

# CDW: A Conceptual Overview

by Margaret Gonsoulin, PhD

January 12, 2015



# Thanks to:

- Richard Pham, BISL/CDW for his mentorship
- Heidi Scheuter and Hira Khan for organizing this session
- Chandrea Culbreath and Dharani Ramanathan for providing feedback in my practice sessions

# Poll #1: Your CDW experience

- What is your level of experience with CDW data?
  - 1- Not worked with it at all
  - 2
  - 3
  - 4
  - 5- Very experienced with CDW data

# Agenda for Today

- Get to the bottom of all of those acronyms!
- Learn to think in “relational data” terms
- Become familiar with the components of CDW
  - Production and Raw Domains
  - Fact and Dimension tables/Views
- Understand how to return to an analytic dataset
  - Primary and Foreign Keys
  - Joining Tables/Views

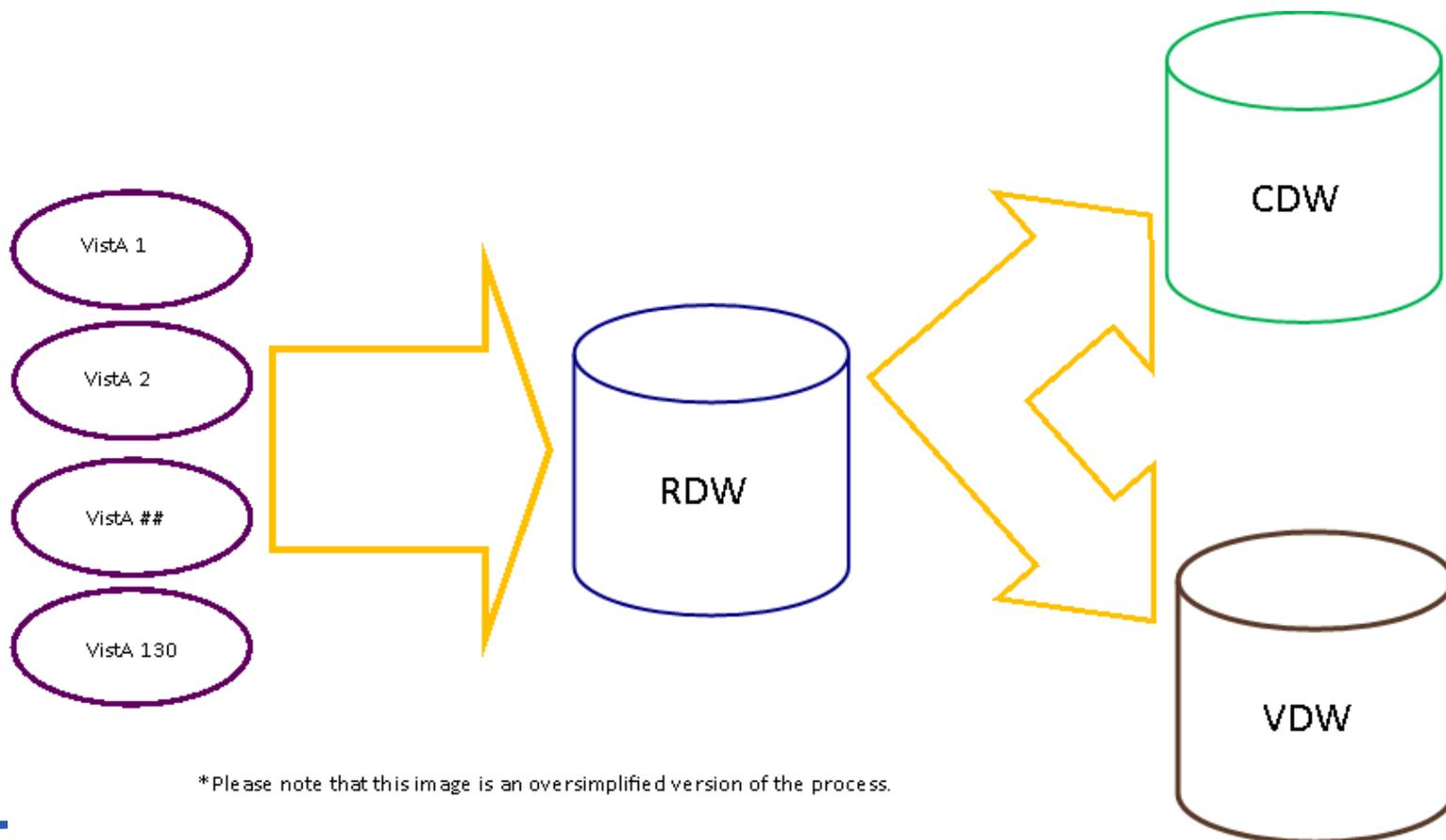
# Agenda for Today

- **Get to the bottom of all of those acronyms!**
- Learn to think in “relational data” terms
- Become familiar with the components of CDW
  - Production and Raw Domains
  - Fact and Dimension tables/Views
- Arriving at an analytic dataset
  - Primary and Foreign Keys
  - Joining Tables/Views

# “C”DW, “R”DW & “V”DW

- Users will see documentation referring to xDW.
- The “x” is a variable waiting to be filled in with either:
  - “V” for VISN,
  - “R” for region or
  - “C” for corporate (meaning “national VHA”)
- Each organizational level of the VA has its own data warehouse focusing on its own population.
- This talk focuses on CDW only.

