Hospital Management
Software Requirements Document
V 1.0

	Prepared By	Reviewed by	Approved By
Name	Arkamita Joarder		
Role	Intern		
Signature	Arkamita Joarder		
Date			

TABLE OF CONTENTS

1. Introduction

Extract, Transform, and Load, is a process for extracting data from one source, transforming it into a format that is compatible with another source, and loading it into the target source. ETL (Extract, Transform, and Load) processes are often used to load data into data warehouses or data lakes for analysis.

The project Building ETL Processes(HOSPITAL MANAGEMENT) in ETL will focus on building ETL processes for vendors. Vendors are companies that provide products or services to other companies. ETL processes for vendors can be used to load data from the vendor's systems into the customer's data warehouse or data lake.

1.1 About this Document

This document serves as a definitive guide to the usage of Talend for implementing the ETL processes and constructing a robust Data Warehouse that caters to the specific business requirements of Cognizant Technology Solutions Ltd. It includes essential details about the system's architecture, table definitions, mappings, and functional requirements.

-Purpose & Scope of the Document

The primary purpose of this document is to clearly articulate the goals and functionality of the ETL- based Data Warehouse project. It aims to provide a reference for project stakeholders, developers, and users to understand the system's capabilities and limitations.

The purpose of the software requirements document is to systematically capture requirements for the project and the system “Enterprise Data Warehouse for Hospital Management” to be developed. Functional requirements of this system are captured in this document. It also serves as the input for the project scoping.

This document's scope covers the complete end-to-end process of building the data warehouse, creating a centralized repository, and implementing ETL processes using Talend. It describes the functionalities that enable Cognizant Technology Solutions Ltd. to generate various reports, forecast sales, and gain valuable insights for making data-driven decisions.

1.2 Intended Audience

The intended audience for this document includes the project team members involved in the development, testing, and maintenance of the ETL processes and the Data Warehouse. It also targets key stakeholders, project managers, business

analysts, and other individuals who need to grasp the project's overall objectives and technical details.

Throughout the document, we will provide a clear and concise explanation of the system's components, Talend's role in the ETL processes, and the impact this system will have on Cognizant Technology Solutions Ltd. 's decision support and reporting capabilities.

2. Software System Overview

The Software System Overview provides a comprehensive understanding of the ETL-based Data Warehouse project developed for Cognizant Technology Solutions Ltd. This section outlines the project's objectives, scope, architecture, and key components, including the table definitions and mappings used in the system.

2.1 About the Software System:

The following section will cover aspects related to Enterprise Data Warehouse for Hospital Management application.

ABC Hospital is a leading multispecialty hospital headquartered in Chennai. It offers comprehensive medical care in more than 40 specialties. It has vast pool of talented and experienced team of doctors who are further supported by team of highly qualified, experienced and dedicated support staff and cutting edge technology.

The following are the modules in this proposed system

a) Patient Module

b) Physician Module

c) Appointment Module

d) Insurance Module

e) Billing Module

-Scope of the System

The scope of the system is explained through its modules as follows

· Patient Module – This Module stores the details on patient ID, name, Date of birth, gender, address, city, country, phone, and insurance id.

· Physician Module -- This Module stores the details on Physician ID, Name, Date of joining, specialty, designation and experience.

· Appointment Module-- This Module maintains the details on Appointment ID, patient ID, physician ID, start date time and examination room.

· Insurance Module -- This module stores the details on Insurance ID, Insurance Name and Maximum coverage Amount

· Billing Module – This module stores the details on Payment ID, Patient ID, Physician ID, Amount to be paid, payment date and Insurance ID.

-Exclusions

While the software system addresses critical aspects of data management and reporting for Cognizant Technology Solutions Ltd. , certain functionalities are explicitly excluded from the project's scope. The following items are not part of the current project:

1. Additional Functionality Beyond Defined Modules

The system will operate only within the confines of the three defined modules: Build Data Warehouse, Build a Centralized Repository, and Create ETL Processes. Any additional functionality outside these modules is considered out of scope for the current project.

2. Integration with Non-Specified Systems

The system will focus on integrating data from specific data sources identified during the project's planning phase. Integration with non-specified or newly introduced systems will be considered as out of scope for the current project.

-System Perspective

The Enterprise Data Warehouse for Hospital Management is an independent software system developed to store the history on all activities happening in five departments of the hospital – Patient details, physician, Patient booking appointment, Billing and health insurance. The data stored could be subsequently used for reporting.

2.2 System Architecture

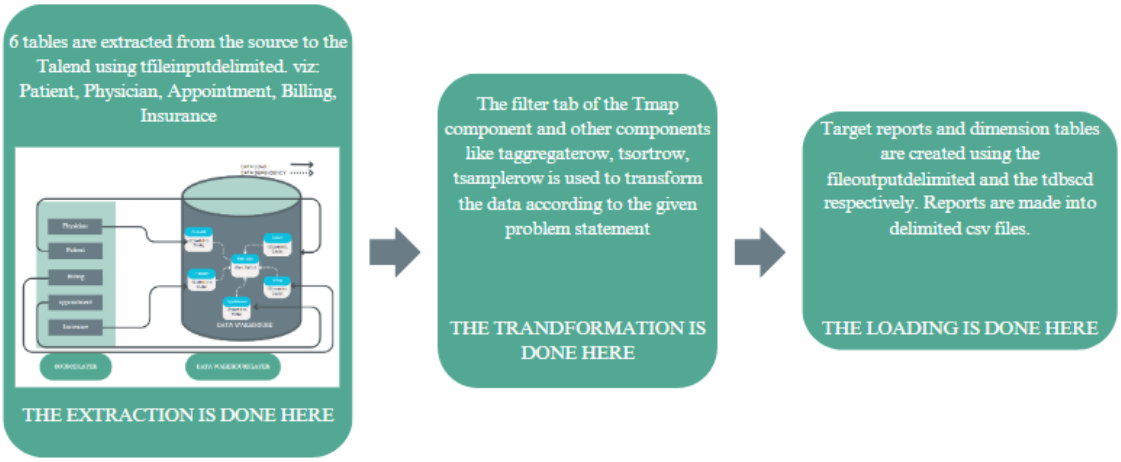
The proposed system follows a well-defined architecture to ensure scalability, maintainability, and performance. The architecture consists of two main layers:

Physical Architecture:

The physical architecture depicts the arrangement of system elements and interfaces. It includes the presentation layer, business logic layer, and data access layer. Talend, as the ETL tool, will be primarily operating in the data access layer, facilitating data movement and transformation.

