

CLOUD WIZARD

¹MURALIKRISHNA B, ²GOTTIPATI SUMANTH, ³POTHUGANTI JATISH,
⁴LINGAREDDY RAVI CHANDRAREDDY, ⁵THANNEERU AVINASH

¹Asst. Professor, Dept. of CSE, PBR VITS, Kavali, A.P, India.

^{2,3,4,5}B.Tech, Dept. of CSE, PBR VITS, Kavali, A.P, India.

Abstract - Cloud Wizard is an intelligent data pipeline engine that allows data to be transformed and integrated so it's AI Ready. Cloud Wizard can process structured, unstructured and IOT data from internal systems or external sources such as websites. Cloud Wizard identifies and enriches any data source in minutes and makes these details available for use by the other product integrates anything, anywhere, on-demand and as-is. The platform combines Tribal Knowledge with IoT, unstructured data and external dark data to create enterprise applications that work together to assess and reduce risk, eliminate unwarranted liabilities, and ensure a high revenue ecosystem. The platform captures Tribal Knowledge and other hidden data across your enterprise and delivers intelligent answers in minutes. And Pipeline in general is to do data processing elements connected in series, where the output of one element is the input of the next one. Every Pipeline starts with Data Source task, ends with Sink task and in between may contain transformer tasks.

Keywords – Cloud Computing, Internet of Things, AI.

I. INTRODUCTION

Cloud Wizard is an intelligent data pipeline engine that allows data to be transformed and integrated so it's AI Ready. Cloud Wizard can process structured, unstructured and IOT data from internal systems or external sources such as websites. Cloud Wizard identifies and enriches any data source in minutes and makes these details available for use by the other product.

If we want to give a solution we need to develop it from the scratch like stepping the project adding authentication and authorization providing the solution and its user interface like graphs and other chart forms as its time taking process and a repetitive task for each solution. The main drawbacks of the project are it requires more manpower, increase time complexity and there is no efficient system.

Using Cloud Wizard platform we can provide solutions with a quick turn around time it has all common functionality available in cloud

wizard core project, Cloud wizard has the AI and Machine learning capabilities which is useful. For example it has dynamic schema generation developed using AI, Cloud wizard is providing the Tribal knowledge capabilities and can interact with data lakes with a minimal configuration

Advantages

- It has optimal storage and efficient data retrieval layer
- It is an efficient system
- It reduces time complexity
- It will reduce manual efforts

This project is used to implement solutions for wide variety of domains and verticals to give a quick solution for complex problems. It supports all structured and unstructured data and supports schema and schema-less databases. Product supports AI integration and Machine learning capabilities. Cloud Wizard is an intelligent data pipeline engine that allows data to be transformed and integrated so its "AI Ready." Cloud Wizard can process structured, unstructured and IoT data from internal systems or external sources such as websites. Cloud Wizard identifies and enriches any data source in minutes and makes these details available for use by the other products.

II. BACKGROUND WORK

A) *Automatic Persistence of The Knowledge Graph*

Knowledge graphs are critical to many enterprises today: They provide the structured data and factual knowledge that drive many products and make them more intelligent and "magical." In general, a knowledge graph describes objects of interest and connections between them. For example, a knowledge graph may have nodes for a movie, the actors in this movie, the director, and so on. Each node may have properties such as an actor's name and age. There may be nodes for multiple movies involving a particular actor. The user can then traverse the knowledge graph to collect information on all the movies in which the actor appeared or, if applicable, directed. Many practical implementations impose constraints on the links in knowledge graphs by defining a *schema* or *ontology*. For example, a link from a movie to its director must connect an object of type *Movie* to an object of type *Person*. In some cases the links themselves might have their own properties: a link connecting an actor and a movie might have the name of the specific role the actor played. Similarly, a link connecting a politician with a specific role in government might have the time period during which the politician held that role.

Knowledge graphs and similar structures usually provide a shared substrate of knowledge within an organization, allowing different products and applications to use similar vocabulary and to reuse definitions and descriptions that others create. Furthermore, they usually provide a compact formal representation that developers can use to infer new facts and build up the knowledge—for example, using the graph connecting movies and actors to find out which actors frequently appear in movies together. This article looks at the knowledge graphs of five diverse tech companies, comparing the similarities and differences in their respective experiences of building and using the graphs, and discussing the challenges that all knowledge-driven enterprises face today. The collection of knowledge graphs discussed here covers the breadth of applications, from search, to product descriptions, to social networks:

B) Optimal Storage Platform

A central theme of new technology requirements surrounding big data is the ability to host data effectively. The foundation of a big data technology stack is the storage layer. Despite the common rhetoric that smart software can efficiently manage storage distributed across many “cheap as possible” commodity servers, the hardware

infrastructure dictates the actual characteristics of the storage and the applications built upon it. Performance, scalability, reliability, availability, accessibility, and manageability are all fundamental operational requirements, but there are often mistaken assumptions about what is needed to get the best results for big data solutions. Additionally, the landscape of data required for analytics is evolving rapidly so how should IT architects approach choosing an optimal storage foundation for all of their data? The good news here is that well-understood traditional enterprise operational requirements aren't changing, only the technologies that can deliver them.

