# BIOTECH CLINICAL TRIALS

## IE6750

## Data Warehousing & Integration

**Group Number: 5**

**Naresh Gajula**

### Introduction

The biotech industry operates in a data-intensive environment where the success of drug development heavily relies on the effective management of clinical trial data. Clinical trials generate vast and diverse datasets, including patient demographics, drug information, trial outcomes, and health metrics, which are essential for assessing drug safety, efficacy, and compliance with regulatory standards. However, the fragmented nature of this data across multiple sources and formats poses significant challenges to its integration and analysis.

Current systems often lack the capacity to consolidate and analyze these datasets effectively, leading to inefficiencies, delayed decision-making, and missed opportunities for critical insights. A centralized data warehouse specifically designed for clinical trials can address these issues. By integrating static data (e.g., patient demographics, drug details) with dynamic, transactional data (e.g., trial outcomes, biomarker trends), a data warehouse enables streamlined data management, advanced analytics, and secure accessibility.

This report focuses on the design and implementation of a data warehouse tailored to meet the unique needs of clinical trials in the biotech industry. With features such as robust data integration, scalability, data quality assurance, and advanced analytical tools, the proposed system aims to revolutionize how biotech companies handle clinical trial data. The system's capabilities will empower researchers and decision-makers with timely and accurate insights, accelerating the drug development cycle, enhancing patient safety, and improving overall efficiency in bringing innovative treatments to market.

By addressing the core challenges of data fragmentation, quality, and scalability, this initiative supports the biotech industry's mission to deliver safer, more effective drugs to patients worldwide. The proposed data warehouse will serve as a cornerstone for the industry's data-driven future, facilitating evidence-based decision-making and fostering innovation.

**Business Problem Introduction:**

Clinical trials are indispensable for evaluating the safety and efficacy of new drugs. However, managing and analyzing clinical trial data poses significant challenges for biotech companies due to the following issues:

1. **Data Fragmentation:** Trial data exists across various systems and formats, such as patient information, drug protocols, and trial outcomes. Consolidating this data for analysis is a complex and time-consuming process.
2. **Inefficient Analysis:** The absence of a centralized repository makes it difficult to assess key metrics like drug efficacy, adverse events, and patient safety, leading to delays in deriving actionable insights.
3. **Delayed Decision-Making:** Without holistic data analysis, decisions regarding drug development and regulatory compliance are often delayed, slowing down time-to-market for critical treatments.
4. **Regulatory Compliance Challenges:** Disparate data systems complicate adherence to data privacy and security standards, especially when dealing with sensitive patient information.
5. **Data Quality Issues:** Ensuring the accuracy, completeness, and consistency of trial data is challenging when it is fragmented across multiple systems and formats.
6. **Limited Analytical Capabilities:** Existing systems often lack the advanced tools necessary for visualizing trends in patient responses, biomarker changes, and adverse events, hindering comprehensive data analysis.
7. **Scalability Limitations:** As the volume of clinical trial data grows, existing systems struggle to manage and process the increasing amounts of information effectively.
8. **Missed Opportunities for Innovation:** The inability to identify patterns and trends in trial data slows the development of personalized therapies and innovative treatment protocols.

A centralized data warehouse tailored for clinical trials can eliminate these inefficiencies. By integrating static and transactional data, it provides a unified platform for data storage, advanced analytics, and secure access. This system empowers researchers and decision-makers with actionable insights, enabling faster, data-driven decision-making and improving the overall efficiency of the drug development lifecycle.

**Problem Statement**

The biotech industry relies on extensive clinical trials to evaluate the safety and efficacy of drugs. However, the data generated from these trials exists in fragmented formats and disparate sources, creating several challenges:

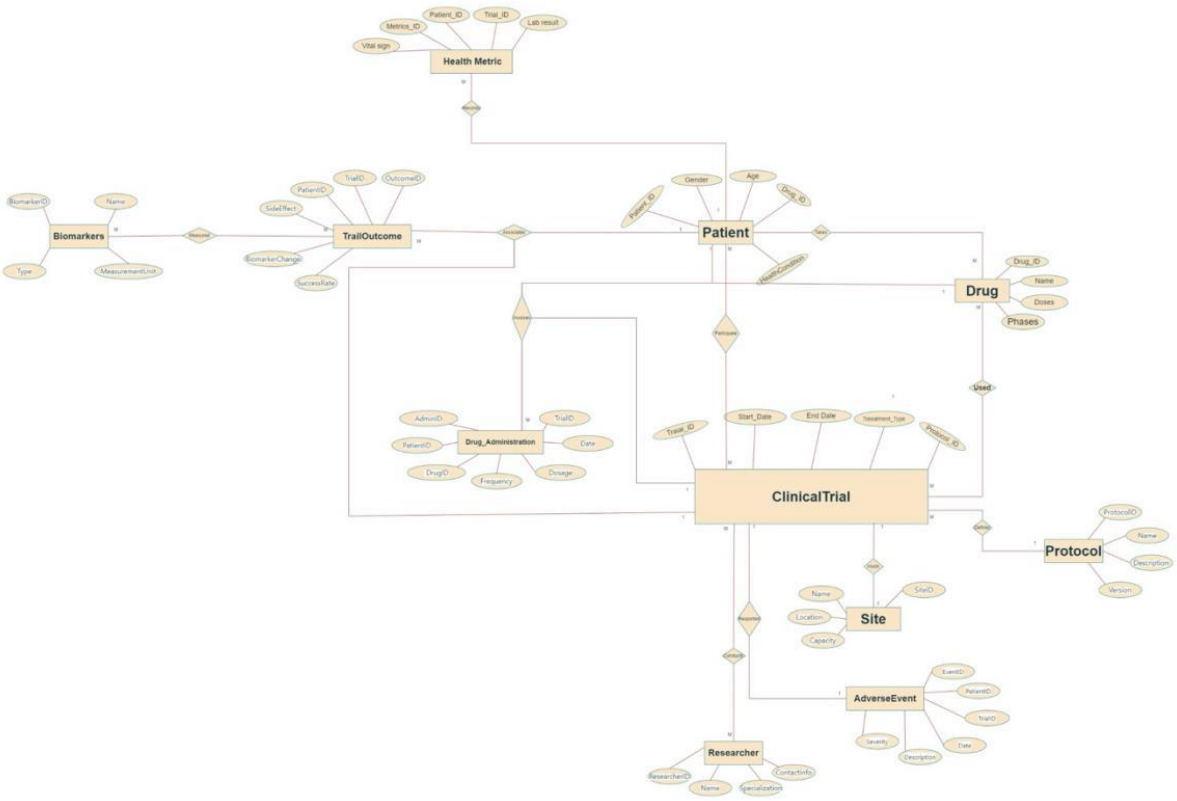
- **Integration Complexity:** Consolidating diverse datasets such as patient demographics, trial protocols, and drug administration records is difficult.
- **Quality Issues:** Ensuring data accuracy, consistency, and completeness is challenging without robust data validation and cleansing processes.
- **Scalability Concerns:** Handling large volumes of trial data over time requires scalable infrastructure.
- **Limited Analytical Capabilities:** Current systems often lack robust tools for visualizing and analyzing trends in patient responses, biomarkers, and adverse events.
- **Data Privacy Risks:** Ensuring compliance with industry standards for sensitive patient data is critical.

### 1. Entity-Relationship Diagram (ERD)

Create an extended ERD with the following entities:

1. Patient
2. Drug
3. ClinicalTrial
4. DrugAdministration
5. TrialOutcome
6. HealthMetric
7. Researcher
8. Site
9. Protocol
10. Biomarker
11. AdverseEvent

**Map the ERD:**




## Relational Model:

1. Patient (PatientID, Name, Age, Gender, HealthCondition)

2. Drug (DrugID, Name, Dosage, Phase)

3. ClinicalTrial (TrialID, ProtocolID, StartDate, EndDate, TreatmentType)

4. DrugAdministration (AdminID, PatientID, DrugID, TrialID, Date, Dosage, Frequency)

5. TrialOutcome (OutcomeID, TrialID, PatientID, SideEffect, BiomarkerChange, SuccessRate)

6. HealthMetric (MetricID, PatientID, Date, VitalSign, LabResult)

7. Researcher (ResearcherID, Name, Specialization, ContactInfo)

8. Site (SiteID, Name, Location, Capacity)

9. Protocol (ProtocolID, Name, Description, Version)

10. Biomarker (BiomarkerID, Name, Type, MeasurementUnit)

11. AdverseEvent (EventID, PatientID, TrialID, Date, Description, Severity)

**Relationships:**

**Patient ↔ ClinicalTrial**

• **Relationship**: A patient can participate in multiple clinical trials, and a clinical trial can have multiple patients.

• **Cardinality**: **Many-to-Many (M to M)**

• **Implementation**: This would require an associative entity, such as **PatientTrial**, with PatientID and TrialID as foreign keys.

**Drug ↔ ClinicalTrial**

• **Relationship**: A drug can be used in multiple clinical trials, and a clinical trial can include multiple drugs.

• **Cardinality**: **Many-to-Many (M to M)**

• **Implementation**: An associative entity, such as **TrialDrug**, can be created with TrialID and DrugID as foreign keys.

**Researcher ↔ ClinicalTrial**

• **Relationship**: A researcher can conduct multiple clinical trials, and a clinical trial can be conducted by multiple researchers.

• **Cardinality**: **Many-to-Many (M to M)• Implementation**: An associative entity, such as **ResearcherTrial**, can be created with ResearcherID and TrialID as foreign keys.

**Site ↔ ClinicalTrial**

• **Relationship**: A site can host multiple clinical trials, and a clinical trial can be hosted at multiple

sites.

• **Cardinality**: **Many-to-Many (M to M)**

• **Implementation**: An associative entity, such as **SiteTrial**, can be created with SiteID and TrialID

as foreign keys.

**Protocol ↔ ClinicalTrial**

• **Relationship**: A protocol defines one clinical trial, but a clinical trial can refer to one protocol.

• **Cardinality**: **One-to-Many (1 to M)**

• **Foreign Key**: ClinicalTrial.ProtocolID references Protocol.ProtocolID.

**DrugAdministration ↔ Patient, Drug, ClinicalTrial**

• **Relationship**: Each drug administration involves one patient, one drug, and one clinical trial.

• **Cardinality**: **Many-to-One (M:1)** for each.

• **Foreign Keys**: DrugAdministration.PatientID, DrugAdministration.DrugID, and

DrugAdministration.TrialID reference their respective tables.

**TrialOutcome ↔ Patient, ClinicalTrial**

• **Relationship**: Each trial outcome is associated with one patient and one clinical trial.

• **Cardinality**: **Many-to-One (M:1)** for each.

• **Foreign Keys**: TrialOutcome.PatientID and TrialOutcome.TrialID reference their respective

tables.

**HealthMetric ↔ Patient**

• **Relationship**: Each health metric is recorded for one patient.

• **Cardinality**: **Many-to-One (M:1)**

• **Foreign Key**: HealthMetric.PatientID references Patient.PatientID.

**Biomarker ↔ TrialOutcome**

• **Relationship**: A biomarker can be measured in multiple trial outcomes, and each trial outcome

can reference multiple biomarkers.

• **Cardinality**: **Many-to-Many (M to M)**

• **Implementation**: An associative entity, such as **TrialOutcomeBiomarker**, can be created with OutcomeID and BiomarkerID as foreign keys.**AdverseEvent ↔ Patient, ClinicalTrial**

• **Relationship**: Each adverse event is reported for one patient in one clinical trial.

• **Cardinality**: **Many-to-One (M:1)** for each.

• **Foreign Keys**: AdverseEvent.PatientID and AdverseEvent.TrialID reference their respective tables.

**Data Population Methodology:**

1. Data Sources: We will be collecting data from clinical trial databases, electronic health records, and drug registries.