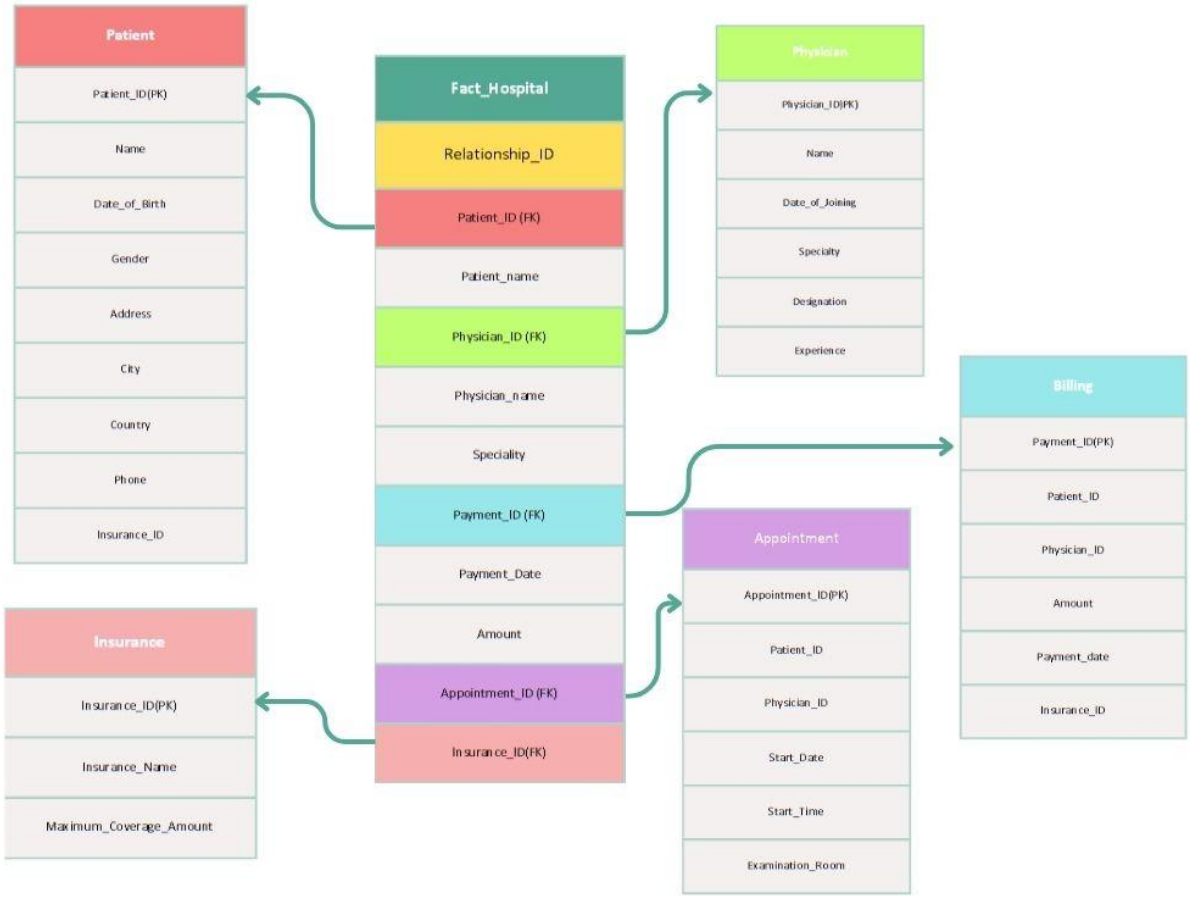
Logical Architecture:

The logical architecture defines the processes required to provide user services. It represents the functional components of the system and how they interact to achieve the project's objectives. Talend's workflows and data integration jobs will be integral parts of the logical architecture, orchestrating the ETL processes seamlessly.



THE ETL WORKFLOW

This will be the workflow of our project.



THIS IS THE TARGET STAR SCHEMA WE NEED TO ACHIEVE

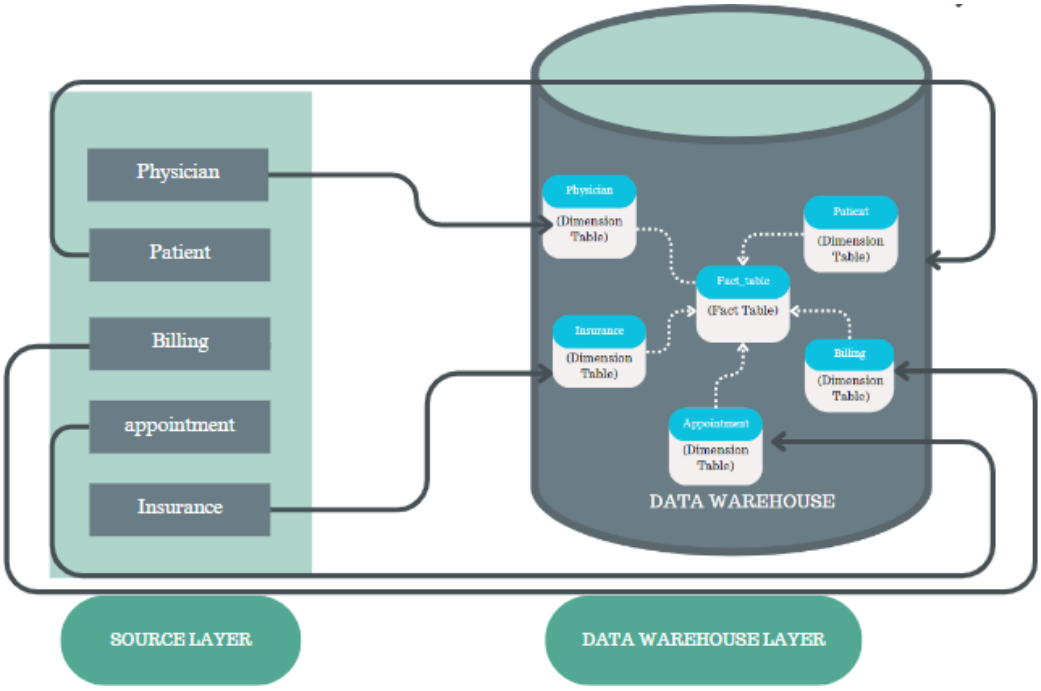
3 Table Definitions & Mappings

● Fact Table: A fact table is a central table in a star schema or snowflake schema of a data warehouse that stores quantitative data and measures.

● Dimension Table: A dimension table is a table in a data warehouse that contains descriptive

attributes used to categorize and filter data in a fact table.

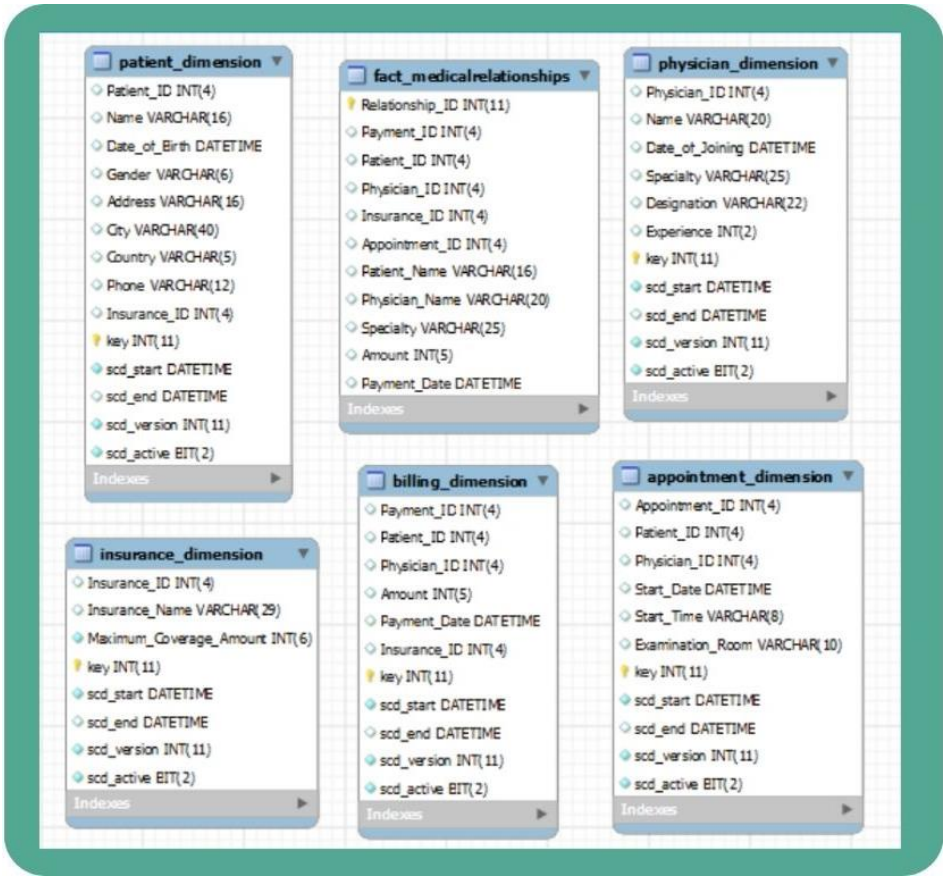
The Data Warehouse will consist of several tables, each serving a specific purpose in the reporting and analysis process. These tables include:



There are 5 main sources of data:

here's a brief description of each of the five tables in a hospital database:

- **Patient Module** – This Module stores the details on patient ID, name, Date of birth, gender, address, city, country, phone, and insurance id.
- **Physician Module** -- This Module stores the details on Physician ID, Name, Date of joining, specialty, designation and experience.
- **Appointment Module**-- This Module maintains the details on Appointment ID, patient ID, physician ID, start date time and examination room.
- **Insurance Module** -- This module stores the details on Insurance ID, Insurance Name and Maximum coverage Amount
- **Billing Module –** This module stores the details on Payment ID, Patient ID, Physician ID, Amount to be paid, payment date and Insurance ID.



