**Development of a Data Warehouse for Coronavirus (COVID- 19) Diseases**

Aysha Siddiqa1 and Siddhaling Urolagin2

1,2 Department of Computer Science,

2APP Centre for AI Research (APPCAIR),

Birla Institute of Technology and Science, Pilani,

Dubai International Academic City,

Dubai, United Arab Emirates.

***ABSTRACT***

*The companies use the most reliable technology Data warehouse for planning, forecasting and management. The healthcare is one of the sectors that deals with huge data and many organizations find it difficult to efficiently and effectively utilize this health data. One of the best techniques of data warehouse is to integrate data of different formats from different source of data*. *This research paper presents a case study depicting how the architecture of data warehouse can be ideal for health care industry specific for COVID- 19. Which could be used by the members of the health care such as administrator of database or executive manager, physicians, nurses, doctors, other healthcare professionals. The entire information including their medical test reports comprising of present and historical health data about patient are stored in the database, which allows the executive manager to make a report by accessing those data that helps the doctor take the case forward. The proposed COVID- 19 DW is relied as a medium to conduct various (OLAP) operations, to obtain analytical results that help decision makers related to COVID-19 infections.*

***Keywords****-* *COVID Data Warehouse, On-Line Analytical Processing (OLAP) Extract- Transform- Load (ETL), Star Schema.*

1. **Introduction**

The world's fastest growing and information rich industries is undoubtedly the healthcare care industry. This industry has led to the adoption of robust healthcare decision support systems due to the massive need for integrated healthcare. A powerful solution for information access problems and data integration is provided by the data warehouse. Data warehouse is an informational environment that renders the organization’s information consistent and supports decision transactions possible without hindering operational systems and gives an integrated insight to the current and historical information of the enterprise that could be available easily for decision making. It is a quite flexible and feasible solution that keeps the strategic information consistent in the industry. Data warehouse is defined as “subject-oriented, integrated, time variant and non-volatile data collection in supportive of management’s decision-making process”[1].

Data Warehouses are designed as they are well driven by aiding the decision-making process free from complexity. For example, medical data for decision support, in contrast, contains a historical record of sales of medicines to patients over specific time periods. If well designed, Subject-oriented data depicts a stable image of medical environment, independent of legacy systems. Data warehouses contain information from various legacy systems that in turn create a stir disorder and inconsistencies among the measuring units. They have to be compiled in a consistent format and thus become integrated. Data should never change once entered. A user analyses what has happened. The data is just appended and not replaced so the database consistently keeps absorbing the new data compiling it with the older data. At the point of variance, there is a difference between Operational data and Informational data. Data warehouse represents data for a prolonged duration thereby making the historical analysis easily perform.

This paper, focuses on the COVID-19 diseases, the cost of treatment for these diseases, Death rate in particular type of COVID-19 and the repercussion of a specific drug on the disease. COVID-19 widely known as the “coronavirus”, which may cause illness in animals or humans.  In humans, by causing respiratory infections ranging from the common cold to more severe diseases. The virus can spread whilst contacting the infected surface. This is why it is advised to maintain more than 1-meter (3 feet) distance from the sick person. The COVID-19 pandemic possesses several unprecedented challenges in different fields to manage, by aiming to face the Covid-19 data collection, this paper presents the design and development of the COVID-WAREHOUSE, to allow their full exploitation and world-wide collaboration for the easy availability that may benefit from computing methods able to integrate the increasing available COVID-19 and related data daily. The orgnaization the paper is as follows. Section 2 presents review of literature on data warehousing. Furthermore, healthcare decision-making challenges and implementation concept of healthcare data warehouse. Section 3 introduces on data warehouse architecture types including details on healthcare information management. In section 4 building the healthcare data warehouse for COVID-19 discussed. In Section 5 the design of the architecture is presented. Section 6 presents the star schema and Section 7 presents the covid-19 data warehouse prototype development. The experimental result analysis in section 8 and conclusion in section 9.

1. **Literature Review**

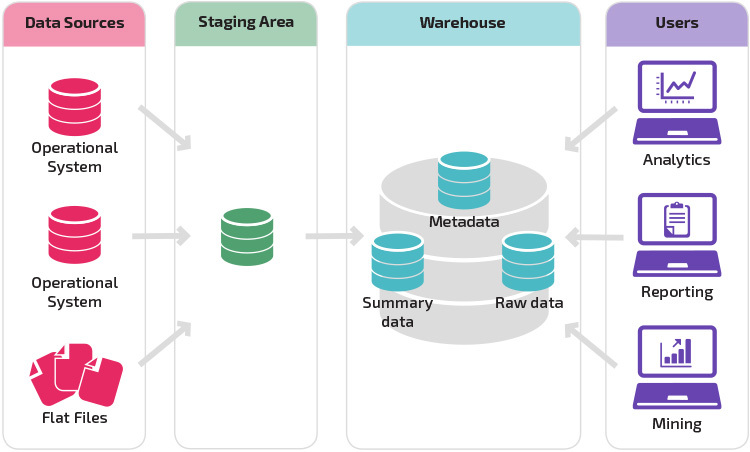
Today, data warehouses are not only widely deployed in banking and finance, consumer goods and retail distribution and manufacturing based on demand, but have also become a hot topic in the non-commercial market, predominantly in the medical fields, government, armed forces, education and research community, etc[1]. Data warehousing originated with pomp in the 1980s. Along with its existence in the fields of finance, consumer goods consumption it has extensively been in prominence in the non-commercial sectors as well. The growth of the application of Electronic Medical Records (EMR) in the health sector has led to the origination of clinical data warehouses[2]. These data warehouses currently exceed 150 terabytes in size. Along with the record of the patient, it also contains Alphanumeric administrative data, x-ray pictures, echography, electrocardiogram, etc[3]. “Research Issues in Clinical Data Warehousing” that clinical data warehouse needs to support for “complex-data modeling features, advanced temporal support, advanced classification structures, continuously valued data, dimensionally reduced data, and the integration of very complex data”. Hence, clinical data warehouse requires advanced data modeling than conventional multidimensional data warehousing approaches[2][4].

A healthcare data warehouse was implemented depending on Electronic Health Records (EHR) by processing them to Electronic Medical Records (EMR) and made them as a source of the proposed DW. OLAP operations are applied to view more analytical results to support clinical decisions[5]. The evaluation model based on the definition of data warehouse is proposed to evaluate the architecture of cancer data warehouse[5]. In radiology-sinology (DWRS) the data warehouse system was proposed to develop and design to assist breast cancer screening in diagnosis, education and research[5]. To incorporate information about a prostate cancer research including demographics, medical history, operation information, laboratory, and quality of life surveys lead to the development of prostate cancer research database system. To produce a comprehensive data base their system includes three different ways of clinical data collection; extracting the data directly from electronic medical record (EMR) system, manual data entry after linking EMR documents like magnetic resonance imaging findings and paper-based data collection for patient surveys[5][6]. A health care data warehouse architecture for diabetes diseases that could be used to control diabetes disease was proposed. In addition, it can also be used by healthcare administrators, clinicians, doctors and other health practitioners to promote the development, healthcare data is processed and analyzed, and to provide the ability to dramatically improve the practice and delivery of healthcare and medical research. They calculate the cost of diseases and identify drug errors[7].

Centered on Electronic Health Records (EHR), the healthcare data warehouse was introduced by processing them into Electronic Medical Records (EMR) and making them the source of the proposed DW. In order to help clinical decisions, OLAP operations are introduced to see further analytical outcomes [8]. To show the centroids and the districts of groups on a map, a population-based cancer data warehouse and a variant of the K-means clustering algorithm were developed. Because of its visual expressiveness, this instrument proved to be particularly useful for analyzing and transmitting the outcomes[9]. Detection of liver disease is the secret to its treatment at its early stage. This results in improved performance of the classification models in terms of their predictive or predictive models. Descriptive precision, decreased computation time required to construct models as they learn faster, and increased model comprehension.  A comparative study of the accuracy of data classification using data on liver disease in various scenarios. The predictive outputs of common classifiers are quantitatively compared[10]. Clinical Decision-Support Systems (CDSS) are designed to support decision-making processes by addressing knowledge-intensive tasks. While several approaches have been proposed to design CDSS, existing solutions are not well-established across numerous organizations due to high implementation costs, as well as the lack of interoperability functionality. Recently, the use of structured formalisms as terminologies for knowledge representation as well as the incorporation of semantically enriched clinical information models as openEHR archetypes and their reuse within CDSS are theoretically regarded as key factors for reusable CDSS[11].

1. **Data Warehouse Architecture**

The data warehouse is the most professional and profound way for providing flexible and interactive source of strategic information for decision making thereby providing an insight to the enterprise's current and historical information that could be easy to access. “It is user-driven, flexible, fluid, immersive, subject oriented, integrated, nonvolatile and time variant ideal environment for data collection in support of management’s decisions and data analysis”[12][13]. As per the organization’s requirement the software and hardware components are arranged in a particular manner that may yield in maximum advantage. As like any project in software development [12] follows series of protocol likewise, data warehouse also ensures the design of working system has fully met the user requirement and is delivered on time. The data warehouse system architecture is designed with two blocks under Data Acquisition Area Namely Source Data components on left side and the next building block presenting the Data Staging component and to manage the data of the data warehouse the data storage component is placed in the middle, along with these components it also comprises of metadata and data mart to keep a track of the data. Information delivery is the architecture's last component that displays various aspects in which the data warehouse information is made accessible to the user.



**Fig.1. Data Warehouse Architecture**

In data acquisition, Data is gathered from various sources, extracting it to the staging area and preparing it to load it in the warehouse. Source data and data staging are the two major architectural components. This source data is categorized into 4 namely- Production Data, Internal Data, Archived Data and External Data[12]. The source data is then converted into the data warehouse with suitable formats done by the staging section. The function and services for this field as follows: The source data in data extraction, may be belonging to several other machines in various formats. It may be in the forms of hierarchical modules, or files from spreadsheets or part of relational database system. So, this is extracted into a separate physical environment (group of flat files, or a data-staging relational database, or a combination of both) and fed into data warehouse[12] [1].

Since the data is entered into the warehouse from various formats there are many tasks that need to be done for the transformation in data transformation stage. The data needs to be cleaned. This needs to be carefully done as it involves an introspection on minute details like checking the state code, zip code. Data transformation also includes combining the different data source. When the function of data transformation ends the data is then cleaned, standardized and summarized in a standard pace once integrated. Data loading includes 2 tasks. Firstly, the design is laid and constructed thereby feeding a large amount of data into the warehouse. Then there is a continuous process of data source extraction as the warehouse keeps functioning. During the data storage phase, data is loaded into the data warehouse in the daily basic and Large amount of historical data are required for analysis[12]. Data repositories contain the essential information in highly normalized form for effective functioning. In data warehouse for quick retrieval of individual data, the data storage is kept separate. These repositories are read-only. The function and services are: Load data to completely update the tables of data warehouse and Optimize the process of loading [1]. In information delivery phase, the data warehouse information is collected by the user. Gathering information from the data warehouse involves many information delivery methods for various users. These information delivery components make it easy to access. The predefined reports are ad hoc reports for casual and novice user[12]. Provision for complex queries, statistical analysis and multidimensional analysis are applicable to the need of the business analysts and power users. This information fed into the system is intended specifically for senior executives. The function and services for this field are: Enables user to acquire information from the data warehouse[1], the aggregate table allows execution of analysis and results faster using queries[1] and it is easy to perform complex analysis by using Online Analytical Processing (OLAP).

1. **Building Health Care Data Warehouse for Covid-19**

The most important thing for any health care organization is both successful health care and financial sustainability. Data on the accuracy of the diagnosis, the efficacy of the procedures, the performance of the physicians and the cost of the service are critical concerns for the health care center. The health care sector is unique should aim to balance efforts to enhance the quality of individuals with efforts to minimize cost for employers at par with the governments. Recently, many forms of COVID-19 viruses have spread in seasonal epidemics The design of a health care center data warehouse for COVID-19 diagnostics is very different from other industries. But much like the software development project, the data warehouse retains a stage-by-step process. Now since this pandemic is the highest issue to be pondered upon, this paper outlines the following two basic stage to build a health care data warehouse for COVID-19 diagnostics.