2. Synthetic Data: Use tools to generate fake data for testing, like patient records and trial outcomes.

3. ETL Process: Use automated scripts to import, clean, and load the data into the database.4. Batch Loading: Insert large datasets at once for efficiency.

5. Data Integrity: Ensure the data is correct by using validation rules and foreign keys.

**1. Static Reference Data:**

- Patient demographics
- Drug information
- Trial protocols
- Researcher information
- Site details

**2. Slowly Changing Dimensional Data:**

- Patient health conditions
- Drug dosage details
- Protocol versions

**3. Transactional Data:**

- Drug administration records
- Trial outcomes
- Patient health metrics

- Adverse events

Potential Dimensions, Hierarchies, and Measures

**1. Dimensions:**

- Time (Date, Month, Quarter, Year)
- Patient (Age group, Gender)
- Drug (Name, Phase)
- Trial (Protocol, Treatment Type)
- Researcher
- Site

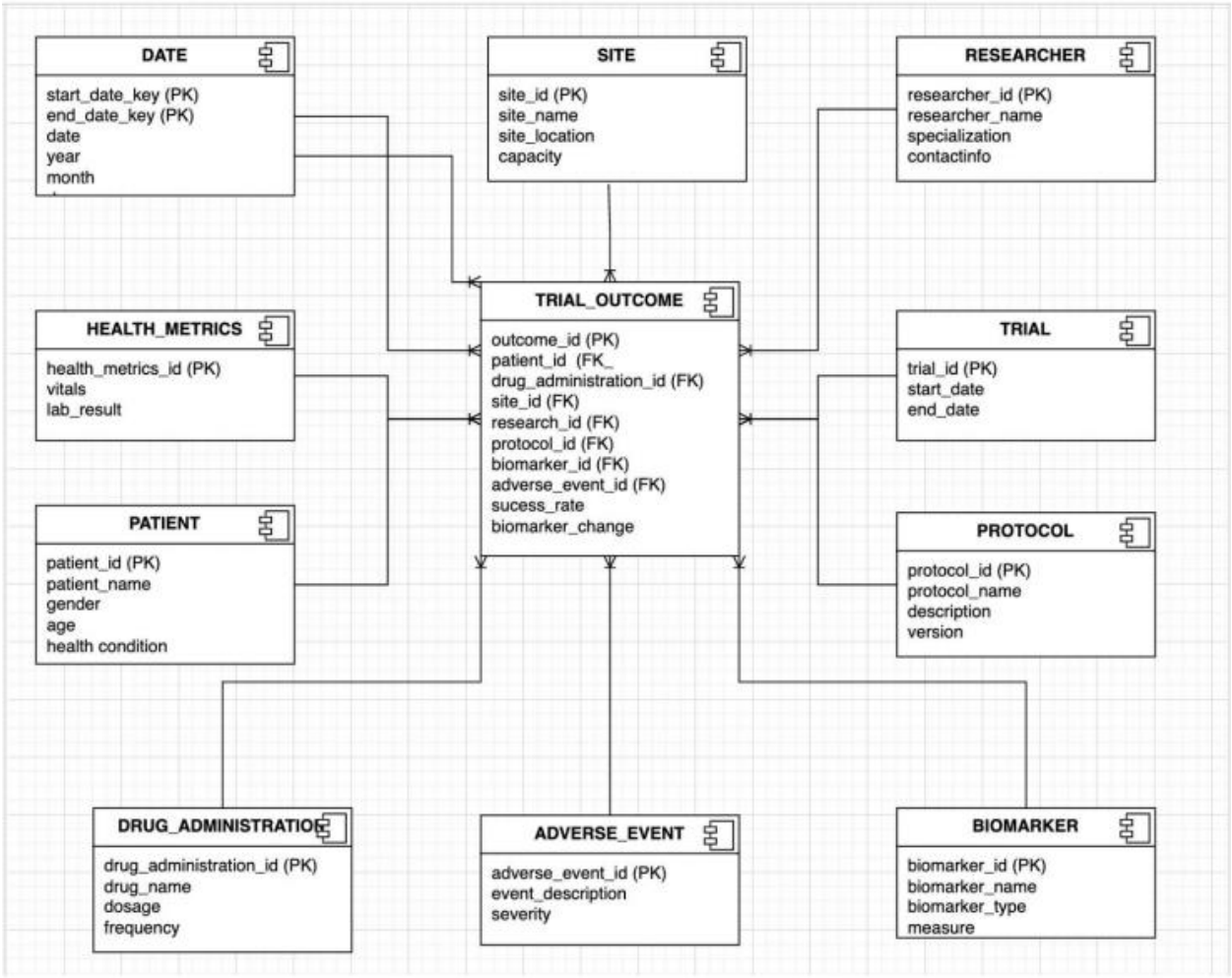
**2. Hierarchies:**

- Time: Date > Month > Quarter > Year• Patient: Individual > Age Group > All Patients
- Drug: Individual Drug > Phase > All Drugs
- Site: Site > Location > All Sites

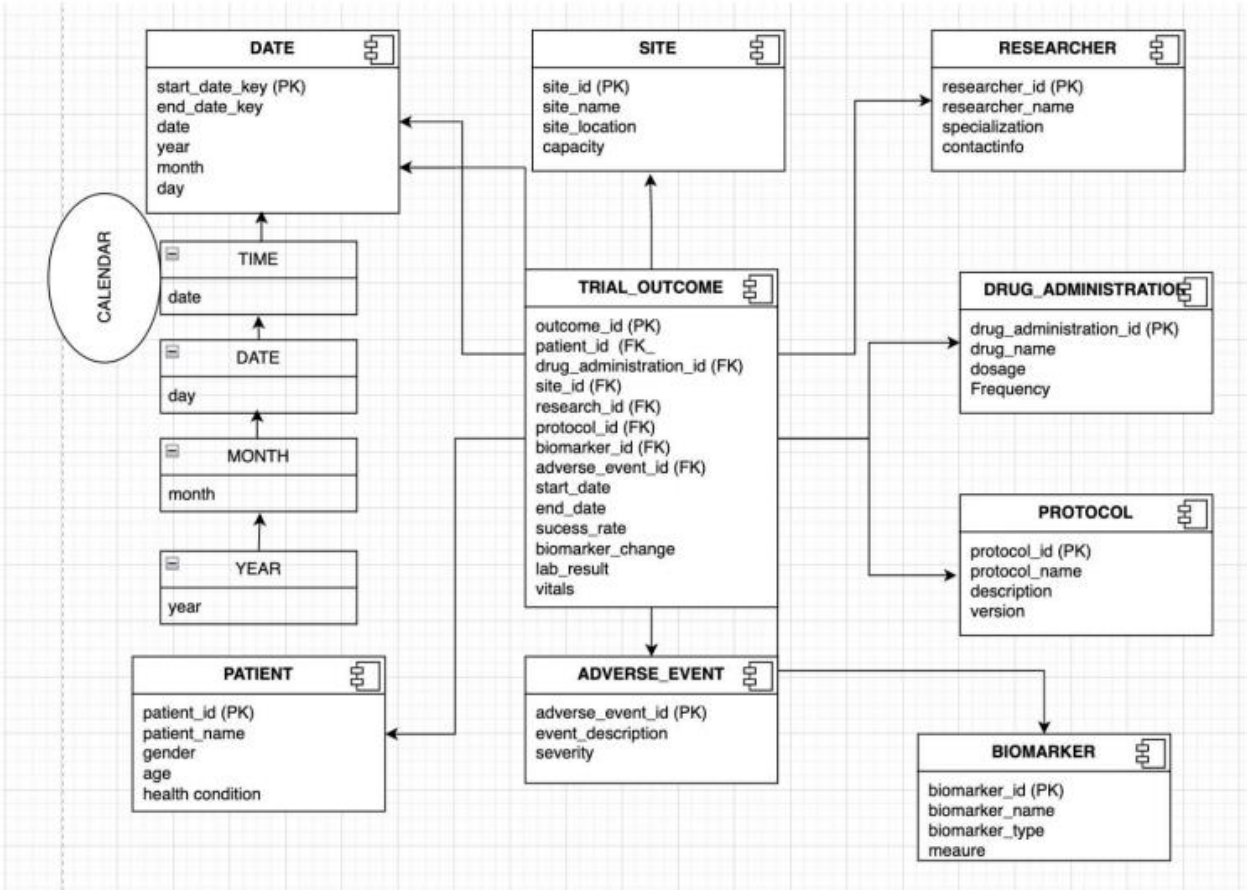
**3. Measures:**

- Success Rate
- Biomarker Change
- Side Effect Frequency
- Drug Dosage
- Vital Sign Measurements
- Number of Adverse Events
- Patient Enrollment Rate

## Conceptual Model :



## Logical Model :




## Database Implementation:

## OLTP:

## 3. DB Implementation:

### 1.Create Database

CREATE DATABASE clinical_trial_db1;

### 2. Create Tables

**DATE Table**

```
CREATE TABLE DATE (
start_date_key SERIAL PRIMARY KEY,
end_date_key INT,
date DATE NOT NULL,
year INT NOT NULL,
month INT NOT NULL,
day INT NOT NULL
);
```

**PATIENT Table:**

```
CREATE TABLE PATIENT (
patient_id SERIAL PRIMARY KEY,
patient_name VARCHAR(255) NOT NULL,
gender VARCHAR(10),
age INT,
health_condition TEXT
);
```

SITE Table:

```
CREATE TABLE SITE (
site_id SERIAL PRIMARY KEY,
site_name VARCHAR(255) NOT NULL,
site_location TEXT,
capacity INT
);
```

**RESEARCHER Table:**

```
CREATE TABLE RESEARCHER (
researcher_id SERIAL PRIMARY KEY,
researcher_name VARCHAR(255) NOT NULL,
specialization VARCHAR(255),
contactinfo TEXT
);
```

**DRUG_ADMINISTRATION Table:**

```
CREATE TABLE DRUG_ADMINISTRATION (
drug_administration_id SERIAL PRIMARY KEY,
drug_name VARCHAR(255) NOT NULL,
dosage VARCHAR(50),
frequency VARCHAR(50)
);
```

**PROTOCOL Table:**

```
CREATE TABLE PROTOCOL (
protocol_id SERIAL PRIMARY KEY,
protocol_name VARCHAR(255) NOT NULL,
description TEXT,
version VARCHAR(50)
);
```

**BIOMARKER Table:**

```
CREATE TABLE BIOMARKER (
biomarker_id SERIAL PRIMARY KEY,
biomarker_name VARCHAR(255) NOT NULL,
biomarker_type VARCHAR(50),
```

measure VARCHAR(50)

);

**ADVERSE_EVENT Table:**

CREATE TABLE ADVERSE_EVENT (

adverse_event_id SERIAL PRIMARY KEY,

event_description TEXT,

severity VARCHAR(50)

);

**TRIAL_OUTCOME Table:**

CREATE TABLE TRIAL_OUTCOME (

outcome_id SERIAL PRIMARY KEY,

patient_id INT REFERENCES PATIENT(patient_id),

drug_administration_idINT REFERENCES DRUG_ADMINISTRATION(drug_administration_id),
site_id INT REFERENCES SITE(site_id),

research_id INT REFERENCES RESEARCHER(researcher_id),

protocol_id INT REFERENCES PROTOCOL(protocol_id),

biomarker_id INT REFERENCES BIOMARKER(biomarker_id),

adverse_event_id INT REFERENCES ADVERSE_EVENT(adverse_event_id),

start_date DATE,

end_date DATE,

success_rate DECIMAL(5, 2),

biomarker_change TEXT,

lab_result TEXT,

vitals TEXT

);

**TIME Table:**CREATE TABLE TIME (

time_id SERIAL PRIMARY KEY,

date DATE NOT NULL,

year INT NOT NULL,

month INT NOT NULL,

day INT NOT NULL

);

