DALHOUSIE UNIVERSITY

CSCI 5408 Data Management, Warehousing, And Analytics

Group Project – Final Report

GROUP 11

Team Members:

1. Sivasubramanian Venkatasubramanian [B00956463]
2. Ashish Kumar Guntipalli [B00981016]
3. Krutikbhai Narendrabhai Kulkarni [B00981839]
4. Bhishman Desai [B00945177]
5. Luv Prakashkumar Patel [B00950942]

Gitlab Link: GDC Program

Table of Contents

Information Table

Table 1: Information Collection Table

Name of Source	**Information Collected**	**URL**
McKinsey& Company	The stakeholders involved in the system	https://www.mckinsey.com/industries/life-sciences/our-insights/
Forbes	Recent Trends in Pharma Supply Chain Management and challenges	https://www.forbes.com/sites/forbestechcouncil/2023/05/11/
The Canadian Encyclopaedia	Learnt about Drug regulation in the pharmaceutical industry	https://www.thecanadianencyclopedia.ca/en/article/pharmaceutical-industry
ProQuest	Information related to the supply chain logistics and structure associated with transportation of the pharmaceutical drugs. The flow of drug between the company producing it, the logistics required to transport it and the relation between the distributor and the retailer was highlighted in this article. This helped in finding the entities related to retailer, wholesaler, and drug company.	https://www.proquest.com/docview/2677576111
Healthline	Phases and procedures involved in testing a medication on test subjects	Clinical Trial Phases: What Happens in Phase 0, I, II, III, and IV (healthline.com)

Background Research Summary:

Pharmacy supply chain management is the process of ensuring the efficient and effective delivery of pharmaceutical products from the manufacturers to the end-users, such as patients, hospitals, and pharmacies. It involves various stakeholders, such as raw material suppliers, drug manufacturers, regulatory agencies, wholesale distributors, pharmacies, pharmacy benefit managers (PBMs), healthcare providers, and patients. Each stakeholder plays a crucial role and proper coordination between them is essential for ensuring the quality, safety, and availability of medications [1][2].

The pharmacy supply chain faces several challenges, such as supply chain visibility, drug counterfeiting, cold-chain shipping, and rising prescription drug prices. These challenges can pose risks to the public health, patient safety, and profitability of the pharmaceutical industry. Therefore, it is important to adopt strategies to make the pharmacy supply chain more resilient, such as greater visibility, rigorous risk management, and newer technologies that help companies better anticipate and respond to shocks [2]. The old model of supply chain included point-to-point connections, which were inefficient and prone to errors.

Whilst some of the reviewed studies have examined how firms use B2B networks as an uncertainty coping mechanism, these studies have focussed exclusively on pharmaceutical firms. In comparison, other types of firms, such as hospitals, have not been examined. It would be interesting to see if hospitals have a different way of organising their resource networks during uncertainties [4].

Clinical trials are essential for evaluating the safety and effectiveness of new drugs or treatments. [5] These trials typically progress through four phases, starting with small-scale tests in Phase 0 to assess initial safety, followed by Phase I to establish dosage and safety parameters. Phase II involves testing on a larger group of participants with the target condition, while Phase III compares the new treatment with existing ones to determine its efficacy and safety. Finally, Phase IV occurs after regulatory approval, focusing on long-term effects and benefits. Each phase plays a crucial role in ensuring that only safe and effective treatments are introduced to the public. This information helped in identification of the drug approval process and trail phases which are helpful in deducing entities for the database design.

All this research on understanding how pharmaceutical supply chains operate and how companies develop new drugs for constantly changing diseases, gaining regulatory approval, manufacturing, and distributing them to pharmacies and ultimately to patients provided insight into the internal workings of the process. This overview also gave us overview of the key entities involved, informing us about the requirements for building a distributed database.

Initial Conceptual Model

Link for clear and complete Image: [Initial_ERD]