**Fig.2. Health care value circle**

A data warehouse is a business intelligence delivery system of information. This provides solutions to the user’s problem with strategic information at phase of defining the requirements, in accordance to what exactly the users need. With the approach of top-down, the warehouse is designed on a third normal form relational data model which in turn forms the data warehouse[12]. This business analysis stage consists of the business requirement and business process analysis. The Business process analysis involves a case diagram of 4 actors namely the Doctors, Patient, Pathologist and an Executive Manager. As the patient comes to treat themselves to the health care centers. the doctor sends them to the pathology according to the symptoms for any medical examination. As per their reports of medical test, doctors diagnose the flu and the treatment begins.

**Fig.3. COVID-19 data warehouse use case diagram.**

Doctor

Pathologist

Patient

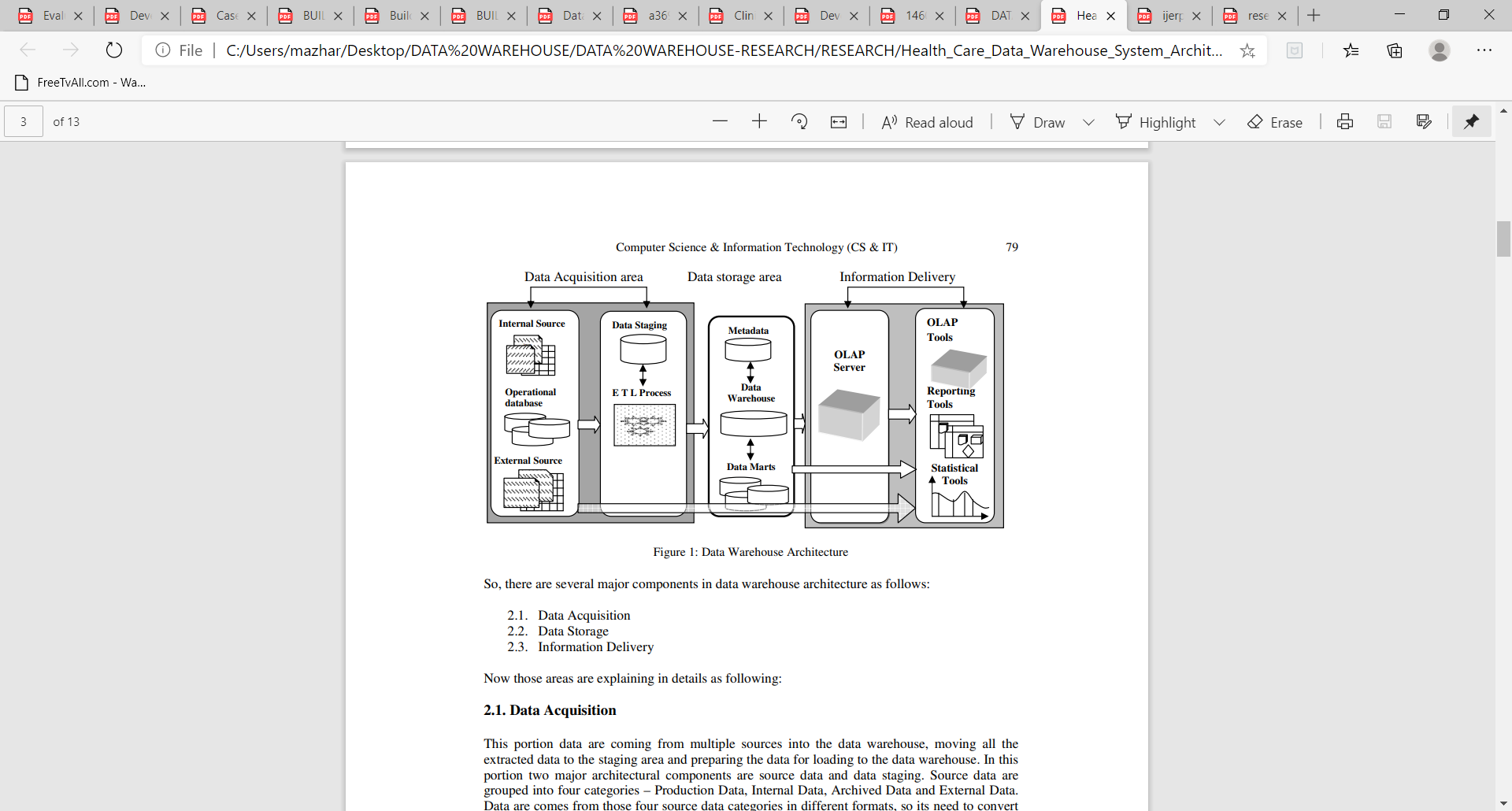
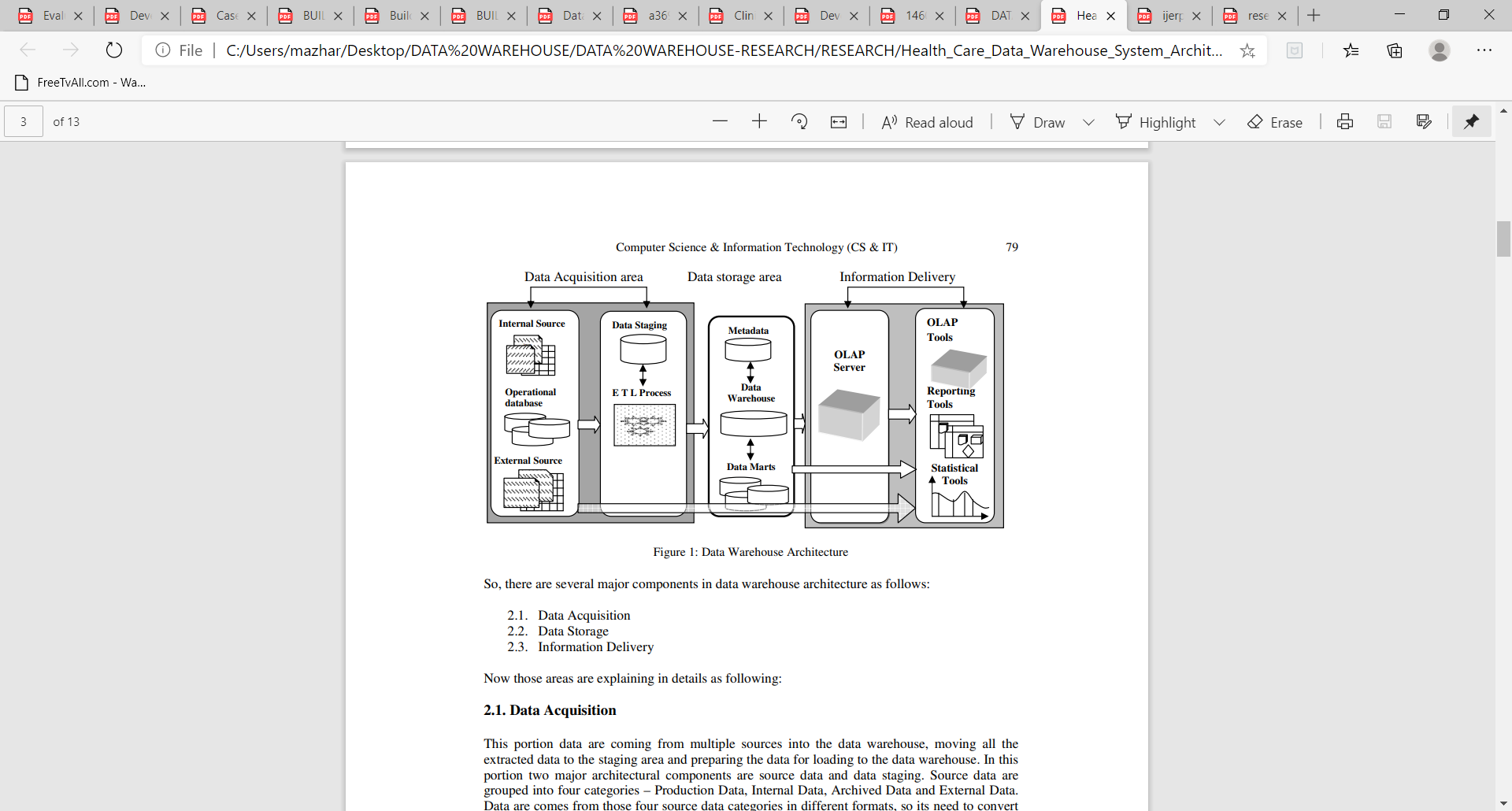
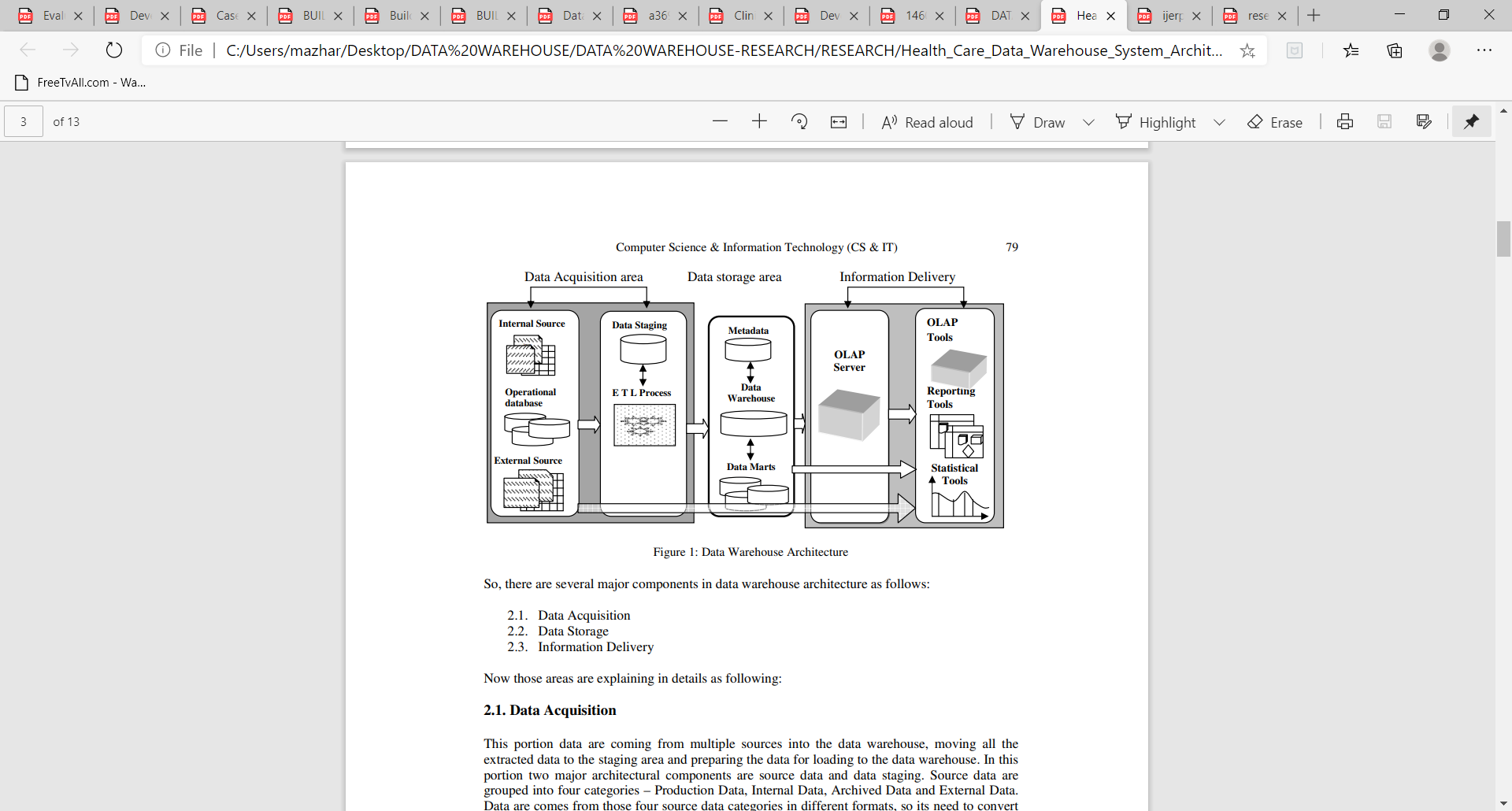
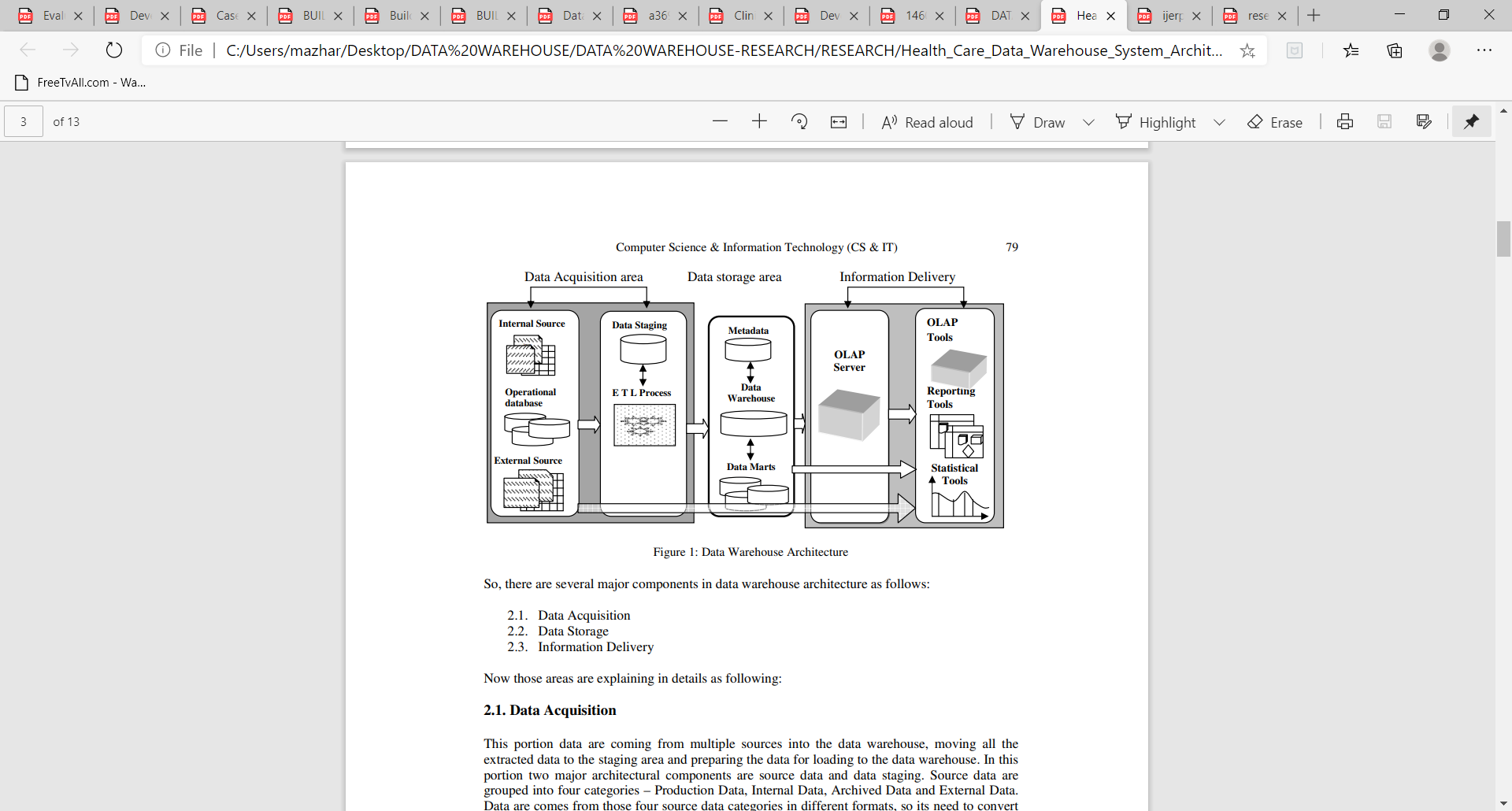
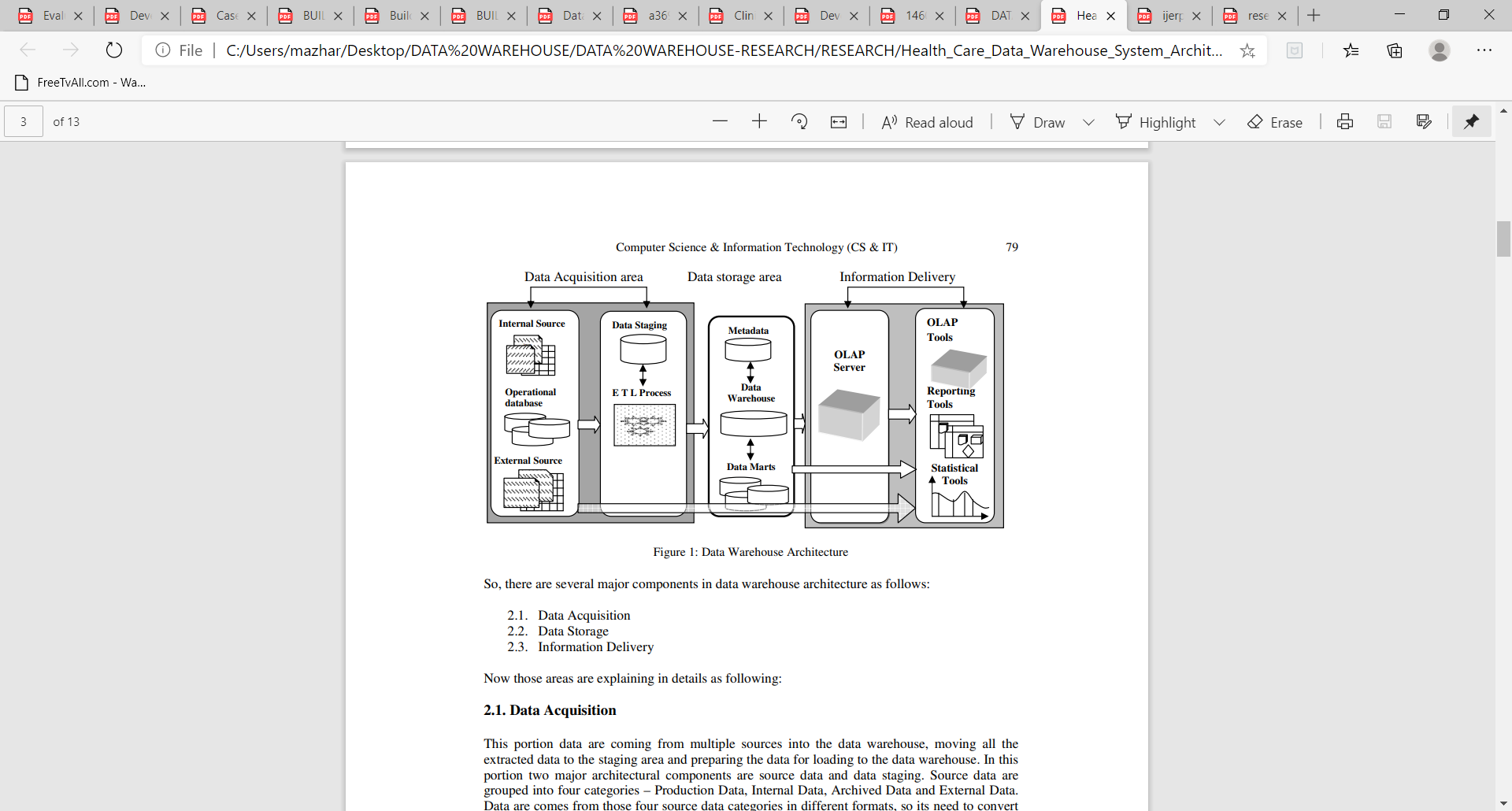
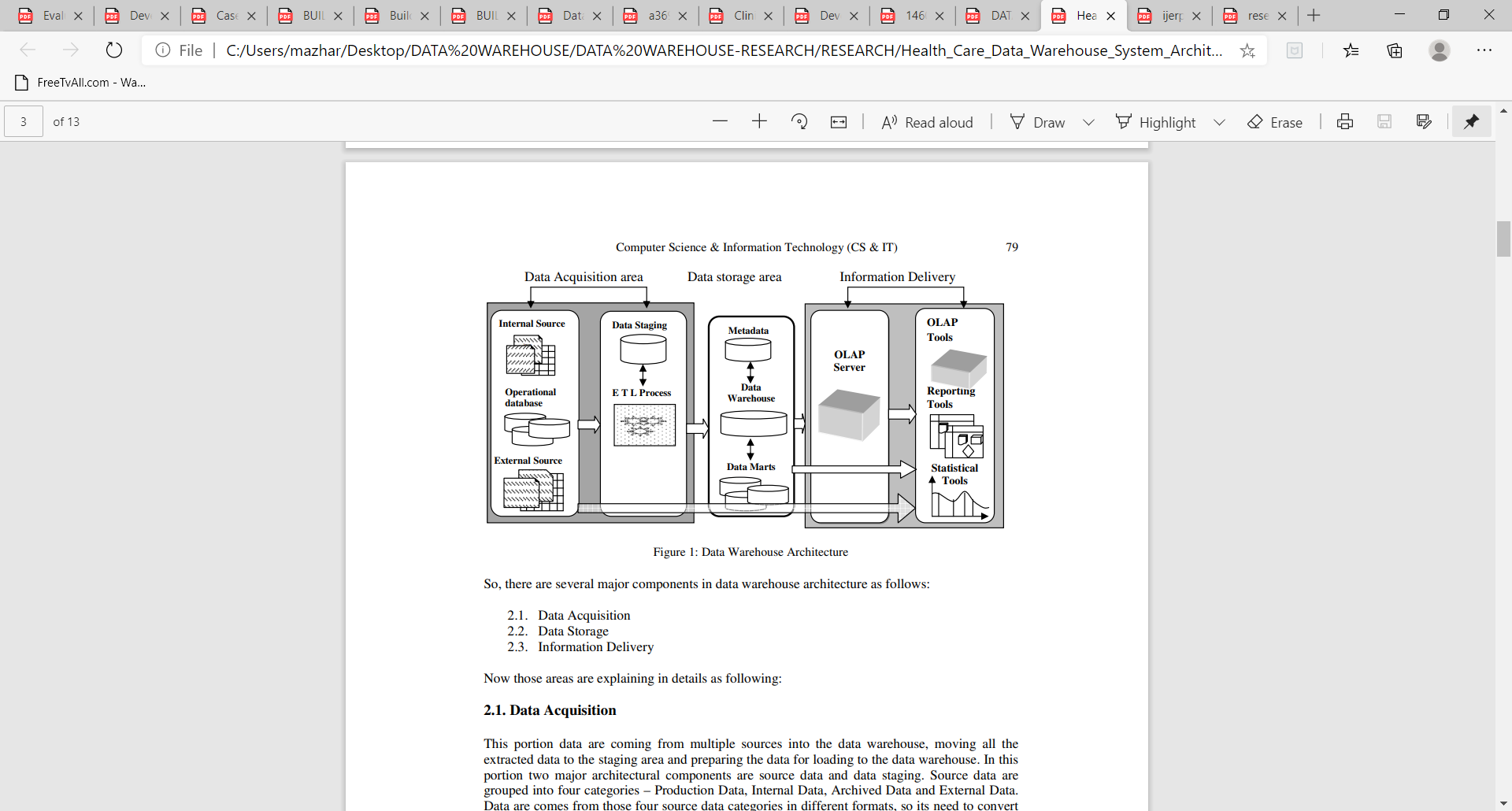
Patients visits the doctor to seek Consultation when certain symptoms are encountered. The doctors and the pathologist carry a series of examinations like Blood tests, Nasopharyngeal culture to determine COVID-19[14]. And to see how the disease has advanced and how far it has spread. Next, to determine the prognosis of the disease the doctor will look into other factors. The doctor then begins the treatment by giving an antiviral medication and advices to stay away from others especially children and aged people. The patients are also advised to take plenty of rest and intake liquids and avoid using liquor and smoking.

