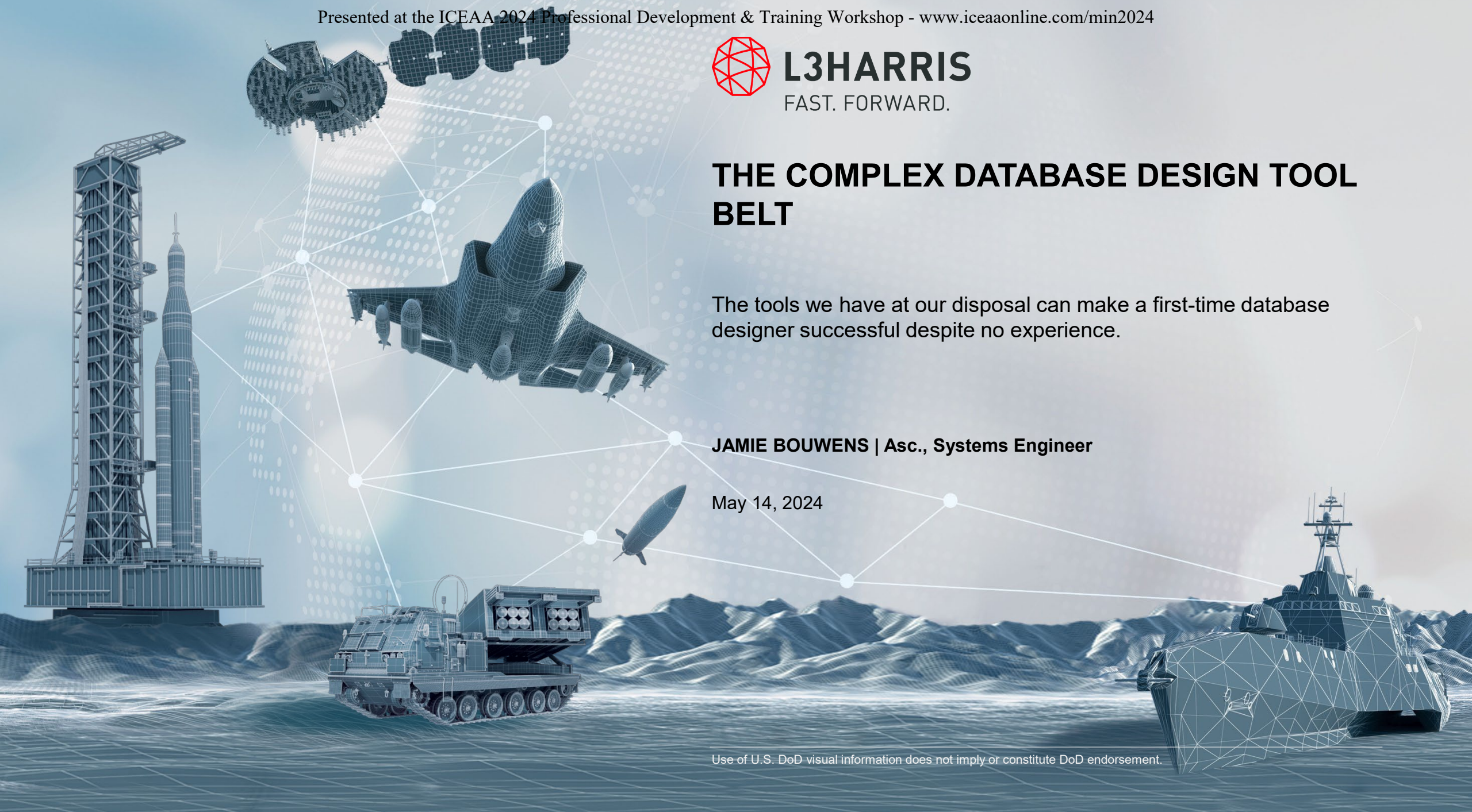


L3HARRIS
FAST. FORWARD.

THE COMPLEX DATABASE DESIGN TOOL BELT

The tools we have at our disposal can make a first-time database designer successful despite no experience.

JAMIE BOUWENS | Asc., Systems Engineer

May 14, 2024

Introduction

Jamie Bouwens – Associate Systems Engineer at L3Harris

- 3.5 years of Industrial/Systems Engineering experience
- B.S. in Industrial Engineering – University of Central Florida
- Yellow belt in Six Sigma – American Society of Quality (ASQ)
- Database Architect for Cost Engineering & Analytics Department (CEAD) with L3Harris

Table of Contents

- **Case Study**
- **Background**
  - Dimensional Database
  - Six Sigma
- **Before Tools**
- **After Tools**
- **Applying Tools**
- **Conclusion**

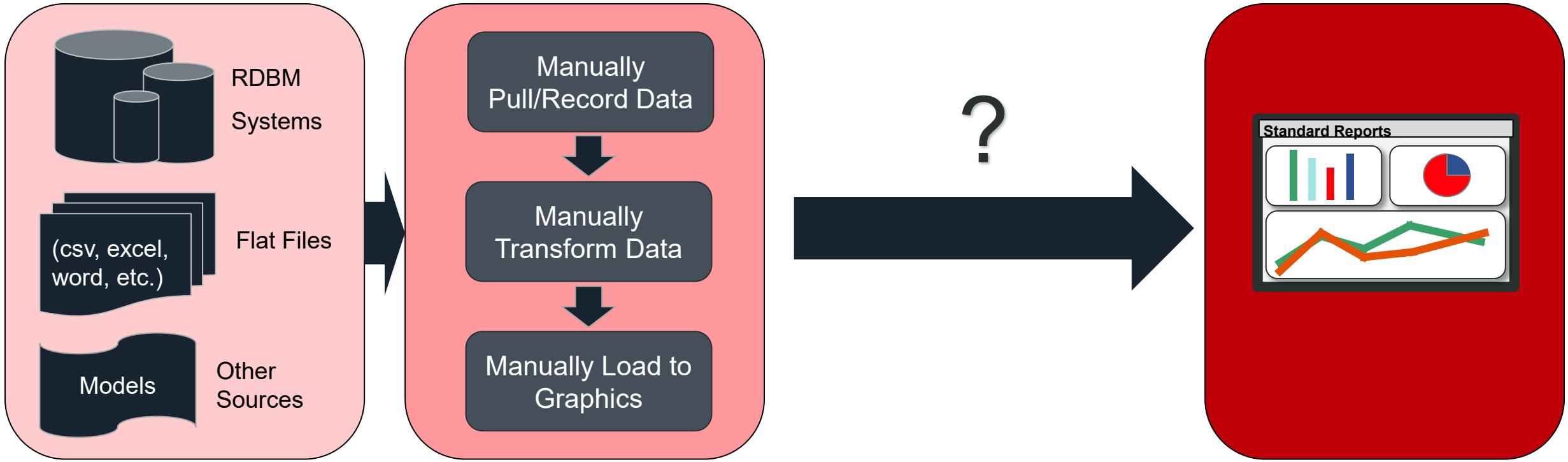
Case Study – Data Pipeline

Subject:

- Team with **no prior database designing experience** was tasked with the project to design a data pipeline.