# Flow of data into the warehouse

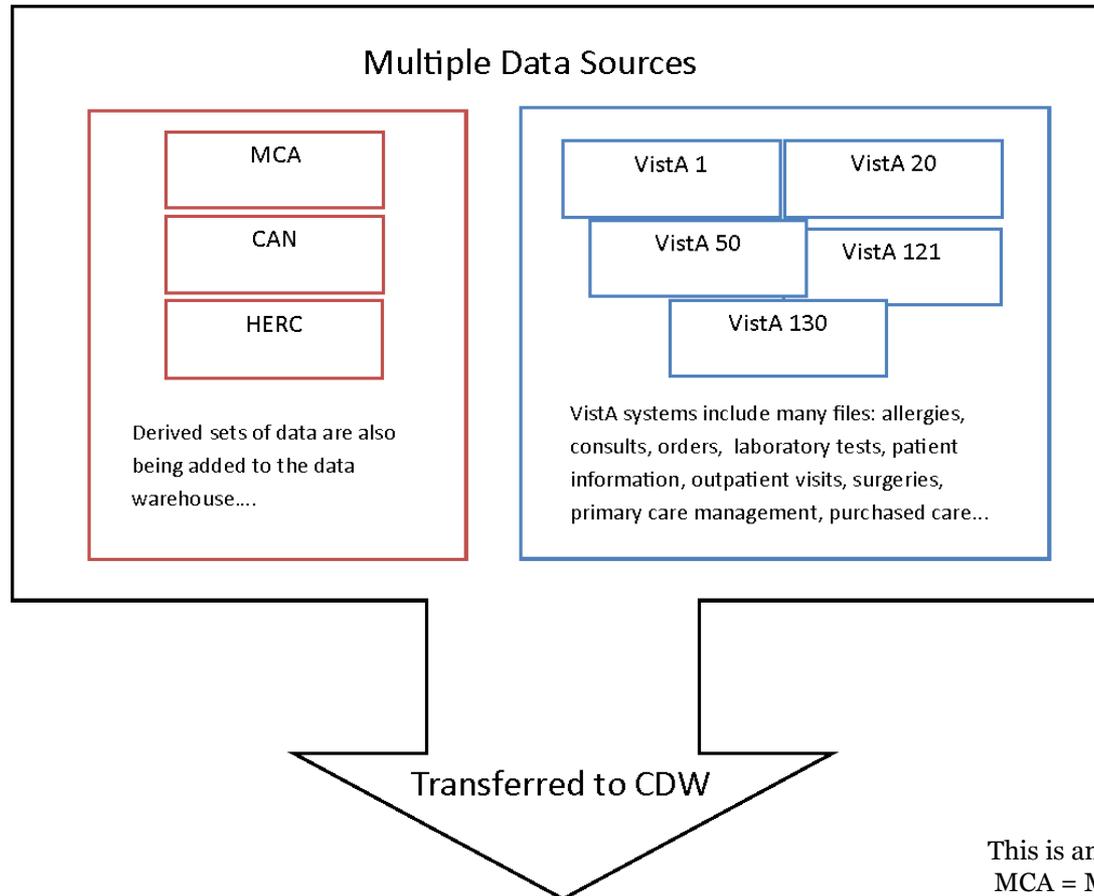


\*Please note that this image is an oversimplified version of the process.

# C“DW”

- The “DW” in CDW stands for “Data Warehouse.”
- Data Warehouse = a data delivery system intended to give users the information they need to support their business decisions.
- In ordinary terms, it is a large storage facility for “big data.”

# What type of data get stored here?



\*notes\*

This is an oversimplified description.