PATIENT TABLE

Patient_module - Notepad

File Edit Format View Help

```
Patient_ID,Name,Date_of_Birth,Gender,Address,City,Country,Phone,Insurance_ID
1001,Johny Doe,1985-03-15,Male,123 Main St,New York,INDIA,555-123-4567,
1002,Alice Johnson,1990-08-21,Female,456 Oak Ave,Los Angeles,INDIA,444-987-6543,1002
1003,Michael Williams,1979-11-10,Male,789 Maple Rd,Chicago,INDIA,333-456-7890,1003
1004,Emily Brown,1995-06-30,Female,101 Elm St,Houston,USA,666-789-4561,1004
1005,Robert Miller,1988-02-25,Male,202 Cedar Ave,Phoenix,INDIA,777-654-3219,1005
1006,Sophia Davis,1980-09-18,Female,303 Pine Rd,Philadelphia,USA,888-321-6540,1006
1007,James Garcia,1976-12-05,Male,404 Birch St,San Antonio,USA,999-987-1236,1007
1008,Lily Martinez,1992-07-08,Female,505 Willow Ave,San Diego,INDIA,222-654-9873,1008
1009,Daniel Brown,1982-04-12,Male,606 Spruce Rd,Dallas,USA,111-852-9637,
1010,Ava Lee,1957-01-03,Female,707 Chestnut St,San Jose,USA,999-741-8528,
1011,Matthew Wilson,1958-10-27,Male,808 Hickory Ave,Austin,INDIA,888-369-8520,
1012,Olivia Taylor,1983-07-14,Female,909 Sycamore Rd,Jacksonville,USA,777-147-2583,1002
1013,William Jones,1993-04-19,Male,1010 Poplar St,San Francisco,INDIA,666-258-3694,1003
1014,Emma White,1991-11-22,Female,1111 Cypress Ave,Indianapolis,INDIA,555-963-1472,1004
1015,Noah Rodriguez,1974-08-11,Male,1212 Juniper Rd,Columbus,USA,444-369-1475,1005
1016,Isabella Thomas,1986-05-26,Female,1313 Mulberry St,Fort Worth,USA,333-741-2580,1006
1017,James Scott,1957-02-09,Male,1414 Aspen Ave,Charlotte,INDIA,222-852-9633,1007
1018,Oliver Young,1989-09-03,Male,1515 Sycamore Rd,Detroit,USA,111-963-8522,1008
1019,Ella Hernandez,1997-12-18,Female,1616 Chestnut St,El Paso,USA,999-123-9873,
1020,Alexander King,1964-03-08,Male,1717 Pine Rd,Seattle,USA,888-987-6544,
1021,Charlotte Moore,1993-05-29,Female,1818 Cedar Ave,Denver,INDIA,777-111-0000,
1022,Henry Jackson,1968-09-02,Male,1919 Willow Ave,Washington D.C.,USA,666-222-3333,1002
1023,Ella Wilson,1999-06-20,Female,2020 Spruce Rd,Boston,INDIA,555-444-5555,1003
1024,William Lee,1979-01-22,Male,2121 Chestnut St,Nashville,USA,444-555-6666,1004
1025,Scarlett Thomas,1965-11-03,Female,2222 Birch St,Memphis,USA,333-666-7777,1005
1026,Henry Davis,1984-07-25,Male,2323 Elm St,Louisville,USA,222-777-8888,1006
1027,Sophia Smith,1992-09-17,Female,2424 Maple Rd,Baltimore,USA,111-888-9999,1007
```

PHYSICIAN TABLE

physician_module - Notepad

File Edit Format View Help

```
Physician_ID,Name,Date_of_Joining,Specialty,Designation,Experience
2001,Dr. John Smith,2010-05-15,Cardiology,Associate Professor,11
2002,Dr. Alice Johnson,2012-09-20,Pediatrics,Department Head,14
2003,Dr. Michael Williams,2008-11-08,Orthopedics,Surgeon,13
2004,Dr. Emily Brown,2015-06-30,Dermatology,Registrar,6
2005,Dr. Robert Miller,2007-02-25,Oncology,Department Head,14
2006,Dr. Sophia Davis,2011-09-18,Neurology,Surgeon,10
2007,Dr. James Garcia,2014-12-05,Gastroenterology,Senior Medical Officer,7
2008,Dr. Lily Martinez,2009-07-08,Obstetrics and Gynecology,Surgeon,12
2009,Dr. Daniel Brown,2018-04-12,Surgery,Registrar,3
2010,Dr. Ava Lee,2016-01-03,Ophthalmology,Senior Medical Officer,5
2011,Dr. William Johnson,2013-08-22,Cardiology,Surgeon,8
2012,Dr. Emma Davis,2017-12-10,Dermatology,Registrar,4
2013,Dr. Noah Wilson,2019-05-14,Orthopedics,Registrar,2
2014,Dr. Olivia Brown,2018-10-29,Neurology,Surgeon,3
2015,Dr. James Miller,2016-02-09,Pediatrics,Senior Medical Officer,5
2016,Dr. Sophia Martinez,2011-07-21,Gastroenterology,Surgeon,10
2017,Dr. Michael Johnson,2020-03-05,Surgery,Medical Practitioner,1
2018,Dr. Ava Williams,2014-09-18,Obstetrics and Gynecology,Department Head,14
2019,Dr. William Garcia,2008-06-17,Oncology,Surgeon,9
2020,Dr. Olivia Anderson,2015-11-03,Ophthalmology,Registrar,6
```

INSURANCE TABLE

insurance_module - Notepad

File Edit Format View Help

```
Insurance_ID,Insurance_Name,Maximum_Coverage_Amount
1002,XYZ Insurance Group,275000
1003,Global Health Insurers,100000
1004,Secure Life Assurance,250000
1005,SafeGuard Insurance Ltd.,150000
1006,United Insurance Corporation,80000
1007,National Health Insurers,120000
1008,FirstClass Insurance Services,90000
1009,Liberty Insurance Solutions,180000
```

BILLING TABLE

billing_module - Notepad

File Edit Format View Help

```
Payment_ID,Patient_ID,Physician_ID,Amount,Payment_Date,Insurance_ID
4001,1001,2001,20000,2023-07-19,
4002,1002,2002,15000,2023-07-20,1002
4003,1003,2003,30000,2023-07-21,
4004,1004,2004,10000,2023-07-22,1004
4005,1005,2005,25000,2023-07-23,1005
4006,1006,2006,17500,2023-07-24,1002
4007,1007,2007,22500,2023-07-25,1002
4008,1008,2008,18000,2023-07-26,1003
4009,1009,2009,21000,2023-07-27,
4010,1010,2010,27500,2023-07-28,1005
4011,1011,2011,19000,2023-07-29,1004
4012,1012,2012,16000,2023-07-30,1002
4013,1013,2013,32000,2023-07-31,1003
4014,1014,2014,11000,2023-08-01,1004
4015,1015,2015,26000,2023-08-02,1005
4016,1016,2016,18500,2023-08-03,1005
4017,1017,2017,23000,2023-08-04,1002
4018,1018,2018,19000,2023-08-05,1003
4019,1019,2019,22000,2023-08-06,1004
4020,1020,2020,29000,2023-08-07,1005
```

