Automated Data Integration of Biodiversity with OLAP and OLTP

Arie Vatresia[1], Asahar Johar[2], Ferzha Putra Utama[3], Sinta Iryani[4]

[1,4]Department of Informatics, Faculty of Engineering, Universitas Bengkulu
[2,3]Department of Information Systems, Fakultas Teknik, Universitas Bengkulu
[1]arie.vatresia@unib.ac.id, [2]asahar.johar@unib.ac.id, [3]fputama@unib.ac.id,
[4]sintairyani1@gmail.com

Abstract— Biodiversity is one of emerging issue over decades; many have performed research to map and to document the data over the world. This issue is very important due to the event of extinction have been accelerating happening because of human extinction. Bengkulu, as one of the province lied in one of 19 hotspots in the world, Sundanese, has experienced the degradation of flora and fauna over the case of forest degradation and habitat loss. Although many application and software has been developed to solve the case, the existences of data standardization still become an issue over this problem. In this research, study of data integration had been developed to make the process of biodiversity data acquisition can be more effective and efficient. The system proposed the integration based on OLAP and OLTP that will be connected to IUCN, as one of the biggest center for monitoring the loss of biodiversity all around the world. This application had been built with web based using UML design and followed SDLC to provide the best fit of the need. This research had also succeeded to build the automated integration to show the record of dynamic number over biodiversity existences in Bengkulu. The application had been tested using black box and has the perfect performance (100%) over the testing that can help the monitoring process over biodiversity data.

Keywords— Bengkulu, Flora & Fauna, *OLAP & OLTP,* Data Integration

I. INTRODUCTION

Bengkulu is a province of Indonesia. Its capital is in Bengkulu city. The province is located in the southwest part of Sumatra Island. The territory of Bengkulu Province established under Law No. 9 of 1967, it covers the area of former Bengkulu residency with an area of 19,813 km2, consisting of four areas of level II, namely the municipality of Bengkulu consisting of two sub-districts, North Bengkulu Regency (capital Argamakmur) consisting of 13 districts, South Bengkulu Regency (capital of Manna) consisting of 11 sub-districts, and Rejang Lebong (capital of Curup) consisting of 10 sub-districts with the following boundaries: The north is bordered by West Sumatra province, east of Jambi Province and South Sumatera Province, the south borders the Indonesian ocean and Lampung province and the west is bordered by the Indonesian ocean[1].

The tropical forest has a rich source of flora and fauna that can be developed into tourist attraction. The wealth of flora that has been famous and has become a tourist attraction in Bengkulu is *rafflesia arnoldi, amorphopalus gigantenum* and *Shorea spp*. While the wealth of fauna potential to be developed into tourist objects such as Sumatran tiger, deer, elephant, tapir, Buffalo Wild. The research on biodiversity in Bengkulu province is still not done much and the data has not been adequate. This is due to lack of public knowledge of the

flora and fauna in Indonesia because it has not integrated data of good flora and fauna and also some human activity that threatens flora and fauna such as "*illegal logging*", clearing land and hunting and if this is left to the rare flora and fauna is anticipated to be extinct. A very strong pressure on biodiversity on a local scale is the destruction of natural and specific habitats [2]

With the advancement of continuous computing technology, it needs the latest implementation of technology in computing that can help store data and make reporting Quickly using THE *OLAP & OLTP* method for flora and fauna data in Bengkulu Province as one of the regions in Indonesia. Therefore, the problem must be assisted with appropriate technology. With the identification of these flora and fauna can help the problems that exist.

II. LITERATURE REVIEW

The IUCN (International Union for Conservation of Nature) assists the world in finding pragmatic solutions for the environment and the most pressing development challenges. IUCN leads natural conservation and sustainable development efforts through global partnerships involving 15,000 experts from 185 countries, generating new knowledge and establishing global standards in their fields, conducting action with hundreds of conservation projects around the world, and giving influence with collective power to more than 1,300 government and non-governmental organisations that are members of the IUCN. IUCN was founded in October 1948, as an international Union for the Protection of Nature (IUPN) following an international conference in Fontainebleau, France. The organization changed its name to the International Institute for Natural Conservation and Natural resources in 1956 with the acronym IUCN and this name remained in use today. IUCN's vision is only a world of value and preserve nature. While the mission of the IUCN is to influence, encourage and assist communities around the world to preserve the integrity and diversity of nature and to ensure that fair and ecological sustainable use of natural resources. IUCN assesses and preserves nature, ensures effective and fair governance in the use of natural resources, and deploys natural-based solutions for climate challenges, food and global development. IUCN supports scientific research, manages field projects around the world, and carries governments, NGOS, UN and joint companies to develop policies, laws and best practices. The Red Data Book is one of the IUCN outputs. The IUCN Red List is a list of endangered animals and plants in the world issued by the IUCN in order to focus the world's attention to endangered species requiring direct conservation efforts. IUCN issued criteria and divides the threat of species into categoriesz[3].