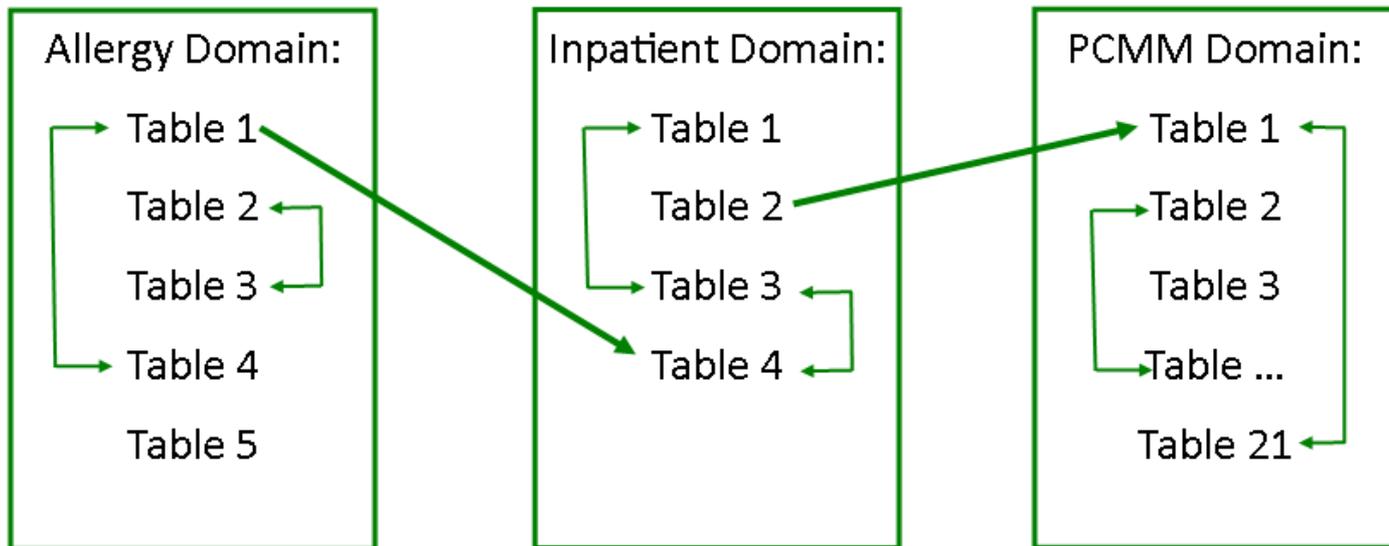
MCA = Managerial Cost Accounting

CAN = Care Assessment Need

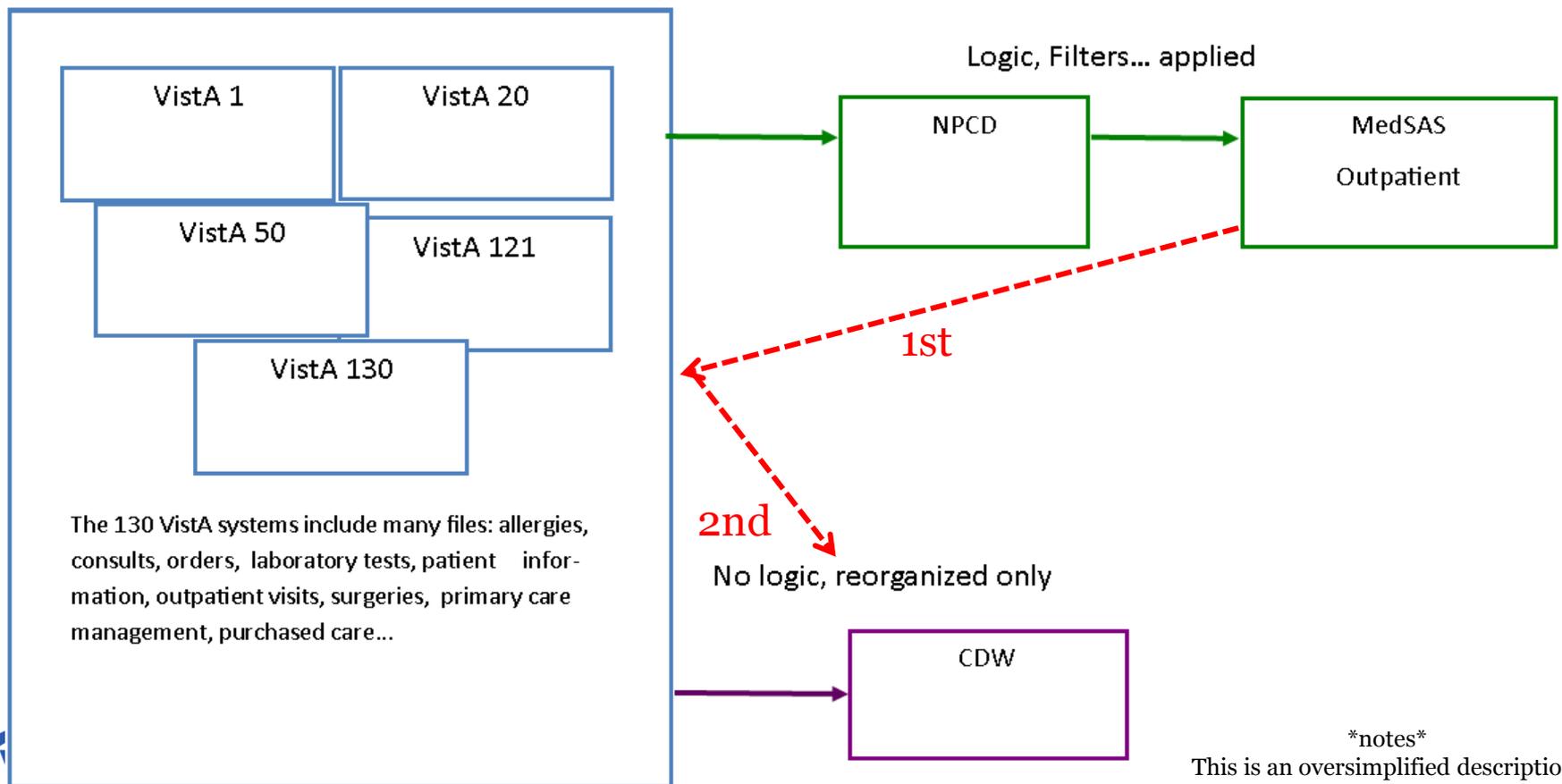
HERC = Health Economics Resource Center

# How the CDW organized?

CDW organizes data into data domains  
with multiple tables in each domain  
and builds linking keys to connect tables and domains:



# How does CDW fit into VA Data?



\*notes\*

This is an oversimplified description.  
 NPCD = National Patient Care Database  
 MedSAS = Medical SAS

# Agenda for Today

- Get to the bottom of all of those acronyms!
- **Learn to think in “relational data” terms**
- Become familiar with the components of CDW
  - Production and Raw Domains
  - Fact and Dimension tables/Views
- Understand how to return to an analytic dataset
  - Primary and Foreign Keys
  - Joining Tables/Views

# C“D”W- The data.

- Data stored in this warehouse are in a “relational” format.
- In ordinary terms, this is data that has been separated out into multiple tables that look like spreadsheets.
- Linking keys are added into these tables so that users may reassemble the tables for analytic use.
- SQL programming language can be used to reassemble the tables.

# A common database, a “flat file”

| LastName | FirstName | Address                | CityState | Zipcode | ICD9   | Sta3n | VisitDate |
|----------|-----------|------------------------|-----------|---------|--------|-------|-----------|
| Jones    | Marianna  | 123 Oak St             | Bee, AR   | 70788   | 110.6  | 578   | 1/2/2014  |
| Frank    | Josie     | 11 Pine Ave            | Flip, OK  | 30032   | 377.75 | 358   | 2/2/2014  |
| Plank    | Bill      | 230 5 <sup>th</sup> St | Miner, TX | 11201   | 202.05 | 402   | 3/3/2014  |

There are three main types of information in this “flat file”

1. Patient information (name and address)
2. Diagnosis the led to the visit (ICD9 code)
3. Information about the visit (station number and date)

# Transformed into a Relational Database

## PatientTable

| PatientKey | LastName | FirstName | Address                | CityState | Zipcode |
|------------|----------|-----------|------------------------|-----------|---------|
| 1          | Jones    | Marianna  | 123 Oak St             | Bee, AR   | 70788   |
| 2          | Frank    | Josie     | 11 Pine Ave            | Flip, OK  | 30032   |
| 3          | Plank    | Bill      | 230 5 <sup>th</sup> St | Miner, TX | 11201   |

## DiagnosisTable

| DxKey | ICD9   |
|-------|--------|
| 1     | 110.6  |
| 2     | 202.05 |
| 3     | 280.8  |
| 4     | 377.75 |

## VisitTable

| VisitKey | DxKey | PatientKey | Sta3n | Date     |
|----------|-------|------------|-------|----------|
| 1        | 1     | 1          | 578   | 1/1/2014 |
| 2        | 4     | 2          | 358   | 2/2/2014 |
| 3        | 2     | 3          | 402   | 3/3/2014 |