Source Data → **Data Preparation** → **Data Storage** → **Analytics & BI**

Current State:



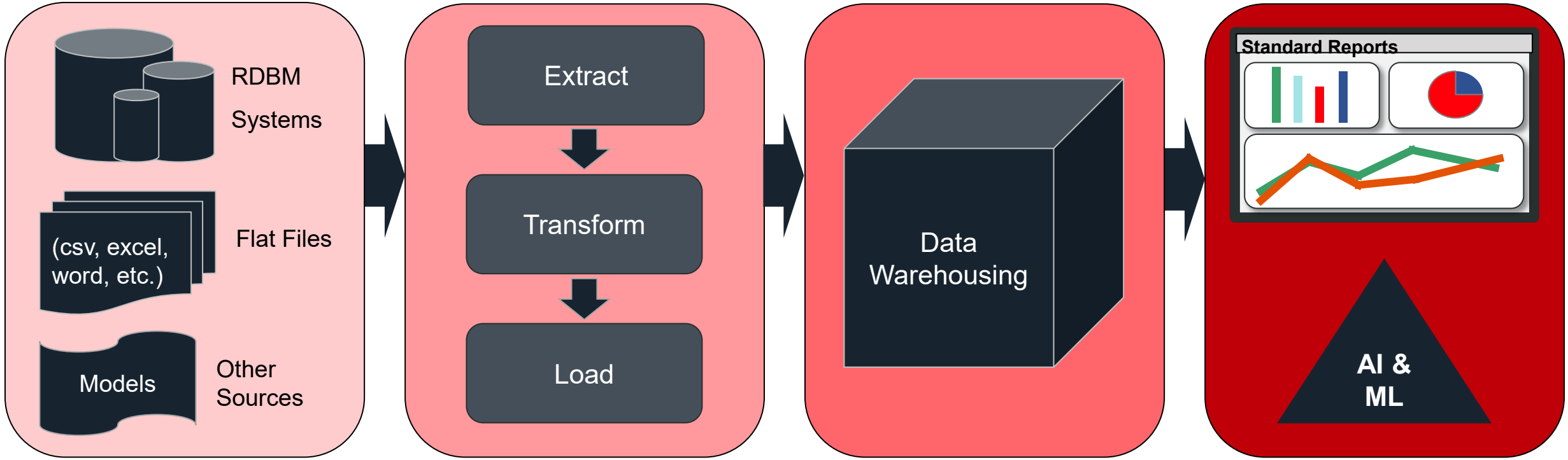
Case Study – Data Pipeline

Subject:

- Team with **no prior database designing experience** was tasked with the project to design a data pipeline.

Source Data | **Data Preparation** | **Data Storage** | **Analytics & BI**

Future State:



Why do we care?

Case Study - Background

Problems

1. Department's current data management method is insufficient for department current and future goals.
2. Team tasked with the data pipeline project has no database experience, dimensional or otherwise.

Goal of Team

- Create a new data management system using a dimensional database as the data storage method.

Goal of Presentation

- Use a case study where six sigma methodologies were implemented to guide the project to success.
- Provide you with a project format and a list of tools you can use to get through this project using Six Sigma methodologies.

Dimensional Database

Context

Dimensional Database:

This is a relational database that uses dimensional data model to organize data. This data model uses the Star Schema architecture.

Star Schema:

Database structure that uses fact tables and dimension tables with foreign-primary key relationships.

Fact Table:

Stores the most basic unit of measurement of a business process. (Usually transactional data of some kind, such as purchases, orders, etc.)

Fact:

Most basic unit of measurement of a business process. (A transaction, an order, a phone call, etc.)

Dimension Table:

Stores the who, when, and where of each business process.

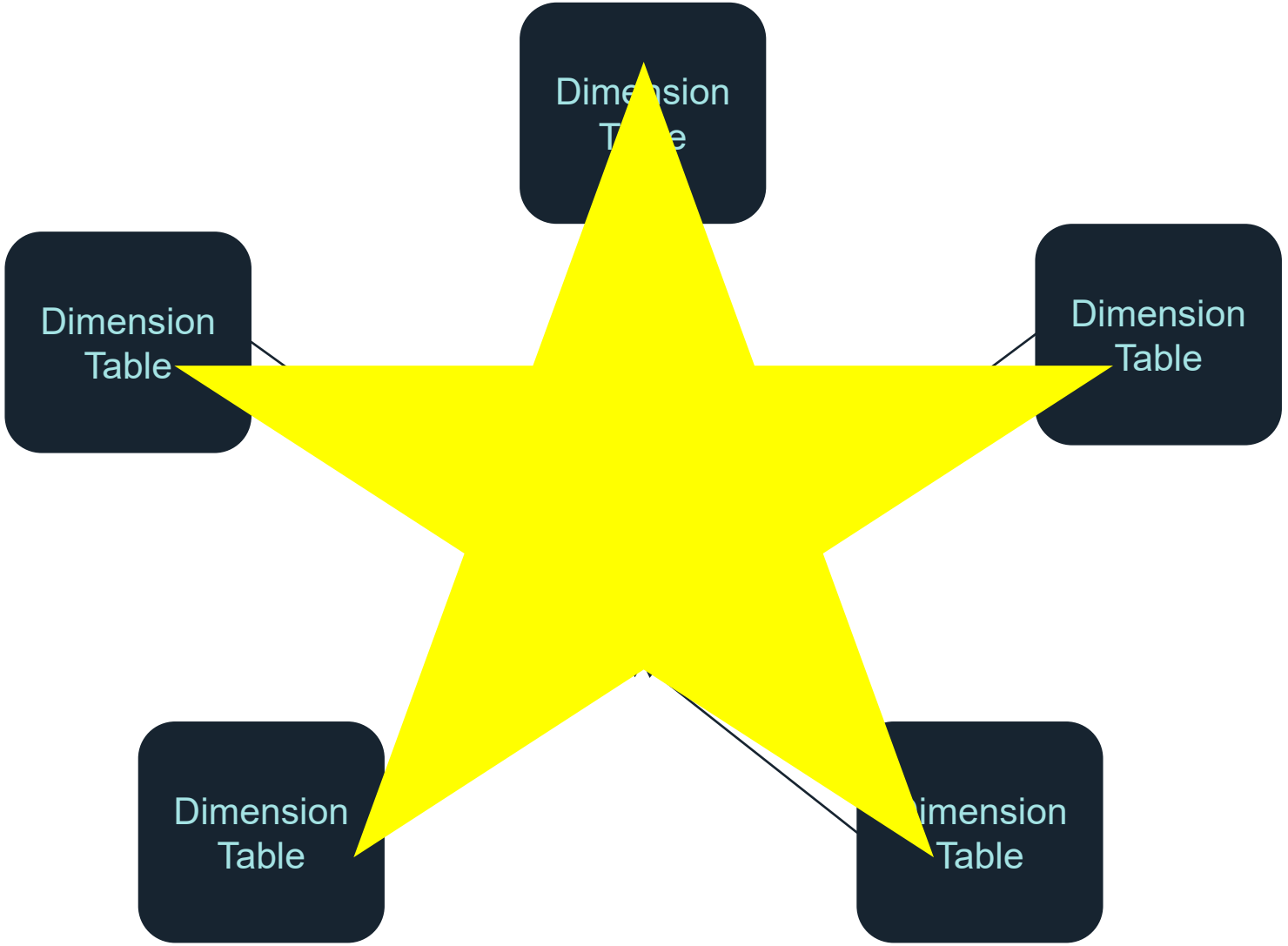
Foreign Key:

A column that references another table (hence the name foreign). It's used to join data from two tables, typically a fact and dimension table.