Data Integration is the process of combining two or more sets of data to make it easier to share and analyze, in order to support information management within a working environment. Data Integration is the process of combining or merging two or more data from different sources of databases into a repository such as a Data Warehouse. As for the reason for the integration of data due to some advantages we can get[4]:

1. Facilitate the process of analyzing for decision making
2. Data Sharing between working environment
3. Avoid duplicate data.

Data Warehouse is defined as a collection of data that has six (6) Characteristics of subject oriented, integrated, process oriented, time variant, accessible and Non Volatile, these six characteristics are useful to support and assist the decision making process in the company or organization. Data Warehouse is a collection of databases that store data now and past data usually contains data history (history data), data is derived from various sources both internal data and external data that are

important part for management in decision making and not to perform transactional processing (operational database) [5].

Multidimensional Data models are designed to facilitate analysis and not transactions. This Model is commonly used in data warehouse. It has an intuitive concept of many dimensions or perspective of business measurement or facts. Multidimensional data forms can be spreadsheets and cubes. Data is more easily manipulated by using cubes [6]. The Dimensional Model often used in the Data warehouse is star or snowflake [7]:

1. Star schema, consisting of one or more fact tables and one or more dimension tables. The fact table is the center of the star schema, because it functions as a binder of the dimension tables.
2. Snowflake Schema is a development of the star Schema. In the snowflake schema, he table dimensions are normalized in part or overall to reduce duplicate values in the table.

A. *Online Analytical Processing (OLAP)*

The first Online Analytical Processing (OLAP) concept was proposed by E. F Codd which is famous as the father of the relational database. The reference of [9], states that "ONLINE Analytical Processing (OLAP) is a technology that allows analysts, managers and executives to simultaneously access data quickly, consistently and interactively with a wide variety of information reviews where each row of data can be transformed to reflect the company's dimensions so it is easy to understand by the user". The special application of OLAP is business reporting for sales, marketing, reporting management, process management, budgeting and forecasting, financial statements and similar fields.

B. *Online Transaction Processing (OLTP)*

Online Transaction process, Or *OLTP*, is an information system class that facilitates and manages transaction-oriented applications, typically for data *entry* and *retrieval*processing. The term is somewhat ambiguous; Some understand "transactions" in the context of a computer or *database*transaction. The methods for analyzing the daily activities of the company include (*Insert, Update, Delete*). The *OLTP* system emphasizes extremely fast *query* processing and maintains data integrity in multi-access environments. For *OLTP*Systems, effectiveness is measured by the number of transactions per second. The *OLTP database* contains the current data details [10].

The data would be process using SQL[11][12] and the help of Code Igneter[13] and display over the web using the language of PHP[14]. To help the visibility of the process in this application, this research use flowchart[15][16] and design method of Unified Modeling Language (UML) [17]. The main goal in the UML design is [18]:

1. Providing users (analysis and system design) a visually expressive language modeling so that they can develop and exchange meaningful data models.
2. Provide mechanisms that specialise in expanding core concepts.
3. Because it is a visual modeling language in the development process, UML is independent of certain programming languages.
4. Provide a formal basis for language understanding modeling.
5. Encouraging market growth against the use of Object oriented system design tools (OO).
6. Support higher-level development concepts such as collaborations, frameworks, patterns and components against a system

III. METHODOLOGY

This research will build a data integration system used to use *OLAP and OLTP* methods. The type of research used is applied research

that builds data integration system of flora and fauna in Bengkulu province which is developed in order to relate to this research, where applied research is a research directed to obtain information to get problem solving of functional research and can be used to overcome practical problems arising or produce a product that has other practical functions.