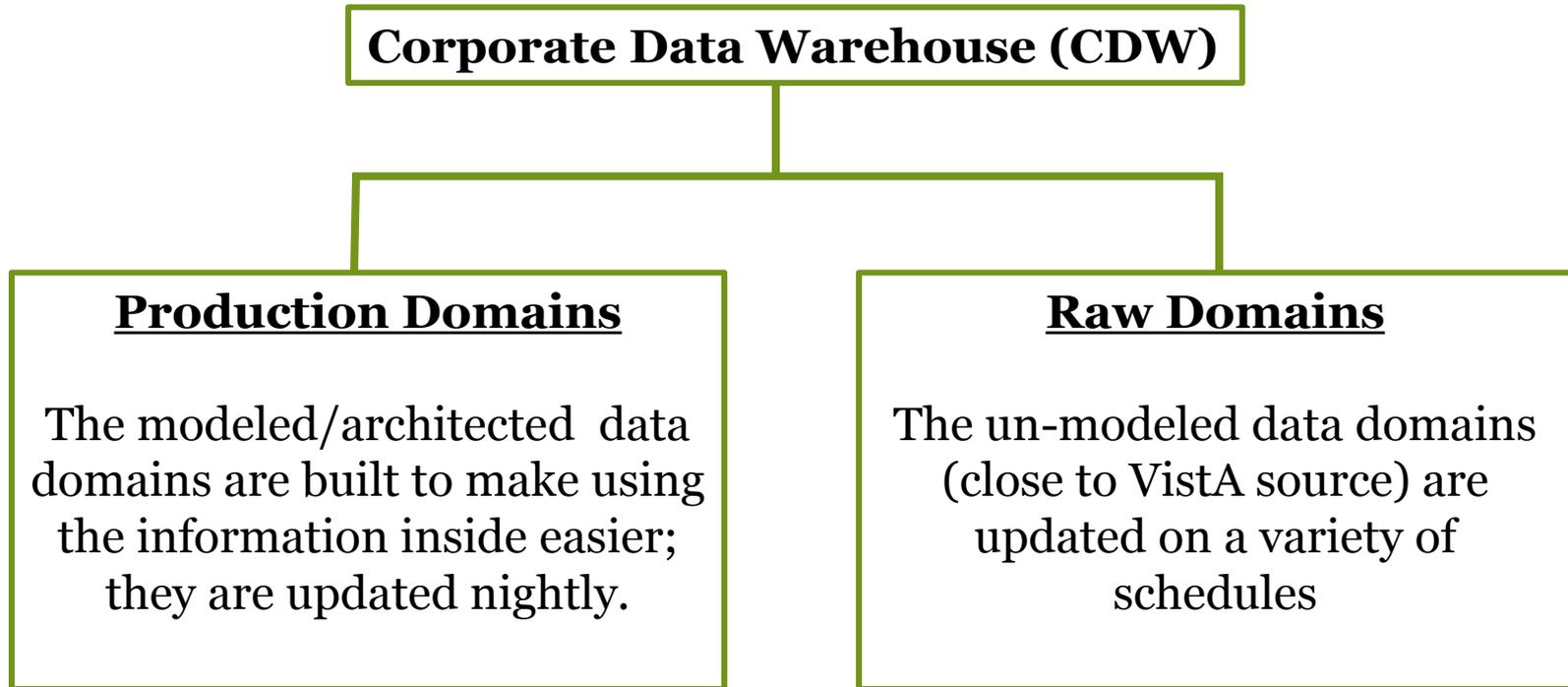
## Poll #2: About you

- What is your role in VA research? (check all that apply)
  - Research investigator/PI
  - Data manager/analyst
  - Project coordinator
  - Other

# Agenda for Today

- Get to the bottom of all of those acronyms!
- Learn to think in “relational data” terms
- **Become familiar with the components of CDW**
  - Production and Raw Domains
  - Fact and Dimension tables/Views
- Understand how to return to an analytic dataset
  - Primary and Foreign Keys
  - Joining Tables/Views

## 2 Parts of CDW (Raw and Production)



# Domains

Groups of tables based on the subject matter found within the group of tables

- Production Domains:
  - Outpatient 2.1
  - Lab Microbiology 1.0
  - Mental Health 1.0
  - Patient 2.0
  - Vital Sign 1.0
  - Patient Insurance 1.0
- Raw Domains
  - Intravenous Meds
  - Oncology
  - Prosthetics
  - Bill Claims
  - Radiology

# Tables v. Views

- A table is a set of columns and rows that contain data elements.
- A view is the result of a procedure that pulls information out of a database into a virtual table; in simple terms it is a “virtual” table.

| Sta3n | State        | County    | Country       | PeriodOfService    | MaritalStatus | InsuranceCoverageFlag | Religion                    |
|-------|--------------|-----------|---------------|--------------------|---------------|-----------------------|-----------------------------|
| 676   | WISCONSIN    | MILWAUKEE | UNITED STATES | PERSIAN GULF WAR   | NEVER MARRIED | N                     | LUTHERAN                    |
| 629   | ARKANSAS     | JEFFERSON | UNITED STATES | VIETNAM ERA        | DIVORCED      | U                     | BAPTIST                     |
| 586   | MISSISSIPPI  | TIPPAH    | UNITED STATES | PERSIAN GULF WAR   | MARRIED       | U                     | UNKNOWN/NO PREFERENCE       |
| 629   | TEXAS        | SMITH     | UNITED STATES | POST-VIETNAM       | SEPARATED     | U                     | BAPTIST                     |
| 636   | NEBRASKA     | LINCOLN   | UNITED STATES | POST-KOREAN        | *Missing*     | U                     | *Missing*                   |
| 623   | OKLAHOMA     | TULSA     | UNITED STATES | PERSIAN GULF WAR   | NEVER MARRIED | Y                     | UNKNOWN/NO PREFERENCE       |
| 676   | PENNSYLVANIA | LANCASTER | UNITED STATES | POST-VIETNAM       | NEVER MARRIED | NULL                  | ROMAN CATHOLIC CHURCH       |
| 623   | OKLAHOMA     | TULSA     | UNITED STATES | VIETNAM ERA        | MARRIED       | Y                     | BAPTIST                     |
| 676   | WISCONSIN    | MONROE    | UNITED STATES | PERSIAN GULF WAR   | MARRIED       | Y                     | BAPTIST                     |
| 636   | CALIFORNIA   | CALAVERAS | UNITED STATES | PERSIAN GULF WAR   | MARRIED       | U                     | CHRISTIAN (NON-SPECIFIC)    |
| 676   | WISCONSIN    | PORTAGE   | UNITED STATES | OTHER NON-VETERANS | *Missing*     | NULL                  | *Missing*                   |
| 674   | TEXAS        | ANDERSON  | UNITED STATES | WORLD WAR II       | WIDOWED       | Y                     | UNKNOWN/NO PREFERENCE       |
| 695   | WISCONSIN    | WAUKESHA  | UNITED STATES | VIETNAM ERA        | MARRIED       | Y                     | *Missing*                   |
| 674   | TEXAS        | HARRIS    | UNITED STATES | VIETNAM ERA        | DIVORCED      | U                     | PROTESTANT, NO DENOMINATION |

# 1. Dimension Tables

- These tables are used as supporting tables; each one holds a specific type of information that is meant to be accessed repeatedly
- These tables
  - do not contain patient information
  - can be viewed with basic read access
  - are relatively small in size

## 2. Fact Tables

- These tables hold measurements
- They tend to be very large (up to billions of records)
- They have patient and staff identifiers in them
- Therefore,
  - It is necessary to request permission to view subsets of these tables
  - It is important to plan your cohort to reduce the size of your request

# Fact tables and Dimension tables

## PatientTable (Fact)

| PatientKey | LastName | FirstName | Address                | CityState | Zipcode |
|------------|----------|-----------|------------------------|-----------|---------|
| 1          | Jones    | Marianna  | 123 Oak St             | Bee, AR   | 70788   |
| 2          | Frank    | Josie     | 11 Pine Ave            | Flip, OK  | 30032   |
| 3          | Plank    | Bill      | 230 5 <sup>th</sup> St | Miner, TX | 11201   |