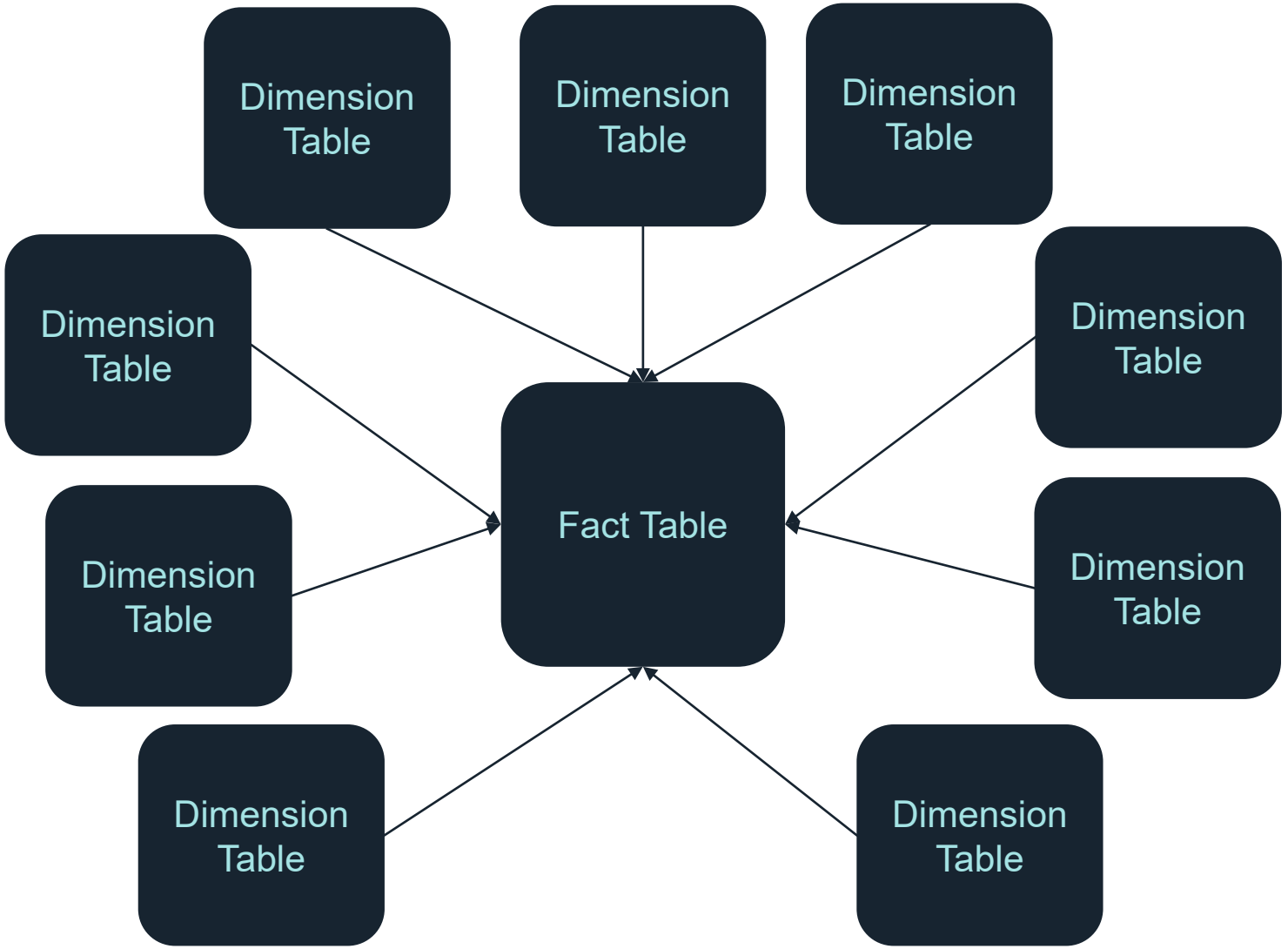
Primary Key:

A column in a dimension table identifying a unique row of data. Primary keys are referenced by foreign keys to join fact and dimension tables

What is Star Schema?



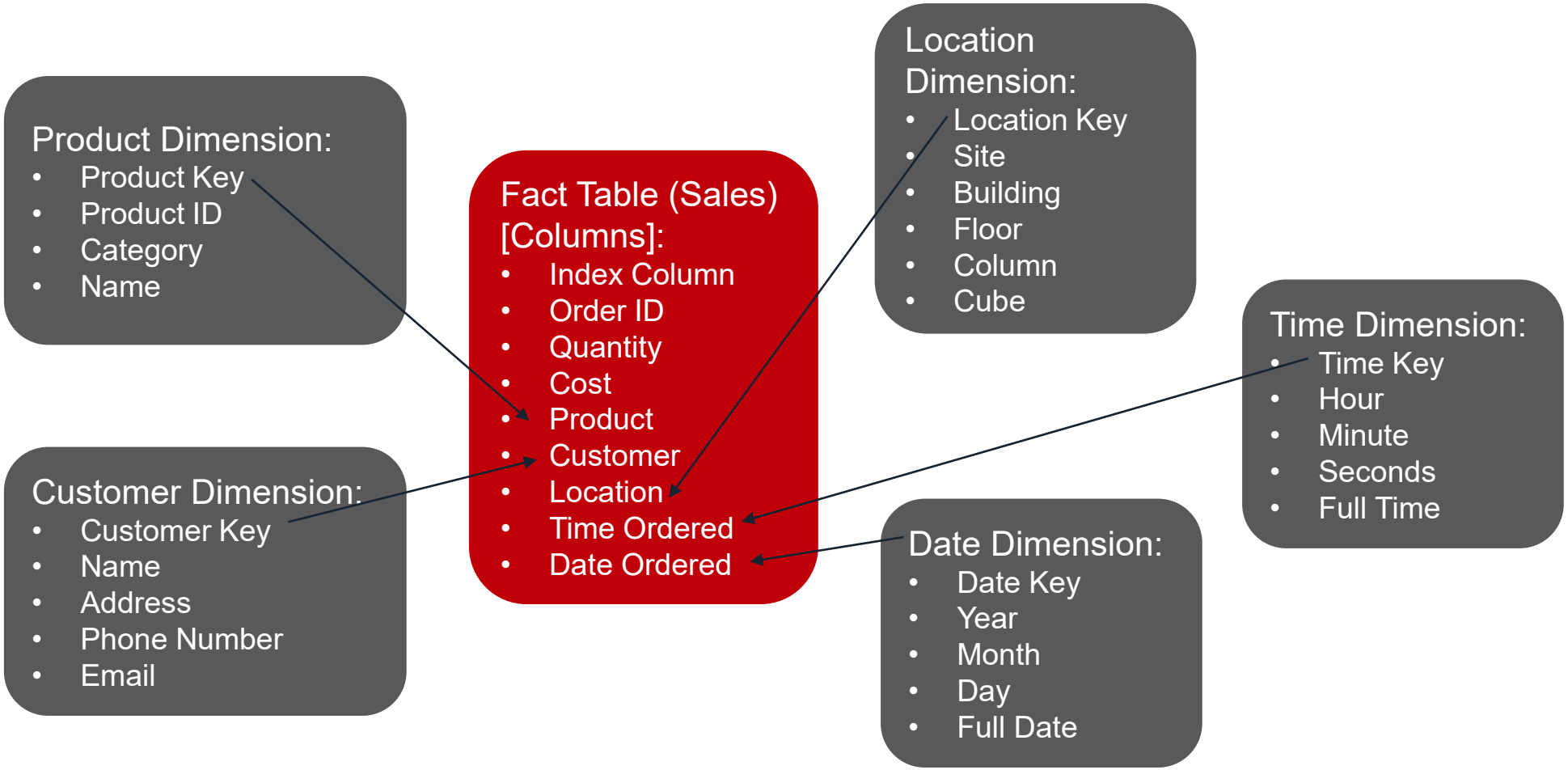
Star Schema is more than just a star…



Dimensional Database – Cost Engineering

Simple Example:

Let's say you have a Cost Engineering team. Your dimensional database would look something like this:



Dimensional Database – Star Schema – Cost Engineering

Complex Example:

Let's say you are the data analyst for a company that has decided to take all its gathered data and store it into a dimensional database. Below are just a FEW of the dimensions you might see, this is not all of them.

Employee Dimension:
- Employee Key
- Name
- ID
- Email
- Location

Business Structure Dimension:
- Business Structure Key
- Organization
- Department
- Team

Fact Table [Columns]:
- Index Column
- Value (fact)
- Business Structure
- Employee
- CES
- Start Time
- End Time
- Charge Code
- Date of Completion
- Start Date
- Program Status
- Training

Time Dimension:
- Time Key
- Hour
- Minute
- Seconds
- Full Time

Date Dimension:
- Date Key
- Year
- Month
- Day
- Full Date

Charge Code Dimension:
- Charge Code Key
- Labor Type
- Project ID
- Project Description
- Project Title

Training Dimension:
- Training Key
- Course Name
- Course ID
- Instructor
- Course Pre-req
- Course Level

