UNTIL THE PAPERWORK IS DONE
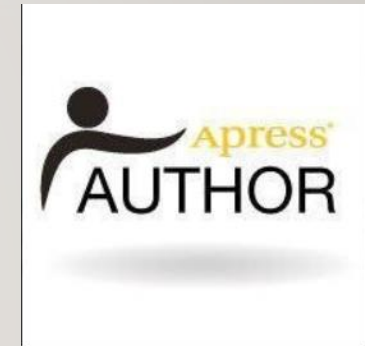
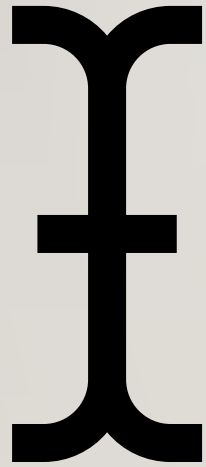DATABASE DOCUMENTATION AND WHY WE NEED IT

About me –

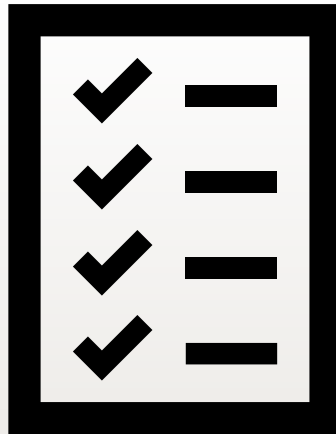❖ **SQL Server DBA/Data Engineer since 1996**

❖ **PASSion Award Winner**

❖ **Co-Lead – Data Platform WIT/TriPASS SQL User Group**

❖ **Book author with Apress**
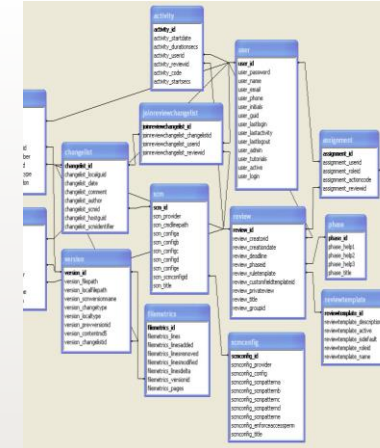
❖ **http://curiousaboutdata.com**

❖ **@sqlmal**

- What constitutes documentation?
- Why document?
- What if there is no documentation?
- What are types/categories of documentation?
- What are uses/purpose of each?
- How do we get started?

What counts as documentation?

Extended Properties   Code Comments

EFFORT

BUT WHAT IT IS…

"Nobody could read his code, so we could not fire him."
Joe Celko's SQL For Smarties, 5th Ed.

" If you can't be replaced, you can't be promoted "

Documentation can get you promoted!!

WHAT IF YOU DON'T DOCUMENT?

- Impact Analysis becomes a bear
  - 'Resident experts' and what happens when they leave
  - 'Deleting just one record'
  - 'Updating just that one value'
  - 'The code change that broke everything'
- Environment of fear/mistrust
  - ('Nobody knows what that is, let's leave it alone')

EXAMPLE 1: THE 'UPDATEDATE' SITUATION

Invoice_Header
- InvoiceNum
- InvoiceDate
- CustomerID
- InvoiceAmount
- CreateDate
- UpdateDate

Is there a timelimit on updating old invoices?

Is UpdateDate populated for new invoices?

Is UpdateDate updated for every update to every field?

Data Dictionary			
Field	DataType	Nullable	What this means
InvoiceNo	INT	No	Unique invoice number
InvoiceDate	DATETIME	No	Date customer is invoiced
CustomerId	INT	No	Unique customer id, foreign keyed to customer table
InvoiceAmount	NUMERIC	No	Amount of invoice, > 0 and less than 10,0000
CreateDate	DATETIME	No	Date invoice was created in system
UpdateDate	DATETIME	No	Date invoice was changed/updated for invoiceamount, only changeable 90 days after creation