APPOINTMENT TABLE

appointment_module - Notepad

File Edit Format View Help

```
Appointment_ID,Patient_ID,Physician_ID,Start_Date,Start_Time,Examination_Room
3001,1001,2001,2023-07-19,09:00:00,"Room 101"
3002,1002,2002,2023-07-20,10:30:00,"Room 102"
3003,1003,2003,2023-07-21,13:15:00,"Room 103"
3004,1004,2004,2023-07-22,11:45:00,"Room 104"
3005,1005,2005,2023-07-23,15:30:00,"Room 105"
3006,1006,2006,2023-07-24,13:00:00,"Room 106"
3007,1007,2007,2023-07-25,16:00:00,"Room 107"
3008,1008,2008,2023-07-26,11:00:00,"Room 108"
3009,1009,2009,2023-07-27,09:45:00,"Room 109"
3010,1010,2010,2023-07-28,14:30:00,"Room 110"
3011,1011,2001,2023-07-29,10:00:00,"Room 101"
3012,1012,2002,2023-07-30,11:30:00,"Room 102"
3013,1013,2003,2023-07-31,15:15:00,"Room 103"
3014,1014,2004,2023-08-01,13:45:00,"Room 104"
3015,1015,2005,2023-08-02,16:30:00,"Room 105"
3016,1016,2006,2023-08-03,14:00:00,"Room 106"
3017,1017,2007,2023-08-04,17:00:00,"Room 107"
3018,1018,2008,2023-08-05,12:00:00,"Room 108"
3019,1019,2009,2023-08-06,09:30:00,"Room 109"
3020,1020,2010,2023-08-07,15:30:00,"Room 110"
3021,1021,2001,2023-08-08,10:15:00,"Room 101"
3022,1022,2002,2023-08-09,11:45:00,"Room 102"
3023,1023,2003,2023-08-10,16:00:00,"Room 103"
3024,1024,2004,2023-08-11,13:30:00,"Room 104"
3025,1025,2005,2023-08-12,17:15:00,"Room 105"
3026,1026,2006,2023-08-13,14:45:00,"Room 106"
```

Hence we are loading the data from the source to the data warehouse.

THE EXTRACTION PROCESS ENDS HERE

4. FILTERING OF DATA TO CREATE THE REPORTS

4.1 Patient Module

Produce report on patient details who are not from India, not covered by any insurance

```
(!row1.Country.equals("INDIA")) &&  row1.Insurance_ID==null
```

tLogRow_9

Patient_ID	Name	Date_of_Birth	Gender	Address	City	Country	Phone
1009	Daniel Brown	1982-04-12	Male	606 Spruce Rd	Dallas	USA	111-852-9637
1010	Ava Lee	1957-01-03	Female	707 Chestnut St	San Jose	USA	999-741-8528
1019	Ella Hernandez	1997-12-18	Female	1616 Chestnut St	El Paso	USA	999-123-9873
1020	Alexander King	1964-03-08	Male	1717 Pine Rd	Seattle	USA	888-987-6544
1029	Ava Anderson	1989-08-08	Female	2626 Pine Rd	Albuquerque	USA	888-666-5555
1031	Emma Brown	1996-07-12	Female	2828 Spruce Rd	Fresno	USA	666-444-3333
1039	Emma Smith	1978-06-06	Female	3636 Spruce Rd	Raleigh	USA	777-666-5555
1049	Olivia Anderson	1990-04-10	Female	4646 Elm St	Tampa	USA	666-666-5555
1050	William Garcia	1976-01-18	Male	4747 Birch St	Honolulu	USA	555-555-4444
1052	Noah Davis	1979-11-12	Male	4949 Cedar Ave	Raleigh	USA	444-555-6666
1055	Sophia Garcia	1992-08-27	Female	5252 Maple Rd	Cleveland	USA	111-222-3333
1059	Emma Johnson	1993-02-28	Female	5656 Elm St	Arlington	USA	666-555-4444

Produce report on patient details above age 50

```
(TalendDate.getCurrentDate().getYear()-row1.Date_of_Birth.getYear())>=50
```

tLogRow_8

Patient_ID	Name	Date_of_Birth	Gender	Address	City	Country	Phone	Insurance_ID	Age
1010	Ava Lee	1957-01-03	Female	707 Chestnut St	San Jose	USA	999-741-8528	null	66
1011	Matthew Wilson	1958-10-27	Male	808 Hickory Ave	Austin	INDIA	888-369-8520	null	65
1017	James Scott	1957-02-09	Male	1414 Aspen Ave	Charlotte	INDIA	222-852-9633	1007	66
1020	Alexander King	1964-03-08	Male	1717 Pine Rd	Seattle	USA	888-987-6544	null	59
1022	Henry Jackson	1968-09-02	Male	1919 Willow Ave	Washington D.C.	USA	666-222-3333	1002	55
1025	Scarlett Thomas	1965-11-03	Female	2222 Birch St	Memphis	USA	333-666-7777	1005	58
1033	Olivia Johnson	1968-09-14	Female	3030 Elm St	Long Beach	USA	444-222-1111	1003	55
1036	Michael Anderson	1971-12-09	Male	3333 Oak Ave	Virginia Beach	USA	111-999-8888	1006	52
1044	Michael Wilson	1970-04-11	Male	4141 Oak Ave	Minneapolis	USA	222-111-0000	1004	53
1054	William Williams	1963-03-14	Male	5151 Willow Ave	Tulsa	USA	222-333-4444	1002	60
1060	Noah Williams	1970-11-21	Male	5757 Cedar Ave	New Orleans	USA	555-444-3333	1003	53
1075	Emma Brown	1969-05-05	Female	7272 Elm St	New Orleans	USA	888-888-8888	1002	54

4.2 Physician Module

Report of physician details who are surgeon who joined hospital between 2000 to 2010 (year)

```
row3.Designation.equals("Surgeon") &&
TalendDate.compareDate(row3.Date_of_Joining,TalendDate.parseDate("yyyy-MM-dd","2000-01-01"))>=0 &&
TalendDate.compareDate(row3.Date_of_Joining,TalendDate.parseDate("yyyy-MM-dd","2010-12-31"))<=0
```

```
.-----------+-------------------+--------------+-----------------------+-----------+----------.
|                                tLogRow_5                                                      |
|=----------+-------------------+--------------+-----------------------+-----------+---------=|
|Physician_ID|Name              |Date_of_Joining|Specialty              |Designation|Experience|
|=----------+-------------------+--------------+-----------------------+-----------+---------=|
|2003        |Dr. Michael Williams|2008-11-08    |Orthopedics            |Surgeon    |13        |
|2008        |Dr. Lily Martinez   |2009-07-08    |Obstetrics and Gynecology|Surgeon  |12        |
|2019        |Dr. William Garcia  |2008-06-17    |Oncology               |Surgeon    |9         |
'-----------+-------------------+--------------+-----------------------+-----------+----------'
```

To display physician details based on

i. Name
ii. Experience

we use the tSortrow component, to sort the Name and Experience in ascending order in alpha and num

Schema column	sort num or alpha?	Order asc or desc?
Name	alpha	asc
Experience	num	asc

```
.-----------+-------------------+--------------+-----------------------+-----------------------+----------.
|                                tLogRow_18                                                                |
|=----------+-------------------+--------------+-----------------------+-----------------------+---------=|
|Physician_ID|Name              |Date_of_Joining|Specialty              |Designation            |Experience|
|=----------+-------------------+--------------+-----------------------+-----------------------+---------=|
|2002        |Dr. Alice Johnson  |2012-09-20    |Pediatrics             |Department Head        |14        |
|2010        |Dr. Ava Lee        |2016-01-03    |Ophthalmology          |Senior Medical Officer|5         |
|2018        |Dr. Ava Williams   |2014-09-18    |Obstetrics and Gynecology|Department Head      |14        |
|2009        |Dr. Daniel Brown   |2018-04-12    |Surgery                |Registrar              |3         |
|2004        |Dr. Emily Brown    |2015-06-30    |Dermatology            |Registrar              |6         |
|2012        |Dr. Emma Davis     |2017-12-10    |Dermatology            |Registrar              |4         |
|2007        |Dr. James Garcia   |2014-12-05    |Gastroenterology       |Senior Medical Officer|7         |
|2015        |Dr. James Miller   |2016-02-09    |Pediatrics             |Senior Medical Officer|5         |
|2001        |Dr. John Smith     |2010-05-15    |Cardiology             |Associate Professor    |11        |
|2008        |Dr. Lily Martinez  |2009-07-08    |Obstetrics and Gynecology|Surgeon              |12        |
|2017        |Dr. Michael Johnson|2020-03-05    |Surgery                |Medical Practitioner  |1         |
|2003        |Dr. Michael Williams|2008-11-08   |Orthopedics            |Surgeon                |13        |
```