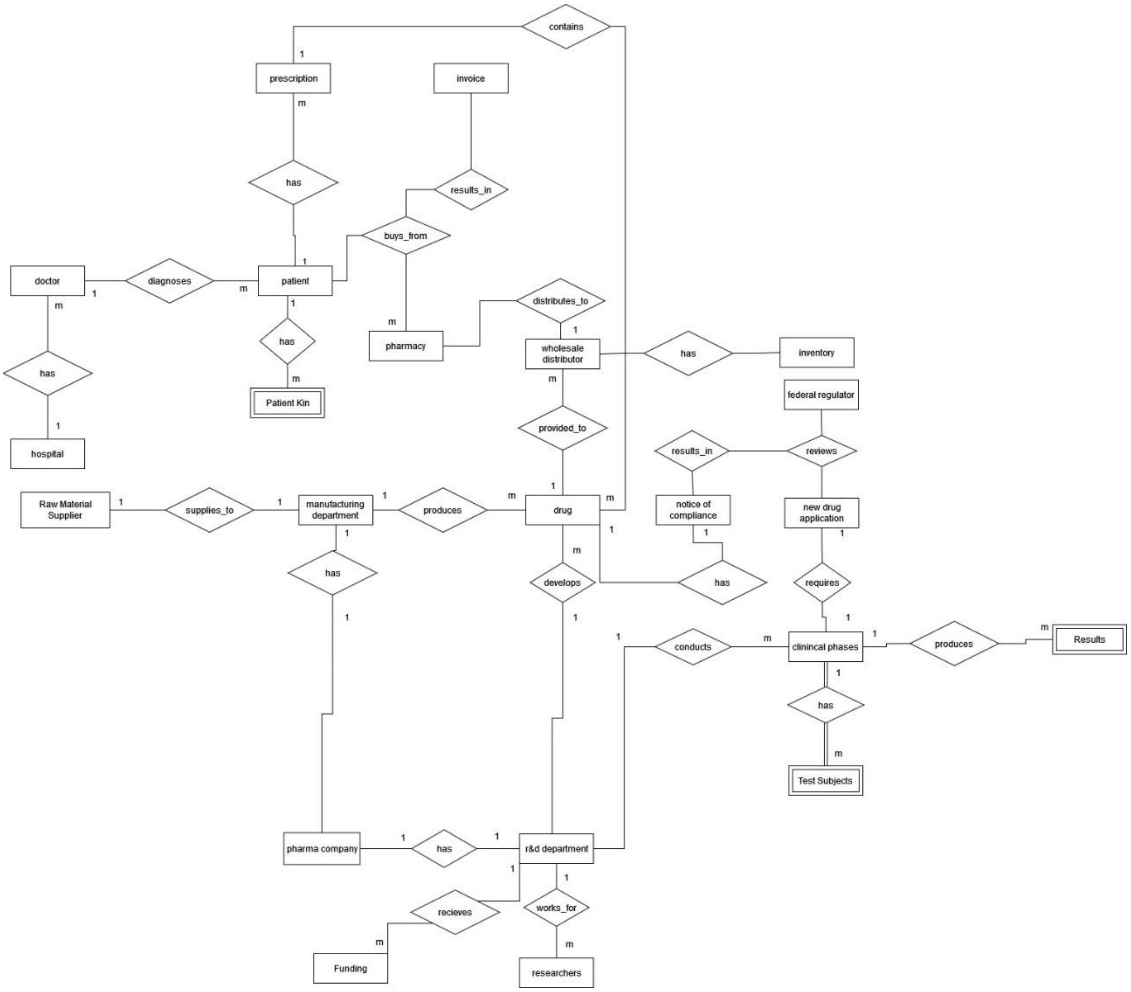

Figure 1: Initial ER Diagram

Design Issues

Two types of design issues were identified from our initial conceptual model which are explained below

Chasm Trap:

Explanation:

The entities “pharma company”, “r&d department”, and “researcher” are involved in a chasm trap. There may be situation when a researcher is not affiliated to the research department, if they are not affiliated then we won't know under which pharmacy company they work for. So, there is an ambiguity or uncertainty about the relationship between researcher and pharma company.

How it was fixed:

To fix this we have established a relationship called “work_for” between pharma company and researchers. Now, no matter what the researcher will always be part of the pharma company and the ambiguity is resolved.

Fan Trap:

Explanation:

The entities “r&d department”, “drug” and “clinical phases” are involved in a fan trap. r&d department develops many drugs and r&d department conducts many clinical phases, but we are not able to specifically say which clinical phase was done for which drug.

How it was fixed:

To solve that we insert “drug” entity between “r&d department” and “clinical phase”. “drug” has “clinical phases” and “r&d department” develops “drug” this way we would be able to deduce which r&d department conducted the clinical phase and consolidate all the clinical phases for a specific drug.

Final ER Model

Link for clear and complete Image: [Final_ERD]

