A picture containing text

Description automatically generatedLogo, company name

Description automatically generated

NJR EHDEN HIPS and KNEES MDS 7.0 ETL

Date 04.10.2024

Version 1.0

# Document History

|  |  |  |
| --- | --- | --- |
| **Version Number** | **Date** | **Changes** |
| First Release V1.0 | 04.10.2024 | Not Applicable |

Table of Contents

[Document History 1](#_Toc178927889)

[Table of Contents 2](#_Toc178927890)

[1. ETL Implementation 3](#_Toc178927891)

[1.1 Introduction 3](#_Toc178927892)

[1.2 ETL Code 3](#_Toc178927893)

[1.3 ETL Mappings 4](#_Toc178927894)

# ETL Implementation

## 1.1 Introduction

The ETL is implemented in a SQL server instance utilising the SQL agent job scheduling feature.

Transact Structured Query Language (T-SQL), the MS tool and extension of SQL language, is then used to execute our Use defined stored procedures (USPS). T-SQL is designed to be used for writing an entire program of block function procedure that defines how things need to be completed and creating an application in which each application sends a transaction query over SQL server and there is no interaction with the database. It executes as a whole block with the extension of SQL language. There are different types of T-SQL functions used in T-SQL that made it ideal for this project.

* Scalar function.
* Ranking function.
* Aggregate function.
* Rowset function.

SQL Server Integration Services (SSIS) was used to load the Athena vocabularies which are executed in a specific order to ensure the OMOP tables are populated with consideration of foreign keys.

## 1.2 ETL Code

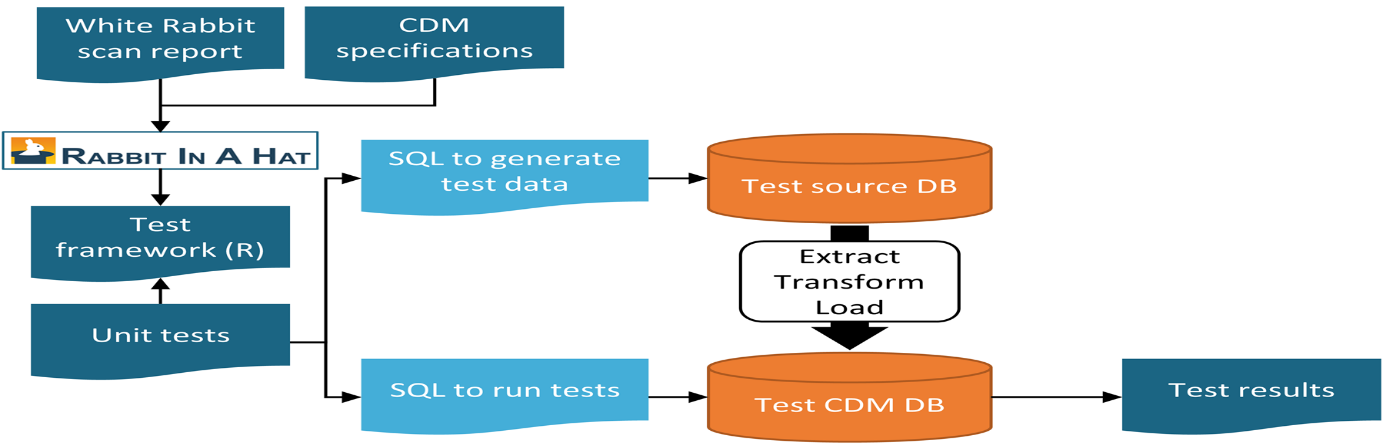
The steps to populate the OMOP CDM with NJR data in the SQL agent job are as follows:

1. Check that the Health Data Warehouse (HDW) is processing.
2. Load Athena files.
3. Drop indexes and keys.
4. Populate NJR specific concepts.
5. Populate CDM\_Source.
6. Populate Person.
7. Populate Death.
8. Populate Location.
9. Populate Care\_Site.
10. Populate Provider.
11. Populate Visit\_Occurrence
12. Populate Measurement
13. Populate Observation\_Period.
14. Populate Observation.
15. Populate Procedure\_Occurrence.
16. Populate Fact\_Relationship.
17. Populate Device\_Exposure and Concept Table.
18. Recreate Indexes and Keys.
19. Update Statistics.

Each of these steps may be described as follows.