The data collection techniques used in this study is done by conducting an analysis of the problem of integrating flora and fauna data using the methods used. The following schematic plot of research is described:
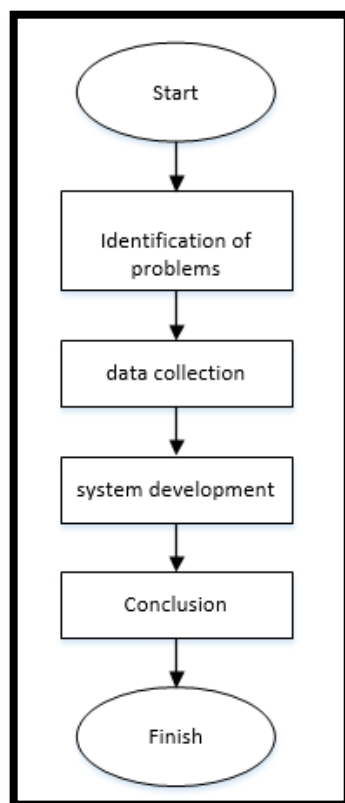

Figure 1. Research Flow Diagram

Here is the complete explanation of each step above:

1. The research process begins with identifying the problems encountered in the process of storing flora and fauna data. The process of identifying the problem surrounds the determination of initial basis of research, namely research background, followed by determining the objectives, benefits and scope of research, while the problem limitation is used to limit the discussion and scope of the research to be not too broad.
2. Data collection. At this stage is the collection of data needed in the research needed to build the system. The collection of data will be obtained from the BKSDA of Bengkulu province and data from *IUCN*.
3. System development. The stage of system development is based on the method of system development used in this research that is *waterfall method*.
4. Conclusion. This stage is done to know the results that have been obtained during the research. At this stage will be inferled from the development of data Integration System using *OLAP* And *OLTP* methods.

IV. RESULT AND DISCUSSION
A. System Analys

System analysis is the parsing of a system intact into parts of its components with a view to identifying and evaluating the problems, opportunities, obstacles, and expected needs, so that it can be proposed Perbaikan-perbaikannya. System analysis on this study includes identifying problems, identifying decision points, and understanding existing systems. Can be seen by the flow diagram method below, which is where the admin will enter the page first, the admin can input the flora fauna data and there will be data processing afterwards the admin can input the query. And the process is complete.
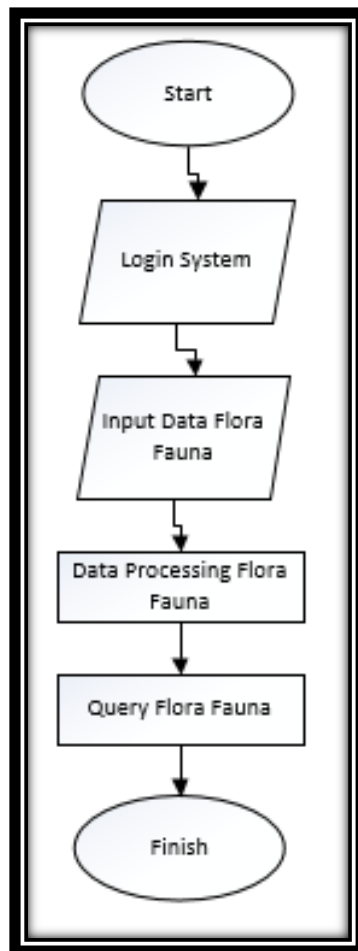
Figure 2. Flow Diagram System

B. **Data Warehouse Designing**

1. **Bussines Process Selection**

In this first process, what is done is analyzing and identifying business processes related to the subject matter. Process selection is done to clarify the limit of data *warehouse* to be created.

The information needed to support the preparation of data integration is information on the numberof species, species location, species conservation status, protection status, ecosystem , habitat type, species threat Type, species Usability, and each species taxon. The Required information will be extracted from the tables in the form of Excel given by the office of BKSDA Bengkulu.

2. **Specify the Grain or Detaild**

In modeled on a dimension, the *Grain* or the details define a line representation in the fact table. As for the *grain* in the draft data warehouse is as follows:

1. Number of Species seen based on recap classification and *taxonomy* on Flora Fauna.
2. The location of Flora fauna is viewed by species, cities/counties and villages found.
3. The number of conservation Status is seen based on the number of species, ecosystem, habitat, threats and usability.
4. Number of protection status viewed by species number, protection status, ecosystem, habitat, threats, and usability.
5. Number of ecosystems viewed by species number, conservation status, protection status, habitat, threats, and usability.
6. Number of Habitat viewed based on species count, conservation status, protection status, ecosystem, threats, and usability.
7. Number of threats seen based on the number of species, conservation status, protection status, habitat, ecosystem, and usability.
8. Number of uses viewed by species number, conservation status, protection status, habitat, ecosystem, and threats.