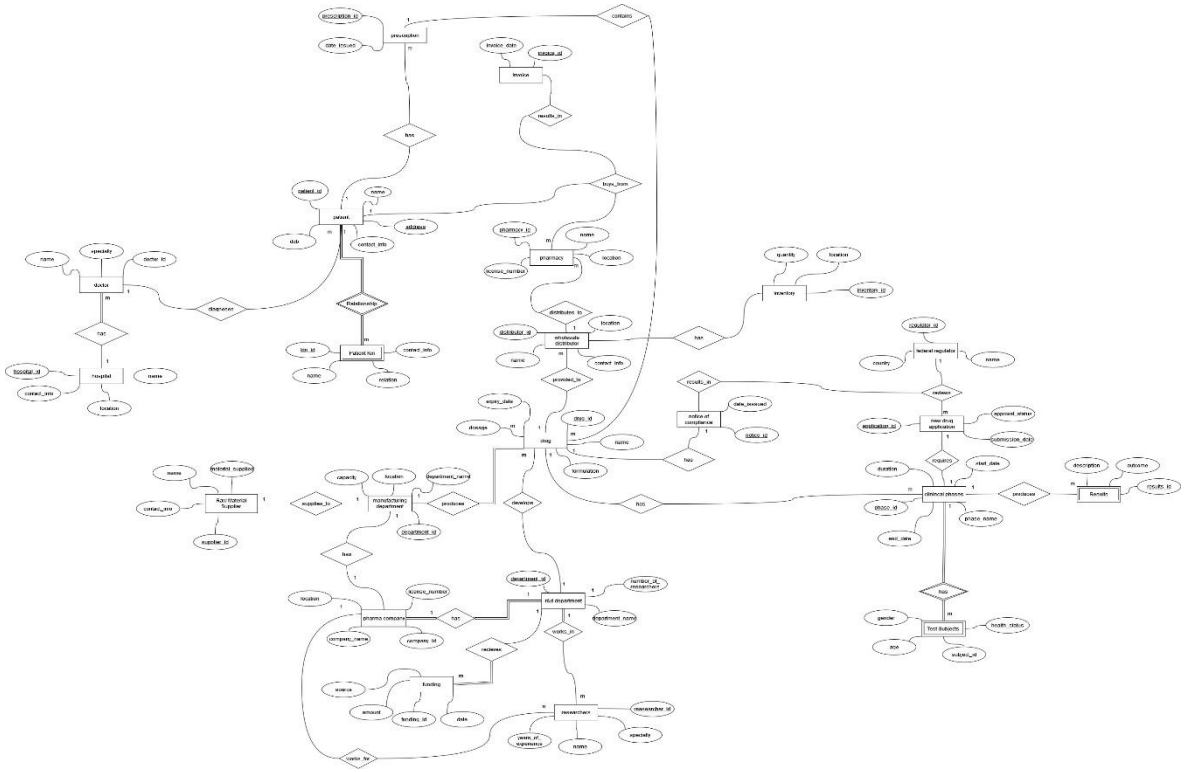

Figure 2: Final ER Diagram with attributes

Logical Phase

There were no partial or transitive dependencies identified in the Final ER Diagram. That's mainly because we built our initial model which is optimized to solve our supply chain visibility problems. Another reason for that is our initial model have a very limited number of attributes for every entity since we wanted to have a simple structure with no redundant data.

Fragmentation Decision:

Pharma companies in India does research on drugs and develop them with the funding received from external entities, these drugs undergo rigorous testing and approval process for safety. Once approval is done, since raw material and labour costs are cheap in India the pharma companies prefer that the drug be manufactured there and exported to USA where wholesale distributors acquire inventory and distribute to retail. Then, patients will get diagnosed by doctors who prescribe them required drug for treatment, which is provided by the pharmacies.

We opted for Database-level fragmentation. Because, drugs are manufactured and developed in India tables relevant to that domain are placed in vmysql1 (Mumbai) instance, where are the drugs are approved, tested and sold in USA so all the related entities to this part of the mini-world are placed in vmysql2 (America) Instance. This approach is chosen to keep related data closer together, making it more efficient to manage and access. For example, having manufacturing-related data stored in the same location as the manufacturing facilities simplifies data management for that part of the process. Likewise, having approval and sales-related data stored together in another location streamlines operations for those activities.

DDL Queries:

VMysql1 Instance

```
Create database phrama;
use phrama;

-- INDIA
-- Raw Material Supplier table
CREATE TABLE RawMaterialSupplier (
    SupplierID INT PRIMARY KEY,
    Name VARCHAR(100) NOT NULL,
    MaterialSupplied VARCHAR(100) NOT NULL,
    ContactInfo VARCHAR(100) NOT NULL
);

-- Manufacturing Department table
CREATE TABLE ManufacturingDepartment (
    DepartmentID INT PRIMARY KEY,
    DepartmentName VARCHAR(100) NOT NULL,
    Location VARCHAR(100) NOT NULL,
    Capacity INT NOT NULL,
    DrugID INT,
    SupplierID INT,
    CompanyID INT,
--  FOREIGN KEY (DrugID) REFERENCES Drug(DrugID),
    FOREIGN KEY (SupplierID) REFERENCES RawMaterialSupplier(SupplierID),
    FOREIGN KEY (CompanyID) REFERENCES PharmaCompany(CompanyID)
);
```

```sql
-- Pharma Company table
CREATE TABLE PharmaCompany (
    CompanyID INT PRIMARY KEY,
    CompanyName VARCHAR(100) NOT NULL,
    Location VARCHAR(100) NOT NULL,
    LicenseNumber VARCHAR(50) NOT NULL
);

-- R&D Department table
CREATE TABLE RDDepartment (
    DepartmentID INT PRIMARY KEY,
    DepartmentName VARCHAR(100) NOT NULL,
    NumResearchers INT NOT NULL,
    CompanyID INT,
    FOREIGN KEY (CompanyID) REFERENCES PharmaCompany(CompanyID)
);

-- Funding table
CREATE TABLE Funding (
    FundingID INT PRIMARY KEY,
    Amount DECIMAL(15, 2) NOT NULL,
    Source VARCHAR(100) NOT NULL,
    DateReceived DATE NOT NULL,
    DepartmentID INT,
    FOREIGN KEY (DepartmentID) REFERENCES RDDepartment(DepartmentID)
);

-- Researchers table
CREATE TABLE Researchers (
    ResearcherID INT PRIMARY KEY,
    Name VARCHAR(100) NOT NULL,
    Specialty VARCHAR(100) NOT NULL,
    YearsOfExperience INT NOT NULL,
    CompanyID INT,
    DepartmentID INT,
    FOREIGN KEY (CompanyID) REFERENCES PharmaCompany(CompanyID),
    FOREIGN KEY (DepartmentID) REFERENCES RDDepartment(DepartmentID)
);
```