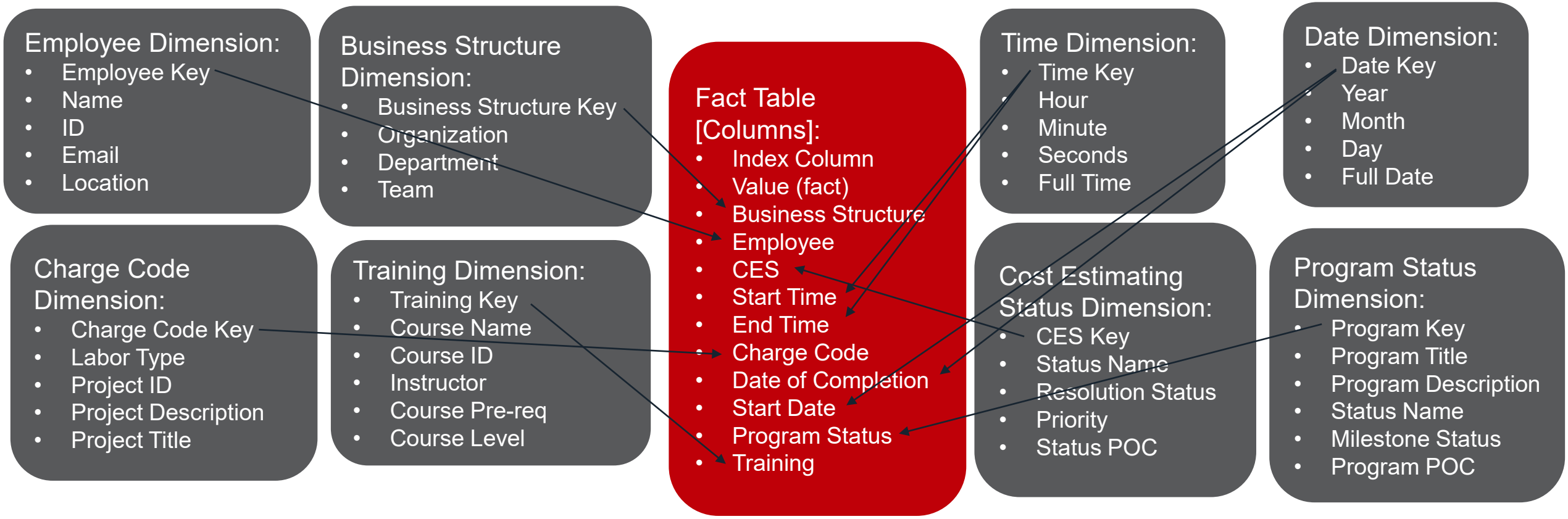
Cost Estimating Status Dimension:
- CES Key
- Status Name
- Resolution Status
- Priority
- Status POC

Program Status Dimension:
- Program Key
- Program Title
- Program Description
- Status Name
- Milestone Status
- Program POC

Dimensional Database – Star Schema – Cost Engineering

Complex Example:

Let's say you are the data analyst for a company that has decided to take all its gathered data and store it into a dimensional database. Below are just a FEW of the dimensions you might see, this is not all of them.



Reference Biehl textbook

Six Sigma

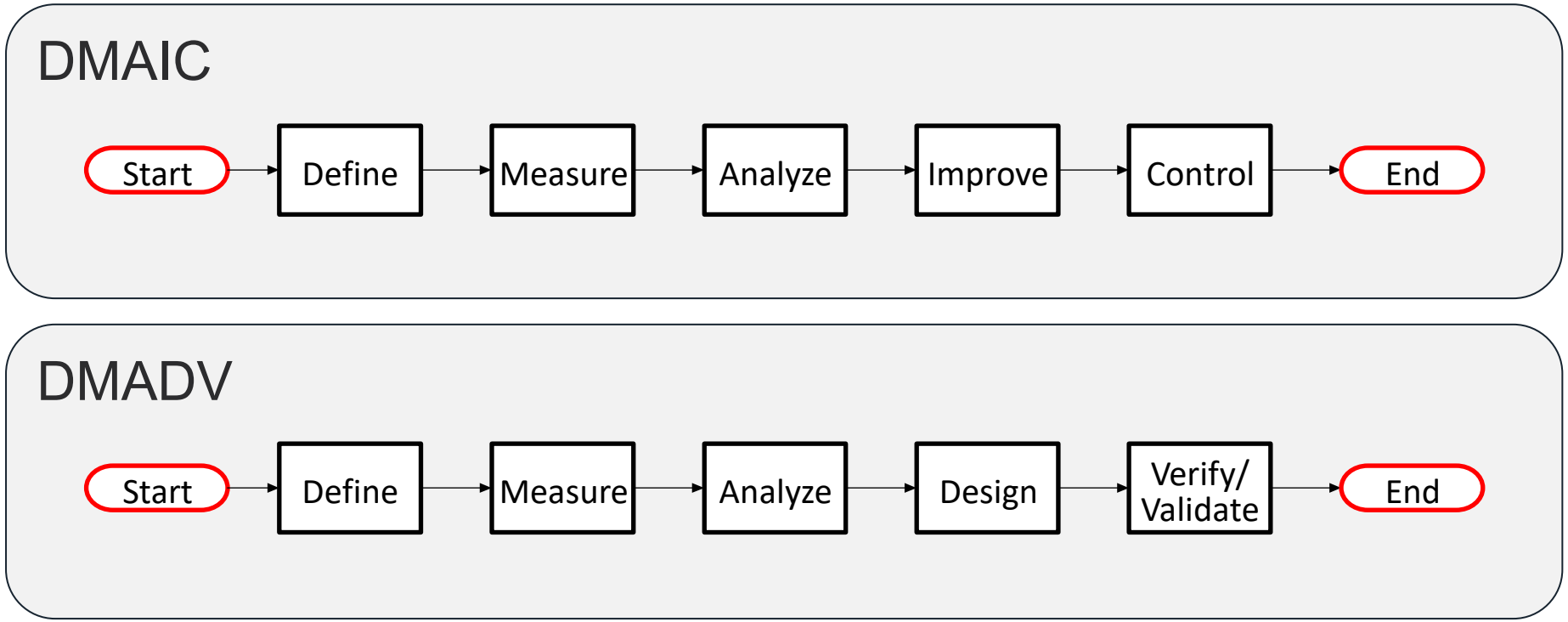
Six Sigma

What is Six Sigma?

- Six Sigma is a set of techniques, methodologies, and tools made for process improvement.
- Think of it like a tool belt

What are we focusing on?

Six Sigma Methodologies and their tools:



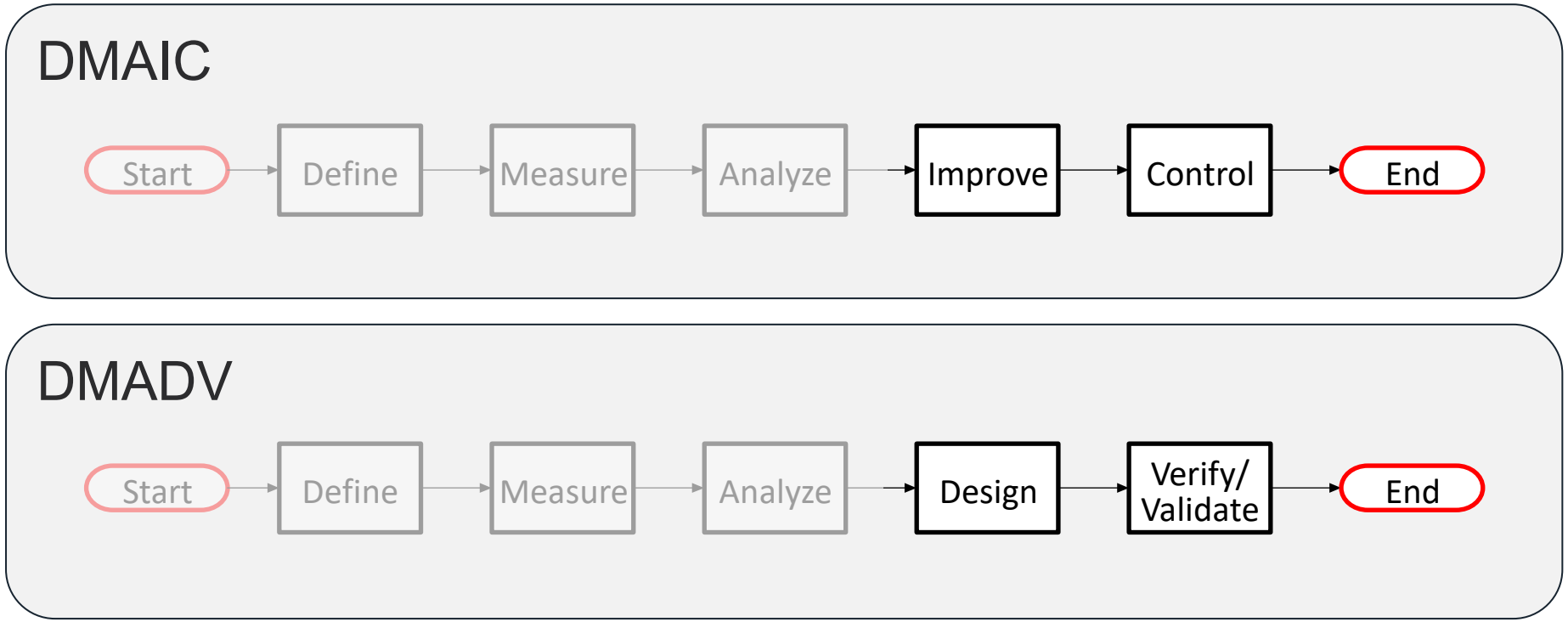
Six Sigma

What is Six Sigma?