C) Efficient Data Retrieval Layer

In a data pipeline, data normally go through 2 stages: Data Processing and Data Access. For any type of data, when it enters an organization (in most cases there are multiple data sources), it is most likely either not clean or not in the format that can be reported or analyzed directly by the eventual business users inside or outside of the organization. Data Processing is therefore needed first, which usually includes data cleansing, standardization, transformation and aggregation.

The finalized data is then presented in the Data Access layer — ready to be reported and used

for analytics in all aspects. Data Processing is sometimes also called Data Preparation, Data Integration or ETL; among these, ETL is probably the most popular name. Data processing and data access have different goals, and therefore have been achieved by different technologies. Data Processing for big data emphasizes “scaling” from the beginning, meaning that whenever data volume increases, the processing time should still be within the expectation given the available hardware. The overall data processing time can range from minutes to hours to days, depending on the amount of data and the complexity of the logic in the processing. On the other hand, data access emphasizes “fast” response time on the order of seconds. On a high level, the scalability of data processing has been achieved mostly by parallel processing, while fast data access is achieved by optimization of data structure based on access patterns as well as increased amounts of memory available on the servers.

D) Business Intelligence

Reporting is a central facet of business intelligence and the dashboard is perhaps the archetypical BI tool. Dashboards are hosted software applications that automatically pull together available data into charts and graphs that give a sense of the immediate state of the company. Although business intelligence does

not tell business users what to do or what will happen if they take a certain course, neither is BI solely about generating reports. Rather, BI offers a way for people to examine data to understand trends and derive insights by streamlining the effort needed to search for, merge and query the data necessary to make sound business decisions.

III. PROPOSED WORK

The proposed work shown in the figure 1.

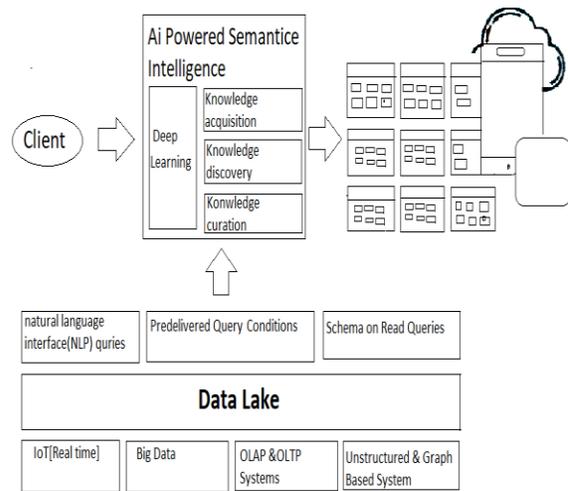


Fig. 1: Proposed Framework.

Implementation Modules

Wide Range of Data Connectors

Out of the box data connectors for structured, unstructured, time series and experimental data sources empowers users to prototype and deploy production grade applications with little effort and governance.

Data Processing Layer

A visually rich data pipeline processor allows

connectors to a wide range of sources, enables a wide range of analytical and AI based transformations and provides for write options to a range of cloud and on-premise data sinks.

Native and Hybrid-Cloud Ready

Native capabilities purpose-built for cloud providers like AWS & Azure, help to exploit the home-grown advantages of popular cloud vendors without sacrificing the cross-purpose abstraction and portability across a hybrid ecosystem.

Data-As-A-Service

Built on top of popular open source options like Apache Dremio, the Data- as-a-service stack provides a multi-modal, SQL-like query framework that cuts across multiple, disparate data sources, providing a homogenous, democratized data analytics layer across large scale heterogenous data silos,

Data Tiering

A Big Data scale, homogenous data tiering architecture that provides a “hot, warm, cold” delineation of storage options that helps optimize technological and economic advantages of various cloud and on-premise storage options.

Auto-Modeling

A cloud ready workbench that uses Machine learning and AI to convert business questions and analytical requirements into semantic,

logical and physical models based on a Natural Language framework that interprets intent and translates it into an optimum execution plan.

Auto-Preparation

A business friendly cloud ready workbench for sampling of large-scale data to help with data preparation for various machine learning and statistical modeling functions. Ready access to hundreds of quality, audit, data-shaping and data profiling transformations that help to convert raw data into business and semantic artifacts.