1. The NJR source database populations are tested to ensure there are no errors. This prevents any corrupt or incomplete NJR data (SQL Test) being loaded into the OMOP CDM.
2. Load the standard Athena vocabularies (SSIS).
3. Drop the Foreign Key on the OMOP database (SQL).
4. Populate the indication and procedure type NJR user vocabularies.
5. Populate the CDM\_Source table.
6. Populate each individual OMOP table in steps 6 to 17 above with NJR data.
7. Recreate indexes and keys.
8. Update statistics.

Any error during the execution of the above steps results in a failure email being sent to the NJR/NEC service team for investigation.



**Figure 1: ETL process using White Rabbit and Rabbit-in-a-hat**

## 1.3 ETL Mappings

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Column** | **Table and Column name** | **Datatype** | **SourceDesc** | **SourceHardcoded** |
| 1 | CARE\_SITE.care\_site\_id | Int | HospitalID | No |
| 2 | CARE\_SITE.care\_site\_name | Varchar | HospitalName | No |
| 3 | CARE\_SITE.place\_of\_service\_concept\_id | Int | OMOP.SOURCE\_TO\_CONCEPT\_MAP | No |
| 4 | CARE\_SITE.location\_id | Int | HospitalID | No |
| 5 | CARE\_SITE.care\_site\_source\_value | Varchar | HospitalName | No |
| 6 | CARE\_SITE.place\_of\_service\_source\_value | Varchar | OrganisationType | No |

| **Column** | **Table and Column name** | **Datatype** | **SourceDesc** | **SourceHardcoded** |
| --- | --- | --- | --- | --- |
| 1 | CDM\_SOURCE.cdm\_source\_name | varchar | National Joint Registry | Yes |
| 2 | CDM\_SOURCE.cdm\_source\_abbreviation | varchar | NJR | Yes |
| 3 | CDM\_SOURCE.cdm\_holder | varchar | HQIP | Yes |
| 4 | CDM\_SOURCE.source\_description | varchar | Hip and knee data for MDS7 | Yes |
| 5 | CDM\_SOURCE.source\_documentation\_reference | varchar | https://reports.njrcentre.org.uk/NJR-Minimum-Data-Set-V7 | Yes |
| 6 | CDM\_SOURCE.cdm\_etl\_reference | varchar | <https://ohdsi.github.io/CommonDataModel/cdm54.html> | Yes |
| 7 | CDM\_SOURCE.source\_release\_date | date | Last createddate of source data table | No |
| 8 | CDM\_SOURCE.cdm\_release\_date | date | select c.valid\_start\_date  from VOCABULARY v  join CONCEPT c on v.vocabulary\_version = c.concept\_code  where v.vocabulary\_id = 'CDM' | No |
| 9 | CDM\_SOURCE.cdm\_version | varchar | select v.vocabulary\_version  from VOCABULARY v  join CONCEPT c on v.vocabulary\_version = c.concept\_code  where v.vocabulary\_id = 'CDM' | No |
| 10 | CDM\_SOURCE.cdm\_version\_concept\_id | int | select c.concept\_id  from VOCABULARY v  join CONCEPT c on v.vocabulary\_version = c.concept\_code  where v.vocabulary\_id = 'CDM' | No |
| 11 | CDM\_SOURCE.vocabulary\_version | varchar | v5.0 31-AUG-23 (select vocabulary\_version from VOCABULARY v where v.vocabulary\_id = 'None') | No |