- Six Sigma is a set of techniques, methodologies, and tools made for process improvement.
- Think of it like a tool belt

What are we focusing on?

Six Sigma Methodologies and their tools:



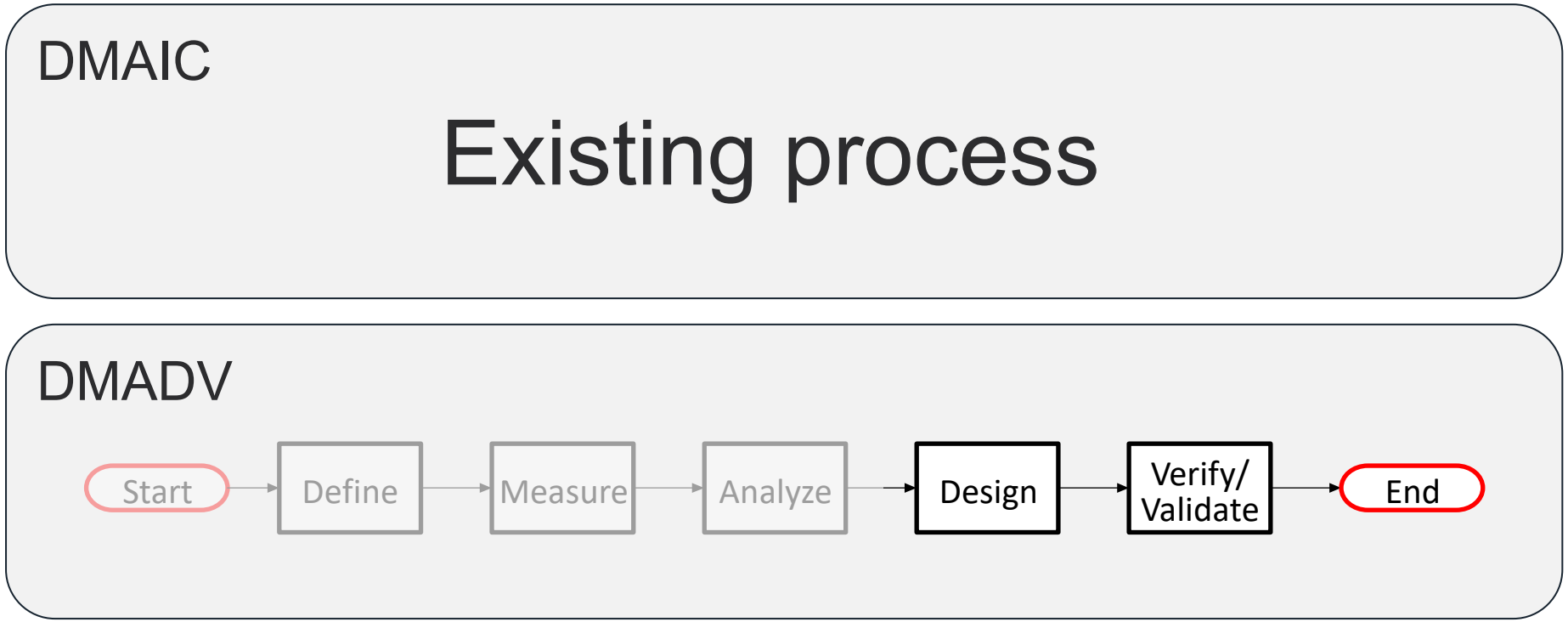
Six Sigma

What is Six Sigma?

- Six Sigma is a set of techniques, methodologies, and tools made for process improvement.
- Think of it like a tool belt

What are we focusing on?

Six Sigma Methodologies and their tools:



Six Sigma

What is Six Sigma?

- Six Sigma is a set of techniques, methodologies, and tools made for process improvement.
- Think of it like a tool belt

What are we focusing on?

Six Sigma Methodologies and their tools:

DMAIC

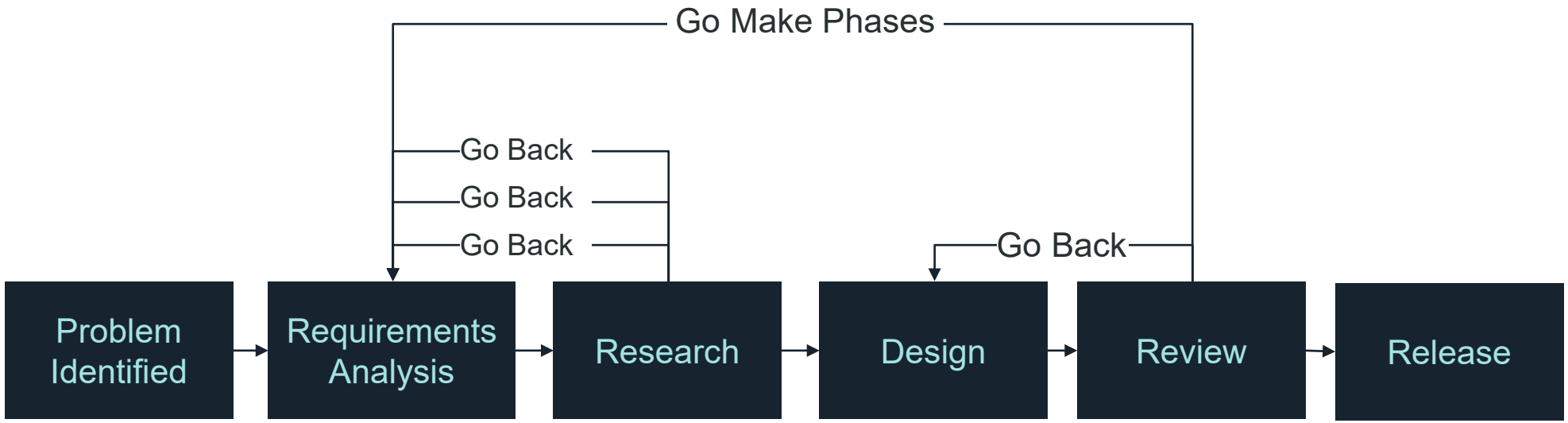
Existing process

DMADV

Entirely New Process

Before/After/Application

Before Method



Data Storage Method Insufficient

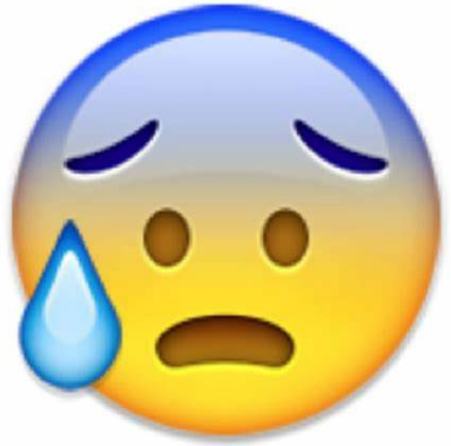
- Store complex Data
- Need Star Shema
- Need Fact Table
- Need Dimensions

- How to store complex data?
- What is star schema?
- What is a fact table?
- What is a dimension?