EXAMPLE 2: 'FIND ALL ORDERS WITH SHORTEST GAP BETWEEN SHIP DATE AND DUE DATE'
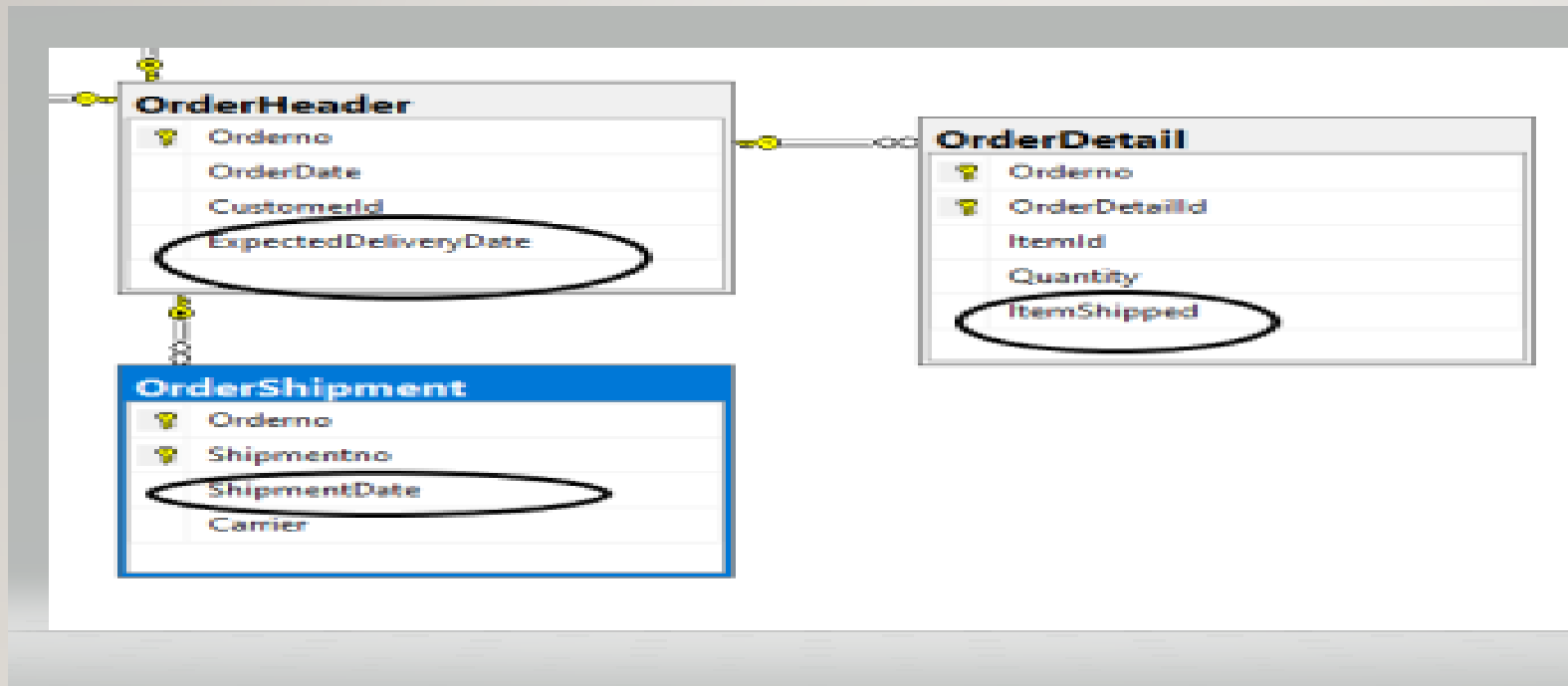

Data Dictionary extract:
- Expected delivery date is also called due date.
- Due date is latest date the order can be completed.
- Each item can ship individually.
- Order is not considered fulfilled until all items have been delivered.

EXAMPLE 3: WEB PAGE IS TIMING OUT!!

- What is feeding this page?

- Is there any programmer around?