## DiagnosisTable (Dim)

| DxKey | ICD9   |
|-------|--------|
| 1     | 110.6  |
| 2     | 202.05 |
| 3     | 280.8  |
| 4     | 377.75 |

## VisitTable (Fact)

| VisitKey | DxKey | PatientKey | Sta3n | Date     |
|----------|-------|------------|-------|----------|
| 1        | 1     | 1          | 578   | 1/1/2014 |
| 2        | 4     | 2          | 358   | 2/2/2014 |
| 3        | 2     | 3          | 402   | 3/3/2014 |

# Prefixes or Schema / Type of table

- Each table name is preceded by a prefix or schema used to indicate whether or not it is a:
  - A fact table (e.g., Schema.TableName)
  - A dimension table (always a “dim.” schema)

| Fact Tables           | Dimension Tables  |
|-----------------------|-------------------|
| Dental.DentalAlerts   | Dim.OrderableItem |
| CPRSOrder.OrderedItem | Dim.AdmitSource   |
| Immun.Immunization    | Dim.Ethnicity     |
| Patient.Patient       | Dim.Antibiotic    |

# Two types of tables/views

## Health Factor 2.1 Domain

HF.HealthFactor

Dim.HealthFactorType

Fact table

Dimension table

# HF and Dim tables in Health Factors

## HF.HealthFactor Content

- Patient identifiers
- Staff identifiers
- Descriptions of the health factor type (such as “colonoscopy” and “flu vaccine done elsewhere”)
- Comments related to the health factor (such as “biopsy of polyp - benign” and “normal”).

## Dim.HealthFactorType Content

- Whether the health factor is a “factor” or a “category”
- Whether or not the health factor is gender specific
- Text describing the factors in words (e.g., “current smoker,” “weight management,” or “abuse/neglect”)

# Agenda for Today

- Get to the bottom of all of those acronyms!
- Learn to think in “relational data” terms
- Become familiar with the components of CDW
  - Production and Raw Domains
  - Fact and Dimension tables/Views
- **Understand how to return to an analytic dataset**
  - Primary and Foreign Keys
  - Joining Tables/Views

# Joining Keys

- Primary Key –A column in every table that uniquely identifies each row.
- Foreign Key –These are column(s) in a table that correspond to or reference a primary key in another table.
  - Values of the foreign key may repeat
  - Names of foreign keys may differ from names of their matching primary keys

# Joining Keys

## PatientTable (Fact)

| PatientKey (PK) | LastName | FirstName | Address                | CityState | Zipcode |
|-----------------|----------|-----------|------------------------|-----------|---------|
| 1               | Jones    | Marianna  | 123 Oak St             | Bee, AR   | 70788   |
| 2               | Frank    | Josie     | 11 Pine Ave            | Flip, OK  | 30032   |
| 3               | Plank    | Bill      | 230 5 <sup>th</sup> St | Miner, TX | 11201   |

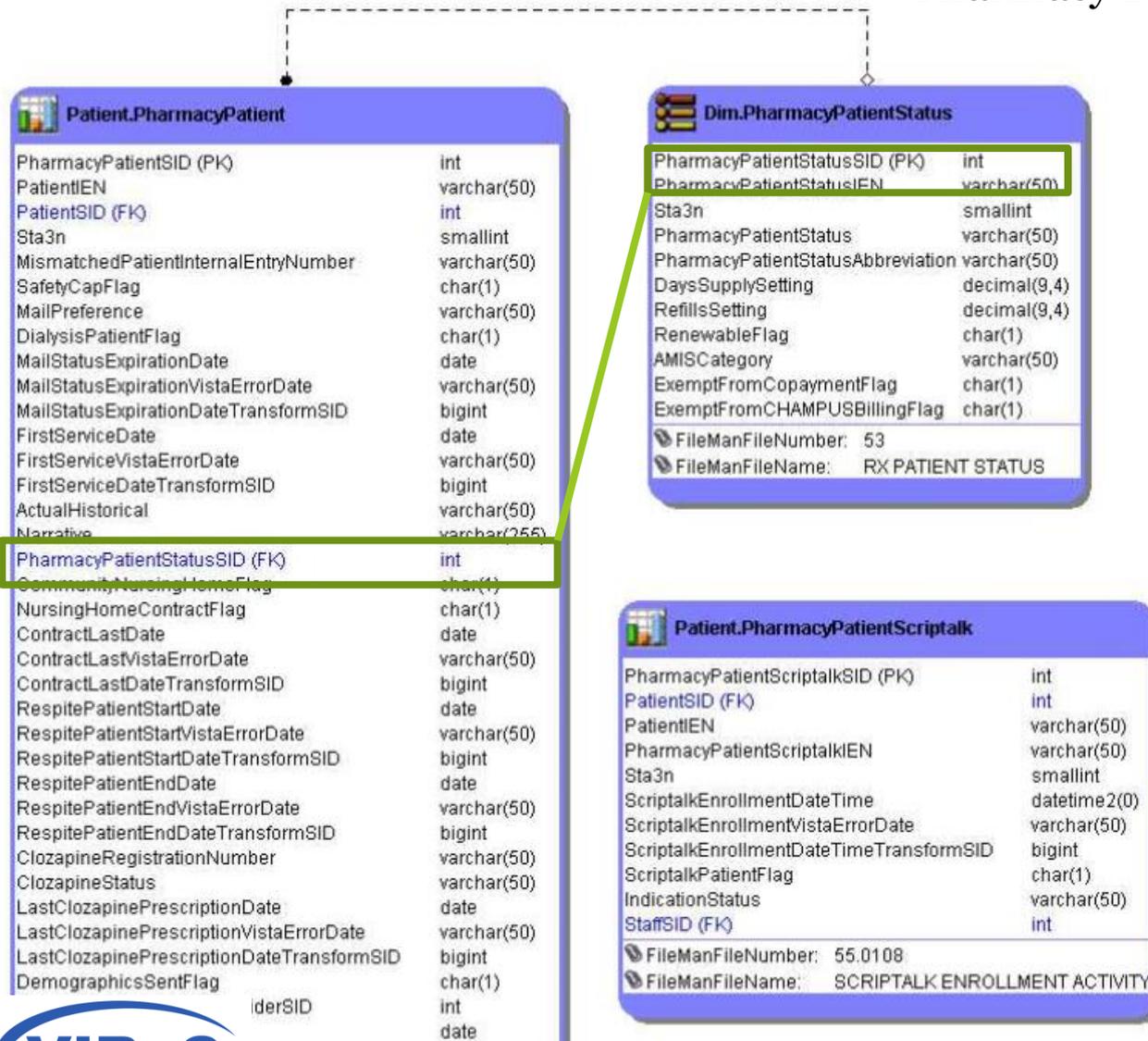
## DiagnosisTable (Dim)

| DxKey (PK) | ICD9   |
|------------|--------|
| 1          | 110.6  |
| 2          | 202.05 |
| 3          | 280.8  |
| 4          | 377.75 |