In business requirement analysis, some significant health care data warehouse proposed requirements are: Minimum level of dimensional nature about the patient business data required where patient details is stored to support COVID-19 diagnosis and treatment recommended by the doctors[12]. The records of the patient like his full name, date of birth, Gender, age, Contact number, Address, Disease details, treatment under which doctor etc.,[1][7] are all fed here. A unique ID number that is also called the primary key is used to recognize the record. This preventing data duplication and is easily accessible while searching. The medical diagnostic function includes updating report of the patient's medical history, symptoms, medication experiences before and after treatment. This allows the system to showcase at both summary and comprehensive levels, by which the use (doctors) enables to have a clear approach towards the diagnosis.

1. **Architecture Design**

The architecture is a description of the warehouse components, with details indicating how the components will fit together. While providing incremental solutions, the Data Warehouse Architecture provides an integrated data warehouse environment. The application of a centralized data warehouse, data marts, individual marts, metadata repositories, and incremental solution architectures is the focus of architectural design. There are various types of data warehouses that have various design and structures. While some have ODS- Operational Data Store, others have Multiple Data Marts. Some may have a small number of data sources, while some may have dozens of data sources. Since a data warehouse aids in decision making, it is important that the data extracted from multiple sources must be corrected as there is a possibility of many errors and anomalies. Therefore, tools are required for data extraction, data cleaning, data integration and finally data load. The overall propose COVID-19 data warehouse architecture before getting integrated, the data is imported from several data sources and stored in the production data warehouse for analysis.

**Fig.4. Data Warehouse Architecture for COVID- 19 Disease.**



Data Acquisition Area

Storage area

Information Delivery

**COVID-19**

**Medical files**

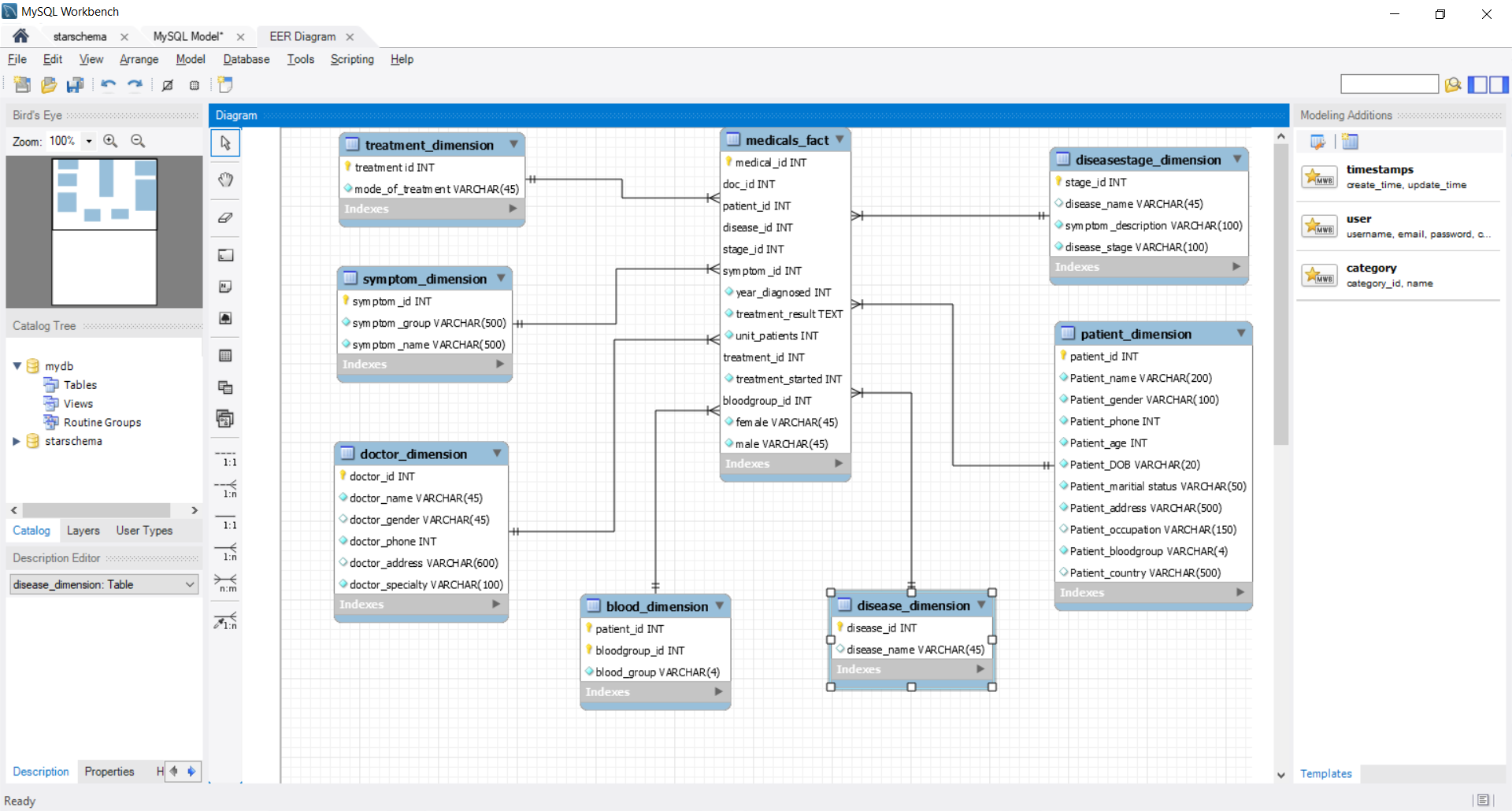
The data acquisitions primary aim is to identify, extract, gather, transform and transport the multiple source data into the data warehouse for the necessary operation. Data acquisition is carried out between several warehouse components, including data warehouse operational and external data sources, data warehouse to data mart, and data mart to individual marts. In the architectural acquisition all data from the database, medical files such as medical records, other tests and results are extracted to the staging area and loaded into the data warehouse repository. This stage involves data extraction and data transformation.

During the data extraction phase, the data is selected from the multiple sources (medical files) and ascertain the types of filters to be applied to COVID- 19 data warehouse. Automatic files extraction from operational systems are Generated by replication and other techniques[12]. Creating intermediary files to store and merge the data selected. Common application codes are created by transporting extracted files from various platforms using automated job control services to overcome inconsistencies. And in data transformation process, the extracted data from databases and medical files is cleaned, de duplicate, merged and denormalized to map this data as required by the dimensional model to data warehouse repository of the COVID- 19 disease data warehouse. Data is consolidated and integrated by fixing missing values and converting Data Types, Calculate and Derive Attribute Values, Verify Referential integrity, aggregate data as appropriate[12].

In data storage stage the transformed and cleansed data is stored. On the basis of scope and functionality there are 3 entities namely the Data warehouse, Data mart, ODS are found. The data from the database and medical files is transformed and integrated from staging area to the data warehouse repository. The varied functions include: Health care data is stored in the data warehouse tables to Optimize the process of loading and Perform incremental loads at regular specified intervals in order to Support loading integrated data into multiple tables. Data presentation stage refers to the information that reaches the users. This is in the format of a graph or a tabular report in a browser and is automatically generated. It is sent via email on a daily basis. An OLAP tool or a Reporting tool is used here. This presentation makes the decision makers and doctors to access information directly from the warehouse. The functions included here are: The contents of data warehouse can be browsed by the doctors and decision makers, Queries of aggregate tables is enabled for faster results and analysis, Multiple levels of data granularity are provided to improve service and for future enhancements monitor doctors and decision-making access.

1. **Star Schema**

Star schema will display the architecture used in dimensional modeling of the data warehouse, as it is simple for analysts to understand and design. A star schema consists of central table that forms the center of the schema called the fact table and may have any number of smaller tables known as dimension tables displayed across the central fact table in radial pattern. The dimension tables include denormalized data that promote the business intelligence and data mining techniques. Users that explicitly query the relational database are better able to navigate the model. Relational queries typically work best against this framework. The flattened table is also easier to handle with fewer tables and associated keys in the ETL process. The dimension table is connected to a fact table of one-to-many relationships. The primary key of the fact table is the concatenation of the primary keys of all dimension’s tables. The dimension tables provide the basis for the aggregation of the measurements in the fact table.



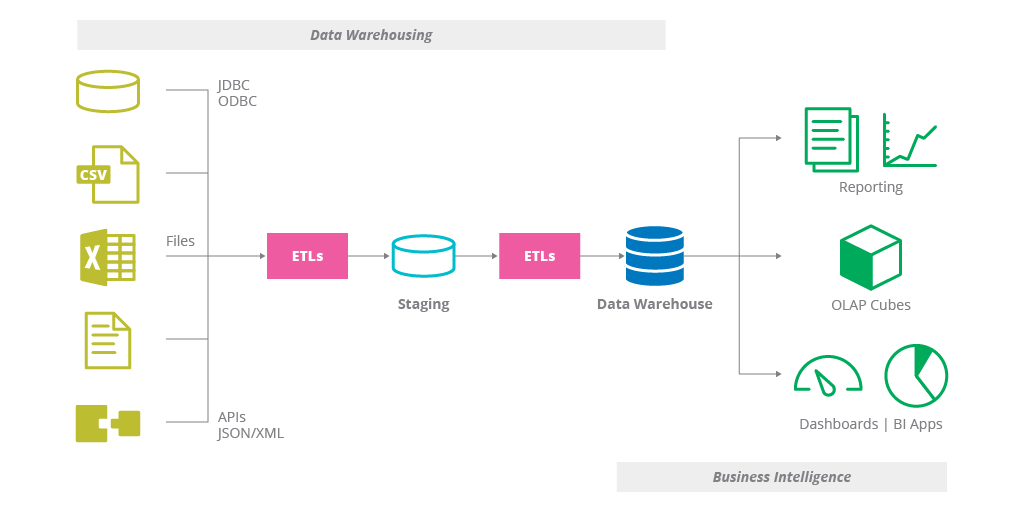
**Fig.5. Star Schema.**

The advantage of using star schemas to represent data is that it minimizes the number of tables in the database, the number of relationships between them and therefore the number of joins needed in user queries [15][16]. "First Principles" methodology focused on user query analysis. It starts with the identification of the appropriate "facts" that need to be aggregated, the dimensional attributes to aggregate by and the creation of star schemas based on these. It results in a data warehouse design which is a set of discrete star schemas.[15] However, this approach has practical disadvantages the requirements for user analysis are extremely unpredictable and open to evolve over time, which gives an inconsistent design basis. If the designer does not understand the underlying data relationships, it may lead to incorrect designs which subsequently leads to information loss by premature aggregation, which restricts the way the data can be analyzed. Therefore, the technique is described through examples rather than an explicit design procedure.

The Patient\_Dimension table holds information about the patient, such as name of the patient, gender, age, disease, phone number, blood group etc. All possible treatment option is stored in the Treatment\_Dimension. A table that stores the entire symptom, the normal condition values and abnormal condition values[12] is stored in the Symptom\_Dimension. Disease\_Type\_Dimension: In this table all the diseases and the types of the diseases are stored. In this case only all types of COVID-19 related data are stored and the diseasestage \_dimension: A table that stores the staging systems which are specific for each type of disease[4]. In this project only information pertaining to the subject matter[2], i.e. COVID- 19 disease will be created. A Blood\_Dimension table comprised of every patients’ blood group. And lastly, the Doctor\_Dimension: doctor information is stored in this table, such as name, gender, phone, address, etc.