**Add Foreign Keys to DATE Table:**

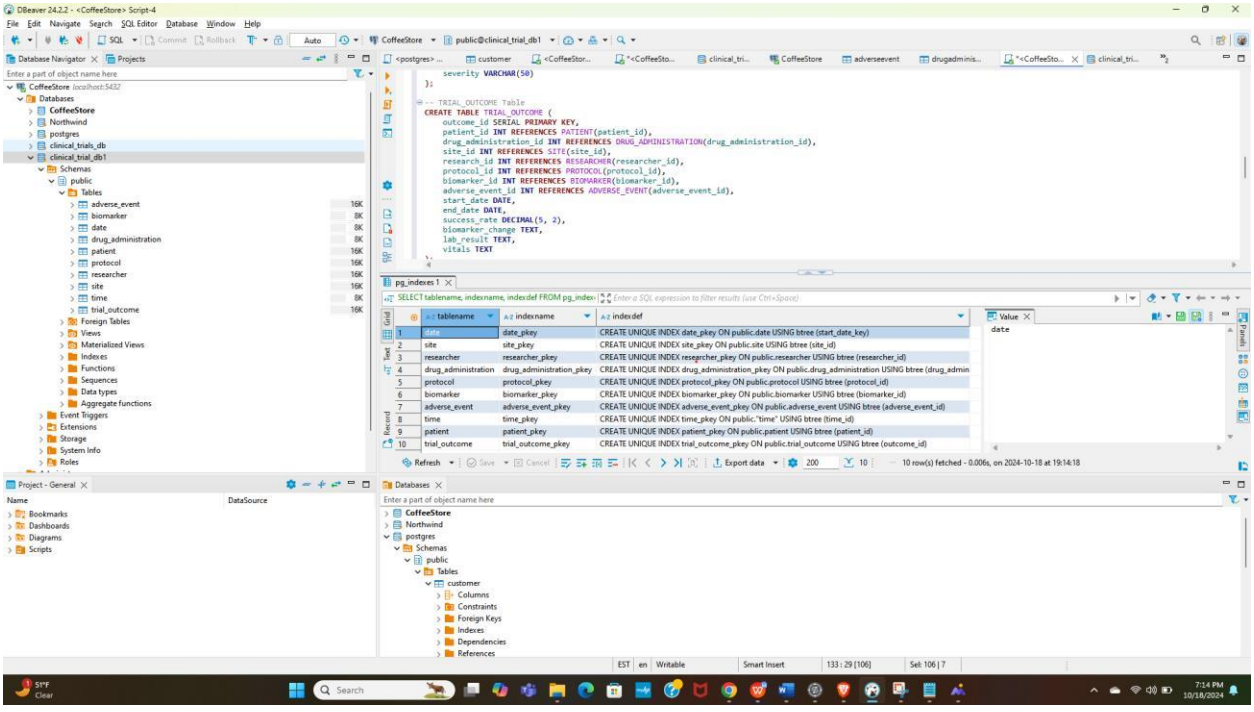
ALTER TABLE DATE ADD CONSTRAINT fk_start_date FOREIGN KEY (start_date_key) REFERENCES

TIME(time_id);

ALTER TABLE DATE ADD CONSTRAINT fk_end_date FOREIGN KEY (end_date_key) REFERENCES

TIME(time_id);

## Implementation Output:



## 1. Primary Events Documentation:

Primary events are key activities or milestones in the trial process that should be tracked in our database. Based on the ER diagram, here are some primary events:

• Drug Administration: Each time a patient receives a drug, it's an important event to track. It involves documenting the drug_name, dosage, and frequency from the DRUG_ADMINISTRATION table.

• Trial Start and End Dates: The start and end dates of each trial in the TRIAL_OUTCOME table are key events.

• Biomarker Changes: The biomarker measurement changes in the TRIAL_OUTCOME table are crucial for evaluating the effect of the trial. This event tracks the progression of the trial based on specific biological markers.

• Adverse Events: Documenting adverse events (from the ADVERSE_EVENT table) is another primary event. This includes the description and severity of any side effects a patient experiences during the trial.

• Patient Health Condition: From the PATIENT table, capturing the patient's health condition before and after the trial is important to assess the trial's success.

## OLAP Operations:

**1. Monthly Drug Administration Volume by Drug Type**

• **Operation**: Roll-Up

• **Description**: This operation groups and counts drug administrations on a monthly basis by drug type.

R1 ← ROLLUP*(DrugAdministration, AdministrationDate → Month, DrugID, COUNT(AdministrationID) AS New_Administrations)

**2. Yearly Trial Success Rate by Site**

• **Operation**: Roll-Up

• **Description**: Summarizes the trial success rate per year at different trial sites.

R2 ← ROLLUP*(TrialOutcome, StartDate → Year, SiteID, AVG(SuccessRate) AS

Avg_SuccessRate)

**3. Adverse Event Severity by Drug Type**

• **Operation**: Dice

• **Description**: Selects and groups the severity of adverse events for a particular drug type.

R3 ← DICE*(AdverseEvent, DrugAdministration.DrugID = X, GROUP BY Severity,

COUNT(AdverseEventID) AS EventCount)

**4. Biomarker Change by Patient Age Group**

• **Operation**: Drill-Down

• **Description**: Analyzes the biomarker changes in trials by drilling down into age groups.

R4 ← DRILLDOWN*(TrialOutcome, PatientID → PatientAgeGroup, BiomarkerChange,

AVG(BiomarkerChange) AS Avg_BiomarkerChange)

**5. Average Dosage by Drug Frequency**

• **Operation**: Roll-Up

• **Description**: Groups and averages the drug dosage by drug frequency.

R5 ← ROLLUP*(DrugAdministration, Frequency, AVG(Dosage) AS Avg_Dosage)

**6. Site Capacity and Trial Volume• Operation**: Pivot

• **Description**: Reorganizes data to compare site capacity against the volume of trials

conducted at each site.

R6 ← PIVOT*(SiteCapacity, TrialOutcome.SiteID, COUNT(TrialOutcomeID) AS Trial_Volume,

Site.Capacity)

**7. Adverse Event Severity by Patient Gender**

• **Operation**: Slice

• **Description**: Slices the dataset to analyze adverse event severity by patient gender.

R7 ← SLICE*(AdverseEvent, Patient.Gender = ‘Female’, COUNT(AdverseEventID) AS

Event_Count)

**8. Success Rate Comparison Across Different Research Protocols**

• **Operation**: Dice

• **Description**: Selects different research protocols and compares the success rates.

R8 ← DICE*(TrialOutcome, ProtocolID IN (X, Y, Z), AVG(SuccessRate) AS Avg_SuccessRate)

**9. Patient Participation by Age Group and Gender**

• **Operation**: Drill-Down and Dice

• **Description**: Analyzes trial participation, first by patient gender, and drills down into different age groups.

R9 ← DRILLDOWN*(Patient, Gender, AGEGROUP(Patient.Age) AS AgeGroup, COUNT(PatientID) AS ParticipationCount)

**10. Lab Results Distribution by Site and Year**

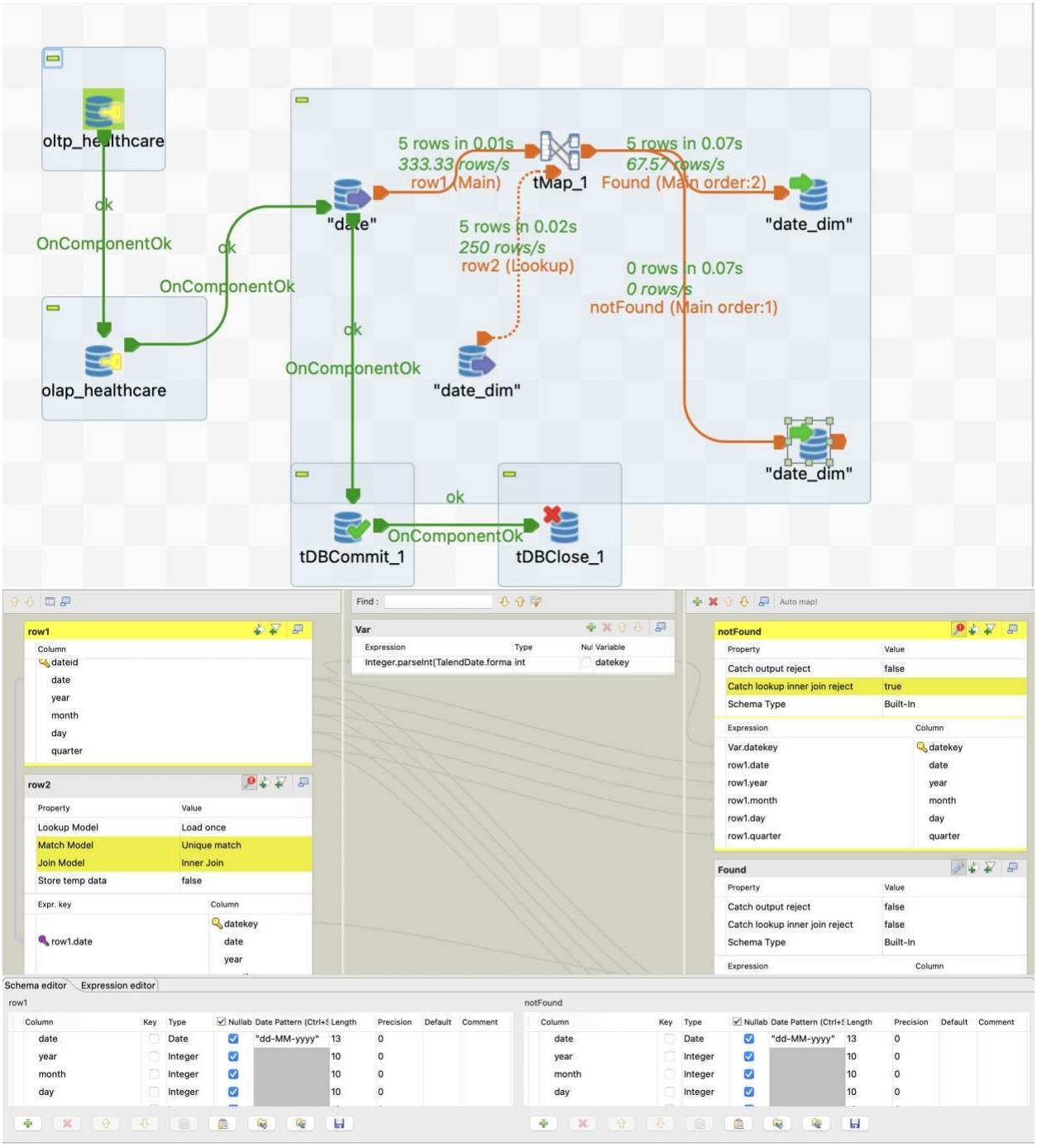
• **Operation**: Roll-Up and Dice

• **Description**: Rolls up the lab results by site and year, filtering for significant biomarkers.