3. **Identify and adjust the dimensions**

In this stage is performed dimension identification for each existing fact table. For simplicity, at this stage you set up the dimensions tables that correspond to the selected *grain* as well as the *records* that will be displayed in the fact table.

Based on the grain selection, the dimensions are determined for each of the facts as shown in Table 1.

Table 1 Grain and Dimensions

No.	Dimensi \ Grain	Klasifikasi Spesies	Persebaran Lokasi	Status Konservasi	Status Perlindungan	Ekosistem	Habitat	Ancaman	Kegunaan
1.	Nama Lokal	x	x	x	x	x	x	x	x
2.	Genus	x							
3.	Class	x							
4.	Kota /Kabupaten		x						
5.	Desa		x						
6.	Tanggal Dijumpai		x						
7.	Status Perlindungan	x							
8.	Status Konservasi			x					
9.	Kingdom	x							
10.	Phylum	x							
11.	Ordo	x							
12.	Nama Saints	x	x	x	x	x	x	x	x
13.	Jenis Habitat						x		
14.	Jenis Ekosistem					x			
15.	Jenis Kegunaan								x
16.	Jenis Ancaman							x	

4. Fact Selection

At this stage are selected facts that will fill the fact table, where the selected facts must conform to the specified grain and dimensions in the previous stages. Here are the facts to be used:

1. Classification species
2. Distribution of species location
3. Species conservation status
4. Species protection status
5. Ecosystem species habitat
6. Species ssage
7. Species threat

5. Save the Pre calculation in the fact table

According to the pre-defined *grain* which is the candidate for the fact. Each fact has a data that can be computed, at which point it is determined the calculation.

1. The Pre-calculation that will be saved into the fact table classification species includes the number of species and the number of taxon.
2. The Pre-calculation that will be deposited into the State Conservation fact table species includes The number of species and the type of status conservation.
3. The Pre-calculation that will be deposited into the habitat species fact table includes the number of species and habitat types.

4. The Pre-calculation that will be deposited into the species ecosystem fact table includes the number of species and ecosystem types.
5. The Pre-calculation that will be deposited into the habitat species fact table includes the number of species and habitat types.
6. The Pre-calculation that will be saved into the fact table of the species location includes The number of species, the date location, the type of detail location and location layout.
7. The Pre-calculation that will be saved into the fact table of the species protection status includes The number of species and types of protection status.
8. The Pre-calculation that will be stored in the Hazard/ *Threats* species fact table covers The number of species and types of *Threats* .
9. Pre-calculations to be saved into the fact table usability/*usability* tables include species number and type of usability.

6. Complete the Dimension table

Complementing the dimension table is useful to complement the dimension information for easy understanding and use. Here are the table dimensions:

Table 2 Dimension

Nama Field	Tipe	Length	Note
Species Dimension			
Species_Id	Varchar	10	Primary Key
Species_Local _Name	Varchar	50	
Species_Engli sh_Name	Varchar	50	
Species_Popu lation_Trend	Enum	-	
Species_Detai l	Text	-	
Species_Last_ Update_UICN	Date	-	
Conservation Status Dimension			
Conservation_ Status_Id	Varchar	10	Primary Key

Conservation_ Status_Name	Varchar	50	
Conservation_ Status_Detail	Text	-	
Ecosystem Dimension			
Ecosystem_Id	Varchar	10	Primary Key
Ecosystem_N ame	Varchar	50	
Habitat Dimension			
Habitat_Id	Varchar	10	Primary Key
Habitat_Name	Text	-	
Location Dimension			
Location_ID	Varchar	10	Primary Key
Location_Na me	Text	-	
Protection Status Dimension			
Protection_St atus_Id	Varchar	10	Primary Key
Protection_St atus_Name	Varchar	50	
Protection_St atus_Detail	Text	-	
Taxon Dimension			
Taxon_Id	Varchar	4	Primary Key
Taxon_Level	Varchar	30	
Taxon_Id_Par ent	Varchar	4	
Threat Dimension			
Threats_Id	Varchar	10	Primary Key
Threats_Nam e	Text	-	
Usability Dimension			
Usability_Id	Varchar	10	Primary Key
Usability_Na me	Text	50	

7. Choosing the duration of the database

The duration of the brilliant Skin Care data inserted into the Data warehouse has 2 years duration data

8. Track dimensional

In this process that is done is observed the change of data from the dimension table. To anticipate any attribute changes that may occur in the dimension table, such as increasing new species in the following year. In this study the attribute changes to that dimension will result in adding a new record to keep the old data in place.