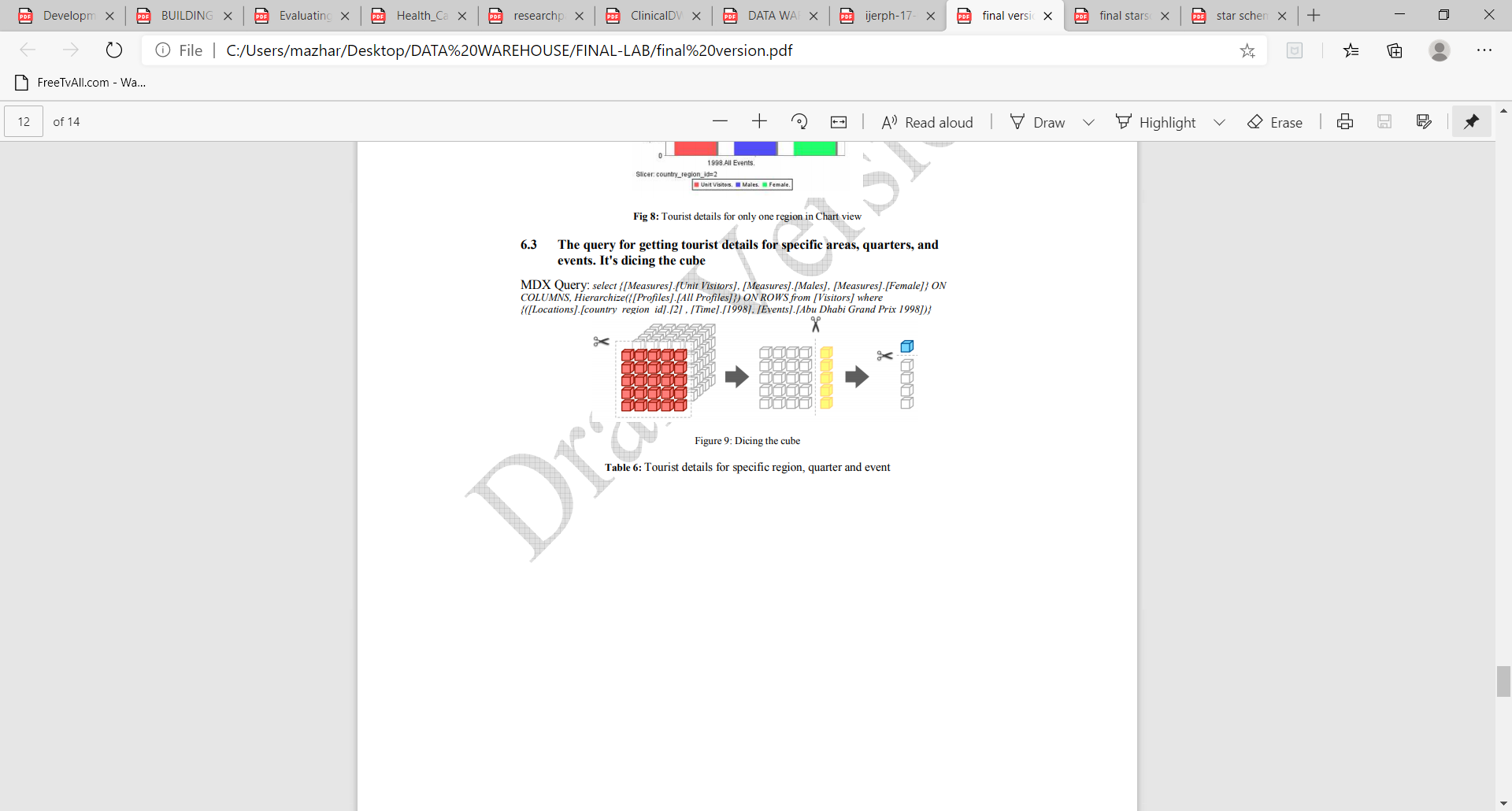
1. **Data warehouse development**

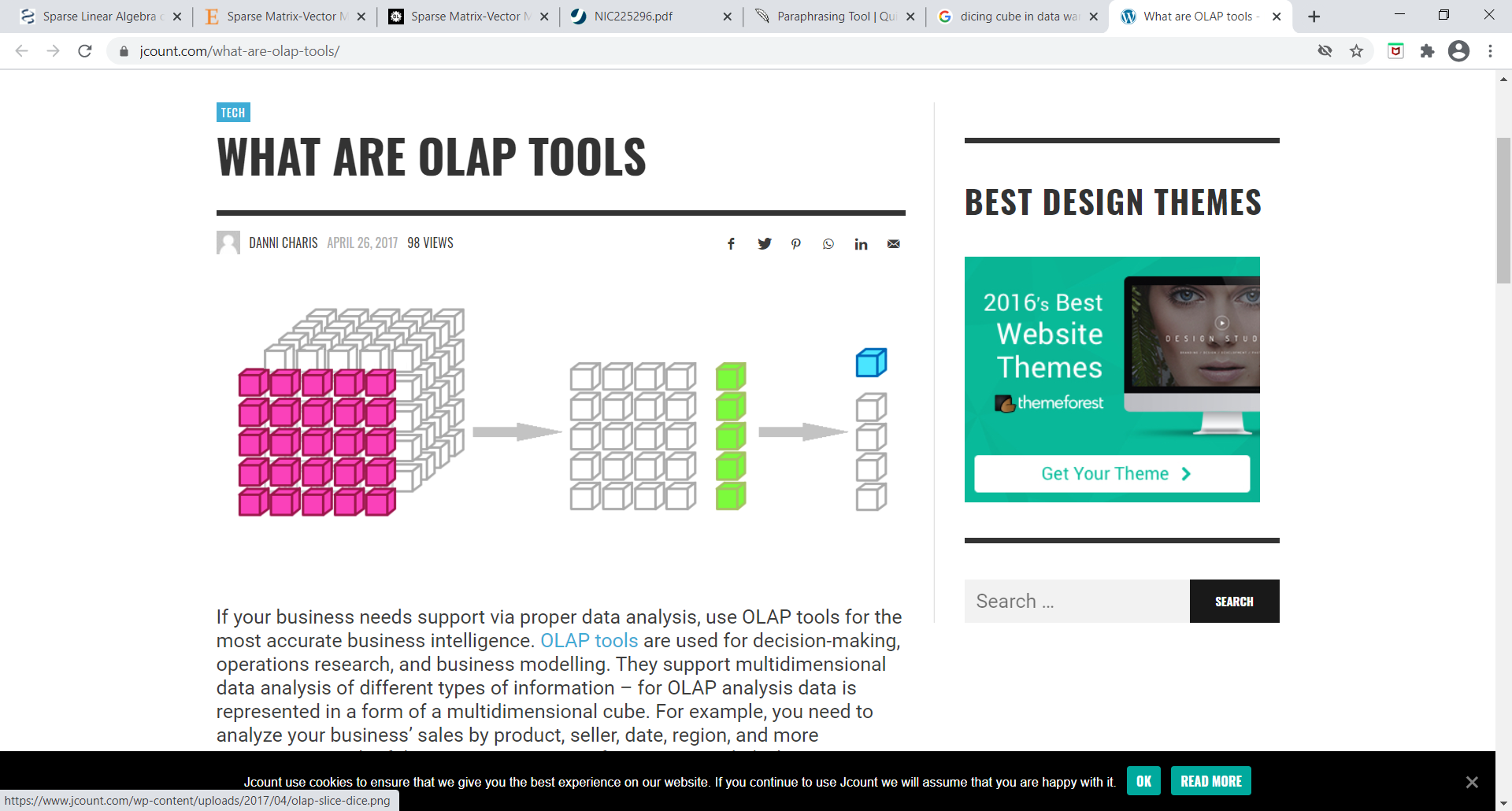
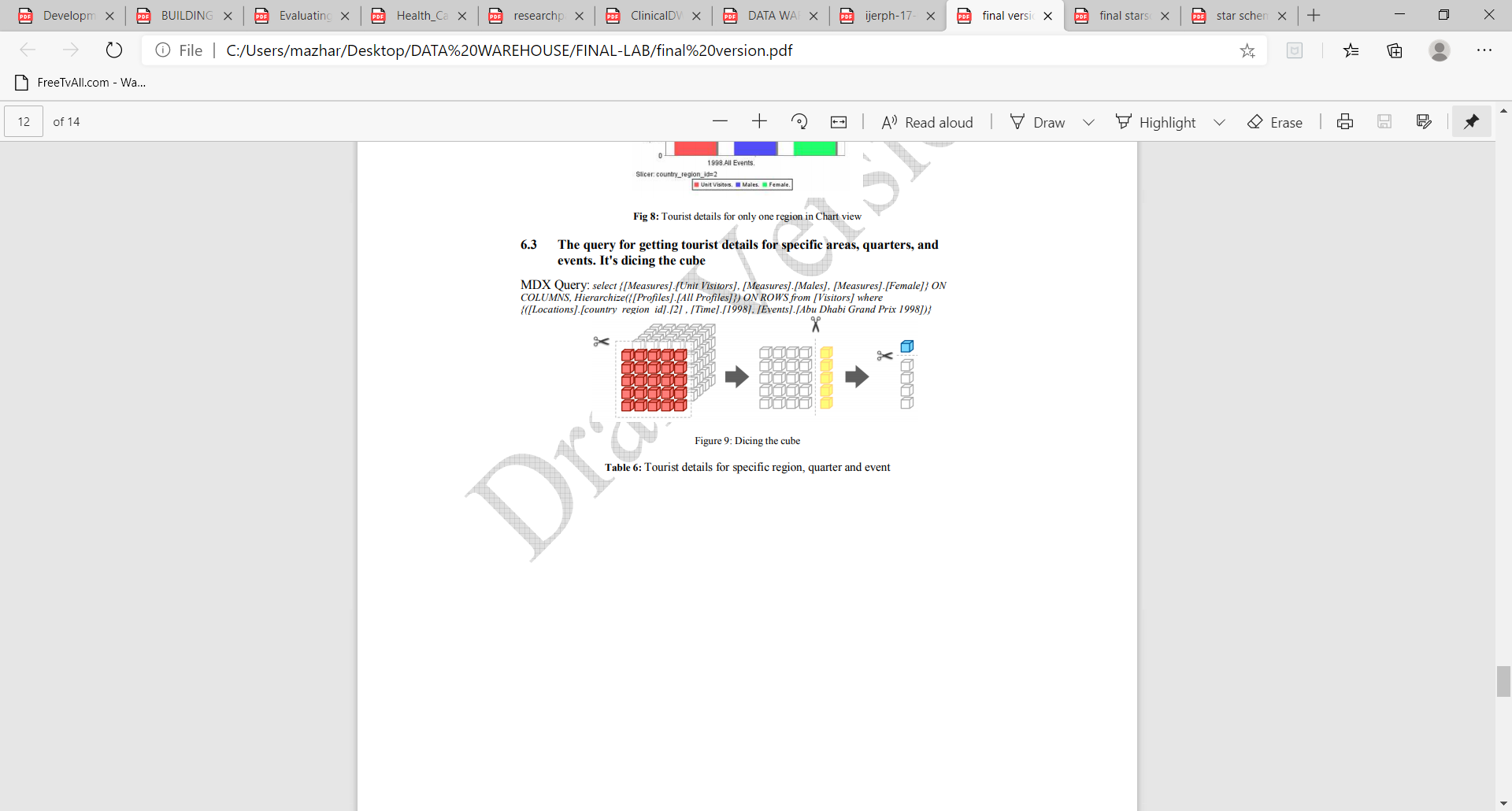
This method consists of all processes for the extraction and transformation of data types and the filling of missing values from the data source of the staging table. It also contains operations for the derivation of new attributes by grouping or classification. Transformation happens by using rules or lookup tables, or by integrating data with other data. Loading is the method of writing the data to the target database. Data is extracted and stored in various excel documents. Each dimension table and fact table data are cleaned and created in separate excel or CSV documents. Imported data of the fact and dimensional to data mart in the COVID 19 data warehouse. Each CSV Excel document will be added one by one. For Mondrian schema the Pentaho workbench tool is used. A Mondrian schema includes a logical model consisting of cube, hierarchy and members and a physical model mapping of this model. The Mondrian schema generated is examined by choosing the new scheme file, then MDX query. The MDX query is the SELECT statement, the most popular query in MDX.   When understanding the MDX SELECT statement, you will have a clear understanding of how to use MDX to request multidimensional data, what the syntax of the SELECT statement is and how to construct a simple query by using the SELECT state.