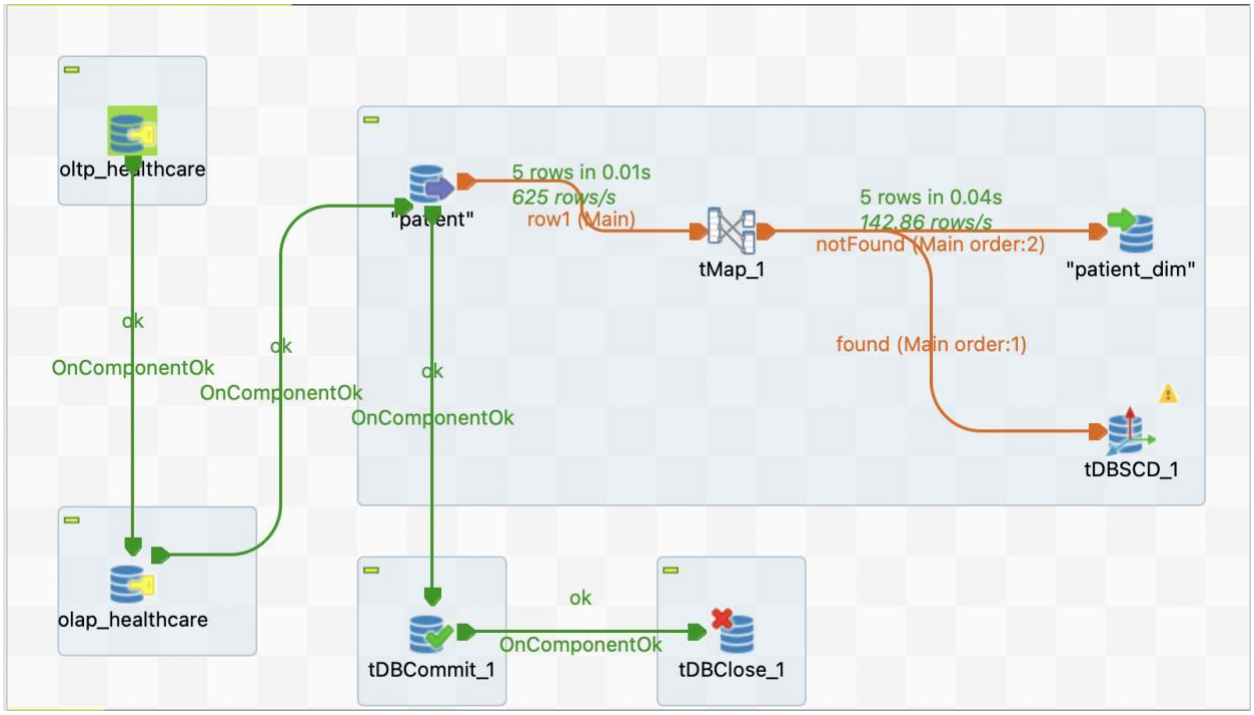
R10 ← ROLLUP*(TrialOutcome, SiteID, StartDate → Year, WHERE BiomarkerChange > X, AVG(LabResult) AS Avg_LabResult)

## Implementation In Talend:

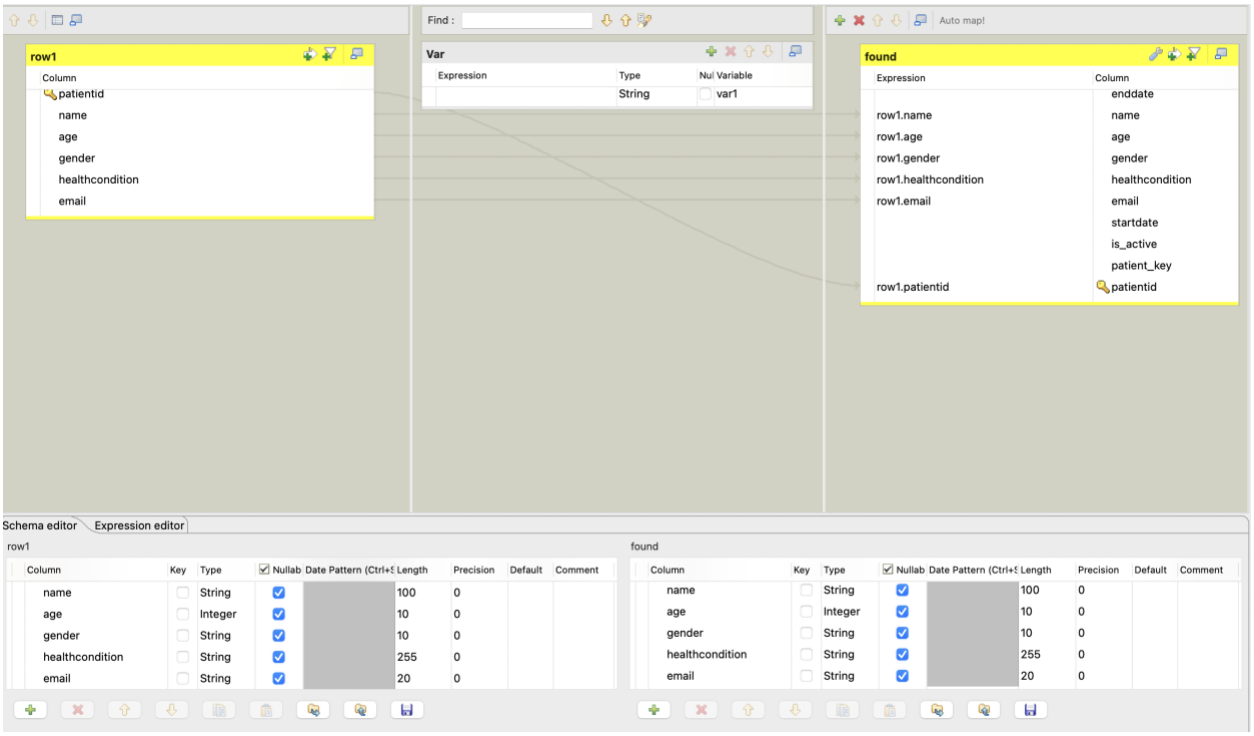
# 1. Date_Dim:


oltp_healthcare
ok
OnComponentOk
ok
OnComponentOk
olap_healthcare
"date"
5 rows in 0.01s
333.33 rows/s
row1 (Main)
tMap_1
5 rows in 0.07s
67.57 rows/s
Found (Main order:2)
"date_dim"
5 rows in 0.02s
250 rows/s
row2 (Lookup)
0 rows in 0.07s
0 rows/s
notFound (Main order:1)
ok
OnComponentOk
"date_dim"
"date_dim"
ok
OnComponentOk
tDBCommit_1
tDBClose_1
row1
Column
dateid
date
year
month
day
quarter
row2
Property
Value
Lookup Model
Load once
Match Model
Unique match
Join Model
Inner Join
Store temp data
false
Expr. key
Column
row1.date
datekey
date
year
Find :
Var
Expression
Type
Nul
Variable
Integer.parseInt(TalendDate.forma
int
datekey
Auto map!
notFound
Property
Value
Catch output reject
false
Catch lookup inner join reject
true
Schema Type
Built-In
Expression
Column
Var.datekey
datekey
row1.date
date
row1.year
year
row1.month
month
row1.day
day
row1.quarter
quarter
Found
Property
Value
Catch output reject
false
Catch lookup inner join reject
false
Schema Type
Built-In
Expression
Column
Schema editor
Expression editor
row1
Column
Key
Type
Nullab
Date Pattern (Ctrl+S
Length
Precision
Default
Comment
date
Date
"dd-MM-yyyy"
13
0
year
Integer
10
0
month
Integer
10
0
day
Integer
10
0
notFound
Column
Key
Type
Nullab
Date Pattern (Ctrl+S
Length
Precision
Default
Comment
date
Date
"dd-MM-yyyy"
13
0
year
Integer
10
0
month
Integer
10
0
day
Integer
10
0

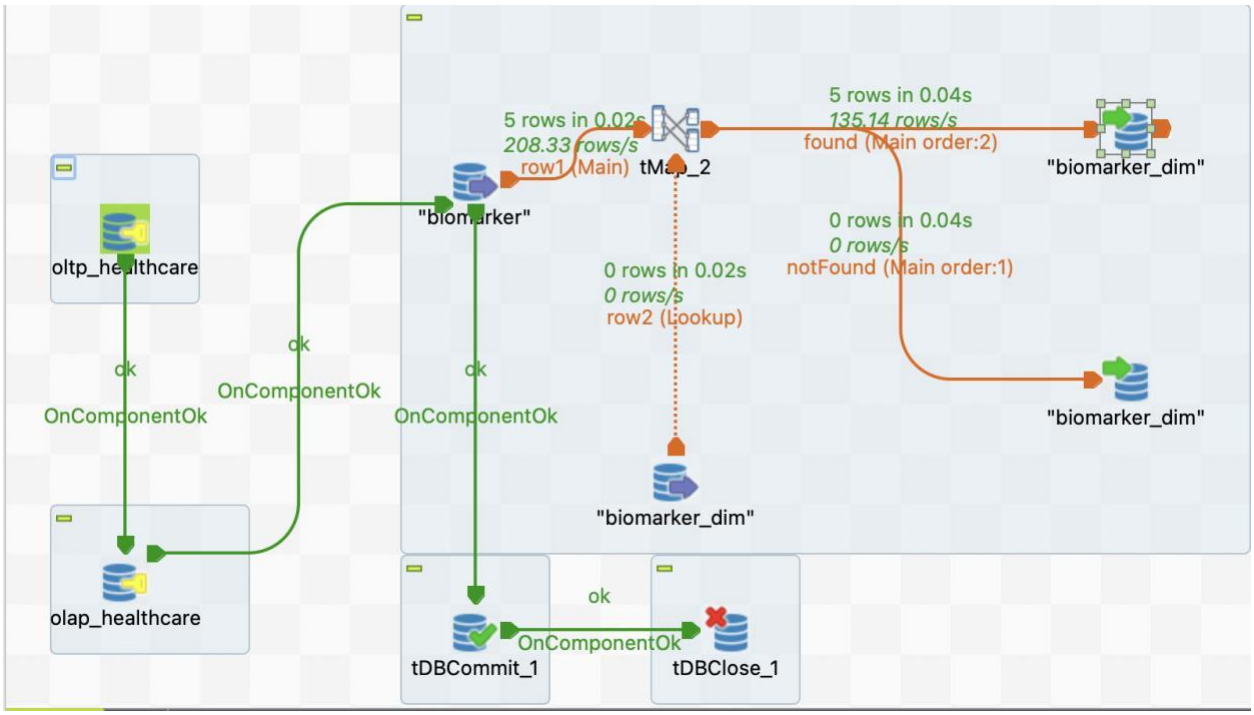
## 2.Patient dim



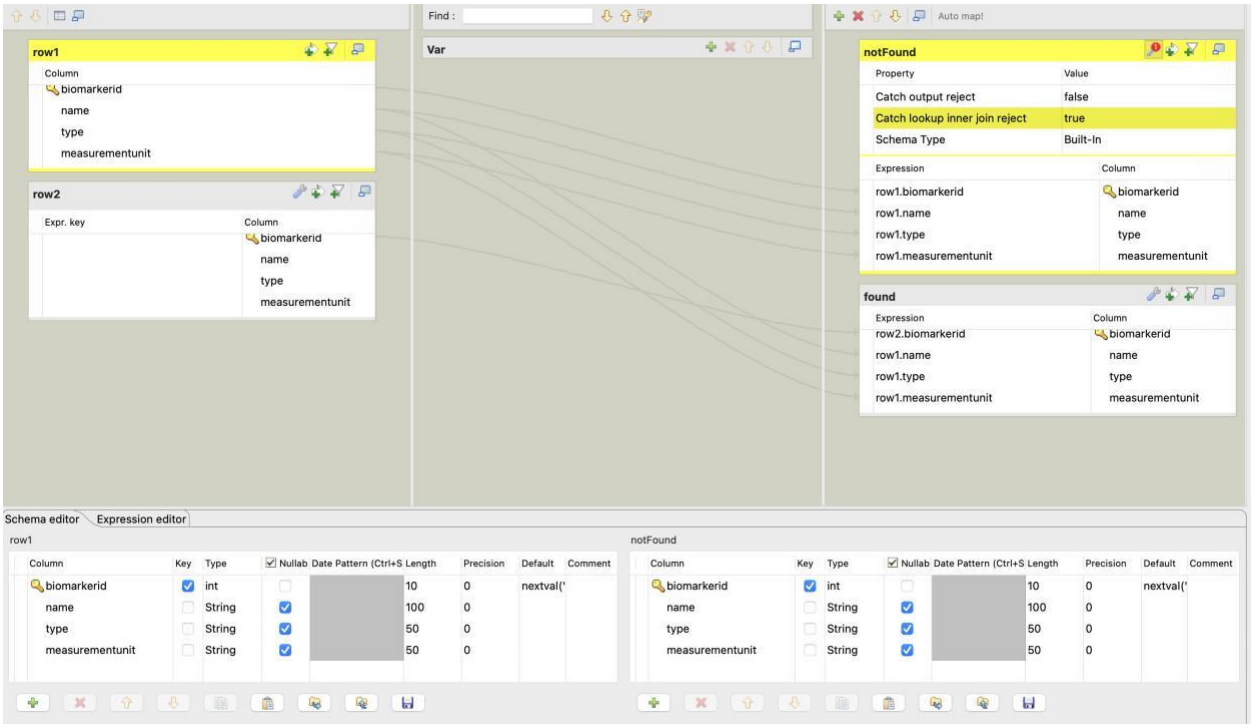

Note: We have incorporated SCD Type 2 in the Supplies dimension table using `tMap` to handle changes in the rate of equipment or medicine on a regular basis.