## VisitTable (Fact)

| VisitKey (PK) | DxKey (FK) | PatientKey (FK) | Sta3n | Date     |
|---------------|------------|-----------------|-------|----------|
| 1             | 1          | 1               | 578   | 1/1/2014 |
| 2             | 4          | 2               | 358   | 2/2/2014 |
| 3             | 2          | 3               | 402   | 3/3/2014 |

# Pharmacy Patient Domain



**Entity Relationship (ER) Diagrams** are drawings of the relationships between the tables in the domain.

(PK) indicates a Primary Key

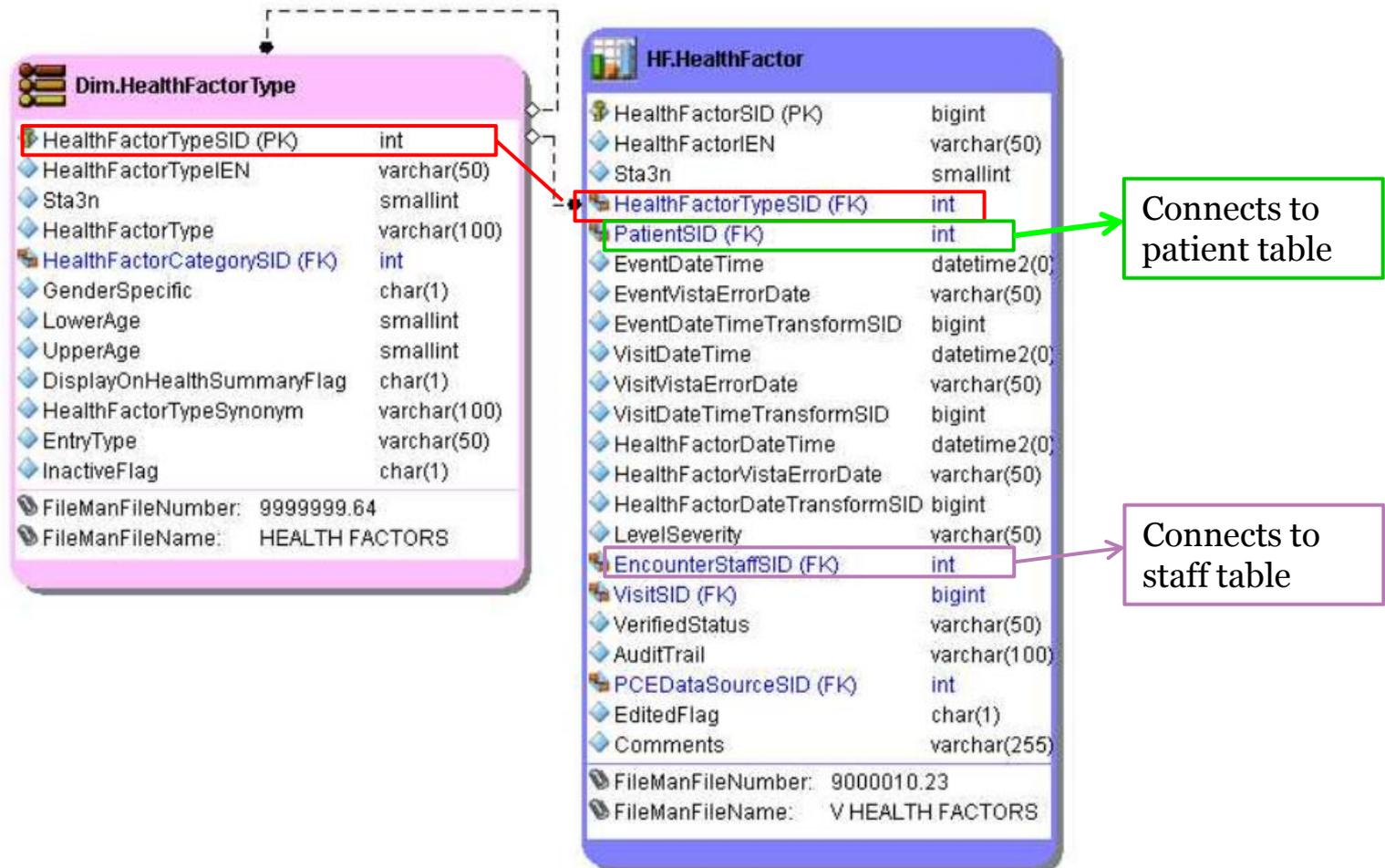
(FK) indicates a Foreign Key

These two join on *PharmacyPatientStatusSID*

# The Patient Domain (connections within a domain)



## The Health Factors Domain (connections inside & outside of domain)



# Another method of identifying joining keys

Department of Veterans Affairs  
Corporate Data Warehouse

CDW Home ▶ CDW HOME

CDW Home | CDW Support ▾ | Community ▾ | Domain Teams ▾ | **MetaData ▾** | TechTeam ▾

### NEW TO CDW?

Are you getting started with the Corporate Data Warehouse (CDW)?