**Fig.6. ETL Process of Data Warehouse for COVID- 19.**

A cube is developed on the basis of COVID- 19 DW dimensions for OLAP operations to execute. It has been designed and carried out by pentaho. The dimensions and Hierarchies are implemented to authorize OLAP operations (slice, dice, drill through, drill up and drill down). The cube is built based on all the dimensions. Thereafter the Cube implementation is done, can be directly accessed by dropping dimensions members and measurement. Special cubes can be developed for Data analysis based on unique dimensions and measurement.



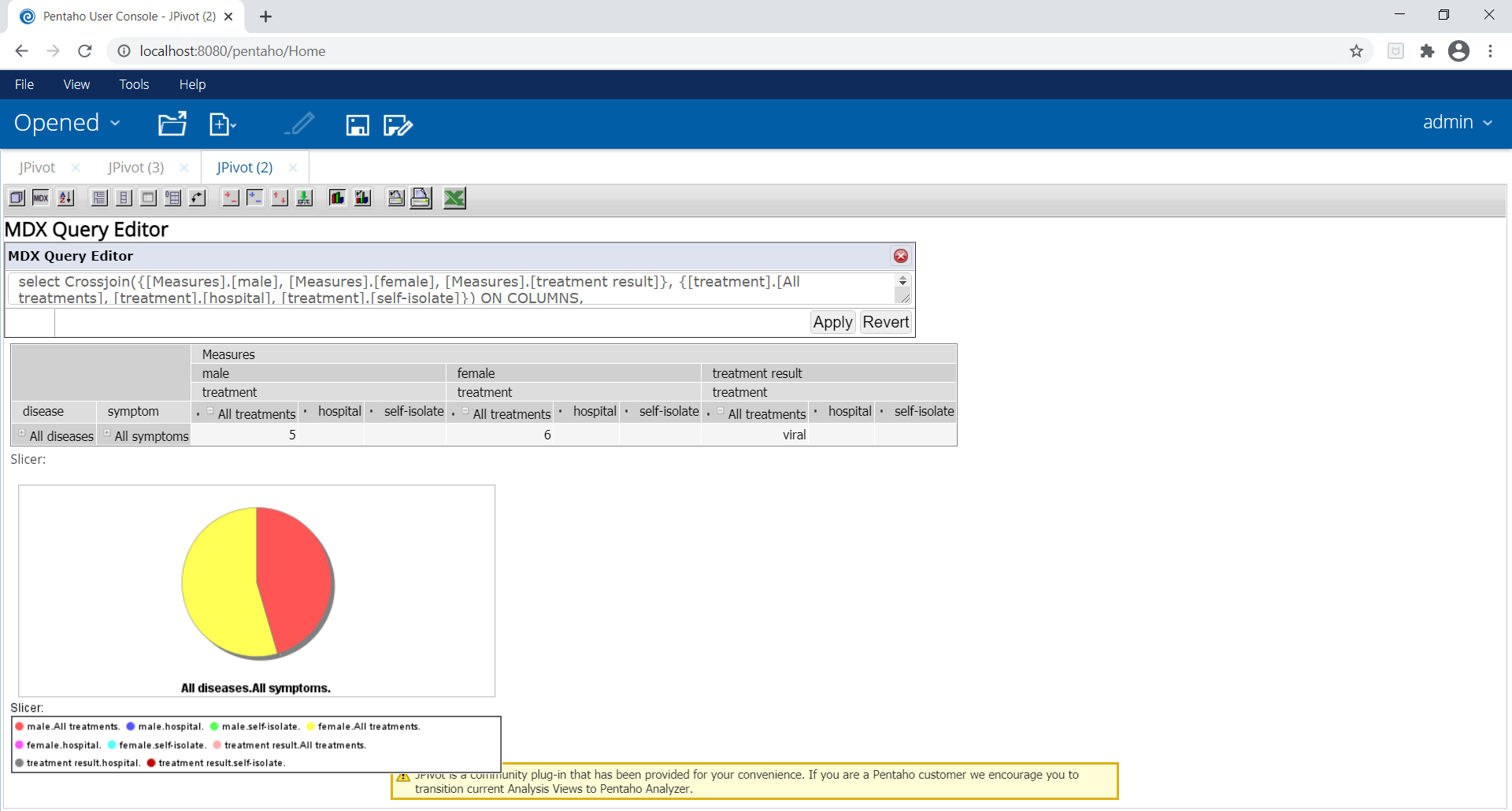


**Fig.7. Dicng the cube for one patient from patient\_dimension**

***MDX QUERY:*** select {[Measures].[unit patients], [Measures].[male], [Measures].[female]} ON COLUMNS, Hierarchize(Union(Union(Crossjoin({[symptom]}, {[patient]}), Crossjoin({[symptom]}, [patient].Children)), Union(Crossjoin([symptom].Children, {[patient]}), Crossjoin([symptom].Children, [patient].Children)))) ON ROWS from [starschema] where [patient].[patient\_id].[18]

1. **Experimental Result**

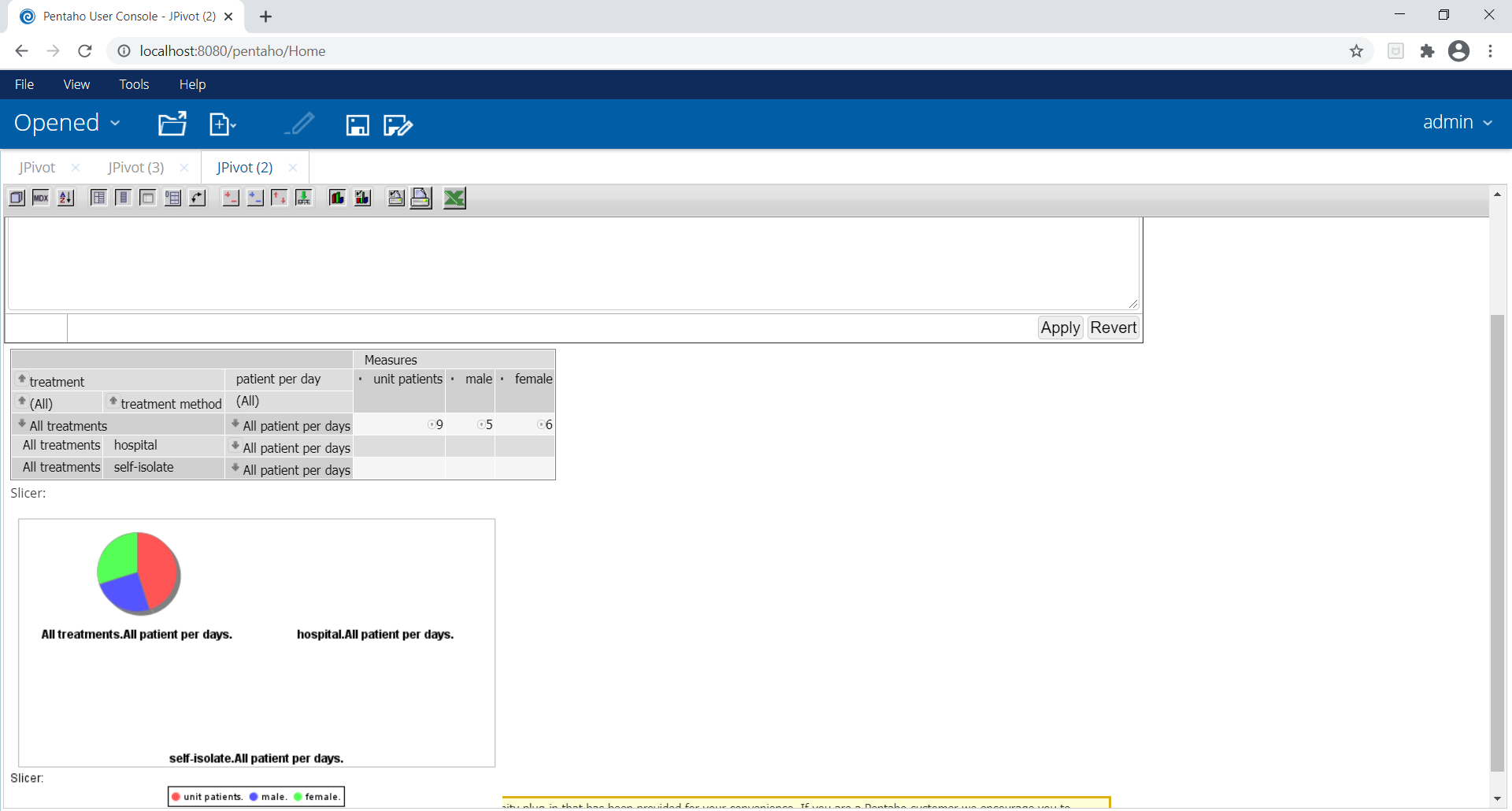
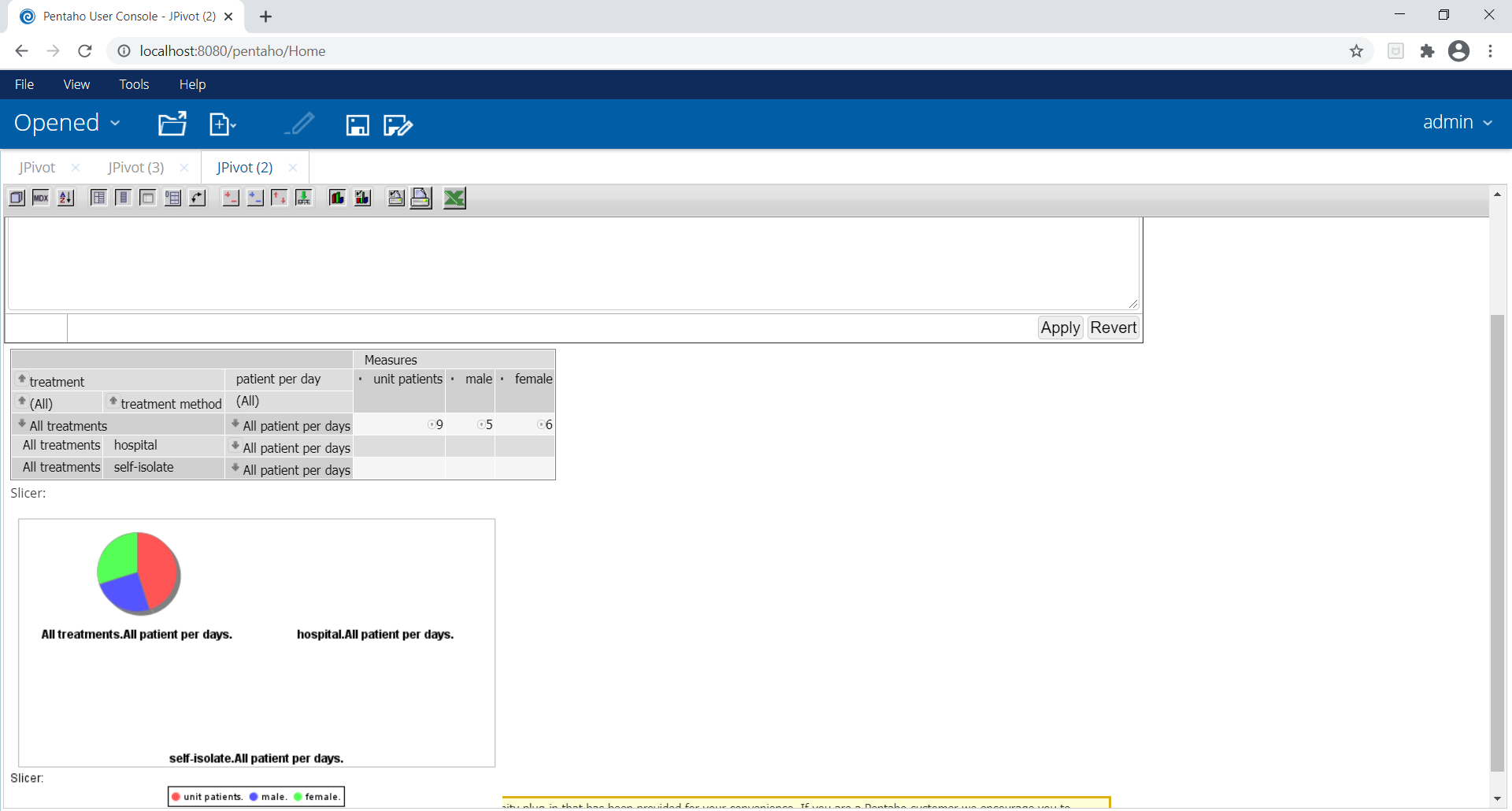
For the purpose of the experiments, the pseudo data is generated in a scientific manner to resemble COVID- 19 data. In Star Schema, the sample data are inserted. The Mondrian schema (Cube) is generated with the Pentaho workbench software. The Pentaho Server Mondrian schema has been validated and published. The MDX (Multidimensional Expression) query is used to query data from the COVID- 19 database of the warehouse to transform and view covid-19 data. For the transformation of cov-19 statistics, a tabular- and chart-format Pentaho data integration or J- Pivot View on the Pentaho BI tool is used. Some MDX queries and results are given below. The proposed prototype has been implemented using JAVA (JDK 1.8.0\_151), MySQL Workbench 8.0, Microsoft Excel, Pentaho BI server and Pentaho workbench.