- Which server should I look at?

https://www.w3computing.com/systemsanalysis/developing-physical-data-flow-diagrams/
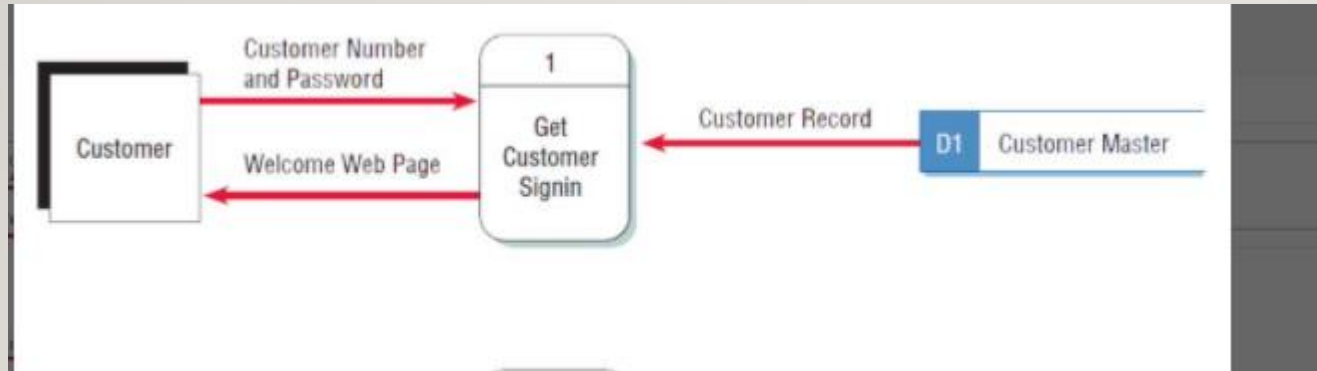
WHAT IF I DO DOCUMENT?

- Impact analysis is easier

- Single source of truth

- Historical context of change (why these values are not in a separate table, why this index hint was added,…)

- New employees get up to speed quickly

# WHY ARE DATABASES DIFFICULT TO DOCUMENT?

- Constant flux/change and difficult to keep up with

- Where to start?!?

- Who will be my audience

- Who will maintain this

- What type of documentation works best for us

TYPES OF DOCUMENTATION

Prescriptive	Descriptive
Logical ER Diagrams	Physical ER Diagrams/Reverse Engineered diagrams
Logical Data Flow Diagrams	Physical Data Flow Diagrams
Data Requirements	Data Dictionary

LOGICAL DATA FLOW DIAGRAMS

- Describes flow of data through system to perform business functionality.