Vmysql2 Instance

```sql
-- create database vmysql2;
use vmysql2;

create table GDC (
id int,
location_name varchar(255),
table_name varchar(255),
```

```
primary key(id)
);

INSERT INTO GDC (id, location_name, table_name)
VALUES
  (1, 'India', 'Pharma company'),
  (2, 'India', 'R&D department'),
  (3, 'India', 'Researchers'),
  (4, 'India', 'Funding'),
  (5, 'India', 'Raw material supplier'),
  (6, 'India', 'Manufacturing department'),
  (7, 'USA', 'Drug'),
  (8, 'USA', 'Wholesale Distributor'),
  (9, 'USA', 'Pharmacy'),
  (10, 'USA', 'Invoice'),
  (11, 'USA', 'Prescription'),
  (12, 'USA', 'Hospital'),
  (13, 'USA', 'Patient'),
  (14, 'USA', 'Doctor'),
  (15, 'USA', 'Patient Kin'),
  (16, 'USA', 'Inventory'),
  (17, 'USA', 'Notice of compliance'),
  (18, 'USA', 'Federal regulator'),
  (19, 'USA', 'New drug application'),
  (20, 'USA', 'Clinical phases'),
  (21, 'USA', 'Test subjects'),
  (22, 'USA', 'Results');

  drop table GDC;



-- Drug table
CREATE TABLE Drug (
    DrugID INT PRIMARY KEY,
    Name VARCHAR(100) NOT NULL,
    Dosage VARCHAR(50) NOT NULL,
    Formulation VARCHAR(100) NOT NULL,
    ExpiryDate DATE NOT NULL,
    DepartmentID INT
);

-- Notice of Compliance table
CREATE TABLE NoticeOfCompliance (
    NoticeID INT PRIMARY KEY,
    DateIssued DATE NOT NULL,
    DrugID INT NOT NULL,
    RegulatorID INT NOT NULL,
    ApplicationID INT,
```

```
    FOREIGN KEY (DrugID) REFERENCES Drug(DrugID),
    FOREIGN KEY (RegulatorID) REFERENCES FederalRegulator(RegulatorID),
    FOREIGN KEY (ApplicationID) REFERENCES NewDrugApplication(ApplicationID)
);

-- New Drug Application table
CREATE TABLE NewDrugApplication (
    ApplicationID INT PRIMARY KEY,
    SubmissionDate DATE NOT NULL,
    ApprovalStatus VARCHAR(50) NOT NULL,
    DrugID INT NOT NULL,
--      FOREIGN KEY (DrugID) REFERENCE Drug(DrugID),
    PhaseID INT,
    RegulatorID INT,
    FOREIGN KEY (PhaseID) REFERENCES ClinicalPhases(PhaseID),
    FOREIGN KEY (RegulatorID) REFERENCES FederalRegulator(RegulatorID)
);

-- Clinical Phases table
CREATE TABLE ClinicalPhases (
    PhaseID INT PRIMARY KEY,
    PhaseName VARCHAR(50) NOT NULL,
    StartDate DATE NOT NULL,
    EndDate DATE NOT NULL,
    DrugID INT,
    FOREIGN KEY (DrugID) REFERENCES Drug(DrugID)
);

CREATE TABLE Results (
    ResultsID INT,
    Description VARCHAR(255) NOT NULL,
    Outcome VARCHAR(255) NOT NULL,
    PhaseID INT,
    PRIMARY KEY (PhaseID, ResultsID),
    FOREIGN KEY (PhaseID) REFERENCES ClinicalPhases(PhaseID)
);


-- Test Subjects table
CREATE TABLE TestSubjects (
    SubjectID INT,
    Age INT NOT NULL,
    Gender VARCHAR(10) NOT NULL,
    HealthStatus VARCHAR(50) NOT NULL,
    PhaseID INT,
    PRIMARY KEY (PhaseID, SubjectID),
    FOREIGN KEY (PhaseID) REFERENCES ClinicalPhases(PhaseID)
);
```

```sql
-- Federal Regulator table
CREATE TABLE FederalRegulator (
    RegulatorID INT PRIMARY KEY,
    Name VARCHAR(100) NOT NULL,
    Country VARCHAR(100) NOT NULL
);

-- Wholesale distributor table
CREATE TABLE WholesaleDistributor (
    DistributorID INT PRIMARY KEY,
    Name VARCHAR(100) NOT NULL,
    Location VARCHAR(100) NOT NULL,
    ContactInfo VARCHAR(100) NOT NULL,
    DrugID INT,
    FOREIGN KEY (DrugID) REFERENCES Drug(DrugID)
);

-- Pharmacy table
CREATE TABLE Pharmacy (
    PharmacyID INT PRIMARY KEY,
    Name VARCHAR(100) NOT NULL,
    Location VARCHAR(100) NOT NULL,
    LicenseNumber VARCHAR(50) NOT NULL,
    InvoiceID INT,
    DistributorID INT,
    FOREIGN KEY (DistributorID) REFERENCES WholesaleDistributor(DistributorID)
);

-- Patient table
CREATE TABLE Patient (
    PatientID INT PRIMARY KEY,
    Name VARCHAR(100) NOT NULL,
    DOB DATE NOT NULL,
    Address VARCHAR(100) NOT NULL,
    ContactInfo VARCHAR(100) NOT NULL,
    InvoiceID INT,
    DoctorID INT,
    FOREIGN KEY (DoctorID) REFERENCES Doctor(DoctorID)
);

CREATE TABLE Invoice (
    InvoiceID INT PRIMARY KEY NOT NULL,
    InvoiceDate DATE NOT NULL,
    PatientID INT,
    PharmacyID INT,
    FOREIGN KEY (PatientID) REFERENCES Patient(PatientID),
    FOREIGN KEY (PharmacyID) REFERENCES Pharmacy(PharmacyID)
);
```

```sql
-- Prescription table
CREATE TABLE Prescription (
    PrescriptionID INT PRIMARY KEY,
    DateIssued DATE NOT NULL,
    PatientID INT NOT NULL,
    DoctorID INT NOT NULL,
    DrugID INT NOT NULL,
    FOREIGN KEY (PatientID) REFERENCES Patient(PatientID),
    FOREIGN KEY (DoctorID) REFERENCES Doctor(DoctorID),
    FOREIGN KEY (DrugID) REFERENCES Drug(DrugID)
);

-- Doctor table
CREATE TABLE Doctor (
    DoctorID INT PRIMARY KEY,
    Name VARCHAR(100) NOT NULL,
    Specialty VARCHAR(100) NOT NULL,
    HospitalID INT NOT NULL,
    FOREIGN KEY (HospitalID) REFERENCES Hospital(HospitalID)
);

-- Hospital table
CREATE TABLE Hospital (
    HospitalID INT PRIMARY KEY,
    Name VARCHAR(100) NOT NULL,
    Location VARCHAR(100) NOT NULL,
    ContactInfo VARCHAR(100) NOT NULL
);



-- Patient Kin table
CREATE TABLE PatientKin (
    KinID INT PRIMARY KEY,
    Name VARCHAR(100) NOT NULL,
    Relation VARCHAR(50) NOT NULL,
    ContactInfo VARCHAR(100) NOT NULL,
    PatientID INT NOT NULL,
    FOREIGN KEY (PatientID) REFERENCES Patient(PatientID)
);

-- Inventory table
CREATE TABLE Inventory (
    InventoryID INT PRIMARY KEY,
    DrugID INT NOT NULL,
    Quantity INT NOT NULL,
    Location VARCHAR(100) NOT NULL,
```

```
    DistributorID INT,
    FOREIGN KEY (DrugID) REFERENCES Drug(DrugID),
    FOREIGN KEY (DistributorID) REFERENCES WholesaleDistributor(DistributorID)
);
```

Distributed Database creation:

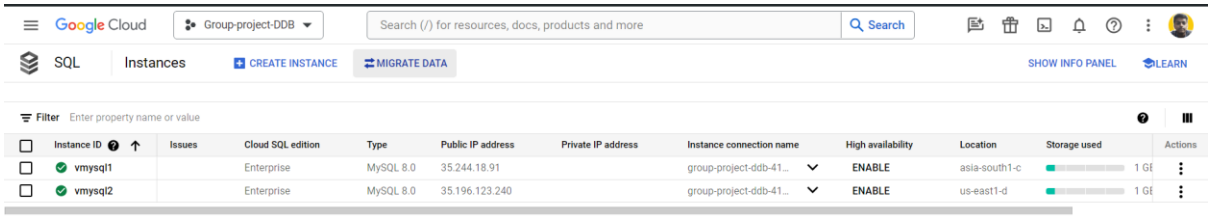

Figure 3: MYSQL Instances created in Google cloud

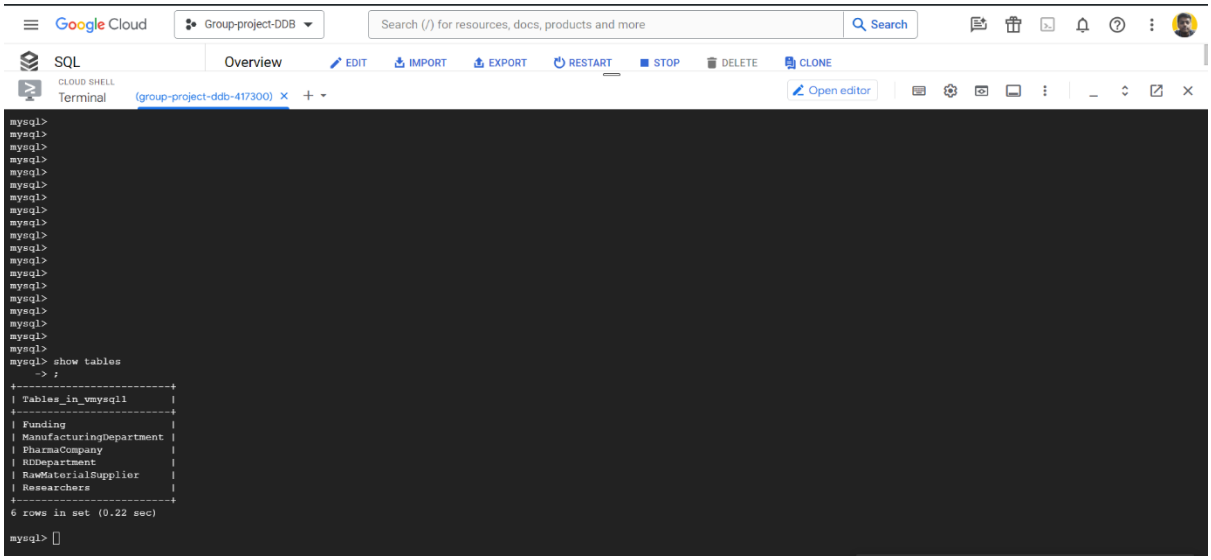

Figure 4: Tables in vmysql1(India) instance

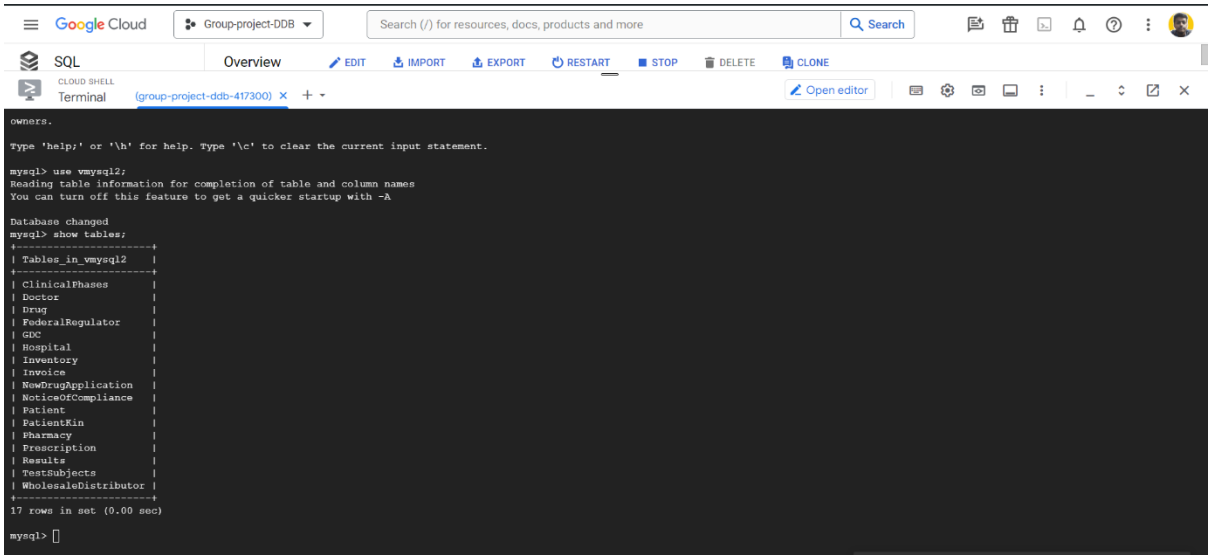

Figure 5: Tables in vmysql2(USA) Instance

DDB Structure:

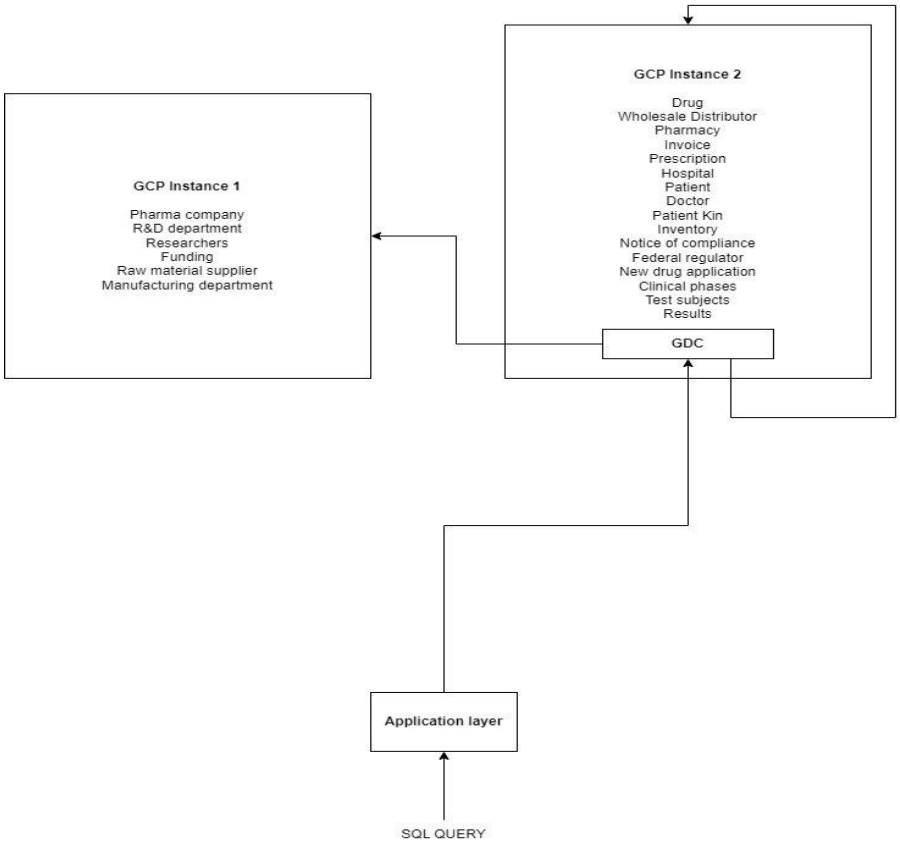

Figure 6: Architecture of Distributed Database Created

The above diagram represents the overall Distributed database structure for our pharma industry mini-world. GCP instance 1 is in India and GCP instance 2 is in USA. SQL query is entered into the application

layer which is a java-based program and it forwards that query to GDC which is In GCP instance 2. The GDC routes the query depending on the table name to respective location.

Global Distributed Catalogue (GDC) Placement:

The GDC table contains all the entities that are in the mini-world. It holds the table names and location names. The GDC is place in the USA instance, because the GDC is more frequently used in the place where the distribution happens rather than the place where it is manufactured. It acts as a bridging table to which all the queries are forwarded. It resolves to which location the query has to be forwarded.

id	location_name	table_name
1	India	Pharma company
2	India	R&D department
3	India	Researchers
4	India	Funding
5	India	Raw material supplier
6	India	Manufacturing department
7	USA	Drug
8	USA	Wholesale Distributor
9	USA	Pharmacy
10	USA	Invoice
11	USA	Prescription
12	USA	Hospital
13	USA	Patient
14	USA	Doctor
15	USA	Patient Kin
16	USA	Inventory
17	USA	Notice of compliance
18	USA	Federal regulator
19	USA	New drug application
20	USA	Clinical phases
21	USA	Test subjects
22	USA	Results

Figure 7: GDC table with tables names and their respective locations

Novelty:

An application layer for the distributed database was developed, in java, that takes user input of table name and searches the GDC table to find in which location is the table stored and redirects the query to that location instance in the GCP. The program fetches the table schema and displays on the console.