9. Specifying query Priority and Type query

In this process things are done is to consider the influence on physical design, such as the existence of the summary (summaries) and summation (aggregate) as well AS THE problems of ETL processes (Extract, tranformation, and Loading), backups,and security thatprovide restrictions to the user is a factor that should be considered.

The Model of the query used IS ETL process (Extract, Transform, Loading). The process of transferring data from the IUCN page Origin database to Data warehouse Integration of Flora Fauna data where the data update process can be done every required by admin by pressing the Apply action of data available to theWeb page.

10. Star Schema Result

The scheme used in this data warehouse design research is the star schema. Star schema is a variant of star schema where the relation / foreign key is cantered on the fact table. In other words one dimension table is joined directly to the fact table. The following are the results of the star schema data warehouse to be built:
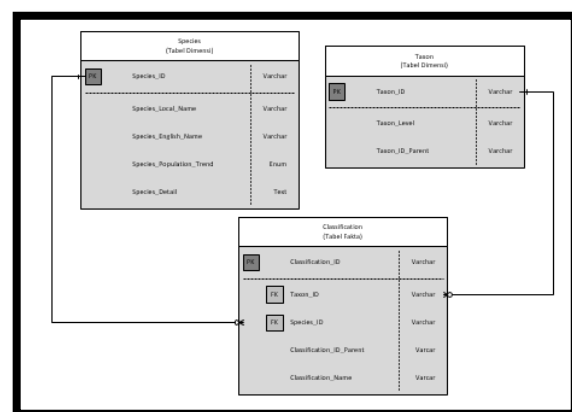

Figure 3. Star Schema Taxon

The figure above shows the results of a star schema consisting of fact classification tables, Species dimension tables, and Taxon dimension tables.
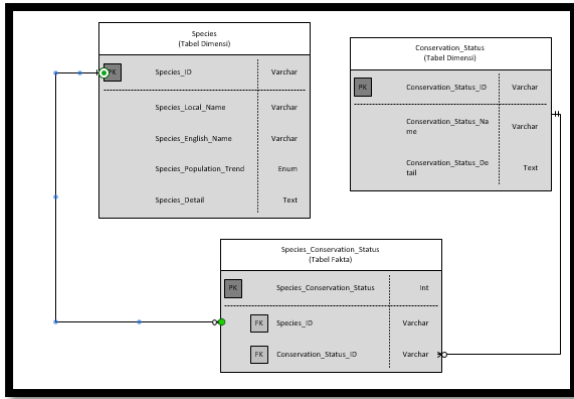
Figure 4. Star Schema Conservation Status

The figure above shows the results of a star schema consisting of a Species Conservation Status fact table, a Species dimension table, and a Conservation Status dimension table..
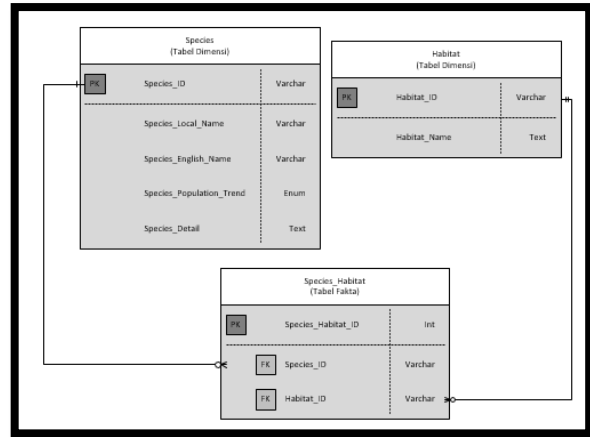