- Intro and Policies
- CDW Support
- CDW Metadata**

### WHAT'S IN THE WORKS?

- General Announcements
- Training Announcements
- CDW Domain Status and Priority (Excel)

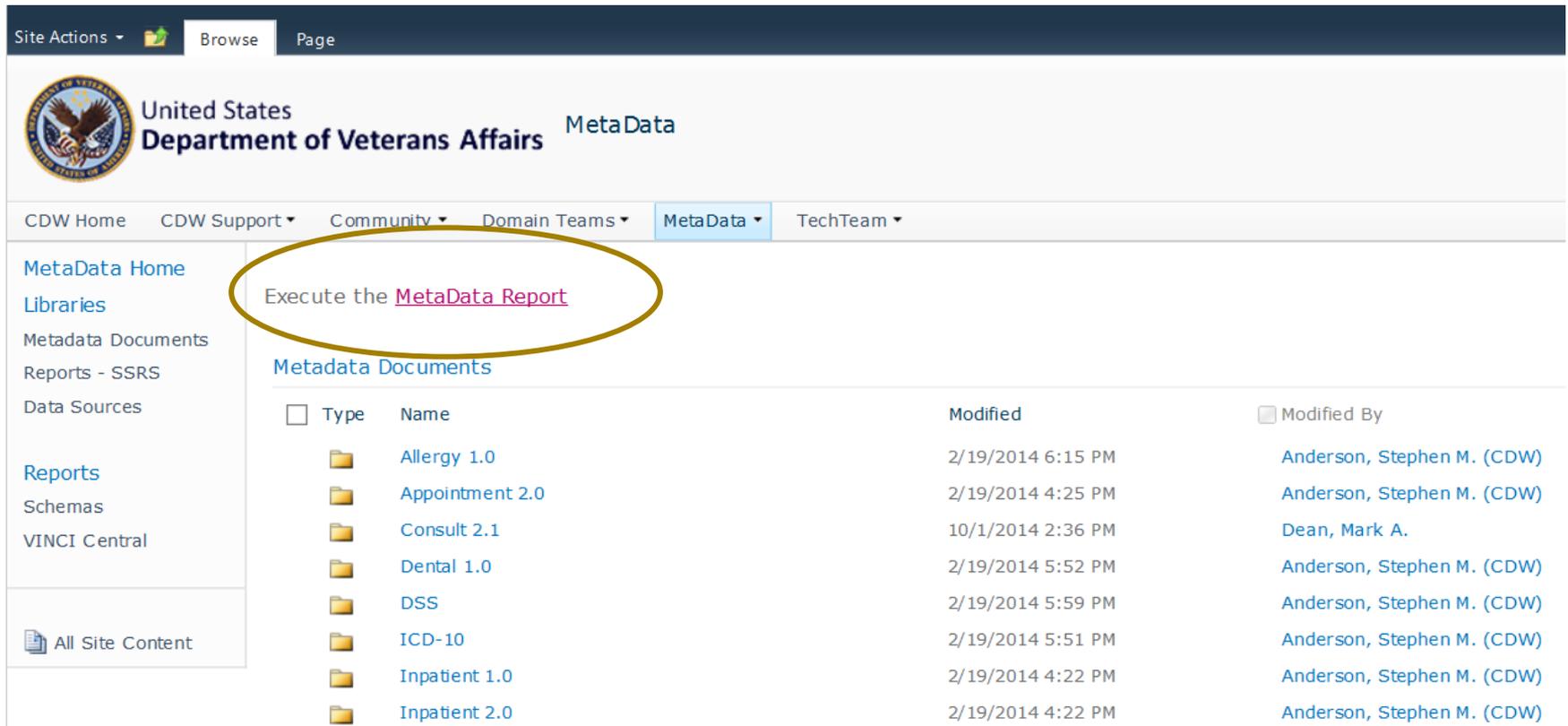
### EXTERNAL LINKS

- HSR&D Listserv
- Data Quality

### DATA ACCESS

**Do you want to get connected to CDW?** There are many types of access to CDW. The best thing to do before you start is to define what kind of access you need.

# Click “execute the metadata report”



Site Actions  Browse Page

 United States Department of Veterans Affairs Metadata

CDW Home CDW Support  Community  Domain Teams  **Metadata**  TechTeam 

Metadata Home  
Libraries  
Metadata Documents  
Reports - SSRS  
Data Sources

Reports  
Schemas  
VINCI Central

 All Site Content

Execute the [Metadata Report](#)

Metadata Documents

| <input type="checkbox"/> | Type  | Name            | Modified          | <input type="checkbox"/> Modified By |
|--------------------------|---|-----------------|-------------------|--------------------------------------|
| <input type="checkbox"/> |    | Allergy 1.0     | 2/19/2014 6:15 PM | Anderson, Stephen M. (CDW)           |
| <input type="checkbox"/> |    | Appointment 2.0 | 2/19/2014 4:25 PM | Anderson, Stephen M. (CDW)           |
| <input type="checkbox"/> |    | Consult 2.1     | 10/1/2014 2:36 PM | Dean, Mark A.                        |
| <input type="checkbox"/> |   | Dental 1.0      | 2/19/2014 5:52 PM | Anderson, Stephen M. (CDW)           |
| <input type="checkbox"/> |  | DSS             | 2/19/2014 5:59 PM | Anderson, Stephen M. (CDW)           |
| <input type="checkbox"/> |  | ICD-10          | 2/19/2014 5:51 PM | Anderson, Stephen M. (CDW)           |
| <input type="checkbox"/> |  | Inpatient 1.0   | 2/19/2014 4:22 PM | Anderson, Stephen M. (CDW)           |
| <input type="checkbox"/> |  | Inpatient 2.0   | 2/19/2014 4:22 PM | Anderson, Stephen M. (CDW)           |

CDW Home > MetaData > Reports

Actions    1 of 1    Find Next 100%  

## CDW Metadata

Contains a grouped list of available CDW ER Diagrams and members.

### ImageDescription

|   |                         |
|---|-------------------------|
| <a href="#">Allergy 1.0</a>                             | Image Date: 01 Feb 2014 |
| <a href="#">Appointment 2.0</a>                         | Image Date: 03 Jun 2015 |
| <a href="#">Consult 2.1</a>                             | Image Date: 24 Sep 2015 |
| <a href="#">CPRSOrder 1.0</a>                           | Image Date: 11 Aug 2014 |
| <a href="#">Data Profiling 1.0</a>                      | Image Date: 21 Feb 2014 |
| <a href="#">Dental 1.0 Diagram 1 of 2</a>               | Image Date: 28 Oct 2015 |
| <a href="#">Dental 1.0 Diagram 2 of 2 for Analytics</a> | Image Date: 28 Oct 2015 |
| <a href="#">Dimensions A Through D 7/8/2015</a>         | Image Date: 08 Jul 2015 |
| <a href="#">Dimensions E Through K 7/8/2015</a>         | Image Date: 08 Jul 2015 |
| <a href="#">Dimensions L Through O 9/24/2015</a>        | Image Date: 24 Sep 2015 |
| <a href="#">Dimensions P Through R 6/30/2015</a>        | Image Date: 30 Jun 2015 |
| <a href="#">Dimensions S Through Z 4/17/2015</a>        | Image Date: 17 Apr 2015 |
| <a href="#">Dimensions, Place</a>                       | Image Date: 09 Oct 2015 |
| <a href="#">Encounter 1.0</a>                           | Image Date: 29 Oct 2013 |
| <a href="#">Health Factor 2.0</a>                       | Image Date: 11 Mar 2015 |
| <a href="#">Health Factor 2.1</a>                       | Image Date: 15 Sep 2015 |
| <a href="#">ICD-9-CM and ICD-10-CM</a>                  | Image Date: 24 Sep 2015 |
| <a href="#">ICD-9-PCS and ICD-10-PCS</a>                | Image Date: 10 Aug 2015 |
| <a href="#">Immunization 2.1</a>                        | Image Date: 03 Jun 2015 |