To verify whether the physicians are head of the department

row3.Designation.equals("Department Head")

```
+----------+----------------+--------------+-------------------------+----------------+----------+
|                                tLogRow_7                                                        |
|=---------+----------------+--------------+-------------------------+----------------+----------=|
|Physician_ID|Name          |Date_of_Joining|Specialty                |Designation     |Experience|
|=---------+----------------+--------------+-------------------------+----------------+----------=|
|2002      |Dr. Alice Johnson|2012-09-20    |Pediatrics               |Department Head|14        |
|2005      |Dr. Robert Miller|2007-02-25    |Oncology                 |Department Head|14        |
|2018      |Dr. Ava Williams |2014-09-18    |Obstetrics and Gynecology|Department Head|14        |
+----------+----------------+--------------+-------------------------+----------------+----------+
```

4.3 Appointment Module

To store patient info who have appointment today

TalendDate.compareDate(row4.Start_Date,TalendDate.getCurrentDate(),"yyyy-MM-dd")==0

```
+----------+-----------+-------------+------+----------+-------+-------+------------+-----------+----------+
|                                tLogRow_3                                                                   |
|=---------+-----------+-------------+------+----------+-------+-------+------------+-----------+----------=|
|Patient_ID|Name       |Date_of_Birth|Gender|Address   |City   |Country|Phone       |Insurance_ID|Start_Date|
|=---------+-----------+-------------+------+----------+-------+-------+------------+-----------+----------=|
|1004      |Emily Brown|1995-06-30   |Female|101 Elm St|Houston|USA    |666-789-4561|1004       |2023-08-31|
+----------+-----------+-------------+------+----------+-------+-------+------------+-----------+----------+
```

To fetch Patient details based on physician and examination room.

created an output table, mapped the patient details and physician ID and the examination room and show the output

appointmentout3

Expression	Column
row5.Patient_ID	Patient_ID
row5.Name	Name
row5.Date_of_Birth	Date_of_Birth
row5.Gender	Gender
row5.Address	Address
row5.City	City
row5.Country	Country
row5.Phone	Phone
row5.Insurance_ID	Insurance_ID
row4.Physician_ID	Physician_ID
row4.Examination_Room	Examination_Room

To display patient who have appointment booked in future and covered by insurance.

row5.Insurance_ID !=null &&
TalendDate.compareDate(row4.Start_Date,TalendDate.getCurrentDate(),"yyyy-MM-dd")>0

tLogRow_6

Patient_ID	Name	Date_of_Birth	Gender	Address	City	Country	Phone	Insurance_ID	Start_Date
1005	Robert Miller	1988-02-25	Male	202 Cedar Ave	Phoenix	INDIA	777-654-3219	1005	2023-09-01
1006	Sophia Davis	1980-09-18	Female	303 Pine Rd	Philadelphia	USA	888-321-6540	1006	2023-09-02
1007	James Garcia	1976-12-05	Male	404 Birch St	San Antonio	USA	999-987-1236	1007	2023-09-03
1008	Lily Martinez	1992-07-08	Female	505 Willow Ave	San Diego	INDIA	222-654-9873	1008	2023-09-04
1032	Noah Davis	1981-03-26	Male	2929 Cedar Ave	Sacramento	USA	555-333-2222	1002	2023-09-27
1033	Olivia Johnson	1968-09-14	Female	3030 Elm St	Long Beach	USA	444-222-1111	1003	2023-09-28
1034	William Wilson	1987-05-01	Male	3131 Birch St	Kansas City	INDIA	333-111-0000	1004	2023-09-29
1035	Sophia Williams	1986-08-23	Female	3232 Maple Rd	Mesa	INDIA	222-000-9999	1005	2023-09-30
1036	Michael Anderson	1971-12-09	Male	3333 Oak Ave	Virginia Beach	USA	111-999-8888	1006	2023-10-01
1037	Ava Garcia	1980-02-29	Female	3434 Pine Rd	Atlanta	USA	999-888-7777	1007	2023-10-02
1038	James Thomas	1995-11-17	Male	3535 Willow Ave	Omaha	USA	888-777-6666	1008	2023-10-03
1040	Noah Johnson	1999-03-15	Male	3737 Cedar Ave	Miami	INDIA	666-555-4444	1004	2023-10-05
1071	Sophia Johnson	1981-09-29	Female	6868 Maple Rd	Minneapolis	USA	333-333-2222	1006	2023-10-25
1072	Michael Smith	1993-08-12	Male	6969 Birch St	Miami	USA	222-222-1111	1005	2023-10-26
1074	James Wilson	1992-04-16	Male	7171 Spruce Rd	Arlington	USA	999-999-9999	1003	2023-10-28
1075	Emma Brown	1969-05-05	Female	7272 Elm St	New Orleans	USA	888-888-8888	1002	2023-10-29
1077	Olivia Williams	1991-10-21	Female	7474 Pine Rd	Tampa	USA	666-666-6666	1008	2023-10-31

4.4 Insurance Module

To store patient info who have insurance amount greater than 2,00,000

row8.Maximum_Coverage_Amount>200000

tLogRow_1

Insurance_ID	Insurance_Name	Maximum_Coverage_Amount	Patient_ID	Name	Date_of_Birth	Gender	Address
1002	XYZ Insurance Group	275000	1002	Alice Johnson	11-01-0027	Female	456 Oak Ave
1004	Secure Life Assurance	250000	1004	Emily Brown	16-11-0035	Female	101 Elm St
1002	XYZ Insurance Group	275000	1013	Olivia Taylor	04-12-0019	Female	909 Sycamore Rd
1004	Secure Life Assurance	250000	1014	Emma White	13-04-0028	Female	1111 Cypress Ave
1002	XYZ Insurance Group	275000	1022	Henry Jackson	20-01-0008	Male	1919 Willow Ave
1004	Secure Life Assurance	250000	1024	William Lee	02-06-0027	Male	2121 Chestnut St
1002	XYZ Insurance Group	275000	1032	Noah Davis	02-08-0031	Male	2929 Cedar Ave
1004	Secure Life Assurance	250000	1034	William Wilson	08-10-0006	Male	3131 Birch St
1004	Secure Life Assurance	250000	1040	Noah Johnson	19-08-0020	Male	3737 Cedar Ave
1002	XYZ Insurance Group	275000	1042	William Moore	05-10-0034	Male	3939 Birch St
1004	Secure Life Assurance	250000	1044	Michael Wilson	21-08-0016	Male	4141 Oak Ave
1002	XYZ Insurance Group	275000	1054	William Williams	15-07-0019	Male	5151 Willow Ave
1002	XYZ Insurance Group	275000	1061	Olivia Brown	29-08-0018	Female	5858 Pine Rd
1002	XYZ Insurance Group	275000	1068	Noah Martinez	07-06-0014	Male	6565 Cedar Ave
1002	XYZ Insurance Group	275000	1075	Emma Brown	20-09-0010	Female	7272 Elm St
1002	XYZ Insurance Group	275000	1082	James Davis	16-02-0027	Male	7979 Cedar Ave
1002	XYZ Insurance Group	275000	1089	Ava Smith	30-10-0020	Female	8686 Cedar Ave
1002	XYZ Insurance Group	275000	1096	Michael Wilson	07-02-0024	Male	9393 Willow Ave

To store and fetch count of patients covered by each insurance.