Figure 6. Star Schema Habitat

The figure above shows the results of a star schema consisting of a Habitat Species fact table, a Species dimension table, and a Habitat dimension table.
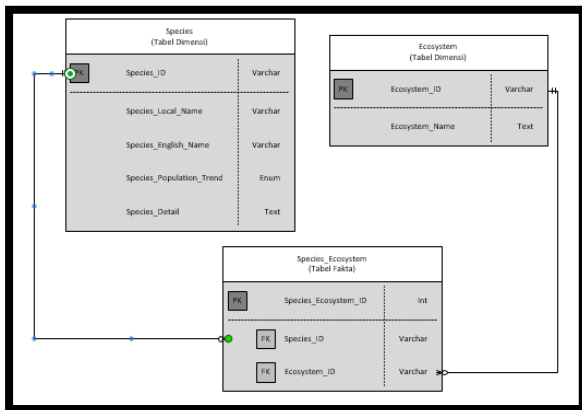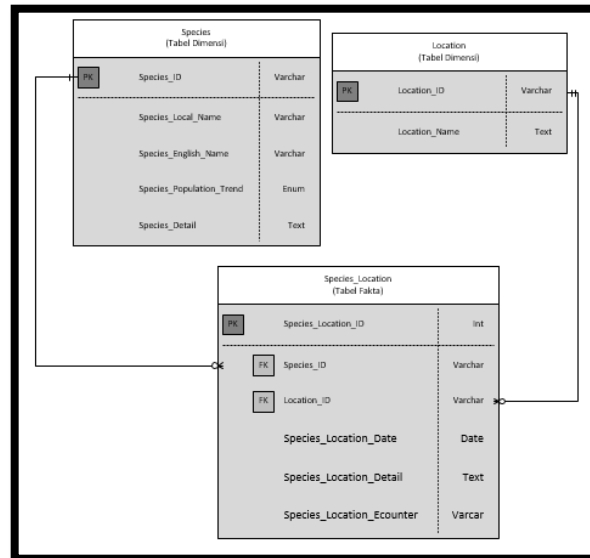

Figure 5. Star Schema Ecosystem

The figure above shows the results of the star schema consisting of Species Ecosystem fact tables, Species dimension tables, and Ecosystem dimension tables.

Figure 7. Star Schema Location

The figure above shows the results of a star schema consisting of a Species Location fact table, a Species dimension table, and a Location dimension table
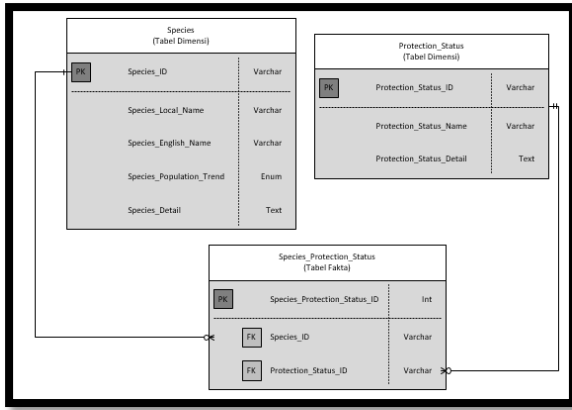
SISFORMA: Journal of Information Systems (e-Journal) Vol.7 | No. 2 |Th. 2020
ISSN 2442-7888 (online) DOI 10.24167/sisforma.v7i2.2817

86

Figure 8. Star Schema Protection Status

.

The picture above shows the results of the star schema which consists of the Species Protection Status fact table, the Species dimension table, and the Protection Status dimension table.
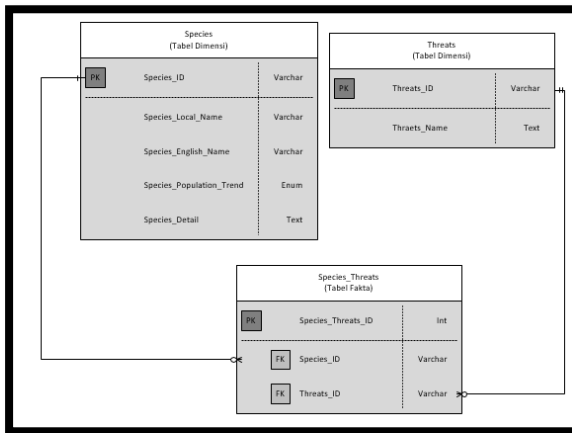

Figure 9. Star Schema Threats

The figure above shows the results of the star schema consisting of Threats Species fact table, Species dimension table, and Threats dimension table.
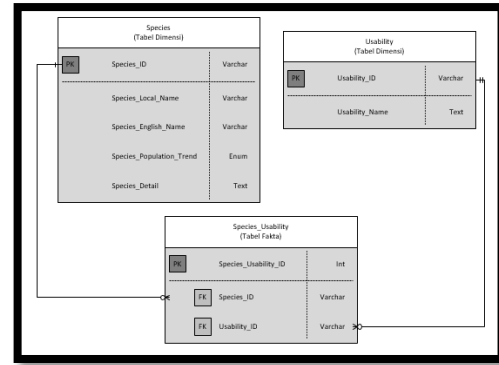

Figure 10. Star Schema Usability

The figure above shows the results of a star schema consisting of a Species Usability fact table, a Species dimension table, and a Usability dimension table.