| **Column** | **Table and Column name** | **Datatype** | **SourceDesc** | **SourceHardcoded** |
| --- | --- | --- | --- | --- |
| 1 | OBSERVATION.observation\_id | int | UniqueRowNumber | No |
| 2 | OBSERVATION.person\_id | int | PersonSpecific UniqueRowNumber | No |
| 3 | OBSERVATION.observation\_concept\_id | int | select concept\_id from concept where [concept\_name] = 'History of event' and domain\_id = 'Observation' and vocabulary\_id ='OMOP Extension'  This returns value “1340204”. | No |
| 4 | OBSERVATION.observation\_date | date | Date of surgery. | No |
| 5 | OBSERVATION.observation\_datetime | datetime | Date of surgery. | No |
| 6 | OBSERVATION.observation\_type\_concept\_id | int | select concept\_id from concept where concept\_name = 'Registry'  and vocabulary\_id = 'Type Concept'  This returns value “32879 Registry” | No |
| 7 | OBSERVATION.value\_as\_number | float | NULL |  |
| 8 | OBSERVATION.value\_as\_string | varchar(60) | Original source Indication name as recorded at the time of the procedure. One row for each indication. | No |
| 9 | OBSERVATION.value\_as\_concept\_id | int | Nearest standard concept\_id to the indication. |  |
| 10 | OBSERVATION.qualifier\_concept\_id | int | NULL |  |
| 11 | OBSERVATION.unit\_concept\_id | int | NULL |  |
| 12 | OBSERVATION.provider\_id | int | ClinicianID + HospitalID | No |
| 13 | OBSERVATION.visit\_occurrence\_id | int | ProcedureFactKey | No |
| 14 | OBSERVATION.visit\_detail\_id | int | NULL |  |
| 15 | OBSERVATION.observation\_source\_value | varchar(50) | NULL |  |
| 16 | OBSERVATION.observation\_source\_concept\_id | int | NULL |  |
| 17 | OBSERVATION.unit\_source\_value | varchar(50) | NULL |  |
| 18 | OBSERVATION.qualifier\_source\_value | varchar(50) | NULL |  |
| 19 | OBSERVATION.value\_source\_value | varchar(50) | NULL |  |
| 20 | OBSERVATION.observation\_event\_id | bigint | NULL |  |
| 21 | OBSERVATION.obs\_event\_field\_concept\_id | int | NULL |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Column** | **Table and Column name** | **Datatype** | **SourceDesc** | **SourceHardcoded** |
| 1 | DEATH.person\_id | int | patientkey | No |
| 2 | DEATH.death\_date | date | DateOfDeathKey | No |
| 3 | DEATH.death\_datetime | datetime | DateOfDeathKey | No |
| 4 | DEATH.death\_type\_concept\_id | int | NULL | Yes |
| 5 | DEATH.cause\_concept\_id | int | NULL | Yes |
| 6 | DEATH.cause\_source\_value | varchar | NULL | Yes |
| 7 | DEATH.cause\_source\_concept\_id | int | NULL | Yes |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Column** | **Table and Column name** | **Datatype** | **SourceDesc** | **SourceHardcoded** |
| 1 | DEVICE\_EXPOSURE.device\_exposure\_id | int | UniqueRowNumber | No |
| 2 | DEVICE\_EXPOSURE.person\_id | int | PersonSpecific UniqueRowNumber | No |
| 3 | DEVICE\_EXPOSURE.device\_concept\_id | int | 0 | No |
| 4 | DEVICE\_EXPOSURE.device\_exposure\_start\_date | date | Date the component was inserted in the patient | No |
| 5 | DEVICE\_EXPOSURE.device\_exposure\_start\_datetime | datetime | Date the component was inserted in the patient | No |
| 6 | DEVICE\_EXPOSURE.device\_exposure\_end\_date | date | NULL | No |
| 7 | DEVICE\_EXPOSURE.device\_exposure\_end\_datetime | datetime | NULL | No |
| 8 | DEVICE\_EXPOSURE.device\_type\_concept\_id | int | 32879 | Yes |
| 9 | DEVICE\_EXPOSURE.unique\_device\_id | varchar | NULL | No |
| 10 | DEVICE\_EXPOSURE.production\_id | varchar | NULL | Yes |
| 11 | DEVICE\_EXPOSURE.quantity | int | 1 (was NULL) | Yes |
| 12 | DEVICE\_EXPOSURE.provider\_id | bigint | ClinicianID + HospitalID | No |
| 13 | DEVICE\_EXPOSURE.visit\_occurrence\_id | int | ProcedureFactKey | No |
| 14 | DEVICE\_EXPOSURE.visit\_detail\_id | int | NULL | Yes |
| 15 | DEVICE\_EXPOSURE.device\_source\_value | varchar | Full component description with joint sub category and catalogue number, but truncated at 50 characters. Full description is in the Concept table. | Yes |
| 16 | DEVICE\_EXPOSURE.device\_source\_concept\_id | int | vocabulary\_id = 'NJR Devices' concept\_id  Each entry in this table is one of the components inserted during an operation. | Yes |
| 17 | DEVICE\_EXPOSURE.unit\_concept\_id | int | NULL | Yes |
| 18 | DEVICE\_EXPOSURE.unit\_source\_value | varchar | NULL | Yes |
| 19 | DEVICE\_EXPOSURE.unit\_source\_concept\_id | int | NULL | Yes |