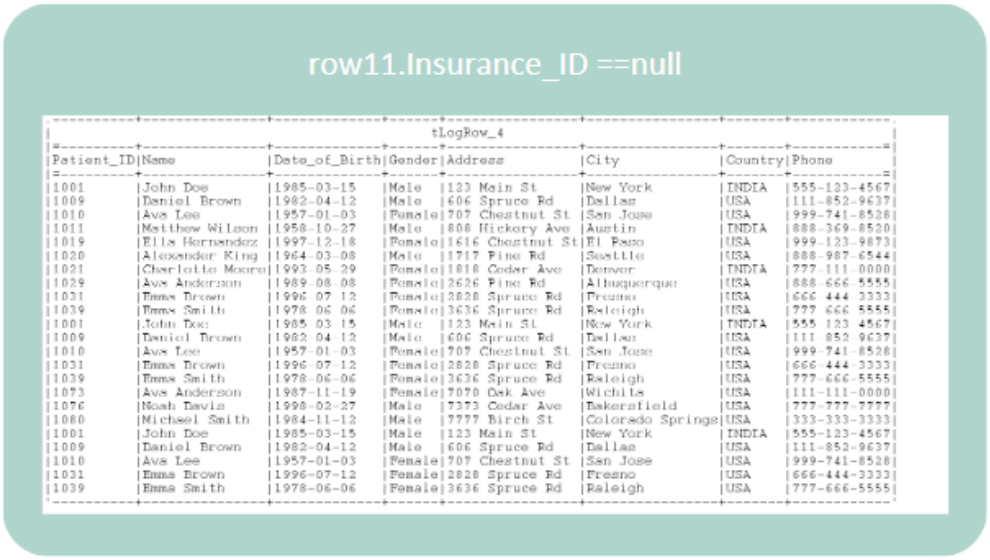
create a new column "no_of_patients"
in no_of_patients use count function and store the count of patients

tLogRow_2

Insurance_ID	Insurance_Name	No_of_Patients
1003	Global Health Insurers	13
1004	Secure Life Assurance	6
1007	National Health Insurers	5
1002	XYZ Insurance Group	12
1006	United Insurance Corporation	12
1005	SafeGuard Insurance Ltd.	13
1008	FirstClass Insurance Services	12

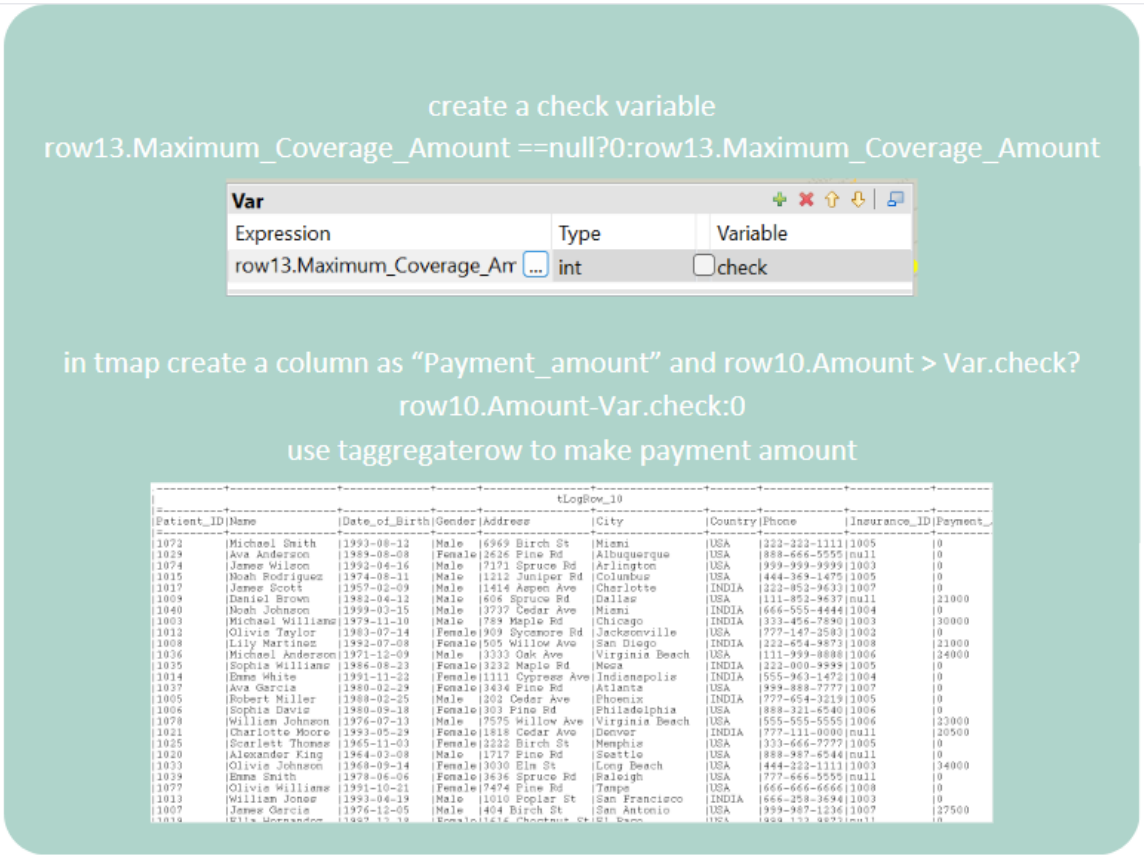
4.5 Billing Module

To store the patient details who don't have insurance coverage and need to pay the complete amount.



Patient_ID	Name	Date_of_Birth	Gender	Address	City	Country	Phone
1001	John Doe	1985-03-15	Male	123 Main St	New York	INDIA	555-123-4567
1009	Daniel Brown	1982-04-12	Male	606 Spruce Rd	Dallas	USA	111-852-9637
1010	Ava Lee	1957-01-03	Female	707 Chestnut St	San Jose	USA	999-741-8528
1011	Matthew Wilson	1958-10-27	Male	808 Hickory Ave	Austin	INDIA	888-369-8520
1019	Ella Hernandez	1997-12-18	Female	1616 Chestnut St	El Paso	USA	999-123-9873
1020	Alexander King	1964-03-08	Male	1717 Pine Rd	Seattle	USA	888-987-6544
1021	Charlotte Moore	1993-05-29	Female	1818 Cedar Ave	Denver	INDIA	777-111-0000
1029	Ava Anderson	1989-08-08	Female	2626 Pine Rd	Albuquerque	USA	888-666-5555
1031	Emma Brown	1996-07-12	Female	2828 Spruce Rd	Fresno	USA	666-444-3333
1039	Emma Smith	1978-06-06	Female	3636 Spruce Rd	Raleigh	USA	777-666-5555
1001	John Doe	1985-03-15	Male	123 Main St	New York	INDIA	555-123-4567
1009	Daniel Brown	1982-04-12	Male	606 Spruce Rd	Dallas	USA	111-852-9637
1010	Ava Lee	1957-01-03	Female	707 Chestnut St	San Jose	USA	999-741-8528
1031	Emma Brown	1996-07-12	Female	2828 Spruce Rd	Fresno	USA	666-444-3333
1039	Emma Smith	1978-06-06	Female	3636 Spruce Rd	Raleigh	USA	777-666-5555
1073	Ava Anderson	1987-11-19	Female	7070 Oak Ave	Wichita	USA	111-111-0000
1076	Noah Davis	1998-02-27	Male	7373 Cedar Ave	Bakersfield	USA	777-777-7777
1080	Michael Smith	1984-11-12	Male	7777 Birch St	Colorado Springs	USA	333-333-3333
1001	John Doe	1985-03-15	Male	123 Main St	New York	INDIA	555-123-4567
1009	Daniel Brown	1982-04-12	Male	606 Spruce Rd	Dallas	USA	111-852-9637
1010	Ava Lee	1957-01-03	Female	707 Chestnut St	San Jose	USA	999-741-8528
1031	Emma Brown	1996-07-12	Female	2828 Spruce Rd	Fresno	USA	666-444-3333
1039	Emma Smith	1978-06-06	Female	3636 Spruce Rd	Raleigh	USA	777-666-5555

To fetch the patient and the actual amount to be paid (Amount to be paid – Covered by insurance).



