A Survey on Python Libraries and Packages for Databases

Rini Shiny M[1], Dr. S. Sarathambekai[2]

[1]PG Scholor, Department of Information Technology, PSG College of Technology, Coimbatore-4, India

[2]Assistant Professor (SG), Department of Information Technology, PSG College of Technology, Coimbatore-4, India

Abstract- **Python is a widely used general-purpose, high level programming language mainly developed for emphasis on code readability, and its syntax allows programmers to express concepts in fewer lines of code. Python scripting language that is scalable, reliable, platform independent is gaining attention because of its ease of use and numerous number of inbuilt modules.**

In this paper, a survey of different papers where Python packages are used to interact with databases are studied. The merits and demerits of each proposed methodology is analysed and presented in a consolidated manner.

Index Terms- python, database, API, scripting, RDBMS, NOSQL

I. INTRODUCTION

Python has been adopted as a programming language across many applications recently. Programmers take advantage of all benefits of Python and implement it in various projects. This paper summarizes the Python packages related to the commonly used databases. We also quote a few papers as example where these packages are being used. Making use of Python libraries drastically increased performance and code readability.

II. SIGNIFICANCE OF PYTHON

The significance of python is described as follows,
a. Python is a high-level, interpreted, interactive and object-oriented scripting language.
b. The readable and clean code base will helps to maintain and update the software without putting extra time and effort.
c. Python also supports several programming paradigm like object oriented and structured programming.
d. Python is compatible with Major Platforms and Systems without recompilation.
e. Python's Robust Standard Library allows you to choose from a wide range of modules according to your precise needs.
f. Several open source Python frameworks, libraries and development tools can be used to curtail development time without increasing development cost.
g. Python even makes it easier to perform coding and testing simultaneously by adopting test driven development (TDD) approach.

III. SIGNIFICANCE OF DATABASES

File systems suffer from several drawbacks like data redundancy, data inconsistency, lesser storage and security. In order to store data more efficiently, databases are used. Databases can hold large amount of data and access mechanisms can be enforced to ensure security, so that only authorized users are able to access the data. Also, the relationships between data is established using many integrity constraints and duplication of data is reduced to a greater extent. Since database systems keeps the backup of data, it is easier to do a full recovery of data in case of a failure. A database management system (DBMS) is a computer software application that interacts with the user and the database to retrieve and analyse data. Popular DBMSs include MySQL, PostgreSQL, Microsoft SQL Server, Oracle, Sybase and IBM DB2.

IV. PYTHON LIBRARIES AND PACKAGES FOR DATABASES

The Python standard for database interfaces is the Python DB-API. Most Python database interfaces adhere to this standard. The Python Database API supports a wide range of database servers such as – GadFly, mSQL, MySQL, PostgreSQL, Microsoft

SQL Server 2000, Informix, Interbase, Oracle and Sybase.

The DB API provides a minimal standard for working with databases using Python structures and syntax wherever possible. This API includes the following steps.

- Importing the API module.
- Acquiring a connection with the database.
- Issuing SQL statements and stored procedures.
- Closing the connection

Python has support for working with databases via a simple API. Modules included with Python include modules for SQLite and Berkeley DB. Modules for MySQL, PostgreSQL, FirebirdSQL and others are available as third-party modules. The latter have to be downloaded and installed before use. Following is a list of a few Python packages related to several databases.

a) python-mysqldb is an interface to the popular MySQL database server for Python. Its features include Compliance with Python database API version 2.0 and also provides thread-friendliness (threads will not block each other).

b) python3-pymssql package contains a Python module allowing direct access to Microsoft SQL server and Sybase databases. It is designed for simplicity and performance, and conforms to Python DB-API 2.0.

c) Psycopg is the most popular PostgreSQL database adapter for the Python programming language. Its key features include complete implementation of the Python DB API 2.0 specification and the thread safety.

d) python-fdb is a Python library package that implements Python Database API 2.0-compliant support for the open source relational database Firebird. In addition to the minimal feature set of the standard Python DB API, FDB also exposes nearly the entire native client API of the database engine including Automatic data conversion, Distributed transactions and Firebird BLOB support.

e) cx_Oracle is a Python extension module that enables access to Oracle Database. It conforms to the Python database API 2.0 specification. To interact with Oracle Database, cx_Oracle uses Oracle Client libraries which makes it fast and reliable.

f) pysqlite is the python module which acts as an interface to powerful SQLite engine compatible with the Python database API specification 2.0. The library implements most of SQL-92 standard, including transactions, triggers and most of complex queries.

g) python-bsddb3 is the Python interface for Berkeley DB which provides complete support for Berkeley DB distributed transactions, Replication Manager, Berkeley DB Base Replication and RPC.

h) pymongo provides an interface to easily access Mongo DB from Python which is a high-performance, open source, schema-free document-oriented data store.

i) python-cassandra is a modern, feature-rich and highly-tunable Python client library for Apache Cassandra using exclusively Cassandra's binary protocol and Cassandra Query Language v3.

V. RELATED WORK

A. An Implementation Approach to Store GIS Spatial Data on NoSQL Database

This paper [1] aims at solving the problem of storage and retrieval of huge amount of GIS (Geographic Information System). Due to several drawbacks from traditional relational databases, the authors have proposed a method of storing GIS data into a NO SQL database, Mongodb and hence parse and query from database using Python programming language. The operations in Mongodb are accessed from Python using the pymongo package. Once the conection is setup, the storage and retrieval functions are written in Python. The results show that the performance has improved to a greater extent with Mongodb and the authors have taken advantage of the robust library in Python.