- Focused on business events

- Easy tool to communicate with users

- Independent of technology/technical terms

- Uses 4 symbols – External entities, Processes, Data Stores and Data Flows
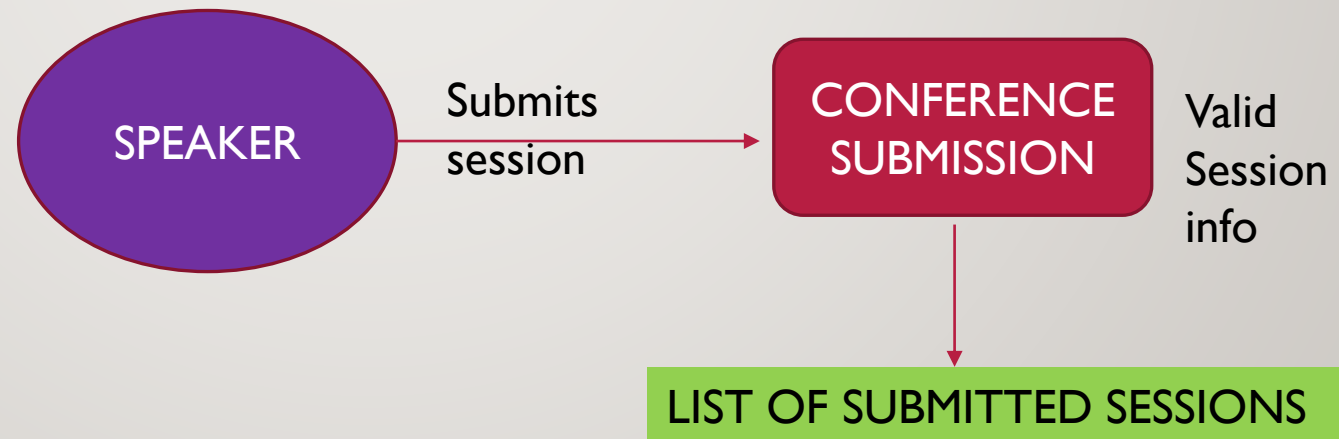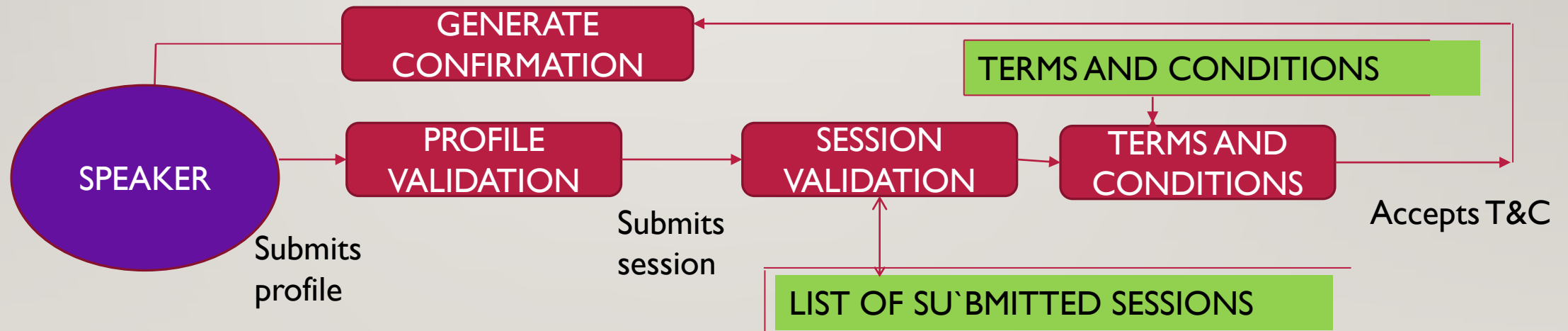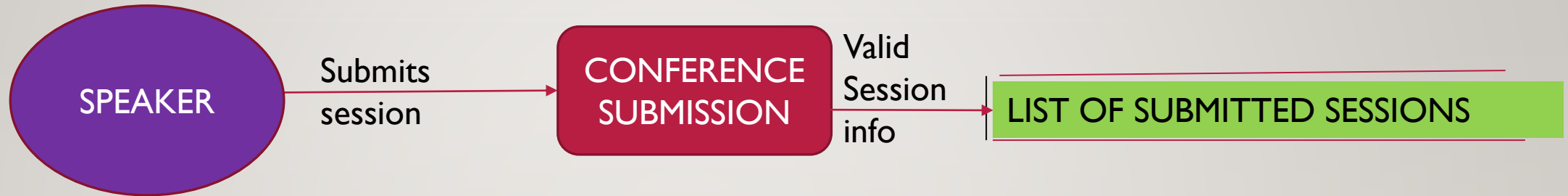
PHYSICAL DATA FLOW DIAGRAMS

- Shows how the system will be implemented – includes data, people, files,sources etc.

- Has a lot more detail than logical data flow diagram

- Shows sequence accurately and temporary data storage if any.