Schema on Read

A dynamic SQL plan fabrication framework that can be containerized in a scalable way in order to translate the query intent into an optimum schema.

IV. RESULT ANALYSIS

Name	Details	Fields	Alerts	Has Discovery	Has Security	Has Lineages	Has Statistics	Has Qualities	Modified By
Address	com.appengine.document	5		No	No	Yes	No	No	system
Article	com.appengine.web	15	Link	No	No	No	No	No	system
Blob	com.appengine.document	4		No	No	Yes	No	No	system
Chat	com.appengine.document	1	Document Entity	No	No	No	No	No	system
Column	com.appengine.document	2	Document Entity	No	No	No	No	No	system
Customer	com.appengine.web	15		No	No	No	No	No	system
Document	com.appengine.document	4	Document Entity	No	No	No	No	No	system
Document Entity	com.appengine.document	12		No	No	Yes	No	No	system
Document Entity attributes	com.appengine.document	7		No	No	No	No	No	system
Document Semantics	com.appengine.document	12	File Metadata	No	Yes	No	No	No	system
Document Search Result	com.appengine.document	20		No	No	No	No	No	system

Fig. 2: Manage Schema Objects

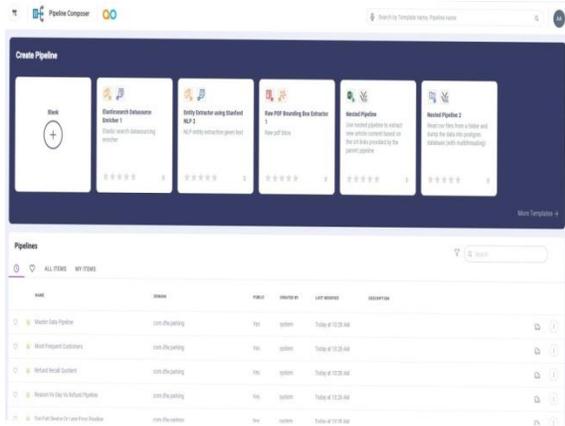


Fig. 3: pipelines

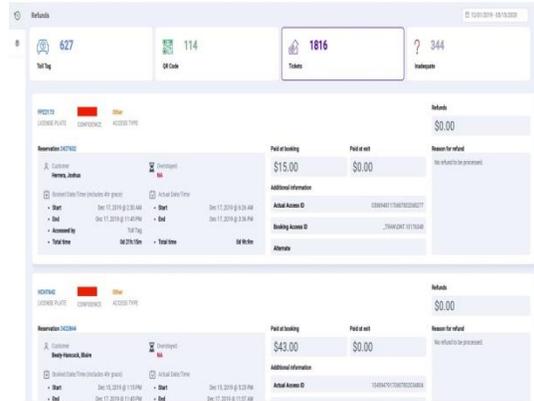


Fig. 6: Refund Ticket

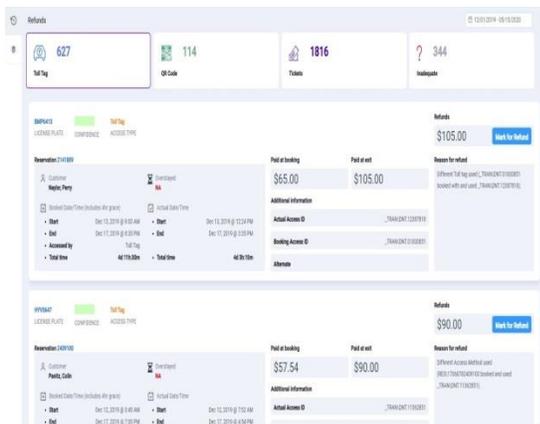


Fig. 4: Refund Toll Tag

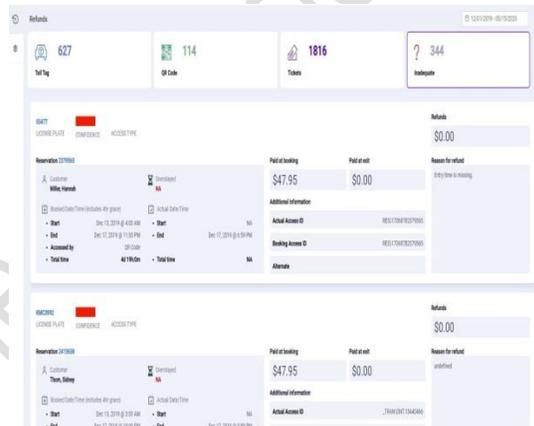


