify your facts:

- How do we measure “sales” (e.g., $, quantity)?
- Does the way in which we measure sales change over time?
- Do we need to measure sales differently for different circumstances (e.g., e-commerce vs. brick-and-mortar retail, geographic differences, store type differences)?

In this case, let’s assume that we measure sales in $USD, and that we don’t measure sales differently under different circumstances (and sales measurements don’t change over time).

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

Our high-level dimensional model



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