- Useful to tech teams

Tools

- Visio

- LucidChart

- Visual-Paradigm.com

EXAMPLE OF LOGICAL VERSUS PHYSICAL DATA FLOW DIAGRAM

LOGICAL ER DIAGRAMS

- Detailed Picture of entities needed for a business

- Information gathered from business requirements

Levels

1. Entity Relationship Diagram

2. Key Based Model

3. Fully Attributed Model

Tools

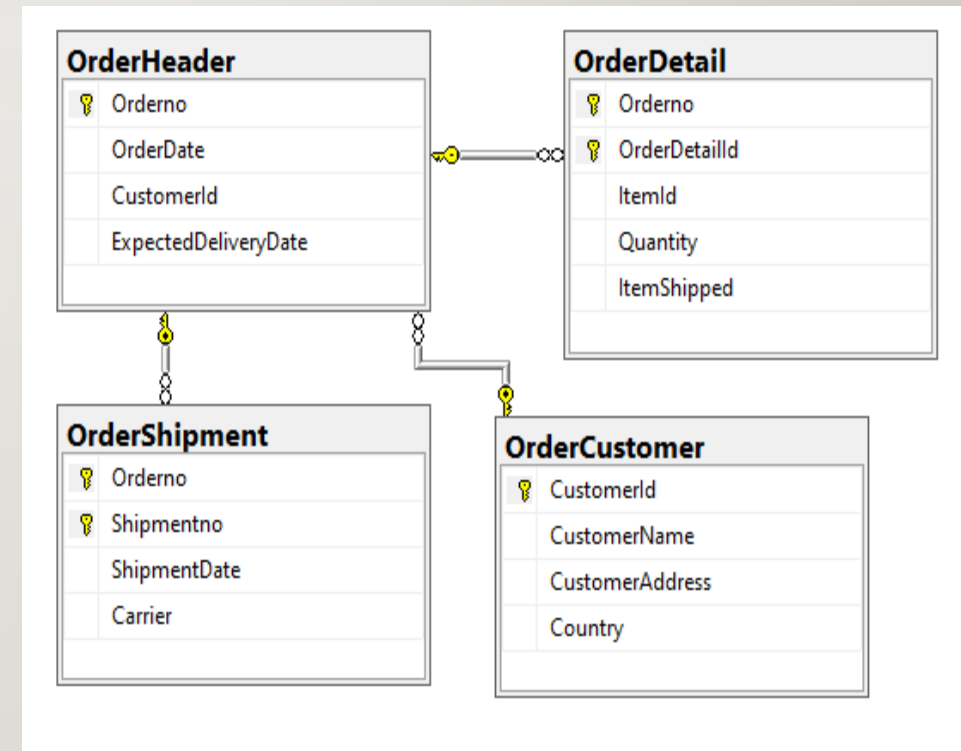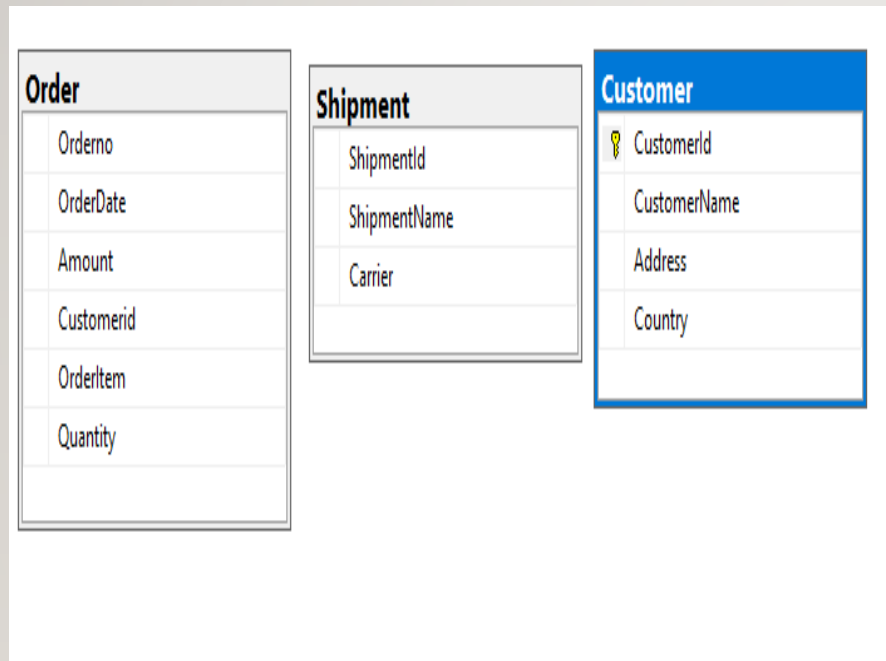- Visual Paradigm

- LucidChart

- Toad Data Modeler

- ERWin

PHYSICAL ER DIAGRAMS

- Design of the physical database

- Includes indexes,constraints and other physical properties

- Can be reverse engineered from an existing database

- Supports the needs of a DBA/Database Engineers

Tools

- ERWin

- Visio

- SQL Server SSMS Database Diagrams

LOGICAL VERSUS PHYSICAL ER DIAGRAMS

DATA DICTIONARY

- Data definition matrix

- Names, definitions and attributes of all data elements

- Helps maintain consistency of data elements

- Easy for in depth analysis of data

Tools

- ApexSQL Doc

- RedGate SQL Doc

- SentryOne DocXPress

DATA REQUIREMENTS

- Identify, formulate and validate data needed for business objectives

- For knowledgeable people to determine what data is needed

- Used to determine ownership of data elements

- Organization specific format

Logical is what we meant to do.
Physical is what *actually* happened.