- Start designing for all complex data
- Chunk complex data to design parts of it

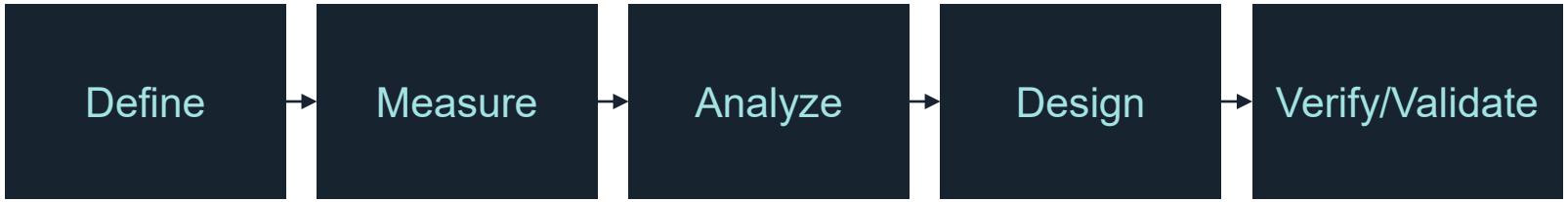
- Review Progress

How are you feeling?



Pause and Take a Breath

After Method (DMADV)



Define

Identify the Problem/Project Goal

- **Tools: Project Charter, Gantt Chart, Stakeholder Analysis Matrix, Voice of Stakeholders Tools, CTQ Trees, SIPOC, Value Stream Map (VSM)**

Problem/Process Analysis

Are you solving the right problem?

What is happening with what at a 3000ft level?

Stakeholder Analysis

Who is involved? What do they provide and receive?

Requirements Analysis

What are the standards for what's going in and what's going out?

Define

Identify the Problem/Project Goal

- **Tools: Project Charter, Gantt Chart, Stakeholder Analysis Matrix, Voice of Stakeholders Tools, CTQ Trees, SIPOC, Value Stream Map (VSM)**

Problem/Process Analysis

- **Voice of Customer/ Employee/ Business (tools)**
- **Project Charter**
- Gantt Chart
- High-level Process map (PMAP)

Stakeholder Analysis

- **Suppliers – Input – Process – Output – Customers (SIPOC)**
- Stakeholder Analysis Matrix

Requirements Analysis

- **SIPOC -> SIRPORC**
- Critical-To-Quality (CTQ) Trees

Define – Example

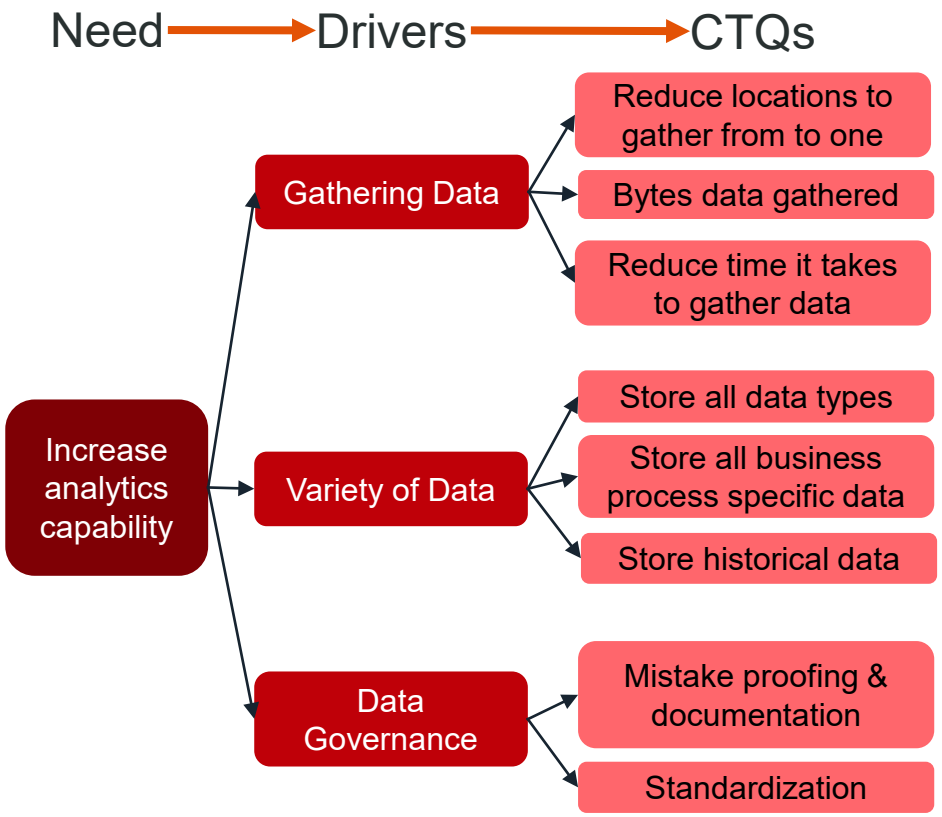
Voice of Customer (VoC):

- Takes too long to produce metrics
- Impossible to do some analytics desired due to the manual workload
- Our methods can't keep up with the increasing work

Project Charter:

- Problem Space: Why are we doing this? How do we judge success? Possible solutions?
- Validation: What do we already know? What do we need to answer?
- Scale & Scope, Why will a customer want this?

SUPPLIERS	REQUIREMENTS	INPUTS	PROCESS	OUTPUTS	Requirements	CUSTOMERS
User	Descriptive and specific syntax	Query	Receive Query → Validate Query → Alter Data → Provide Altered Data	Queried Data	Must match query's actions	User



Measure

Define the Critical-To-Quality's and determine the methods you will use to measure them

- **Tools: Data Collection Plan, CTQs cont'd**

Data Collection Analysis

Prioritize your customer requirements.

How do you plan to gather your measurements?

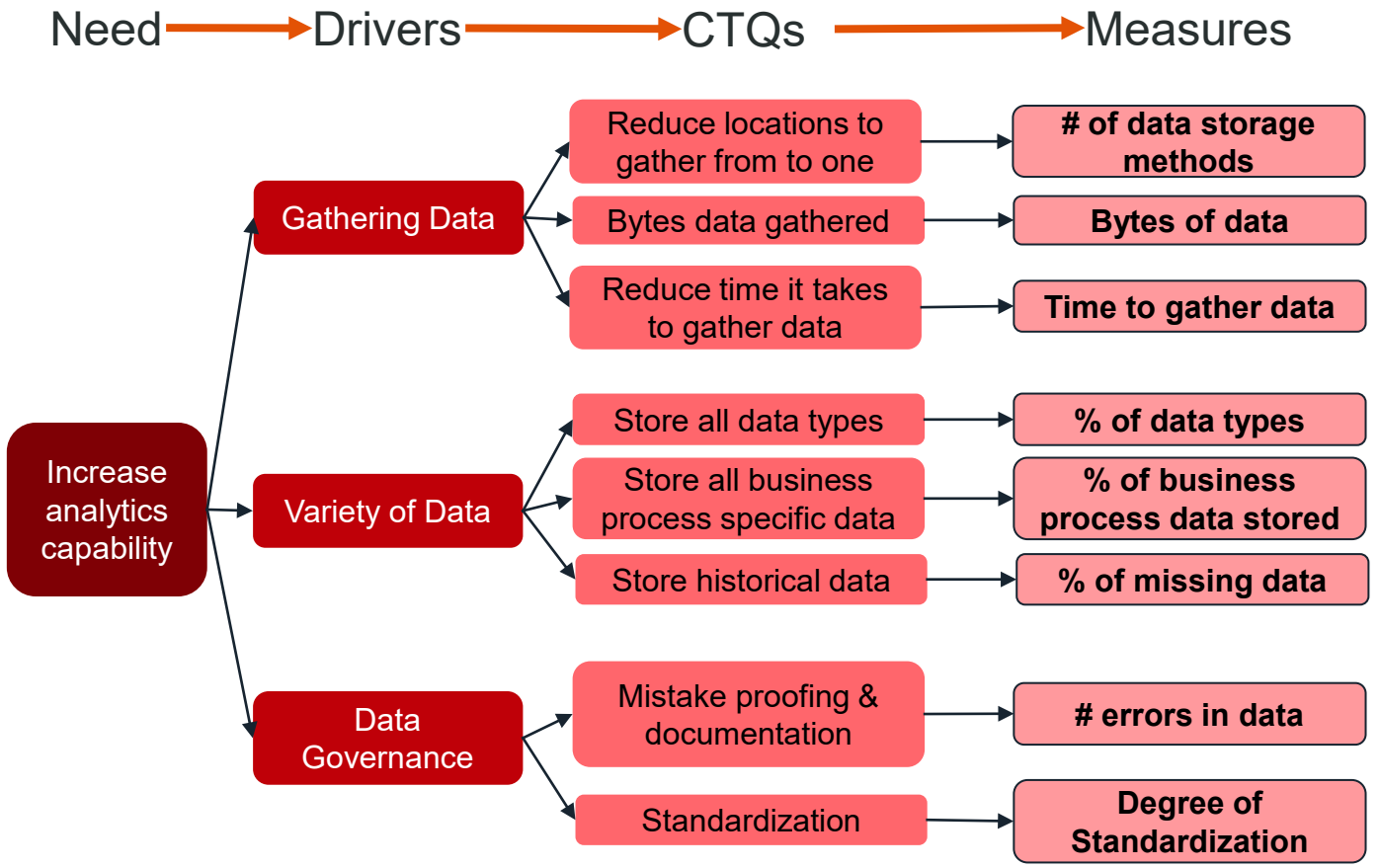
Measure – CTQ+ Example

Define the Critical-To-Quality's and determine the methods you will use to measure them