| **Column** | **Table and Column name** | **Datatype** | **SourceDesc** | **SourceHardcoded** |
| --- | --- | --- | --- | --- |
| 1 | FACT\_RELATIONSHIP.domain\_concept\_id\_1 | int | SELECT relationship\_concept\_id FROM dbo.RELATIONSHIP where relationship\_id = 'Unit of' | No |
| 2 | FACT\_RELATIONSHIP.fact\_id\_1 | int | HospitalID | No |
| 3 | FACT\_RELATIONSHIP.domain\_concept\_id\_2 | int | SELECT relationship\_concept\_id FROM dbo.RELATIONSHIP where relationship\_id = 'Unit of' | No |
| 4 | FACT\_RELATIONSHIP.fact\_id\_2 | int | TrustID | No |
| 5 | FACT\_RELATIONSHIP.relationship\_concept\_id | int | SELECT domain\_concept\_id FROM domain WHERE domain\_id = 'Unit' | No |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Column** | **Table and Column name** | **Datatype** | **SourceDesc** | **SourceHardcoded** |
| 1 | LOCATION.location\_id | int | HospitalID | No |
| 2 | LOCATION.address\_1 | varchar | HospitalAddress1 | No |
| 3 | LOCATION.address\_2 | varchar | HospitalAddress2 | No |
| 4 | LOCATION.city | varchar | hospitalTown | No |
| 5 | LOCATION.state | varchar | NULL | Yes |
| 6 | LOCATION.zip | varchar | HospitalPostcode | No |
| 7 | LOCATION.county | varchar | HospitalCounty | No |
| 8 | LOCATION.location\_source\_value | varchar | HospitalName | No |
| 9 | LOCATION.country\_concept\_id | int | concept\_id country map | No |
| 10 | LOCATION.country\_source\_value | varchar | country | No |
| 11 | LOCATION.latitude | float | NULL | Yes |
| 12 | LOCATION.longitude | float | NULL | Yes |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Column** | **Table and Column name** | **Datatype** | **SourceDesc** | **SourceHardcoded** |
| 1 | OBSERVATION\_PERIOD.observation\_period\_id | int | ProcedureFactKey | No |
| 2 | OBSERVATION\_PERIOD.person\_id | int | PersonSpecific UniqueRowNumber | No |
| 3 | OBSERVATION\_PERIOD.observation\_period\_start\_date | date | Earliest NJR MDS7 date | No |
| 4 | OBSERVATION\_PERIOD.observation\_period\_end\_date | date | Latest NJR MDS7 date | No |
| 5 | OBSERVATION\_PERIOD.period\_type\_concept\_id | int | 32879 | Yes |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Column** | **Table and Column name** | **Datatype** | | **SourceDesc** | | **SourceHardcoded** |
| 1 | PERSON.person\_id | int | | Patientkey | | No |
| 2 | PERSON.gender\_concept\_id | int | | concept\_id gender map | | No |
| 3 | PERSON.year\_of\_birth | int | | DateOfBirthKey year | | No |
| 4 | PERSON.month\_of\_birth | int | | DateOfBirthKey month | | No |
| 5 | PERSON.day\_of\_birth | int | | DateOfBirthKey day | | No |
| 6 | PERSON.birth\_datetime | datetime | | DateOfBirthKey | | No |
| 7 | PERSON.race\_concept\_id | int | | 0 | | Yes |
| 8 | PERSON.ethnicity\_concept\_id | int | | 0 | | Yes |
| 9 | PERSON.location\_id | int | | NULL | | Yes |
| 10 | PERSON.provider\_id | bigint | | ClinicianID + HospitalID | | No |
| 11 | PERSON.care\_site\_id | int | | HospitalID | | No |
| 12 | PERSON.person\_source\_value | varchar | | NULL | | Yes |
| 13 | PERSON.gender\_source\_value | varchar | | Male/Female | | No |
| 14 | PERSON.gender\_source\_concept\_id | int | | NULL | | Yes |
| 15 | PERSON.race\_source\_value | varchar | | NULL | | Yes |
| 16 | PERSON.race\_source\_concept\_id | int | | NULL | | Yes |
| 17 | PERSON.ethnicity\_source\_value | varchar | | NULL | | Yes |
| 18 | PERSON.ethnicity\_source\_concept\_id | int | | NULL | | Yes |
| Column | Table and Column name | | Datatype | | SourceDesc | SourceHardcoded |
| 1 | PROCEDURE\_OCCURRENCE.procedure\_occurrence\_id | | int | | UniqueRowNumber | No |
| 2 | PROCEDURE\_OCCURRENCE.person\_id | | int | | Person UniqueRowNumber | No |
| 3 | PROCEDURE\_OCCURRENCE.procedure\_concept\_id | | int | | 0 | No |
| 4 | PROCEDURE\_OCCURRENCE.procedure\_date | | date | | SurgeryDateKey | No |
| 5 | PROCEDURE\_OCCURRENCE.procedure\_datetime | | datetime | | SurgeryDateKey | No |
| 6 | PROCEDURE\_OCCURRENCE.procedure\_end\_date | | date | | SurgeryDateKey | No |
| 7 | PROCEDURE\_OCCURRENCE.procedure\_end\_datetime | | datetime | | SurgeryDateKey | No |
| 8 | PROCEDURE\_OCCURRENCE.procedure\_type\_concept\_id | | int | | 32879 | Yes |
| 9 | PROCEDURE\_OCCURRENCE.modifier\_concept\_id | | int | | NULL | Yes |
| 10 | PROCEDURE\_OCCURRENCE.quantity | | int | | 1 | Yes |
| 11 | PROCEDURE\_OCCURRENCE.provider\_id | | bigint | | ClinicianID + HospitalID | No |
| 12 | PROCEDURE\_OCCURRENCE.visit\_occurrence\_id | | int | | ProcedureFactKey | No |
| 13 | PROCEDURE\_OCCURRENCE.visit\_detail\_id | | int | | NULL | Yes |
| 14 | PROCEDURE\_OCCURRENCE.procedure\_source\_value | | varchar | | Procedure Type Key (NJR’s key value for the procedure type) | No |
| 15 | PROCEDURE\_OCCURRENCE.procedure\_source\_concept\_id | | int | | [vocabulary\_id] = 'NJR Procedures' concept\_id | Yes |
| 16 | PROCEDURE\_OCCURRENCE.modifier\_source\_value | | varchar | | NULL | Yes |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Column** | **Table and Column name** | **Datatype** | **SourceDesc** | **SourceHardcoded** |
| 1 | PROVIDER.provider\_id | bigint | ClinicianID + HospitalID | No |
| 2 | PROVIDER.provider\_name | varchar | ConsultantSurgeonKey DisplayName | No |
| 3 | PROVIDER.npi | varchar | na | Yes |
| 4 | PROVIDER.dea | varchar | na | Yes |
| 5 | PROVIDER.specialty\_concept\_id | int | 38004465 | Yes |
| 6 | PROVIDER.care\_site\_id | int | HospitalID | No |
| 7 | PROVIDER.year\_of\_birth | int | NULL | Yes |
| 8 | PROVIDER.gender\_concept\_id | int | concept\_id Gender map Male/Female | No |
| 9 | PROVIDER.provider\_source\_value | varchar | ClinicianID | No |
| 10 | PROVIDER.specialty\_source\_value | varchar | Trauma and Orthopaedics | Yes |
| 11 | PROVIDER.specialty\_source\_concept\_id | int | NULL | Yes |
| 12 | PROVIDER.gender\_source\_value | varchar | Male/Female | No |
| 13 | PROVIDER.gender\_source\_concept\_id | int | NULL | Yes |