# CDW Metadata Report

# Getting a list of joining keys

| <a href="#">Dimensions S Through Z 4/17/2015</a> | Image Date: 17 Apr 2015 |                               |   |                                     |   |  |
|--|-------------------------|-------------------------------|---|-------------------------------------|---|--|
| <a href="#">Dimensions_Place</a>                 | Image Date: 09 Oct 2015 |                               |   |                                     |   |  |
| <a href="#">Encounter 1.0</a>                    | Image Date: 29 Oct 2013 |                               |   |                                     |   |  |
| <a href="#">Health Factor 2.0</a>                | Image Date: 11 Mar 2015 |                               |   |                                     |   |  |
| <a href="#">Health Factor 2.1</a>                |                         |                               |   |                                     |   |  |
| DWViewName                                       | Field Count             | FileMan File Data Source      | View Version                                | Relevant Dates                      | Relationships   |  |
| <a href="#">Dim_HealthFactorType</a>             | 12                      | HEALTH FACTORS (9999999.64)   | DWViewDeployed: xDWWork<br>View Version: 3  |                                     |  |  |
| <a href="#">HF_HealthFactor</a>                  | 21                      | V HEALTH FACTORS (9000010.23) | DWViewDeployed: SPVNext<br>View Version: 34 | Partition Key: HealthFactorDateTime |  |  |
| <a href="#">ICD-9-CM and ICD-10-CM</a>           | Image Date: 24 Sep 2015 |                               |   |                                     |   |  |
| <a href="#">ICD-9-PCS and ICD-10-PCS</a>         | Image Date: 10 Aug 2015 |                               |   |                                     |   |  |
| <a href="#">Immunization 2.1</a>                 | Image Date: 03 Jun 2015 |                               |   |                                     |   |  |

## Reading the “CDW foreign keys” table...

| <b>CDW Foreign Keys</b>   |              |                     |               |              |                  |                     |
|---|--------------|---------------------|---------------|--------------|------------------|---------------------|
| Provides a listing of the foreign and primary keys for CDW views. |              |                     |               |              |                  |                     |
| FKSchemaName  | FKViewName   | FKViewFieldName     | FKViewVersion | PKSchemaName | PKViewName       | PKViewFieldName     |
| HF  | HealthFactor | EncounterStaffSID   | 9             | Staff        | Staff            | StaffSID            |
| HF  | HealthFactor | EncounterStaffSID   | 34            | Staff        | Staff            | StaffSID            |
| HF  | HealthFactor | HealthFactorDateSID | 34            | Dim          | Date             | DateSID             |
| HF  | HealthFactor | HealthFactorTypeSID | 9             | Dim          | HealthFactorType | HealthFactorTypeSID |
| HF  | HealthFactor | HealthFactorTypeSID | 34            | Dim          | HealthFactorType | HealthFactorTypeSID |
| HF  | HealthFactor | PatientSID          | 9             | Patient      | Patient          | PatientSID          |
| HF  | HealthFactor | PatientSID          | 34            | Patient      | Patient          | PatientSID          |
| HF  | HealthFactor | PCEDDataSourceSID   | 9             | Dim          | PCEDDataSource   | PCEDDataSourceSID   |
| HF  | HealthFactor | PCEDDataSourceSID   | 34            | Dim          | PCEDDataSource   | PCEDDataSourceSID   |
| HF  | HealthFactor | VisitSID            | 9             | Outpat       | Visit            | VisitSID            |
| HF  | HealthFactor | VisitSID            | 34            | Outpat       | Visit            | VisitSID            |

Report Date: 11/18/2015 10:08:17 AM

The foreign key EncounterStaffSID in HF.HealthFactor connects to the primary key StaffSID in Staff.Staff.

# Connecting the HF to a patient

## CDW Foreign Keys

Provides a listing of the foreign and primary keys for CDW views.

| FKSchemaName | FKViewName   | FKViewFieldName     | FKViewVersion | PKSchemaName | PKViewName       | PKViewFieldName     | PKViewVersion |
|--------------|--------------|---------------------|---------------|--------------|------------------|---------------------|---------------|
| HF           | HealthFactor | EncounterStaffSID   | 9             | Staff        | Staff            | StaffSID            | 9             |
| HF           | HealthFactor | EncounterStaffSID   | 34            | Staff        | Staff            | StaffSID            | 34            |
| HF           | HealthFactor | HealthFactorDateSID | 34            | Dim          | Date             | DateSID             | 34            |
| HF           | HealthFactor | HealthFactorTypeSID | 9             | Dim          | HealthFactorType | HealthFactorTypeSID | 9             |
| HF           | HealthFactor | HealthFactorTypeSID | 34            | Dim          | HealthFactorType | HealthFactorTypeSID | 34            |
| HF           | HealthFactor | PatientSID          | 9             | Patient      | Patient          | PatientSID          | 9             |
| HF           | HealthFactor | PatientSID          | 34            | Patient      | Patient          | PatientSID          | 34            |
| HF           | HealthFactor | PCEDataSourceSID    | 9             | Dim          | PCEDataSource    | PCEDataSourceSID    | 9             |
| HF           | HealthFactor | PCEDataSourceSID    | 34            | Dim          | PCEDataSource    | PCEDataSourceSID    | 34            |
| HF           | HealthFactor | VisitSID            | 9             | Outpat       | Visit            | VisitSID            | 9             |
| HF           | HealthFactor | VisitSID            | 34            | Outpat       | Visit            | VisitSID            | 34            |

Report Date: 11/18/2015 10:08:17 AM

The foreign key PatientSID in HF.HealthFactor connects to the primary key PatientSID in Patient.Patient

# Summary/Conclusion

- CDW contains a variety of data.
- Currently, most of it comes from VistA.
- It is broken into many domains, fact tables and dimension tables that must be brought back together to create an analytic set.
- SQL and linking keys are necessary tools for creating that analytic set of data.

# Questions?

Margaret Gonsoulin, PhD

[VIReC@va.gov](mailto:VIReC@va.gov)

708-202-2413