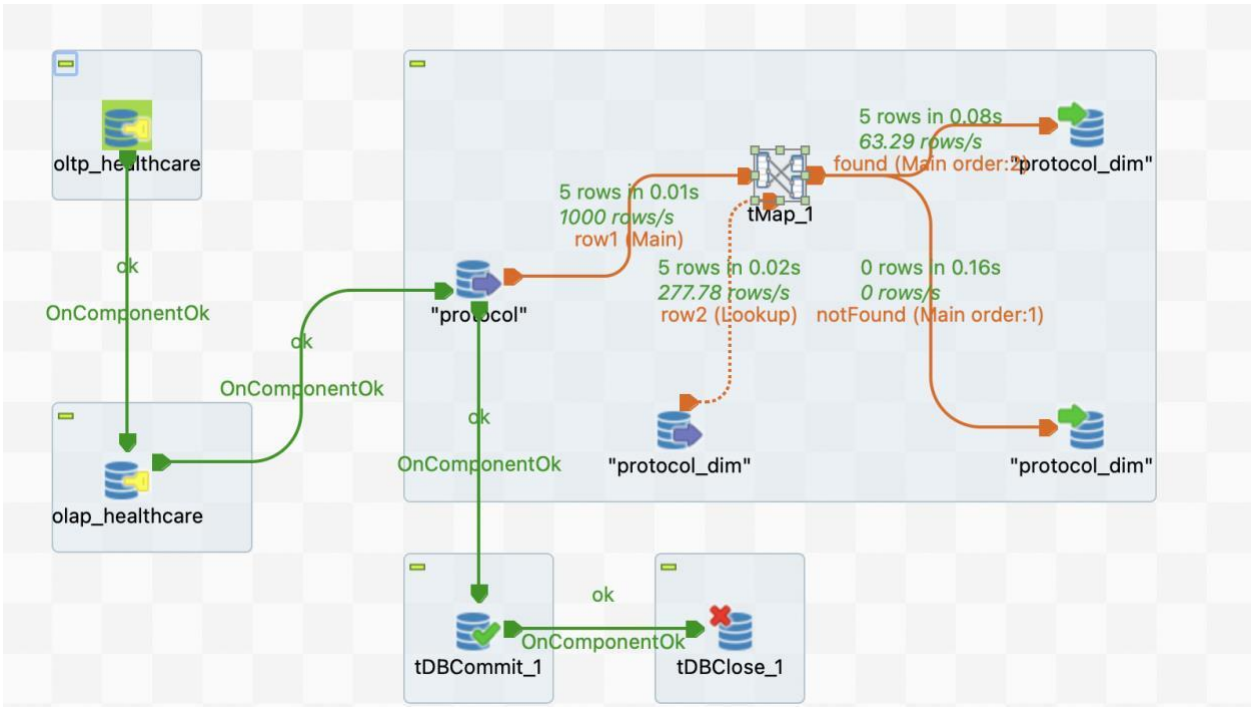

## 2. Biomarker dim

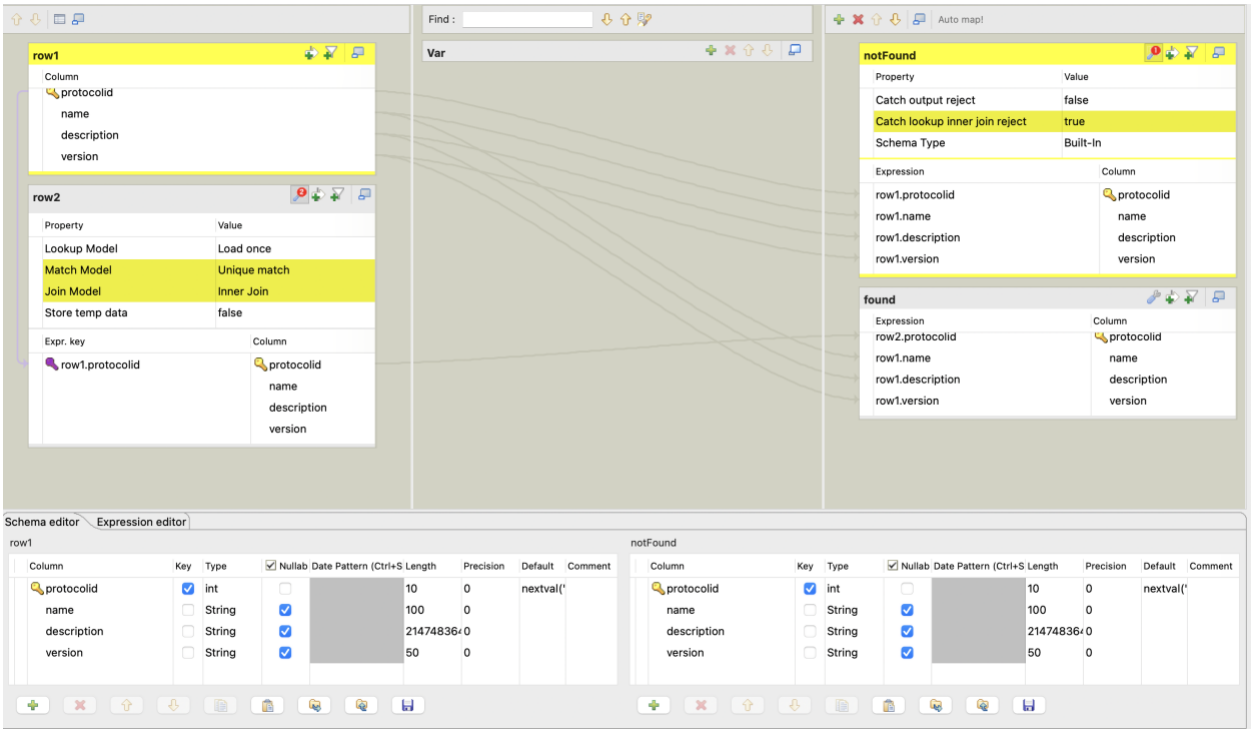



**Tmap connection:**

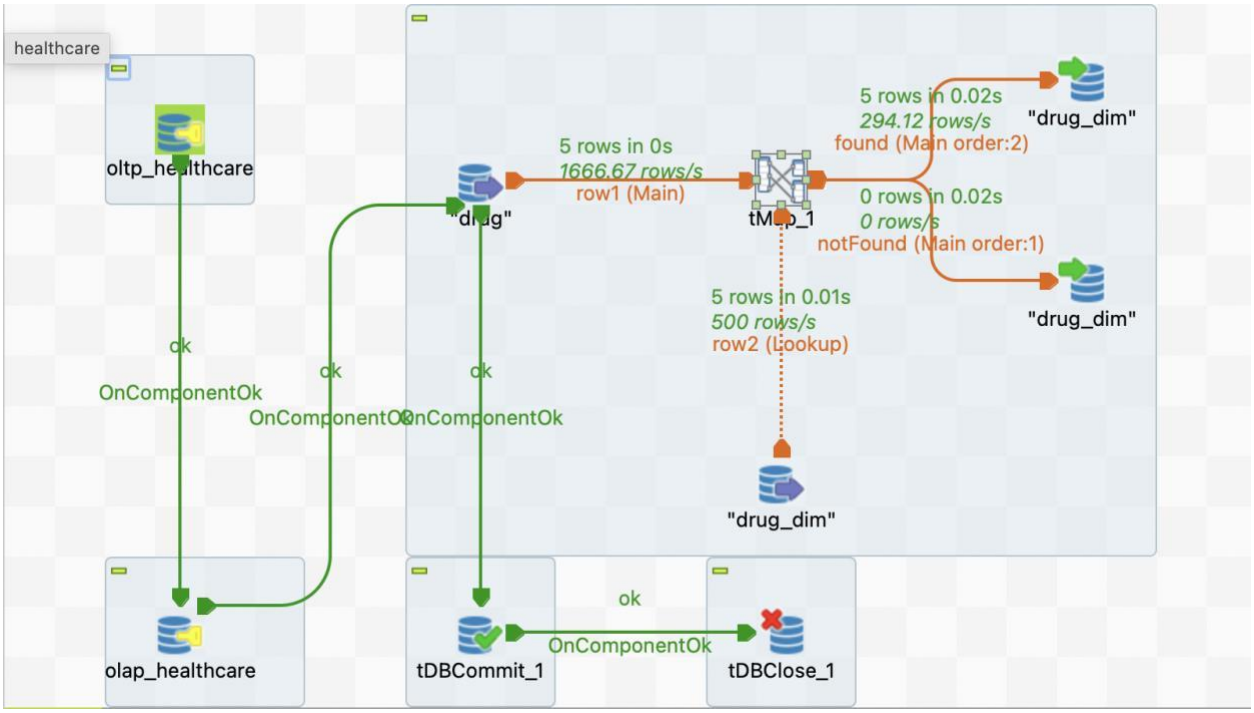


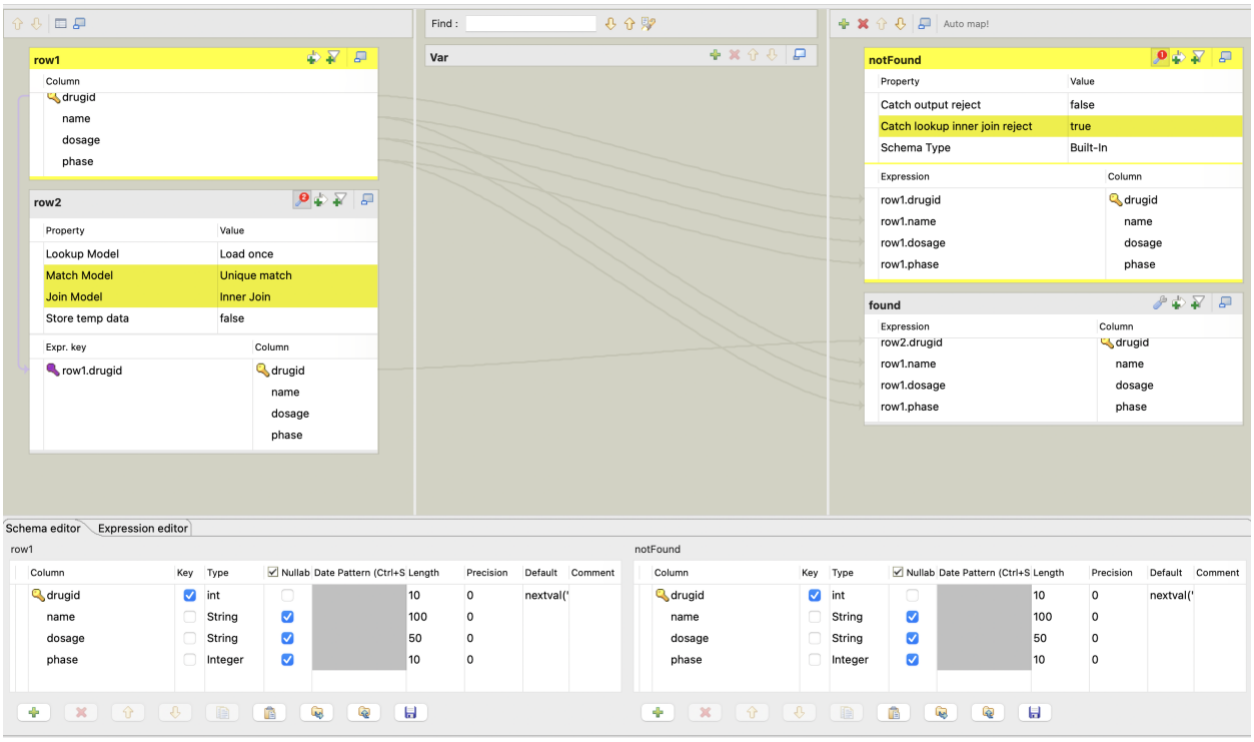

# 4.Protocol Dim:


oltp_healthcare
ok
OnComponentOk
olap_healthcare
ok
OnComponentOk
"protocol"
5 rows in 0.01s
1000 rows/s
row1 (Main)
tMap_1
5 rows in 0.08s
63.29 rows/s
found (Main order:2)
"protocol_dim"
5 rows in 0.02s
277.78 rows/s
row2 (Lookup)
0 rows in 0.16s
0 rows/s
notFound (Main order:1)
"protocol_dim"
"protocol_dim"
ok
OnComponentOk
tDBCommit_1
ok
OnComponentOk
tDBClose_1