**Fig.8.Patients diagnosed with all symptoms (male/female)**

***MDX QUERY*:** select Crossjoin({[Measures].[male], [Measures].[female], [Measures].[treatment result]}, {[treatment].[All treatments], [treatment].[hospital], [treatment].[self-isolate]}) ON COLUMNS, {([disease].[All diseases], [symptom].[All symptoms])} ON ROWS from [starschema]

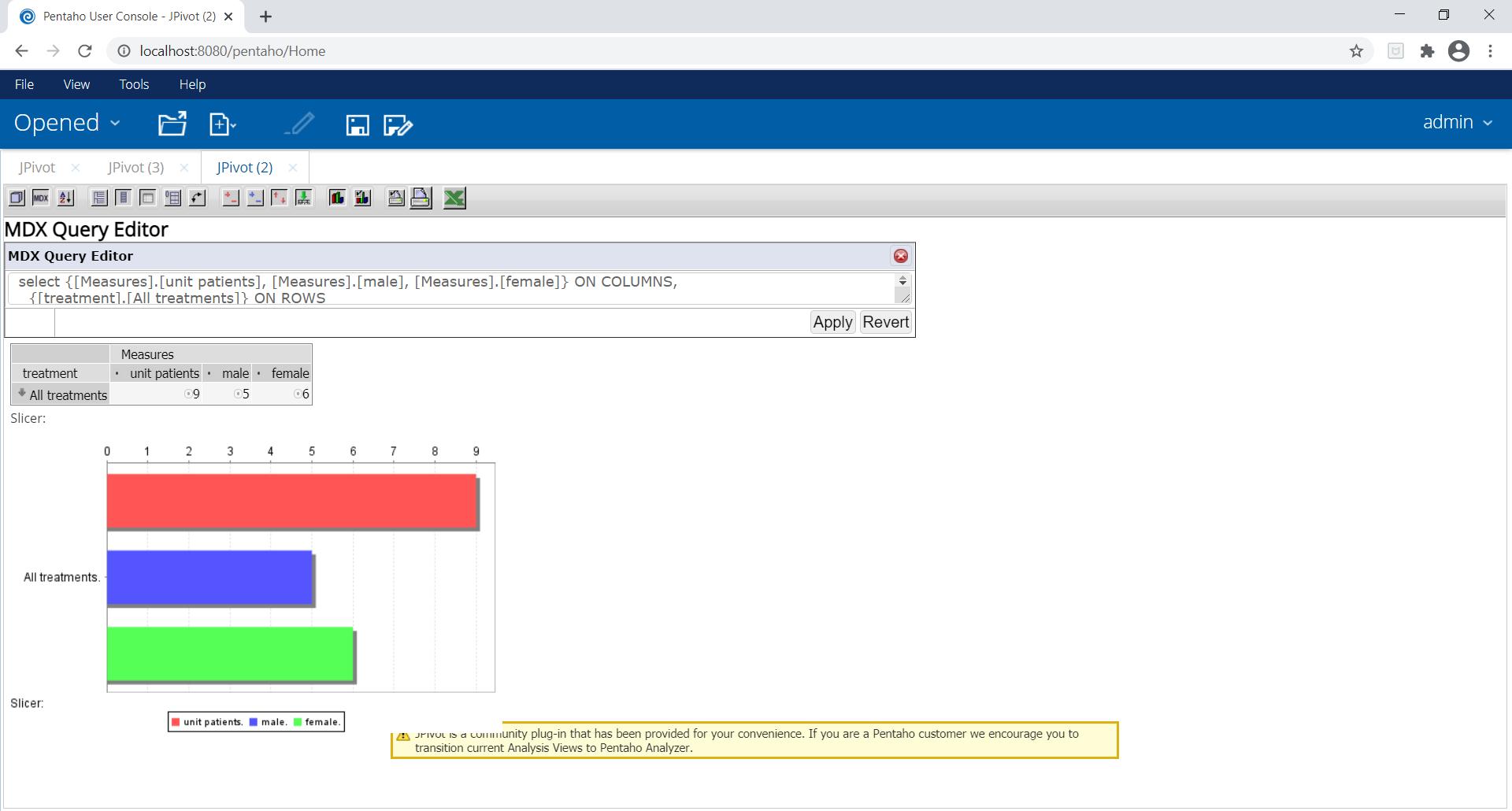
The above graph depicts the Patients categorized gender namely male and female diagnosed with all the seven symptoms such as SARs-COV, MER-COV, etc.



**Fig.9. total patients each day**

The graph indicates the total number of patients diagnosed with disease every day and are treated in hospital/are self-isolated 51% of women in comparison with men are diagnosed with coronavirus.

***MDX QUERY***: select {[Measures].[unit patients], [Measures].[male], [Measures].[female]} ON COLUMNS,Hierarchize(Union(Union(Crossjoin({[treatment]}, {[patient]}), Crossjoin({[treatment]}, [patient].Children)),Union(Crossjoin([treatment].Children,{[patient]}),Crossjoin([treatment].Children, [patient].Children)))) ON ROWS from [starschema]



COVID- 19

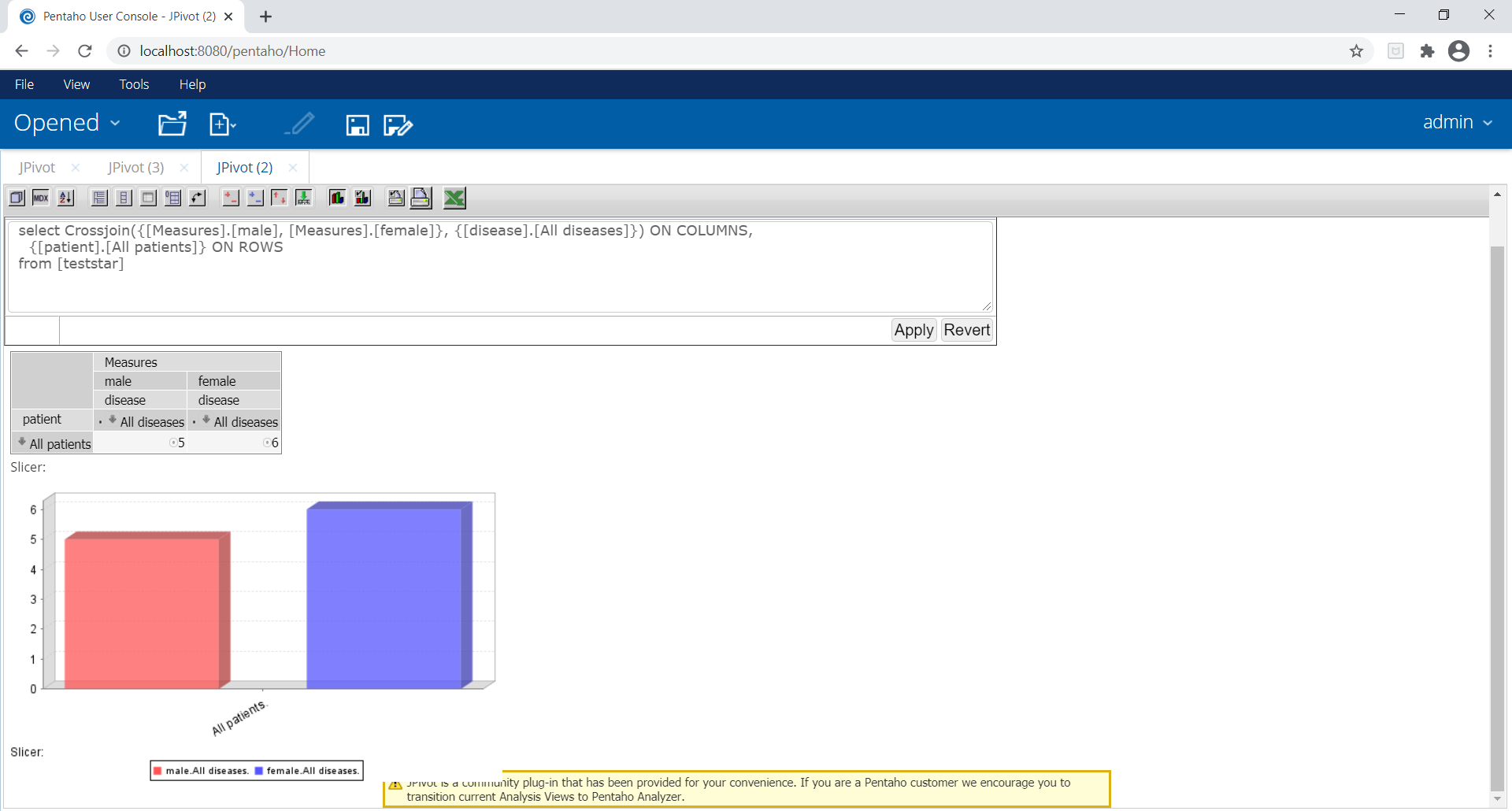
Self- isolate

hospital

**Fig.10. query for getting patients treatment details**

The above graph shows the query for getting total patients’ details (hospitalized/self-isolated) diagnosed with COVID-19 each day and therefore the maximum number of patients are hospitalized due to chronic illness whereas the patients with mild symptoms are self-isolated.