Meeting Logs

Table 2: Meeting Logs of the team

Meeting Date	Meeting Time	Meeting place	Discussion	Outcomes	Meeting Duration
Feb 17	7:11 PM	Online	Intro with team members	Got to know each other	19 Mins
Feb 18	8:35 PM	Online	Discussed about project specification	Got an idea of how to proceed	12 Mins
Feb 21	11:00 AM	Online	We split up tasks for sprint 1	Each member got clarity	24 Mins

				about their task	
Feb 22	11:11 AM	Online	Every member came up with few research/news articles	Got some knowledge about pharma industry	1hr 14 Mins
Feb 22	8:18 PM	Online	Considered most relevant news articles/ research papers	Finalised some entities	36 Mins
Feb 22	10:05 PM	Online	We made initial conceptual model and identified design issues	Initial ER Diagram was created	52 Mins
Feb 24	7:17 PM	Online	Modified ER diagram to solve design issues and added attributes	Final ERD was created	52 Mins
Feb 27	1:05 Pm	Online	We created initial draft of the report for sprint 1	Initial draft was created	37 Mins
Feb 27	8:40 PM	Online	Modified the initial draft to prepare for submission	Completed report for sprint 1 and submitted	14 Mins
Mar 6	9:08 PM	Online	Distributed work to identify any dependencies in ER model	Assignment of individual tasks	15 Mins
Mar 11	9:52 PM	Online	Discussed about our findings regarding dependencies	No partial or transitive dependencies were found	40 Mins
Mar 12	9:42 PM	Online	Discussed about DDL statements required for table creation	Completed DDL statements	1hr
Mar 13	6:44 PM	Online	Discussed about how to Configure GCP based on our mini world	Created GCP MySQL instances	1hr 59 mins
Mar 13	10:17 PM	Online	Discussed about type of fragmentation suitable for our mini world	Decided on performing DB level fragmentation	1hr 27 Mins
Mar 14	9:32 Pm	Online	Split up tables and created them in their respective MySQL instances on GCP	Remote instances were setup and ready to use	2hrs 15 Mins
Mar 15	11 AM	Online	Discussed about GDC placement and its structure	Created GDC table in one of GCP instance	1hr 3 Mins
Mar 15	12:09 PM	Online	Wrote an application layer that routes input queries to GDC	Created an application layer for our	1hr

				Distributed database	
Mar 15	10:07 PM	Online	Testing the final distributed system	Verified that DB was working as intended	1hr
Mar 20	6 PM	Online	Created initial draft of Final project report	Initial draft of final report was created	23 Mins
Mar 26	7:20 PM	Online	Finalized and completed the project report	Successfully prepared report for submission	2 hrs

References:

[1] T. Foster, P. Patel and K. Skiba, “Four ways pharma companies can make their supply chains more resilient”, “*McKinsey & Company*”, [Online] , Available: https://www.mckinsey.com/industries/life-sciences/our-insights/four-ways-pharma-companies-can-make-their-supply-chains-more-resilient [Accessed: Feb 22, 2024].

[2] M. Walker, “Top 10 Pharma Supply Chain Trends For 2023: Part One”, “*Forbes*”, [Online], Available: https://www.forbes.com/sites/forbestechcouncil/2023/05/11/top-10-pharma-supply-chain-trends-for-2023-part-one/?sh=40609d5d16a6 [Accessed: Feb 23, 2024]

[3] J. Lexchin, “Pharmaceutical Industry”, “*The Canadian Encyclopaedia*”, [Online], Available: https://www.thecanadianencyclopedia.ca/en/article/pharmaceutical-industry [Accessed: Feb 25, 2024]

[4] Madanaguli, A. Thirumalesh, Dhir, Amandeep, T. Shalini, G. Singh, O. Escobar, “Business to business (B2B) alliances in the healthcare industry: a review of research trends and pertinent issues”, “*The Journal of business & industrial marketing*” [Online], Vol. 37, no. 8, pp. 1688-1705, Available: https://www.proquest.com/docview/2677576111?pq-origsite=primo&parentSessionId=DoTR9U6Q%2FXiy63Je4yVP4xfglfyulpvDbCy5mF6f%2BCg%3D&sourcetype=Scholarly%20Journals [Accessed: Feb 23, 2024]

[5] Jill Seladi-Schulman, “What Happens in a Clinical Trial?”, “*Healthline*”, [Online], Available: https://www.healthline.com/health/clinical-trial-phases [Accessed: Feb 27, 2024]