IS IT WORTH THE TIME?

WHERE DO I START?

WHO IS YOUR PRIMARY AUDIENCE?

WHAT ARE THEIR GOALS?

IDENTIFY QUICK WINS

IDENTIFY RESOURCES

WHO IS YOUR PRIMARY AUDIENCE?

Audience	Goals	Documentation
'Just me'		
Business side	Identify how data flows through system/do better impact analysis	Logical Data Flow Diagram
Developers	Find out how tables are connected/do better impact analysis	Physical ERD/Data Dictionary
DBAs	Find what is feeding application-web page/do better impact analysis	Physical Data Flow Diagram/Physical ERD/Data Dictionary

QUICK WINS - DEVELOPING A CYCLE OF GROWTH

GATHERING QUICK WINS

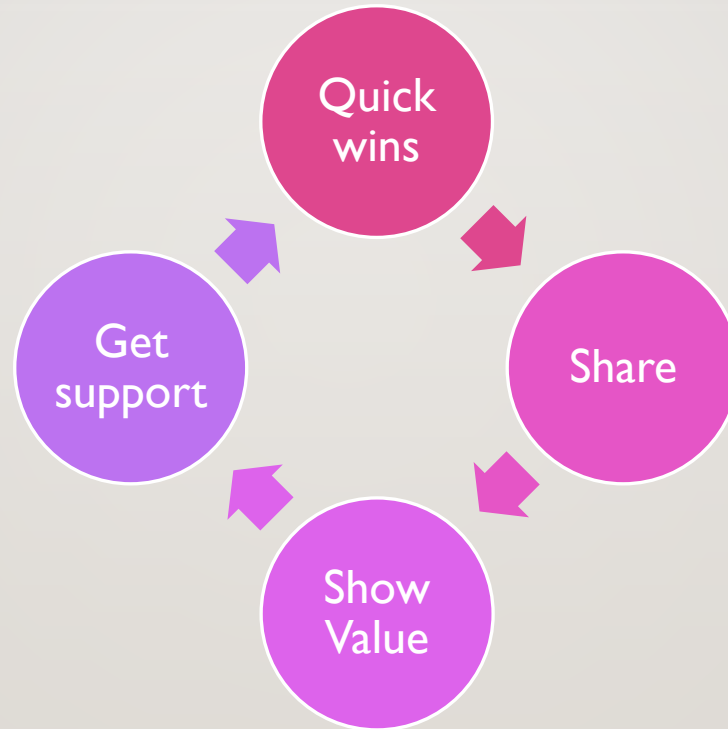|																							
Frequent																							
Semi-frequent																							
Critical																							
Vacation																							
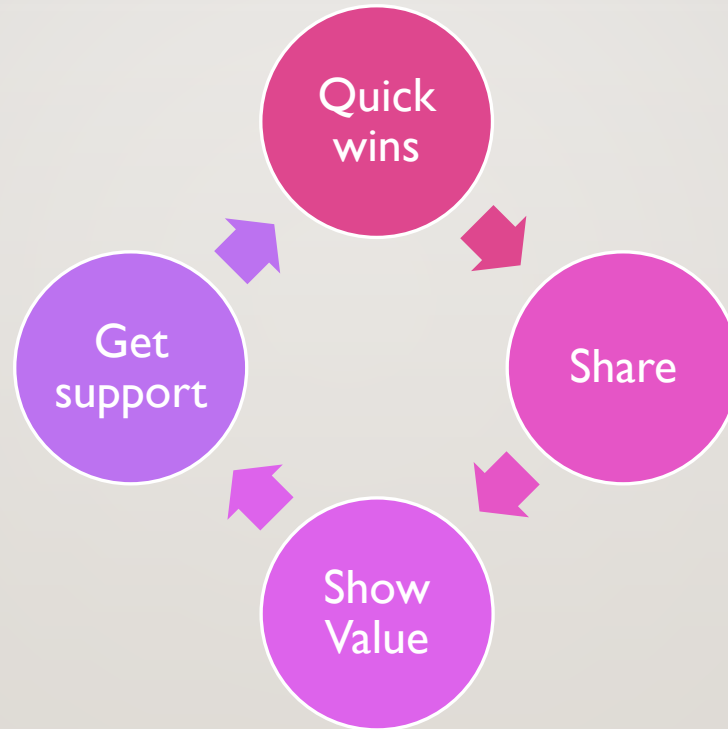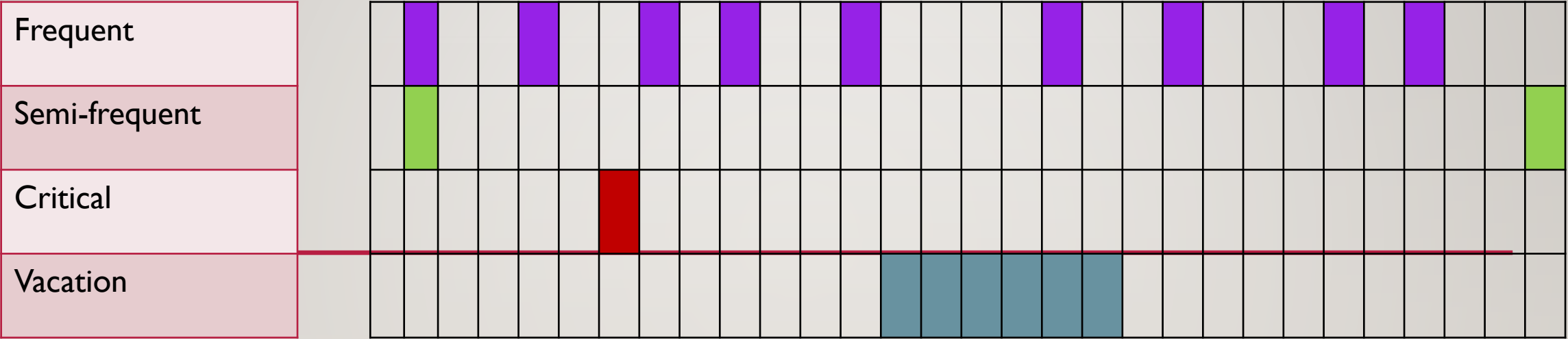
QUICK WINS - DEVELOPING A CYCLE OF GROWTH