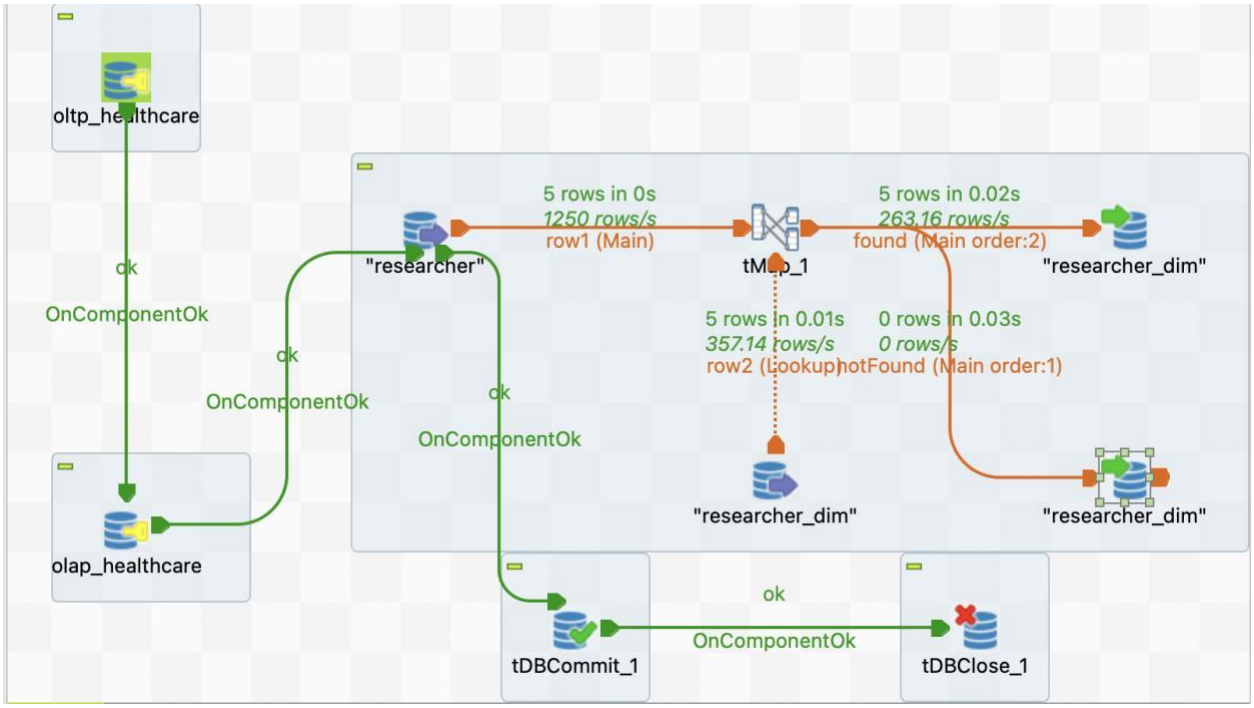

row1
Column
protocolid
name
description
version
row2
Property
Value
Lookup Model
Load once
Match Model
Unique match
Join Model
Inner Join
Store temp data
false
Expr. key
Column
row1.protocolid
protocolid
name
description
version
Find :
Var
Auto map!
notFound
Property
Value
Catch output reject
false
Catch lookup inner join reject
true
Schema Type
Built-In
Expression
Column
row1.protocolid
row1.name
row1.description
row1.version
found
Expression
Column
row2.protocolid
row1.name
row1.description
row1.version
Schema editor
Expression editor
row1
notFound
Column
Key
Type
Nullab
Date Pattern (Ctrl+S
Length
Precision
Default
Comment
protocolid
int
10
0
nextval('
name
String
100
0
description
String
2147483640
version
String
50
0

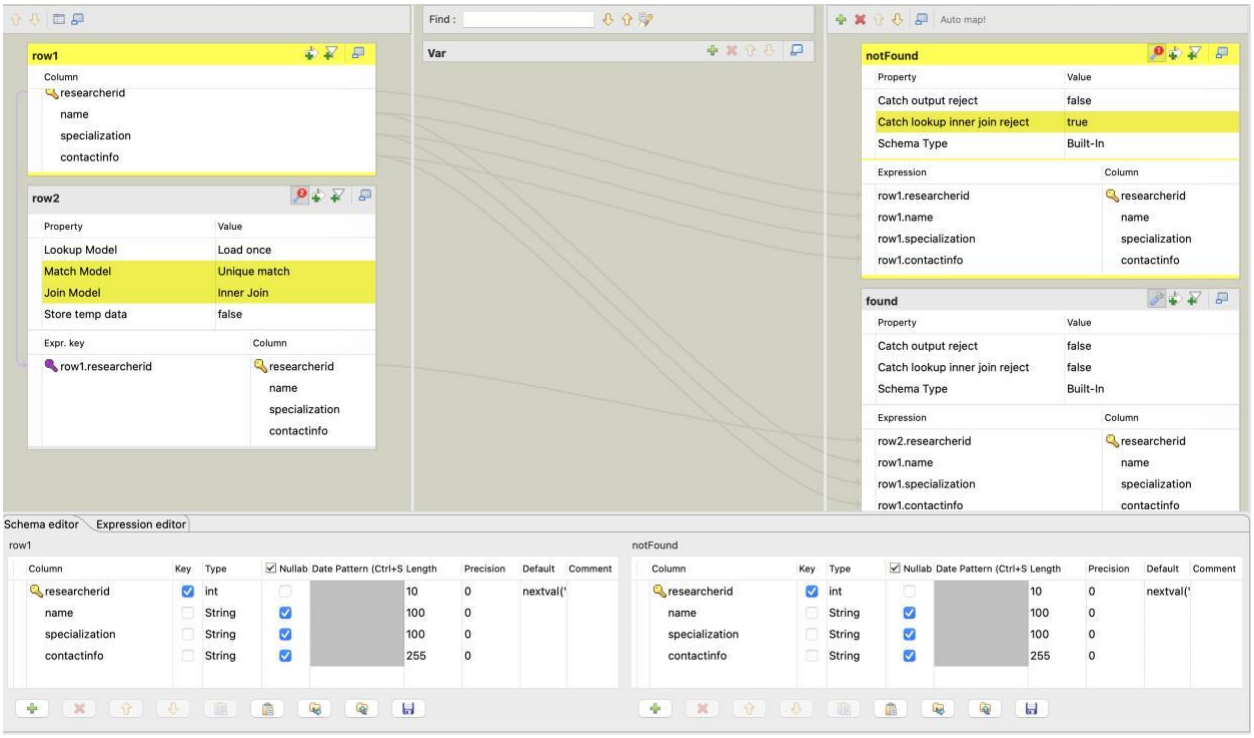
# 5.Drug Admin:



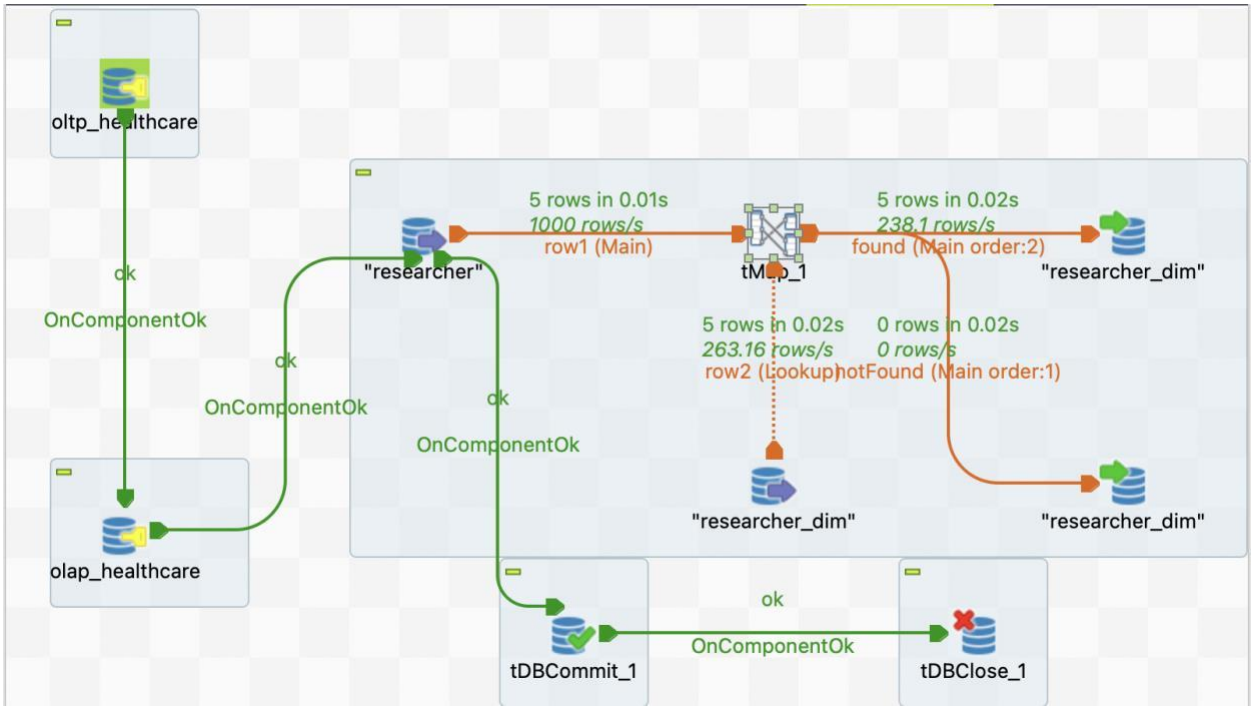



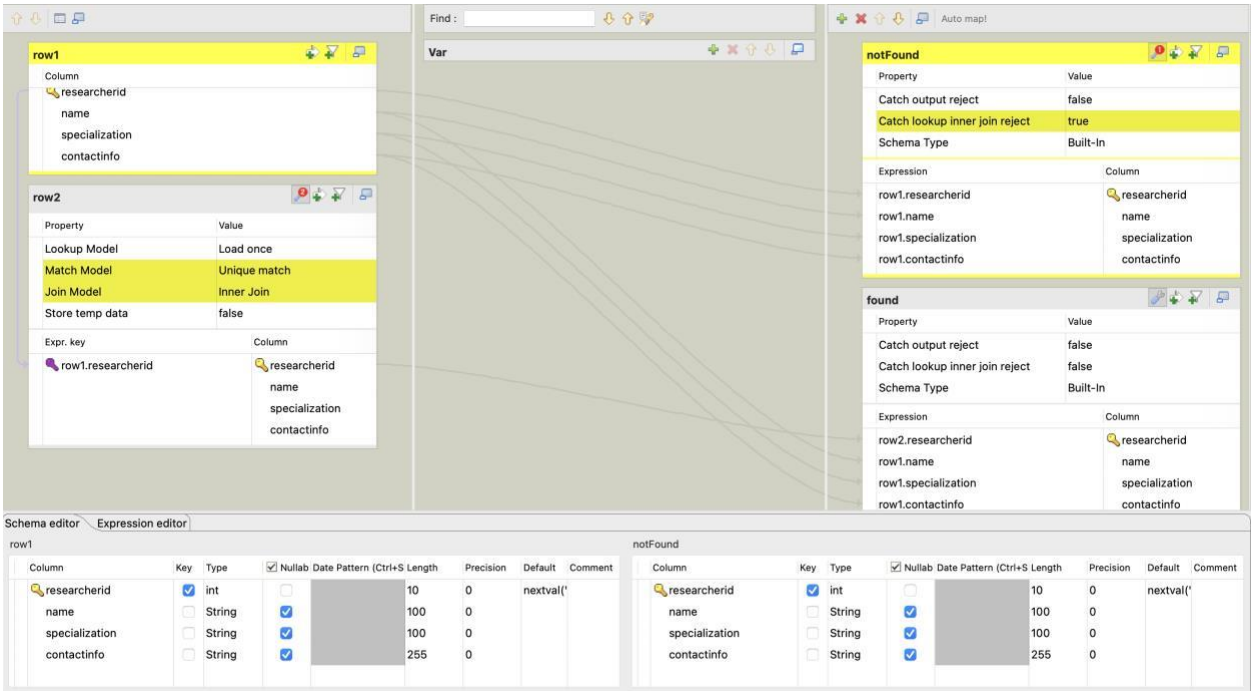

# 6.Researcher Dim:


oltp_healthcare
ok
OnComponentOk
"researcher"
5 rows in 0s
1250 rows/s
row1 (Main)
tMap_1
5 rows in 0.02s
263.16 rows/s
found (Main order:2)
"researcher_dim"
5 rows in 0.01s
357.14 rows/s
row2 (Lookup)
0 rows in 0.03s
0 rows/s
notFound (Main order:1)
ok
OnComponentOk
ok
OnComponentOk
"researcher_dim"
"researcher_dim"
olap_healthcare
ok
OnComponentOk
tDBCommit_1
tDBClose_1

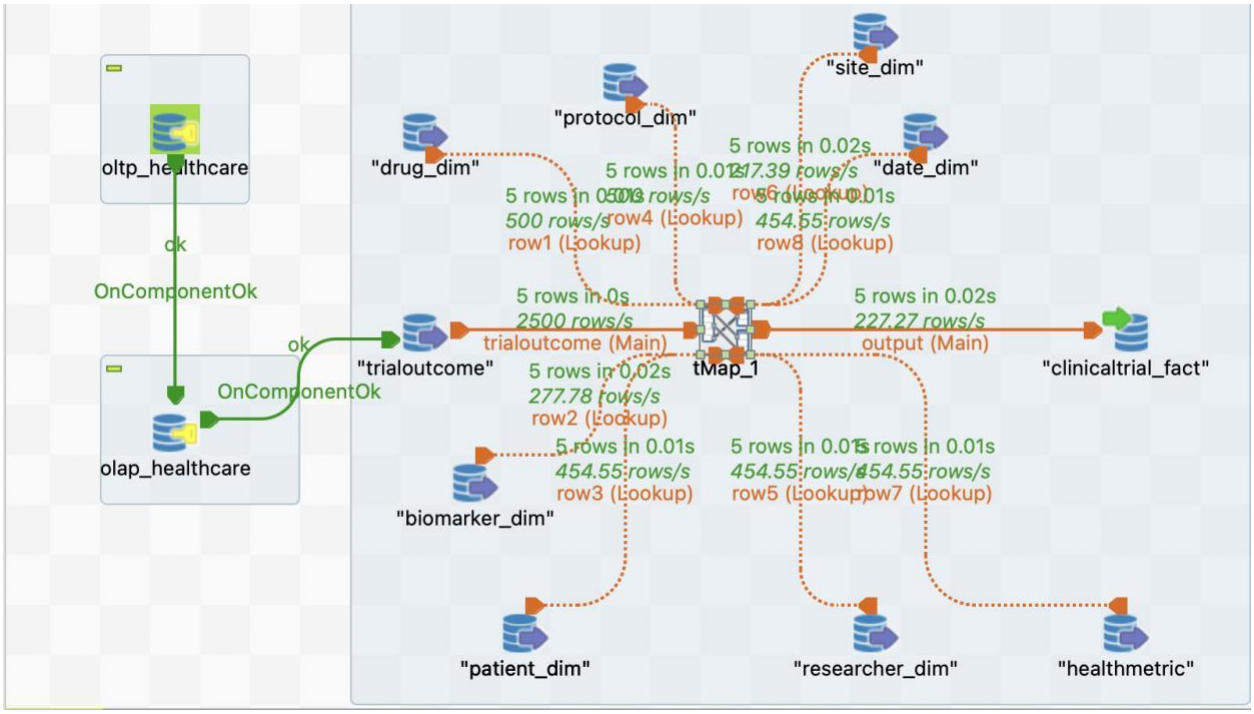

row1
Column
researcherid
name
specialization
contactinfo
row2
Property
Value
Lookup Model
Load once
Match Model
Unique match
Join Model
Inner Join
Store temp data
false
Expr. key
Column
row1.researcherid
researcherid
name
specialization
contactinfo
Find :
Var
notFound
Property
Value
Catch output reject
false
Catch lookup inner join reject
true
Schema Type
Built-In
Expression
Column
row1.researcherid
row1.name
row1.specialization
row1.contactinfo
found
Property
Value
Catch output reject
false
Catch lookup inner join reject
false
Schema Type
Built-In
Expression
Column
row2.researcherid
row1.name
row1.specialization
row1.contactinfo
Schema editor
Expression editor
row1
Column
Key
Type
Nullab
Date Pattern (Ctrl+S
Length
Precision
Default
Comment
researcherid
int
10
0
nextval('
name
String
100
0
specialization
String
100
0
contactinfo
String
255
0
notFound
Column
Key
Type
Nullab
Date Pattern (Ctrl+S
Length
Precision
Default
Comment
researcherid
int
10
0
nextval('
name
String
100
0
specialization
String
100
0
contactinfo
String
255
0

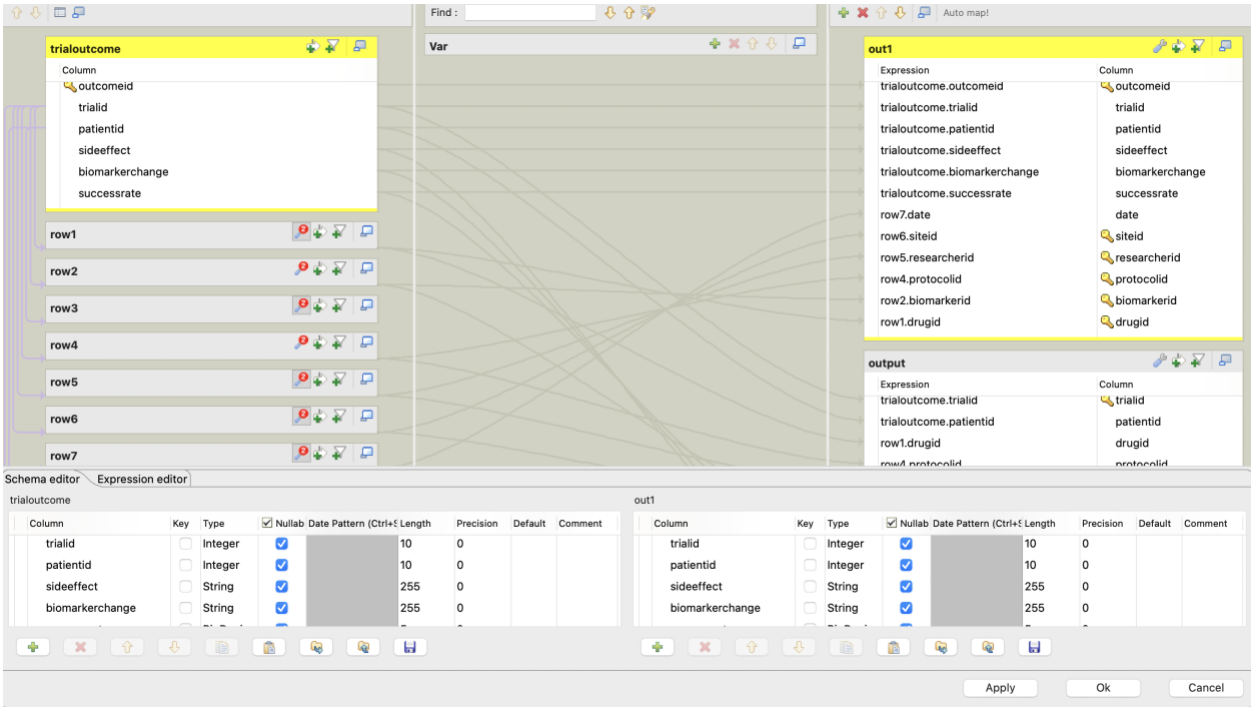
## 7.Site Dim:



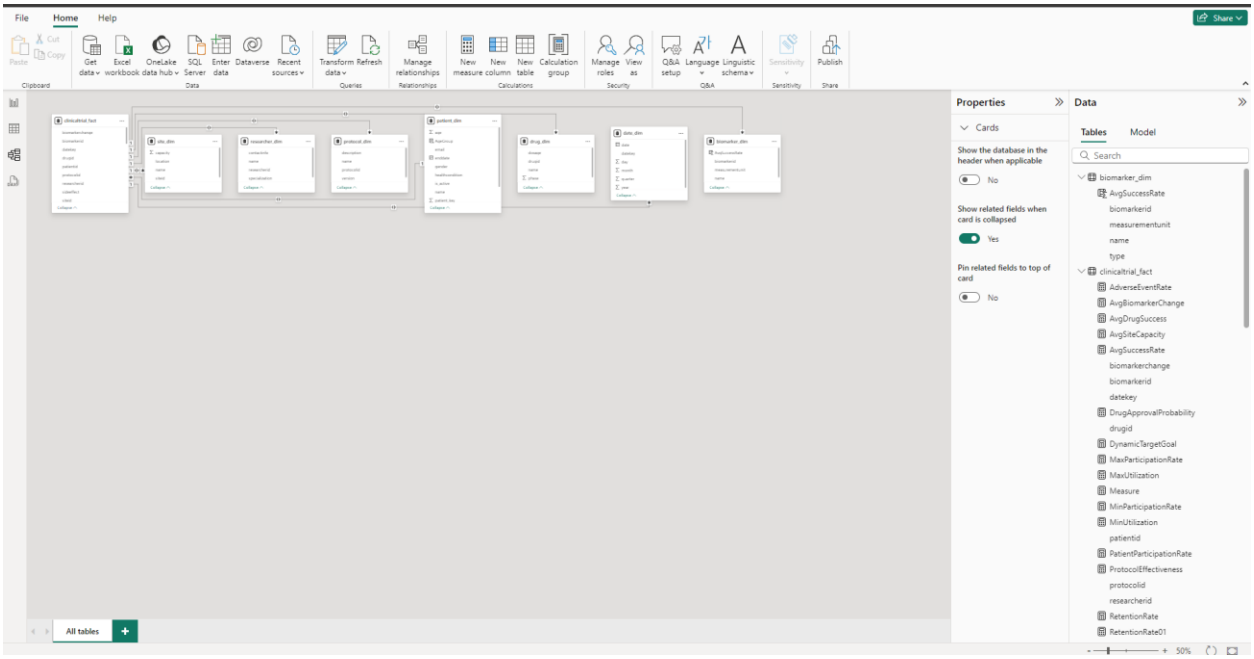

## 9. Trial Outcome Fact:

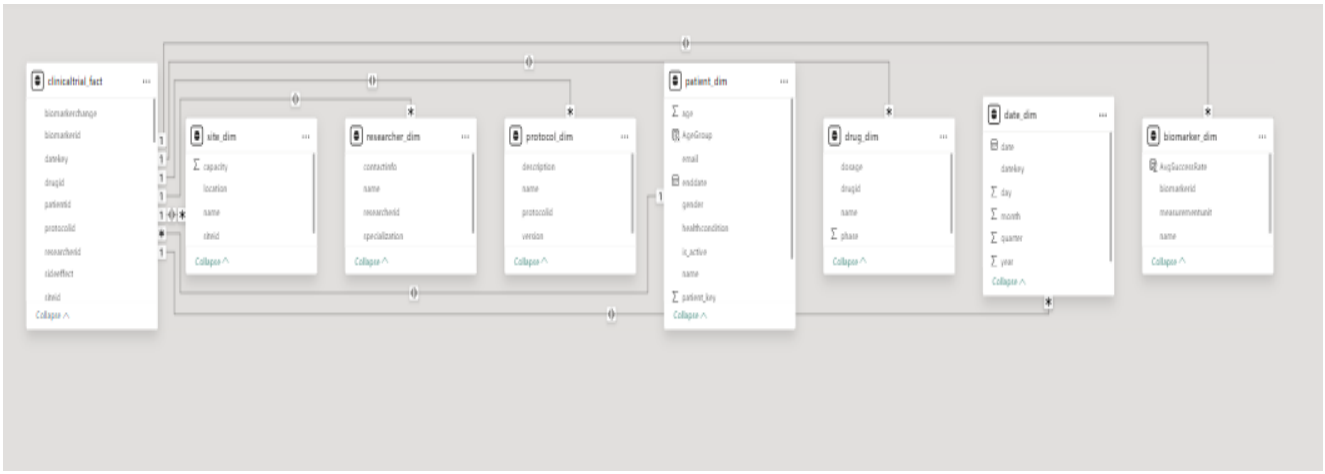



## KPIs, METRICS AND DASHBOARDS FOR ON PREMISE PROJECT:

## Data Model for Clinical Trial Data Warehouse to Power Bi(Entity-Relationship)




## Figure: Relation between the fact table and dimension tables

**KPIs:**

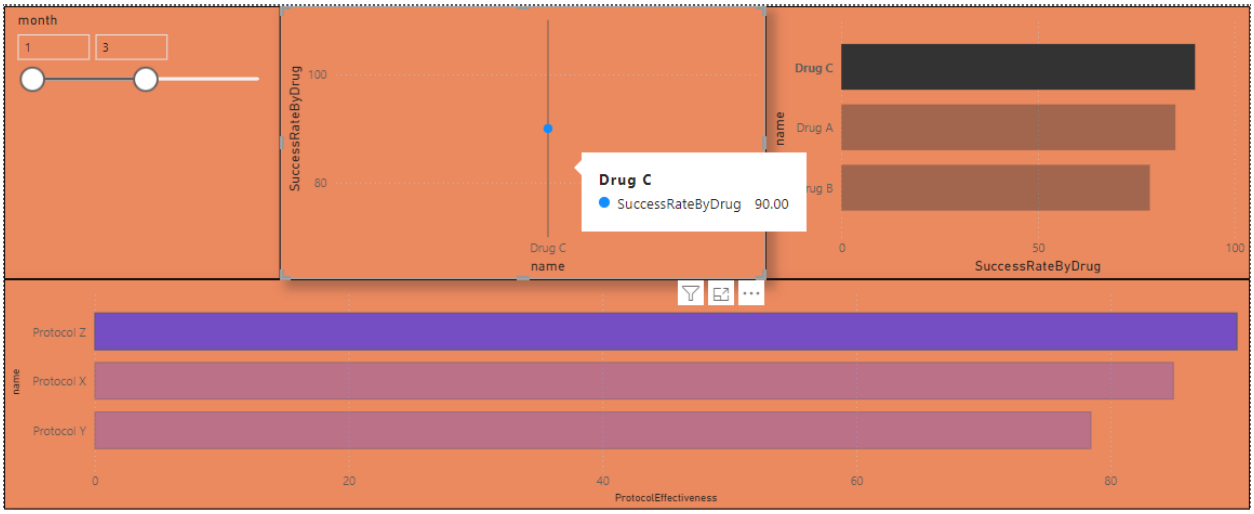

## 1. Gauge Chart

- **KPis:** Site Utilization
- Purpose:
  - Tracks the percentage utilization of site capacity against a defined goal (e.g., target utilization of 80%).
- Insight:
  - Allows stakeholders to monitor whether site resources are being effectively utilized.

## 2. Success Rate Over Time

**Visuals:** Gauge and Line Chart

- KPIs:
  - Average Success Rate by Date: Displays success rate trends over time with a comparison against a defined target.
- Purpose:
  - Provides insights into how success rates have evolved across different time periods.
- Insight:
  - Identifies whether clinical trials are improving over time and meeting expected benchmarks.




### 3. Drug Metrics

**Visuals:** Clustered Bar Chart and Scatter Plot

- KPIs:
  - Success Rate by Drug: Compares the success rates of different drugs.
- Purpose:
  - Identifies high-performing drugs and helps prioritize them in trials.
- Insight:
  - Enables data-driven decisions regarding which drugs to advance to the next trial phase.

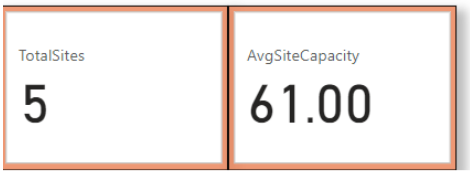
### 4. Protocol Effectiveness

**Visuals:** Bar Chart

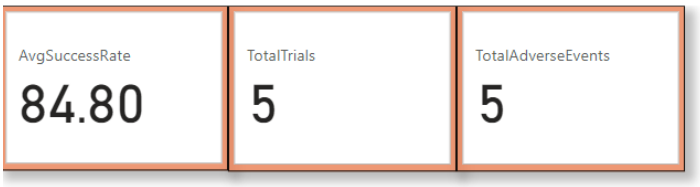
- KPIs:
  - Protocol Effectiveness: Compares the average success rates across various protocols.
- Purpose:
  - Evaluates the effectiveness of different clinical trial protocols.
- Insight:
  - Helps refine and standardize successful protocols to improve overall trial outcomes.

## Metrics:

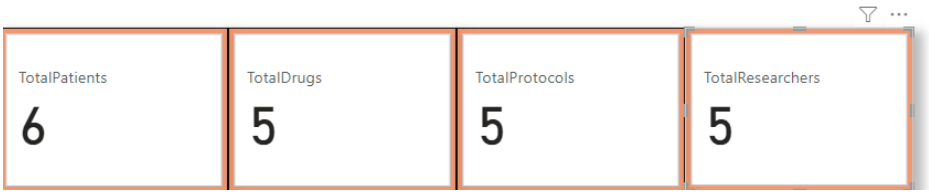
**Site Metrics**




**Clinical Trial Metrics**




**All Other Metrics**




### 1. Site Metrics

Visuals: Card Visuals

- Metrics:
  - Total Sites: Shows the total number of sites involved in clinical trials.
  - Average Site Capacity: Displays the average capacity of all sites.
- Purpose:
  - Provides a quick snapshot of the infrastructure and resource capacity available for conducting clinical trials.
- Insight:
  - Helps evaluate whether the site capacity is sufficient to meet the demands of ongoing and upcoming trials.

### 2. Clinical Trial Metrics

**Visuals:** Card Visuals

- Metrics:
  - Average Success Rate: Displays the overall average success rate of trials across all sites.
  - Total Trials: Shows the total number of trials conducted.
  - Total Adverse Events: Indicates the total number of adverse events reported during the trials.
- Purpose:
  - Gives a high-level overview of trial performance, including successes and potential safety issues.
- Insight:
  - Enables stakeholders to monitor trial success trends and identify the occurrence of adverse

### 3. All Other Metrics

**Visuals:** Card Visuals

- Metrics:
  - Total Patients: The total number of patients participating in trials.
  - Total Drugs: The total number of drugs being tested.
  - Total Protocols: The total number of distinct protocols used.
  - Total Researchers: The total number of researchers contributing to the trials.
- Purpose:
  - Highlights the scale of participation, drug variety, and research activity within clinical trials.
- Insight:
  - Assesses the engagement level of patients, variety in drug testing, and the overall research effort.

## Dashboards:


### Explanation of the Dashboard Components

This dashboard provides insights into clinical trial performance and site/patient metrics. Each section contributes to monitoring and optimizing trials, focusing on site capacity, success rates, trial data, and protocol effectiveness.

### 1. Metrics Overview (Top Row)

#### a. AvgSiteCapacity (61.00):

- **Description**: Average capacity of clinical trial sites, reflecting how many participants each site can handle effectively.
- **Insight**: If this number is low, it may indicate limited resources or smaller trial sizes.

#### b. TotalSites (5):

- **Description**: The total number of clinical trial sites participating in the study.
- **Insight**: A small number of sites might suggest localized trials, while a larger number could indicate multi-site studies.

#### c. AvgSuccessRate (84.80):

- **Description**: The average success rate across trials or sites.
- **Insight**: This is a high success rate, signaling that most trials are meeting their objectives. However, individual outliers should be analyzed.

#### d. TotalTrials (5):

- **Description**: The total number of trials being tracked in the system.

- **Insight**: The data suggests this is a focused analysis on a limited number of trials.

**e. TotalPatients (6):**

- **Description**: Total number of patients participating across all trials.
- **Insight**: A low patient count may indicate early-phase trials or limited recruitment.

**2. AvgSuccessRate, Dynamic by Date**

- **Metric**: 88.50% (with a goal of 75%, +13% improvement).
- **Description**: Tracks the overall success rate over time, comparing it to a predefined goal.
- **Insight**: The success rate exceeds the goal, which is a positive outcome. Continuous monitoring can ensure this trend sustains.

**3. Success Rate by Drug (Line Chart)**

- **Description**: A line chart showing success rates for different drugs (Drug C, Drug A, Drug B).
    - **Drug C**: Highest success rate.
    - **Drug A and Drug B**: Success rates decline progressively.
- **Insight**: Indicates variability in drug effectiveness. Focus may be needed on drugs with lower success rates to identify and address issues.

**4. Protocol Effectiveness (Bar Chart)**

- **Description**: Shows the effectiveness of various protocols (measured in %).
- **Insight**: Protocols with lower effectiveness can be revised for better outcomes, while highly effective protocols can serve as benchmarks.

**5. MinUtilization, MaxUtilization (Gauge Chart)**

- **Metric**: 5.00
- **Description**: Displays the utilization of resources, possibly reflecting the average or target for both minimum and maximum values.
- **Insight**: A balanced utilization ensures efficient resource management. If the value trends toward the extremes, resources may be over- or under-utilized.

## Key Dashboard Insights:

1. **High-Level Summary**: The top-row metrics provide an at-a-glance view of the clinical trial landscape, ensuring stakeholders can quickly assess trial and site performance.
2. **Drug Success**: The success rate by drug chart highlights variability, which could guide prioritization of drugs for further testing or improvement.
3. **Protocol Optimization**: The protocol effectiveness data ensures that the best practices are being followed and can identify areas for improvement.
4. **Resource Efficiency**: The utilization gauge focuses on resource allocation, ensuring neither overburdening nor underutilization of assets.
5. **Goal Achievement**: The AvgSuccessRate exceeding the goal reflects positive outcomes but requires sustained efforts to maintain.