***MDX QUERY***: select {[Measures].[unit patients], [Measures].[male], [Measures].[female]} ON COLUMNS, {[treatment].[All treatments]}, {[treatment].[hospital]}, {[treatment].[self- isolate]} ON ROWS from [starschema] where [disease].[disease name].[covid- 19]

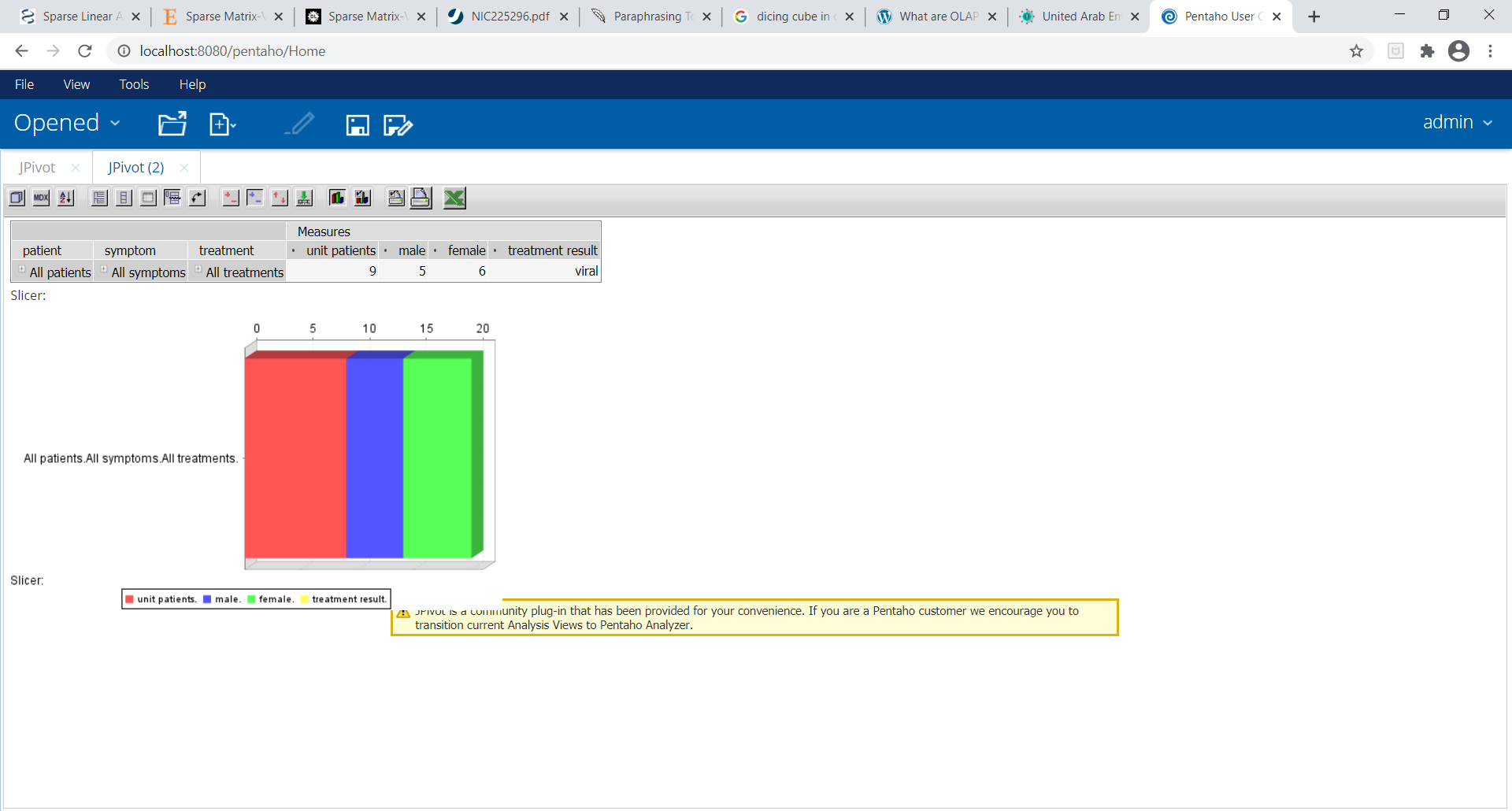


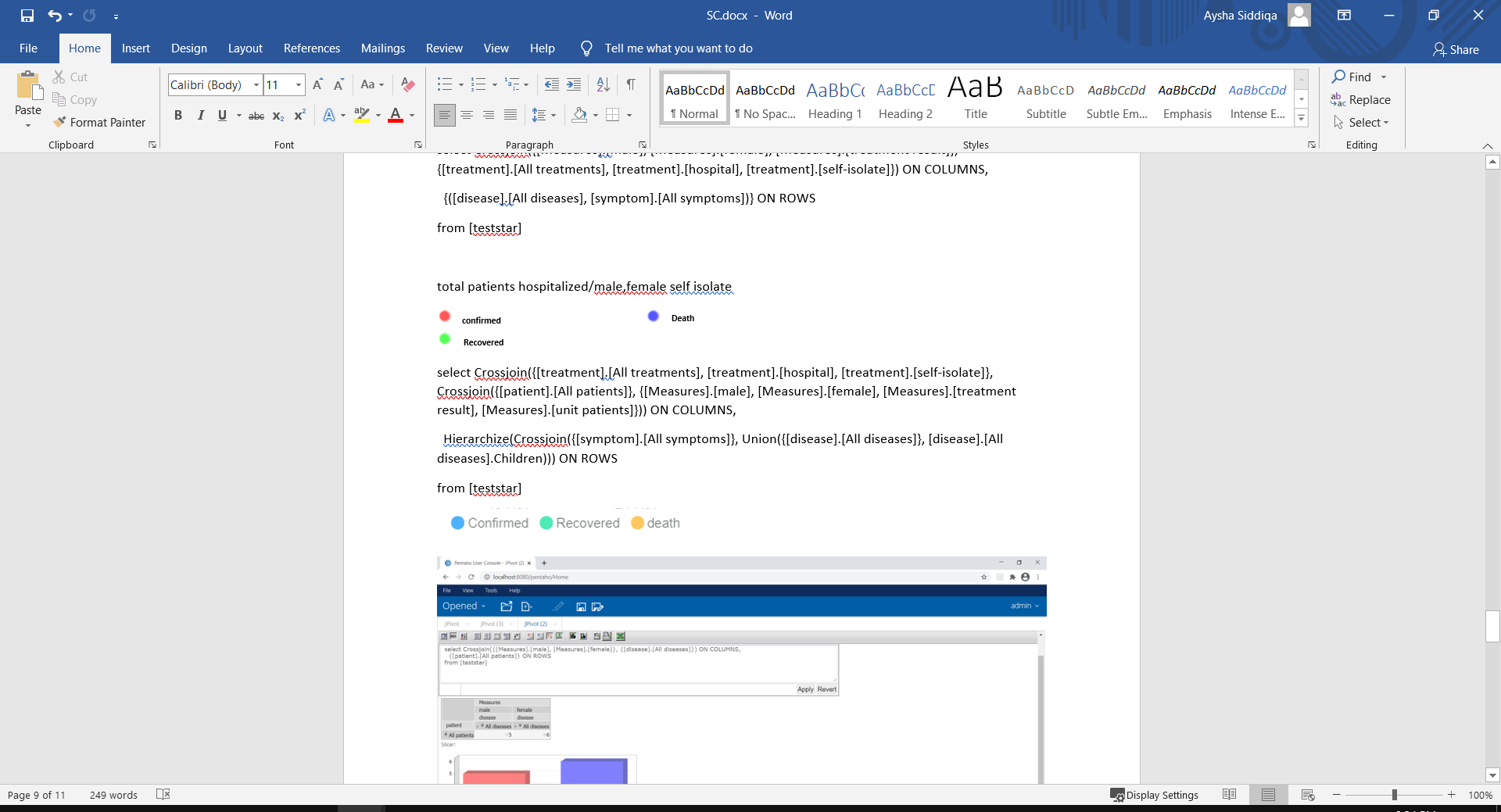
**Fig.11. comparison of covid-19 with other disease in 2020**

The chart above compares between COVID-19 disease to other diseases in the year 2020. The rise of coronavirus is 60-95% per month.

***MDX QUERY:*** select Crossjoin({[Measures].[male], [Measures].[female]}, {[disease].[All diseases].Children}) ON COLUMNS, {[patient].[All patients]} ON ROWS from [starschema]

The graph below describes the daily incidence chart, there have been almost 75,110,651 confirmed cases of COVID-19, including 1,680,395 deaths.





**Fig.12. daily incidence chart**

***MDX QUERY:*** select Crossjoin({[treatment].[All treatments], [treatment].[hospital], [treatment].[self-isolate]}, Crossjoin({[patient].[All patients]}, {[Measures].[male], [Measures].[female], [Measures].[treatment result], [Measures].[unit patients]})) ON COLUMNS, Hierarchize(Crossjoin({[symptom].[All symptoms]}, Union({[disease].[All diseases]}, [disease].[All diseases].Children))) ON ROWS from [starschema]

1. **Conclusion**

Developing health care data warehouses, lays quality of data high on the agenda. This health care industry is the fast developing, most data sufficient industry. In this paper, we proposed a new approach for building a data warehouse from a limited volume of data. COVID-WAREHOUSE allows users to import, clean, merge, and aggregate heterogeneous data, to provide the central repository (data warehouse) from which to assemble multidimensional cubes. Data warehousing is essential tool in support critical decisions rather than understanding the behavior of COVID- 19 virus. By applying OLAP operations, the analysts can get valuable information related to each coronavirus case. The goal of this work is to help clinicians involved in studying critical medical cases to understand deeply the relationship among the factors of medical records (members on dimensions). The COVID-19 data warehouse will aid executive managers and doctors by providing reliable and consistent health care decision-making information. This will help the health care provider enhance the care they give to their patients. It will also enable the health care provider to recognize and diagnose the issue ahead of time and prevent issues later.

1. **References**

[1] O. E. Sheta and A. N. Eldeen, “B UILDING A H EALTH C ARE D ATA W AREHOUSE F OR C ANCER D ISEASES,” vol. 4, no. 5, pp. 39–46, 2012.

[2] T. E. H. Y. Wah and O. N. G. S. Sim, “Development of a Data Warehouse for Lymphoma Cancer Diagnosis and Treatment Decision Support,” no. August, 2014.

[3] A. Arigon, “Handling Multiple Points of View in a Multimedia Data Warehouse,” vol. 2, no. 3, pp. 199–218, 2007.

[4] V. Appidi, S. Umar, and S. Vallamkonda, “Development of a Data Warehouse for Cancer Diagnosis andTreatment Decision Support,” vol. 5, no. 3, pp. 22–26, 2014.

[5] A. Khalaf and T. Obaid, “Design and Implementing Cancer Data Warehouse to Support Clinical Decisions,” no. October, 2016.

[6] I. Y. Choi, S. Park, B. Park, B. H. Chung, C. Kim, and H. M. Lee, “Development of prostate cancer research database with the clinical data warehouse technology for direct linkage with electronic medical record system,” no. June, 2013.

[7] A. Ado, A. Aliyu, S. A. Bello, and A. S. Gezawa, “Building a Diabetes Data Warehouse to Support Decision making in healthcare industry Building a Diabetes Data Warehouse to Support Decision making in healthcare industry,” no. December 2016, 2014.

[8] A. K. Hamoud, “Building Data Warehouse for Diseases Registry : First step for Clinical Data.”

[9] A. Rodolfo, “Spatial Data Mining of a Population-Based Data Warehouse of Cancer in Mexico . Redalyc Scientific Information System Spatial Data Mining of a Population-Based Data Warehouse of Cancer in Mexico International Journal of Combinatorial Optimization Problems and Informatics , vol . 1 ,” no. January, 2010.

[10] T. Ranjan and S. Kumar, “Analysis of Data Mining Techniques For Healthcare Decision Support System Using Liver Disorder Dataset,” *Procedia - Procedia Comput. Sci.*, vol. 85, no. Cms, pp. 862–870, 2016.

[11] A. Wul, B. Haarbrandt, E. Tute, M. Marschollek, P. Beerbaum, and T. Jack, “Arti fi cial Intelligence In Medicine An interoperable clinical decision-support system for early detection of SIRS in pediatric intensive care using openEHR,” vol. 89, no. September 2017, pp. 10–23, 2018.

[12] R. Dutta, “H EALTH C ARE D ATA W AREHOUSE S YSTEM A RCHITECTURE FOR I NFLUENZA ( F LU ),” pp. 77–89, 2013.

[13] E. Hoyo-barbolla and D. Lees, “The use of data ware- houses in the health- care sector,” pp. 43–46, 2002.

[14] G. Agapito and C. Zucco, “COVID-WAREHOUSE : A Data Warehouse of Italian COVID-19 , Pollution , and Climate Data,” 2020.

[15] D. Warehouse, “CHAPTER 3 BUILDING ARCHITECTURAL DATA WAREHOUSE FOR,” pp. 32–85.

[16] M. Karami and H. Services, “An Effective Tool to Create Intelligence in,” no. October, 2017.

[17] *UNITED ARAB EMIRATES MINISTRY OF HEALTH AND PREVENTION*. [Online]. Available: https://www.mohap.gov.ae/en/AwarenessCenter/Pages/COVID-19.aspx%0A.

[18] [Online]. Available: https://www.medicalnewstoday.com/articles/types-of-coronavirus#types.