A. Implementation Interface

(1) Menu *Login*

The login page on this system is the first page to run. This page is based on the design in Figure 11.

Figure 11. Form Login

(2) Menu *Home*

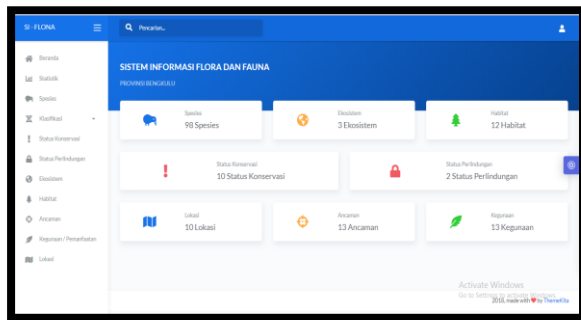The home page on this system is the first page to run. This page is based on the design in Figure 12.

Figure 12. Menu Home

(3) Menu Statistic

The Statistics menu is the main process of this application. Where the application user must fill in data such as the display to be shown in Figure 13.
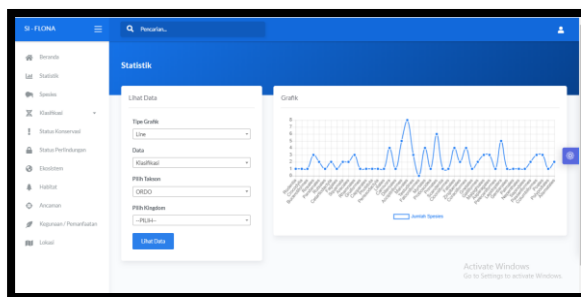

Figure 13. Diagnose Menu

(4) Menu Data Species *IUCN*

The IUCN Species Data menu is a menu that displays data contained on the IUCN page, on this page the user can use the data that is on the IUCN. Following is the appearance of the handling menu shown in Figure 14.

V. CONCLUSIONS

Based on the results and discussion previously described, the conclusions that can be drawn are as follows.
1. The research has successfully implemented the integration of flora and fauna data using Web-based OLAP

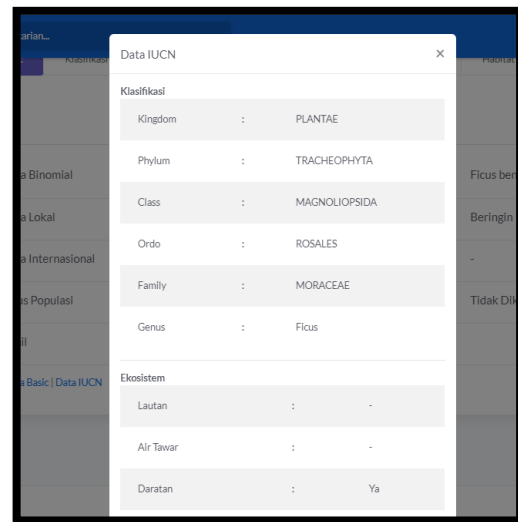methods using the Warehouse data concept.

Figure 14. IUCN integration

2. The research has successfully implemented the integration of flora and fauna data using Web-based OLTP methods using the Warehouse data concept
3. Data is successfully integrated by retrieving data from the IUCN by implementing the existing data on the system if the data contained in the system is incomplete in any species, and the resulting data can be made.

REFRENCE

[1] Diskominfotik, "Sekilas Bengkulu," 2020. [Online]. Available: https://bengkuluprov.go.id/sekilas-bengkulu/.

[2] W. Darwis, "Konservasi Hayati," *J. Ilim.*, vol. 08, no. 01, pp. 1–55, 2012.

[3] I. Status, "Mendukung Penelitian Doktor UCSD," no. September 2015, pp. 1–14, 2015.