- **Tools: Data Collection Plan, CTQs cont'd**

Data Collection Analysis

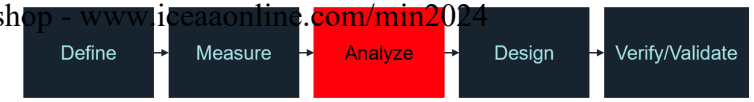
- **CTQs cont'd**
  - Take those requirements from the CTQs and turn them into measures
- **Data Collection Plan**
  - What's your measurement plan?
  - Gather As-Is process metrics



Measure – Data Collection Plan Example

Measure Name	Data Type	Operational Definition	Stratification Factors	Sampling Notes	Who & How?
What would you call it?	*Discrete/ Continuous*	*What is the airtight description?*	*Will you slice data by who, what, where, and when?*	*How much data will you collect? How often will you collect?*	*Who will collect and what method will they use?*
Time to gather Data	Continuous	This is the time (in minutes) it takes from the moment of gathering data to update metrics to the moment the presentation of metrics is uploaded, and every graphic is updated & validated.	By business purpose	Sample every update made to department metrics for the span of 3 months starting 1/1	Team member will observe operations manager, timing with the clock.
Number of Errors in Data	Discrete	This is the number of errors found in the data that has been uploaded into the database after ETL and before BI.	By data upload by business purpose by data source	Sample every update made to department metrics for the span of 3 months starting 1/1	Team member will perform tests on gathered data to see if there are any discrepancies.
Degree of Standardization	Continuous	This is the level of variability or deviation from standards that can be found in the Data Governance procedures/processes.	By procedure/process by employee	Sample every update made to department metrics for the span of 3 months starting 1/1	Team member will calculate the variability in the data by pulling the data for certain processes uploaded by different employees.

Analyze



Convert critical to quality trees to process features. Develop and evaluate design concepts

- Tools: Pugh Matrix, FMEA, Evaluate design options from measure, Detailed Process Mapping

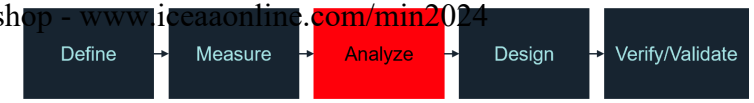
Design Analysis

What variety of solutions are there?

Solution pros/cons?

Risk assessment of features

Analyze



Convert critical to quality trees to process features. Develop and evaluate design concepts

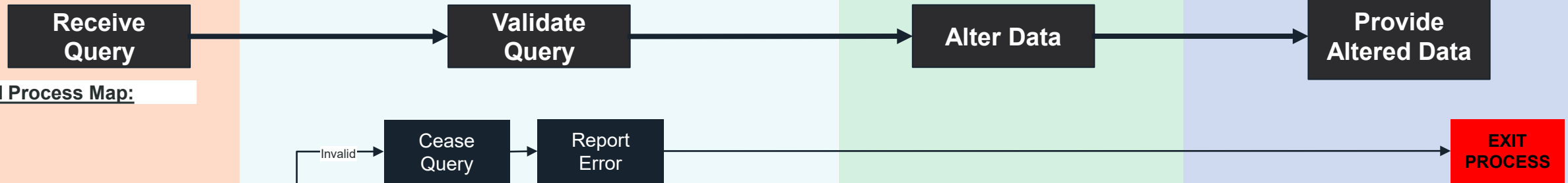
- Tools: Pugh Matrix, FMEA, Evaluate design options from measure, Detailed Process Mapping

Design Analysis

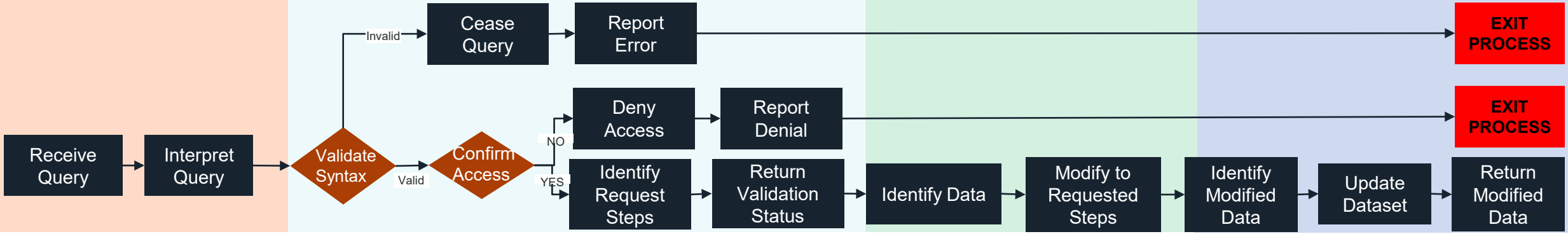
- **Detailed Process Mapping**
- Decision Tree Analysis
- Pugh Matrix
- **FMEA**
- Evaluate design options from measure

Analyze – Example

High-Level Process Map:



Detailed Process Map:



Process Step	Potential Failure Mode	Potential Effects of Fail	SEV	Potential Cause(s) of Failure	OCC	Current Process Controls to Prevent Failure Mode	Current Process Controls to Detect Failure Mode	DET	RPN	Recommended Actions	POC	Target Complete Date	Actions Taken	SEV	OCC	DET	RPN
Identify Request Steps	Non-specific syntax	Incorrect data gathered	8	Understanding of data source insufficient	7	None	Visual Inspection	4	224	Consolidate data sources	N/A	*mm/dd/yyyy*	Developed Database with known structure	8	2	1	16
Confirm Access	False Positive Access	Unauthorized access to sensitive data	10	Insufficient Data Governance	2	Manually give access as required to manually verified people.	None	8	160	Increase Degree of Standardization	N/A	*mm/dd/yyyy*	Developed standard approval process and regular access checks	10	1	2	20

Analyze – Example

Measures	Current State	Design 1 – Dimensional Database – SCD Method 1 – Type 1	Design 2 – Dimensional Database – SCD Method 2 – Type 2	Design 3 – Dimensional Database – SCD Method 2 – Mixed Type 1 & 2
# of Data Storage Methods	50+	1	1	1
Bytes of Data Stored	20 GB	1 GB	25 GB	16 GB
Bytes of Data Accessible at once	2GB	Microsoft SQL Server: 180GB	Microsoft SQL Server: 180GB	Microsoft SQL Server: 180GB
Time to gather data	Several Hours	< 5min	< 10min	< 10min
% of Business Process Data Stored	100%	10%	100%	70%
% of Missing Data	0%	90%	0%	30%
Degree of Standardization	No Standards	Standard Procedures	Standard Procedures	Standard Procedures

Design

Create a prototype of the design and complete trial runs. Finalize the design.

- Tools: Quality Function Deployment (QFD), Prototyping, Tolerance Analysis, Standard Operating Procedures, Standard Work

Develop Detailed Plans/Procedures

What happens if you don't meet those requirements?

Is your design reliable?