WHAT RESOURCES DO I HAVE?

Free

Resources available with SQL Server

DMV Queries to build Data Dictionary

Comments in Code

Paid but part of bigger suite $$

Visio for Reverse Engineering/Flow Diagrams

RedGate SQL Doc – part of RedGate Toolbelt

Exclusively Documentation $$$

Apex SQLDoc

Dataedo

SentryOne Doc XPress

SQL SERVER RESOURCES – PROS AND CONS

Resource	Pros	Cons
Extended Properties	Free	Clunky, difficult to browse
	Metadata and data reside together	
	Easy to access	
SSMS Database Diagrams	Free	Limited in size
	Can Reverse Engineer	Difficult to print
	Auto update	Cannot export

DEMO: DATA DICTIONARY SCRIPT

- https://curiousaboutdata.com/2021/03/12/data-dictionary-script/

SUMMARY

- Start small – use maintainable, free tools to begin with

- Be consistent in recording and showing results

- Be careful with labels ('they do documentation') – it is everyone's job

- Make it part of the CI process.

REFERENCES

- https://www.visual-paradigm.com/tutorials

- https://erwin.com/bookshelf/public_html/2020R2/Content/User%20Guides/erwin%20Help/Comparison_of_Logical_and_Physical_Model_Objects.html

- https://www.writethedocs.org/

- https://www.red-gate.com/simple-talk/sql/database-delivery/scripting-description-database-tables-using-extended-properties/

THANK YOU!