[4] Softbless, "Data Integration," *PT Softbless Solutions,* 2016. [Online]. Available: https://www.softbless.com/Data-Integration.

[5] M. Hasanudin *et al.*, "PERANCANGAN MODEL DATA WAREHOUSE PENJUALAN," pp. 34–41, 2019.

[6] A. Johar, A. Vatresia, and L. Martasari, "Aplikasi Business Intelligence (Bi) Data Pasien Rumah Sakit M .Yunus Menggunakan Metode Olap (Online Analytical Processing)," *Rekursif*, vol. 3, no. Bisnis Intelijen, pp. 12–22, 2015.

[7] S. P. Adithama, "RANCANG BANGUN APLIKASI BUSINESS INTELLIGENCE BERBASIS WEB UNTUK SUBJEK KEGIATAN AKADEMIK PADA UNIVERSITAS," vol. 2014, no. Sentika, 2014.

[8] Rianto and Cucu Hadis, "Perancangan Data Warehouse Pada Rumah Sakit (Studi Kasus : BLUD RSU Kota Banjar)," *Sains dan Teknolgi*, vol. 3, no. 2, pp. 214–221, 2017.

[9] P. Ponniah, *DATA WAREHOUSING FUNDAMENTALS Data Warehousing Fundamentals: A Comprehensive Guide for IT Professionals. Paulraj Ponniah DATA WAREHOUSING FUNDAMENTALS A Comprehensive Guide for IT Professionals PAULRAJ PONNIAH*, vol. 6. 2001.

[10] Y. Firdaus, S. B. Premapasha, and S. Yulias, "Analisis dan Perancangan Sistem Online Transaction Processing (OLTP) Menggunakan SCRUM (Studi Kasus Rumah Sakit Puti Bungsu) Analytics and Development Online Transaction Processing (OLTP) System Using SCRUM (Case Study in Puti Busngsu Hospital)," *e-Proceeding Eng.*, vol. 4, no. 2, pp. 3130–3137, 2017.

[11] E. Iswandy, D. S. T. M. I. Komputer, and S. J. Padang, "Sistem Penunjang Keputusan Untuk Menentukan Dan Santunan Sosial Anak Nagari Dan Penyaluran Bagi Mahasiswa Dan Pelajar Kurang Mampu," *J. TEKNOIF*, vol. 3, no. 2, 2015.

[12] E. Z. Henry Februariyanti, "Rancang Bangun Sistem Perpustakaan untuk Jurnal Elektronik," *J. Teknol. Inf. Din.*, vol. 17, no. 2, pp. 124–132, 2012.

[13] E. W. Hidayat, "Pengembangan Sistem Manajemen Jurusan Dan Laboratorium TI Universitas Siliwangi berbasis Framework," *Semin. Nas. Apl. Teknol. Inf.*, vol. 2011, no. Snati, pp. 19–23, 2011.

[14] S. Tyowati and R. Irawan, "Implementasi Framework Codeignter Untuk Pengembangan Website Pada Dinas Perkebunan Provinsi Kalimantan Tengah," *J. SAINTEKOM*, vol. 7, no. 1, p. 67, 2017.

[15] O. T. K, M. Irfan, and A. Nurpianti, "Pdf (218 Kb)," *Ed. Juli 2013 Vol.*, vol. VII, no. 1, pp. 33–52, 2013.

[16] S. Santoso and R. Nurmalina, "Perencanaan dan Pengembangan Aplikasi Absensi Mahasiswa Menggunakan Smart Card Guna Pengembangan Kampus Cerdas (Studi Kasus Politeknik Negeri Tanah Laut)," *J. Integr.*, vol. 9, no. 1, pp. 84–91, 2017.

[17] Suendri, "Implementasi Diagram UML (Unified Modelling Language) Pada Perancangan Sistem Informasi Remunerasi Dosen Dengan Database Oracle (Studi Kasus: UIN Sumatera Utara Medan)," *J. Ilmu Komput. dan Inform.*, vol. 3, no. 1, pp. 1–9, 2018.

[18] Haviluddin, "Memahami Penggunaan UML (Unified Modelling Language)," *Memahami Pengguna. UML (Unified Model. Lang.*, vol. 6, no. 1, pp. 1–15, 2011.

SISFORMA: Journal of Information Systems (e-Journal) Vol.7 | No. 2 |Th. 2020
ISSN 2442-7888 (online) DOI 10.24167/sisforma.v7i2.2817

89