Building Design

Pick the design of preference from Analyze phase

Build that design in detail

Design

Create a prototype of the design and complete trial runs. Finalize the design.

- Tools: Quality Function Deployment (QFD), Prototyping, Tolerance Analysis, Standard Operating Procedures, Standard Work

Develop Detailed Plans/Procedures

- Procedure Design
- Management Systems Standard (MSS)
- Standard Work

Building Design

- Quality Function Deployment (QFD)
- Tolerance Analysis
- **Prototyping**

Design – Example

Complex Example:

Let's say you are the data analyst for a company that has decided to take all its gathered data and store it into a dimensional database. Below are just a FEW of the dimensions you might see, this is not all of them.

Employee Dimension:
- Employee Key
- Name
- ID
- Email
- Location

Business Structure Dimension:
- Business Structure Key
- Organization
- Department
- Team

Fact Table [Columns]:
- Index Column
- Value (fact)
- Business Structure
- Employee
- CES
- Start Time
- End Time
- Charge Code
- Date of Completion
- Start Date
- Program Status
- Training

Time Dimension:
- Time Key
- Hour
- Minute
- Seconds
- Full Time

Date Dimension:
- Date Key
- Year
- Month
- Day
- Full Date

Charge Code Dimension:
- Charge Code Key
- Labor Type
- Project ID
- Project Description
- Project Title

Training Dimension:
- Training Key
- Course Name
- Course ID
- Instructor
- Course Pre-req
- Course Level

Cost Estimating Status Dimension:
- CES Key
- Status Name
- Resolution Status
- Priority
- Status POC

Program Status Dimension:
- Program Key
- Program Title
- Program Description
- Status Name
- Milestone Status
- Program POC

Validate/Verify

Implement the final design and confirm that it achieves the desired results

- Tools: DVP&R, Check Sheets, Standard Operating Procedures (SOP), Standard Work, Change Management, Control Plan, Implementation Plan

Testing Plans/Reports

How are you going to test your design meets your KPI requirements?

Performance Evaluation

How did your design perform?

Final Project Evaluation

Lessons Learned?

How do you plan to implement?

Validate/Verify

Implement the final design and confirm that it achieves the desired results

- Tools: DVP&R, Check Sheets, Standard Operating Procedures (SOP), Standard Work, Change Management, Control Plan, Implementation Plan

Testing Plans/Reports

- Test Plan
- **Design Verification plan & Report (DVP&R)**

Performance Evaluation

- Pilot
- Control Charts
- Check Sheets

Final Project Evaluation

- Lessons Learned
- Change Management
- **Implementation Plan**

Validate/Verify – Example

Design Verification plan & Report (DVP&R):

Test ID	Test Name/ Description	Objective	Steps	Expected Result	Actual Result	Severity / Priority	Comments	Assigned To:	Execution Date:	Bug ID / Defect ID
1	Time to Query	Measure the time it takes to query data once request received	1. Receive Query 2. . 3. .	<10 min	4 min	Medium Priority	Exceeded expectations	Database Manager	Mm/dd/yyyy	N/A
2	# errors in data	Find errors in data	1. . 2. .	0	2	High Priority	Unique naming error located	ETL Manager	Mm/dd/yyyy	Bug A
3	Access Restrictions	Catch unauthorized access	1. . 2. .	4	4	Medium Priority	Met Expectations	Database Manager	Mm/dd/yyyy	N/A

Implementation Plan:

Action Item (List steps required to implement)	Responsible (List person(s) responsible for action steps)	Due Date (When action items must be completed)
Pull all the data into Database	ETL Manager	Mm/dd/yyyy
Regularly pull data weekly for 3 months	ETL Manager	Mm/dd/yyyy
Transition queries to database	Database Manager	Mm/dd/yyyy

Conclusion

Motto

“You are not making a system more complex; you’re just realizing how complex the system really is”

-Richard Biehl

You made it!

Thank you

Backup Slides

Slowly Changing Dimensions

oType 0: Retain Original
- •No Changes
- •Date Table (Except holidays)
- •"Original" (e.g. original product name)
- •Very simple & easy to maintain

oType 1: Overwrite
- •Old attributes just overwritten
- •Only current state reflected
- •Very simple
- •No Fact Table needs to be modified

oType 1 Problems:
- •History is lost
- •Might affect/break existing queries

oWhen to use Type 1:
- •When making insignificant changes

oType 2: Additional Row
- •Perfectly partitions history
- •Changes reflected in history
- •NO updates in fact table
- •From that moment, new foreign key

oType 2 Problems:
- •Difficult to do queries such as "All current products?"
- •Not distinguishable so additional strategies much be implemented (SEE SOLUTIONS)

...	Category	Prev_Category
...	Sweets	Sweets
...	Biscuit	Sweets

oType 2 Solutions:
- •No. of products?: Still can use natural keys that exist and count distinct ID's to get # of products
- •Add columns "Effective Date", "Expired Date", and "Is_Current"
  1. e.g. 1/1/2020, 9/8/2023, No
  2. Steps: Add new Row, lookup in Dimension using the natural key and ef_/Ex_Date

oMixing Type 1 & 2
- •Use Type 1 for the columns that contain insignificant attributes, like product name, where history is not important.
- •Use Type 2 for the columns that contain majorly significant attributes, such as category, where history is important.

oWhen to mix Type 1 & Type 2
- •Note that the crossed out Chocolate would've been fully replaced with Candy but to show the replacement it is left in this example.

oType 3: Additional Attribute
- •In-between: You might do this if you want to be able to switch between versions
- •Instead of adding row - add a column

oType 3 Limitations:
- •Not suitable for frequent or unpredictable changes
  1. USE TYPE 2 FOR THAT
- •Not suitable for minor changes
  1. USE TYPE 1 FOR THAT

Primary Key	ID	Product Name	Category	Ef_Date	Ex_Date
1	SG	Glasses	A	1/1/2022	12/31/2100
2	CH	~~Chocolate~~ Candy	S	1/1/2022	12/31/2100
3	OT	Biscuits	S	1/1/2022	6/1/2022
4	OT	Biscuits	B	6/1/2022	12/31/2100

Dimensional Database

Product Dimension					
Key	ID	Category	Name	Selling Price	Cost to Make
1	1000	Skin Care	Retinol	$20	$10
2	2000	Jewelry	Earrings	$15	$5
3	3000	Art	Sunset Painting	$50	$10

Time Dimension				
Key	Hour	Minute	Seconds	Full Time
1	15	30	16	15:30:16
2	9	29	32	09:29:32

Fact Table							
Index	Order ID	Quantity	Revenue	Product	Customer	Time Ordered	Date Ordered
1	12001	2	$80	3	1	2	1
2	12002	1	$10	1	2	1	2
3	12002	1	$10	2	2	1	2

Customer Dimension				
Key	Name	Address	Phone Number	Email
1	John Smith	101 Smith St. Orlando, Fl 123456	321-123-9090	J.smith@gmail.com
2	Doe Richards	1301 Doe St. Rochester, NY 123456	509-321-7890	d.richards@gmail.com

Date Dimension				
Key	Year	Month	Day	Full Date
1	2023	11	15	11/15/2023
2	2024	01	20	01/20/2024

Dimensional Database

Product Dimension

Key	ID	Category	Name	Selling Price	Cost to Make
1	1000	Skin Care	Retinol	$20	$10
2	2000	Jewelry	Earrings	$15	$5
3	3000	Art	Sunset Painting	$50	$10

Time Dimension

Key	Hour	Minute	Seconds	Full Time
1	15	30	16	15:30:16
2	9	29	32	09:29:32

Fact Table

Index	Order ID	Quantity	Revenue	Product	Customer	Time Ordered	Date Ordered
1	12001	2	$80	3	1	2	1
2	12002	1	$10	1	2	1	2
3	12002	1	$10	2	2	1	2

Customer Dimension

Key	Name	Address	Phone Number	Email
1	John Smith	101 Smith St. Orlando, Fl 123456	321-123-9090	J.smith@gmail.com
2	Doe Richards	1301 Doe St. Rochester, NY 123456	509-321-7890	d.richards@gmail.com

Date Dimension

Key	Year	Month	Day	Full Date
1	2023	11	15	11/15/2023
2	2024	01	20	01/20/2024