Patient_ID	Name	Date_of_Birth	Gender	Address	City	Country	Phone	Insurance_ID	Payment_
1073	Michael Smith	1993-08-12	Male	6969 Birch St	Miami	USA	222-222-1111	1005	0
1029	Ava Anderson	1989-08-08	Female	2626 Pine Rd	Albuquerque	USA	888-666-5555	null	0
1074	James Wilson	1992-04-16	Male	7171 Spruce Rd	Arlington	USA	999-999-9999	1003	0
1015	Noah Rodriguez	1974-08-11	Male	1212 Juniper Rd	Columbus	USA	444-369-1475	1005	0
1017	James Scott	1957-02-09	Male	1414 Aspen Ave	Charlotte	INDIA	222-852-9633	1007	0
1009	Daniel Brown	1982-04-12	Male	606 Spruce Rd	Dallas	USA	111-852-9637	null	21000
1040	Noah Johnson	1999-03-15	Male	3737 Cedar Ave	Miami	INDIA	666-555-4444	1004	0
1003	Michael Williams	1979-11-10	Male	789 Maple Rd	Chicago	INDIA	333-456-7890	1003	30000
1012	Olivia Taylor	1983-07-14	Female	909 Sycamore Rd	Jacksonville	USA	777-147-2583	1002	0
1008	Lily Martinez	1992-07-08	Female	505 Willow Ave	San Diego	INDIA	222-654-9873	1008	21000
1036	Michael Anderson	1971-12-09	Male	3333 Oak Ave	Virginia Beach	USA	111-999-8888	1006	24000
1035	Sophia Williams	1986-08-23	Female	3232 Maple Rd	Mesa	INDIA	222-000-9999	1005	0
1014	Emma White	1991-11-22	Female	1111 Cypress Ave	Indianapolis	INDIA	555-963-1472	1004	0
1027	Ava Garcia	1980-02-29	Female	3434 Pine Rd	Atlanta	USA	999-888-7777	1007	0
1005	Robert Miller	1988-02-25	Male	202 Cedar Ave	Phoenix	INDIA	777-654-3219	1005	0
1006	Sophia Davis	1980-09-18	Female	303 Pine Rd	Philadelphia	USA	888-321-6540	1006	0
1078	William Johnson	1976-07-13	Male	7575 Willow Ave	Virginia Beach	USA	555-555-5555	1006	23000
1021	Charlotte Moore	1993-05-29	Female	1818 Cedar Ave	Denver	INDIA	777-111-0000	null	20500
1025	Scarlett Thomas	1965-11-03	Female	2222 Birch St	Memphis	USA	333-666-7777	1005	0
1020	Alexander King	1964-03-08	Male	1717 Pine Rd	Seattle	USA	888-987-6544	null	0
1033	Olivia Johnson	1968-09-14	Female	3030 Elm St	Long Beach	USA	444-222-1111	1003	34000
1039	Emma Smith	1978-06-06	Female	3636 Spruce Rd	Raleigh	USA	777-666-5555	null	0
1077	Olivia Williams	1991-10-21	Female	7474 Pine Rd	Tampa	USA	666-666-6666	1008	0
1013	William Jones	1993-04-19	Male	1010 Poplar St	San Francisco	INDIA	666-258-3694	1003	0
1007	James Garcia	1976-12-05	Male	404 Birch St	San Antonio	USA	999-987-1236	1007	27500

To fetch physician who had maximum amount to be paid.

There are a few steps we need to follow:
Step 1: use taggregaterow to sum the amount
Step 2: use tsortrow for decesnding order of amount
Step 3: create a context variable "due_payment" initialize it with 0
Step 4: Use the tJavarow component

```
output_row.Physician_ID=input_row.Physician_ID;
output_row.Name=input_row.Name;
output_row.Amount=input_row.Amount;

if (input_row.Amount.compareTo(context.due_payment)==1)
{
context.due_payment=input_row.Amount;
}
```

Step 5: use tfilterrow and then compare it with the context variable(which now it has the max amount)

```
.------------+-----------+------.
|           tLogRow_11             |
|=-----------+-----------+-----=|
|Physician_ID|Name       |Amount|
|=-----------+-----------+-----=|
|2010        |Dr. Ava Lee|232000|
'------------+-----------+------'
```

To fetch the list of patient who needs to pay this month

```
TalendDate.getPartOfDate("MONTH",row10.Payment_Date
)==TalendDate.getPartOfDate("MONTH",TalendDate.getCurrentDate() ) &&
TalendDate.getPartOfDate("YEAR",row10.Payment_Date
)==TalendDate.getPartOfDate("YEAR",TalendDate.getCurrentDate() )
```

```
.----------+--------------.
|         tLogRow_12         |
|=---------+--------------=|
|Patient_ID|Name            |
|=---------+--------------=|
|1005      |Robert Miller   |
|1006      |Sophia Davis    |
|1007      |James Garcia    |
|1008      |Lily Martinez   |
|1009      |Daniel Brown    |
|1010      |Ava Lee         |
|1031      |Emma Brown      |
|1032      |Noah Davis      |
|1033      |Olivia Johnson  |
|1034      |William Wilson  |
|1035      |Sophia Williams |
'----------+---------------'
```

THE TRANSFORMATION PROCESS ENDS HERE

5. Loading of Tables into database whilst assigning SCD

Datas are loaded through tDBSCD for implementing type 2 Dimension

For tDBSCD, the schema was first edited where the “key”(primary key) was chosen and inside the component the required columns were specified for the source keys, surrogate keys (for the Type 2), the “creation” was chosen as “Table max + 1” and the variables for the Type 0 fields (variables which will retain the old values), Type 1 fields (overwrites the older value), Type 2 fields (the history of the variable will be maintained) and Type 3 fields (where we have a current value and previous value)
Followed by this, scd_version (Integer data type) and scd_active (Boolean data type) fields were checked and their corresponding sizes were again specified inside the schema.
If we have “scd_version as 1” and “scd_active as True” for any row, then that row has the current value and if we have scd_version as some other value and “scd_active as False” then that row has it’s older value.

Let suppose we assign the PHYSICIAN table into type 2 dimension
Lets take a small snippet