B. Using python inbuild modules for Effective Pharmacy Billing Application

This paper [2] uses Python's inbuilt modules in order to be build a pharmacy billing application. Since medicines get expired after the expiry date, pharmacies face loss every year. To address this problem, the authors developed this application which retrieves the medicines which are closer to the expiry date, so that the medicines can be sold based on customer requirements. MySQL database is used

for storing the data and Python imports the MySQL connector module for connectivity. This application is proven to be effective because it predicts the expiry date and reduce the loss occurred at a pharmacy level. The authors have chosen Python programming language because of its readability, portability and its huge collection of inbuilt module support for various databases.

C. Design of VMware vSphere Automatic Operation and Maintenance System Based on Python

Cloud computing and Big data are emerging areas which demands arduous operation and maintenance work. This paper [3] aims at automating this process by analysing VMware vSphere. The system is developed in Java as its front end and Python is used as backend because of its rich and a smart functionality, thereby increasing the development efficiency. The modules Pysphere and pyVmomi are used to pull data from the host and track its configuration. Based on the information collected it performs a start or shut down operations and transfers the documents to MySQL database. MySQL extension module is used to store the data in MySQL database. This model significantly reduced the burden of maintenance and brings a revolution in the cloud computing paradigm and improves the efficiency instantly.

D. Application Oriented Sensor Database System

To develop a smart city application, sensor network consisting of sensor nodes communicate with each

other. Sensor database systems were introduced in order to store the information from various sensors. This paper [4] analyses this information for fire detection. A client server model is developed with socket programming done is Python and Python is also used to feed this data into MySQL database which runs on a local computer. python-mysqldb is used to establish connectivity from python to MySQL database. This application thus helps in detecting fire in the nearby surrounding by analysing the data stored in the database.

E. Mysql to Cassandra conversion engine

With advancement of technology, huge amount of data is generated each and every day. Considering the benefits of NoSQL databases over relational databases, data migration is crucial to mount this huge amount of data from RDMS to NoSQL database.

This paper [5] proposes a data migration system designed on Python environment to migrate data from MySQL to Cassandra database. A GUI is developed where the source tables are displayed and enables the user to select specific data to be migrated. The connectivity is established using respective python libraries - python-mysqldb and python-cassandra. Hence, this system can be used for migrating data into NoSQL so that large volumes of data can be handled easily.

Summary of the Research Related Works – Table 1

S. no	Title	Author Name	Description	Python packages used	Merits	Demerits
A	An Implementation Approach to Store GIS Spatial Data on NoSQL Database	Xiaomin Zhang et al.,	Storing GIS information in Mongodb using Python.	pymongo package	Provides an efficient method to store and retrieve data from NoSQL database.	The proposed method focusses only on GIS data, rather than all unstructured data.
B	Using python inbuild modules for Effective Pharmacy Billing Application	Jayaprakash. D And Fancy.C	An application which retrieves the medicines closer to the expiry date, so that the medicines can be sold based on customer requirements.	python-mysqldb	Prediction based on expiry date helps to reduce loss in pharmacies.	The loss incurred also depends on varying customer requirements.
C	Design of VMware vSphere Automatic Operation and Maintenance System Based on Python	Fangtao Liu and Zhiyong Yang	Aims to automate the operation and maintenance in cloud environments and documents in MySQL database.	python-mysqldb	Based on the configuration information, it automatically starts or shuts down servers and stores in MySQL.	The enterprise must give continuous attention and investment in automatic operations.
D	Application Oriented Sensor Database System	G S R Satyanarayana., et.al,	Storing data collected from sensors in MySQL database using Python.	python-mysqldb	The stored data is used for analysis in fire detection application.	Using RDMS to store huge amount of sensor data might lead to performance issues.
E	Mysql to Cassandra conversion engine	Giridhara Gopalan M, et al.,	Migrating data from MySQL database to Cassandra database in Python environment.	python-mysqldb and python-cassandra	Takes advantage of NoSQL databases by migrating data from RDMS.	Implementation restricted to specific databases.

VI. CONCLUSION

This paper consists of various implementations which involves the usage of Python's inbuilt packages. Since Python has a huge collection of libraries, the code size and the development effort has reduced drastically. This in turn reduces the memory used by the code and enhances throughput. The code provides readability and so it gives a flexible and comfortable environment for the developers. With respect to database, Python supports most of the popular and emerging databases.

Using the inbuilt libraries, the database engineers can easily import and use them to store and retrieve the data from the databases.

Importing the various python debian libraries for databases, makes the coding simpler and efficient. As python is platform independent, it is supported by many operating systems, and can be deployed in all types of platforms. Additionally, it also can be deployed in the virtual environment.

REFERENCES

[1] Zhang, X., Song, W., & Liu, L. (2014, June). An implementation approach to store GIS spatial data on NoSQL database. In 2014 22nd international conference on geoinformatics (pp. 1-5). IEEE.

[2] Jayaprakash, D., & Fancy, C. (2017, May). Using python inbuild modules for effective pharmacy billing application. In 2017 International Conference on IoT and Application (ICIOT)(pp. 1-6). IEEE.

[3] Liu, F., & Yang, Z. (2018, August). Design of VMware vSphere Automatic Operation and Maintenance System Based on Python. In 2018 International Conference on Advanced Mechatronic Systems (ICAMechS) (pp. 283-286). IEEE.

[4] Satyanarayana, G. S. R., Chakraborty, D., & Das, S. K. (2017, October). Application Oriented Sensor Database System. In 2017 International Conference on Networking and Network Applications (NaNA) (pp. 141-146). IEEE.

[5] Gopalan, M. G., Prasanna, C., Krishna, Y. S., Shanthini, B., & Arulkumar, A. (2017, May). MYSQL to cassandra conversion engine. In 2017 Third International Conference on Sensing, Signal Processing and Security (ICSSS) (pp. 503-508). IEEE.