| **Column** | **Table and Column name** | **Datatype** | **SourceDesc** | **SourceHardcoded** |
| --- | --- | --- | --- | --- |
| 1 | VISIT\_OCCURRENCE.visit\_occurrence\_id | int | ProcedureFactKey | No |
| 2 | VISIT\_OCCURRENCE.person\_id | int | Person UniqueRowNumber | No |
| 3 | VISIT\_OCCURRENCE.visit\_concept\_id | int | 9201 | Yes |
| 4 | VISIT\_OCCURRENCE.visit\_start\_date | date | SurgeryDateKey | No |
| 5 | VISIT\_OCCURRENCE.visit\_start\_datetime | datetime | SurgeryDateKey | No |
| 6 | VISIT\_OCCURRENCE.visit\_end\_date | date | SurgeryDateKey | No |
| 7 | VISIT\_OCCURRENCE.visit\_end\_datetime | datetime | SurgeryDateKey | No |
| 8 | VISIT\_OCCURRENCE.visit\_type\_concept\_id | int | 32879 | Yes |
| 9 | VISIT\_OCCURRENCE.provider\_id | bigint | ClinicianID + HospitalID | No |
| 10 | VISIT\_OCCURRENCE.care\_site\_id | int | HospitalID | No |
| 11 | VISIT\_OCCURRENCE.visit\_source\_value | varchar | Inpatient | No |
| 12 | VISIT\_OCCURRENCE.visit\_source\_concept\_id | int | NULL | Yes |
| 13 | VISIT\_OCCURRENCE.admitted\_from\_concept\_id | int | NULL | Yes |
| 14 | VISIT\_OCCURRENCE.admitted\_from\_source\_value | varchar | NULL | Yes |
| 15 | VISIT\_OCCURRENCE.discharged\_to\_concept\_id | int | NULL | Yes |
| 16 | VISIT\_OCCURRENCE.discharged\_to\_source\_value | varchar | NULL | Yes |
| 17 | VISIT\_OCCURRENCE.preceding\_visit\_occurrence\_id | int | NULL | Yes |