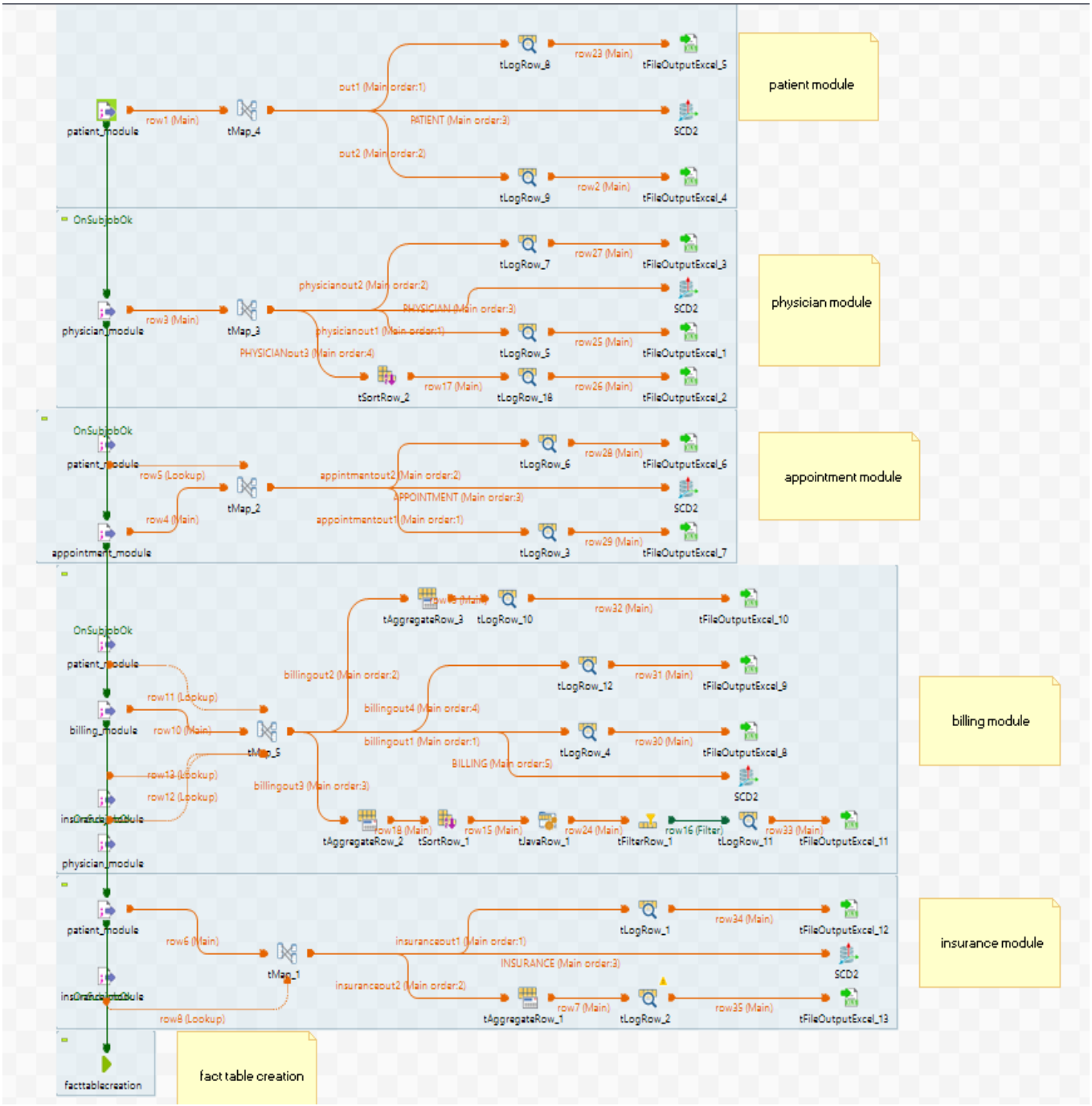
Physician_ID	Name	Date_of_Joining	Specialty	Designation	Experience	Key	SCD Start	SCD End	SCD Version	SCD Active
2001	Dr. John Smith	2010-05-15	Cardiology	Associate Profes	11	1	02/09/2023		1	TRUE
2002	Dr. Alice Johnso	2012-09-20	Pediatrics	Department Hea	14	2	02/09/2023		1	TRUE
2003	Dr. Michael Willi	2008-11-08	Orthopedics	Surgeon	13	3	02/09/2023		1	TRUE
2004	Dr. Emily Brown	2015-06-30	Dermatology	Registrar	6	4	02/09/2023		1	TRUE
2005	Dr. Robert Miller	2007-02-25	Oncology	Department Hea	14	5	02/09/2023		1	TRUE

“Dr. John Smith” changes his name to “Dr. John Alice Smith”

Physician_ID	Name	Date_of_Joining	Specialty	Designation	Experience	Key	SCD Start	SCD End	SCD Version	SCD Active
2001	Dr. John Smith	2010-05-15	Cardiology	Associate Professor	11	1	2023-02-09	05-09-2023	1	FALSE
2002	Dr. Alice Johnson	2012-09-20	Pediatrics	Department Head	14	2	2023-02-09		1	TRUE
2003	Dr. Michael Williams	2008-11-08	Orthopedics	Surgeon	13	3	2023-02-09		1	TRUE
2004	Dr. Emily Brown	2015-06-30	Dermatology	Registrar	6	4	2023-02-09		1	TRUE
2005	Dr. Robert Miller	2007-02-25	Oncology	Department Head	14	5	2023-02-09		1	TRUE
2001	Dr. John Alice Smith	2010-05-15	Cardiology	Associate Professor	11	6	05-09-2023		2	TRUE

THE TABLES ARE NOW IN THE DATABASE .
THE LOADING PROCESS ENDS HERE

THE JOB CREATION IN TALEND



6. Annexure:

6.1 Terms & Conditions:

-Licensing Information:

The ETL project using Talend is subject to specific licensing terms and conditions. These terms govern the usage, distribution, and modification of the project deliverables. The licensing information is outlined as follows:

1. The ETL project's source code and documentation are the intellectual property of Cognizant Technology Solutions Ltd. and are protected under applicable copyright laws.

2. The project's deliverables, including Talend job designs, ETL workflows, and documentation, are solely for internal use by Cognizant Technology Solutions Ltd. and may not be distributed or shared with external parties without prior written consent.

3. Any modifications or enhancements to the ETL project must comply with Cognizant Technology Solutions Ltd. 's change management process and be approved by the designated project authorities.

4. Cognizant Technology Solutions Ltd. shall not hold Talend responsible for any issues or challenges arising from the usage of the Talend software, as per the terms and conditions set forth by Talend's licensing agreement.

-Data Privacy and Security Policies:

As part of the ETL project's implementation, data privacy and security policies are paramount to safeguard sensitive information. The following policies are to be adhered to:

1. Data Encryption: All sensitive data transmitted between systems and during ETL processes must be encrypted to prevent unauthorized access.

2. Access Controls: Access to the Data Warehouse and related systems shall be granted based on the principle of least privilege, ensuring that only authorized personnel can access sensitive data.

3. Data Anonymization: Personally identifiable information (PII) and other sensitive data must be anonymized or pseudonymized when not required for specific reporting or analysis.

4. Data Retention: Data retention policies shall be defined to manage the storage and archival of data in compliance with legal and regulatory requirements.

5. Audit Logging: Comprehensive audit logs shall be maintained to track data access, changes, and user activities for accountability and troubleshooting purposes.

6. Disaster Recovery: Robust data backup and disaster recovery

mechanisms shall be established to ensure data availability and continuity in case of unforeseen events.

7 Appendix

Source files: All the data used as source to test this project has been provided in this file,.

Patient.csv physician_module.csv appointment_module.csv billing_module.csv insurance_module.csv

Target Reports: All the target reports are mentioned in this file

Patient Module Report

Conditions	Report files
1) **Produce report on patient details who are not from India, not covered by any insurance**	notfromindianoinsurance.xlsx
2) **Produce report on patient details above age 50**	agemorethan50.xlsx

Physician Module Report

Conditions	Report Files
1) **Report of physician details who are surgeon who joined hospital between 2000 to 2010 (year)**	surgeon2000to2010.xlsx
2) **To display physician details based on** **i) Name** **ii) Experience**	nameandexperience.xlsx
3) **To verify whether the physicians are head of the department**	hod.xlsx

Appointment Module Report

Conditions	Report Files
1) **To store patient info who have appointment today**	appointmenttoday.xlsx
2) **To display patient who have appointment booked in future and covered by insurance.**	withinsuranceandfutureappointment.xlsx

Billing Module Report

Conditions	Report files
1) **To store the patient details who do not have insurance coverage and need to pay the complete amount.**	noinsurance.xlsx
2) **To fetch the patient and the actual amount to be paid (Amount to be paid – Covered by insurance).**	actualamountobepaid.xlsx
3) **To fetch physician who had maximum amount to be paid.**	highestpayingphysician.xlsx
4) **To fetch the list of patients who needs to pay this month**	needtopaythismonth.xlsx

Insurance Module Report

Conditions	Report Files
1) **To store patient info who have insurance amount greater than 2,00,000**	greaterthan200000.xlsx
2) **To store and fetch count of patients covered by each insurance.**	noofpeoplehavinginsurance.xlsx

Thank You