| **Column** | **Table and Column name** | **Datatype** | **SourceDesc** | **SourceHardcoded** |
| --- | --- | --- | --- | --- |
| 1 | MEASUREMENT.measurement\_id | int | UniqueRowNumber | no |
| 2 | MEASUREMENT.person\_id | int | PersonSpecific UniqueRowNumber | no |
| 3 | MEASUREMENT.measurement\_concept\_id | int | Multiple rows, one for each measure within each visit\_id, possible entries are:-   1. 4178502 Body weight 2. 4177340 Body height 3. 44783982 BMI 4. 4159411 ASA grade 5. 1761860 Flexion contracture range of Knee | no |
| 4 | MEASUREMENT.measurement\_date | date | Surgery Date | no |
| 5 | MEASUREMENT.measurement\_datetime | datetime | Surgery Date | no |
| 6 | MEASUREMENT.measurement\_time | varchar(10) | NULL |  |
| 7 | MEASUREMENT.measurement\_type\_concept\_id | int | 32879 Registry | Yes |
| 8 | MEASUREMENT.operator\_concept\_id | int | 1 above = 4172703 Equal  2 above = 4172703 Equal  3 above = 4150298 Range  4 above = 4172703 Equal  5 above = <70 = 4171756 Less  >110 = 4172704 Greater  Else null | no |
| 9 | MEASUREMENT.value\_as\_number | float | Weight,  Height,  BMI,  ASAGradeID,  KneeFlexionRange | No |
| 10 | MEASUREMENT.value\_as\_concept\_id | int | NULL |  |
| 11 | MEASUREMENT.unit\_concept\_id | int | 1 above = 9529 kilogram  2 above = 8582 centimeter  3 above = 9531 Kilogram per square meter  4 above = 44777566 score  5 above = 9484 degree | no |
| 12 | MEASUREMENT.range\_low | float | 1,2,4 above = null  3 above = bmiRange  5 above = KneeFlexionRange | no |
| 13 | MEASUREMENT.range\_high | float | 1,2,4 above = null  3 above = bmiRange  5 above = KneeFlexionRange | no |
| 14 | MEASUREMENT.provider\_id | int | ClinicianID + HospitalID | no |
| 15 | MEASUREMENT.visit\_occurrence\_id | int | ProcedureFactKey | no |
| 16 | MEASUREMENT.visit\_detail\_id | int | NULL |  |
| 17 | MEASUREMENT.measurement\_source\_value | varchar(50) | Original text source values | no |
| 18 | MEASUREMENT.measurement\_source\_concept\_id | int | NULL |  |
| 19 | MEASUREMENT.unit\_source\_value | varchar(50) | Original text source values  1=”kilogram”,2=”centimeter”  3=”Kilogram per square meter”  4= “Score”, 5= “Degree” | no |
| 20 | MEASUREMENT.unit\_source\_concept\_id | int | NULL |  |
| 21 | MEASUREMENT.value\_source\_value | varchar(50) | Original text source values | no |
| 22 | MEASUREMENT.measurement\_event\_id | bigint | NULL |  |
| 23 | MEASUREMENT.meas\_event\_field\_concept\_id | int | NULL |  |

**END OF DOCUMENT**