Fig. 7: Refund Inadquate

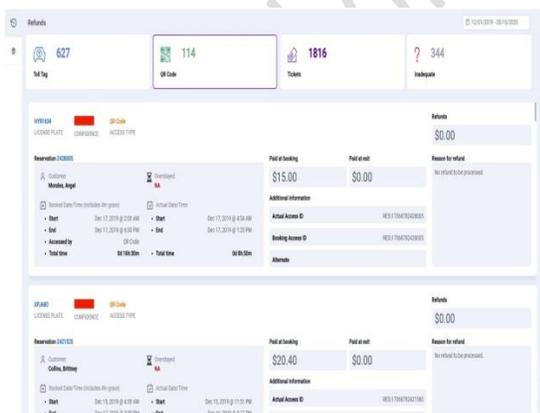


Fig. 5: Refund QR code

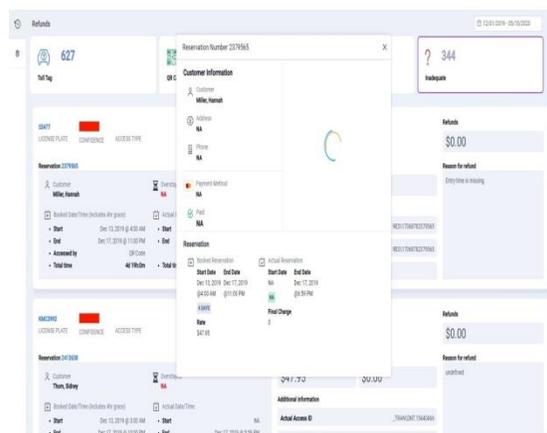


Fig. 8: Inadquate Data

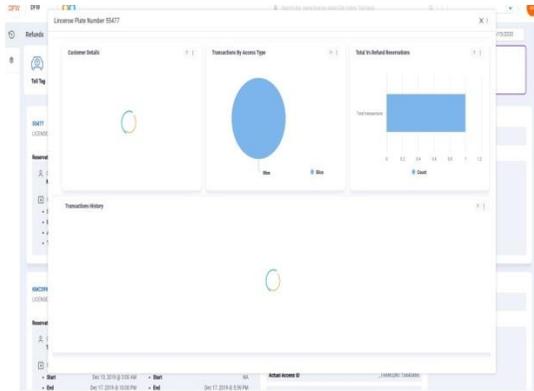


Fig. 9:Customer Data

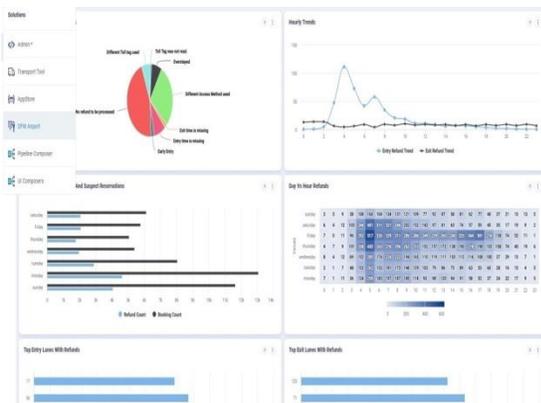


Fig. 10:Transaction Data in Chat

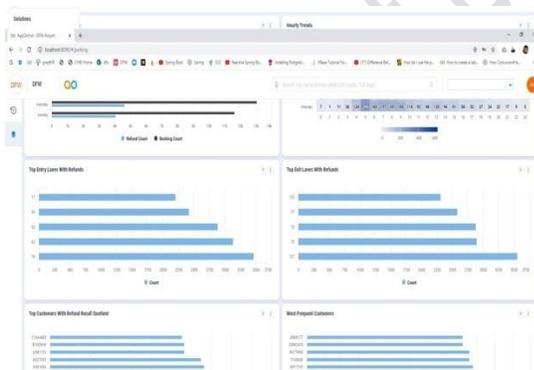


Fig. 11: Transaction Data in Graphs

V. CONCLUSION

In conclusion, Cloud Wizard is recently new technological development that has the potential to have a great impact on the world.

It has many benefits that it provides to its users and businesses. For example, some of the benefits that it provides to businesses, is that it reduces operating cost by spending less on maintenance and software upgrades and focus more on the businesses itself. And we can provide solutions with a quick turn around time it has all common functionality available in cloud wizard core project, Cloud wizard has the AI and Machine learning capabilities, which is useful optimal storage and efficient data retrieval layer, efficient system, reduces time complexity, reduce manual efforts.

REFERENCES

- [1] D.Wu, G. Zhang, and J. Lu, "A fuzzy preference tree-based recommender system for personalized business-to-business E-services," *IEEE Trans.Fuzzy Syst.*, vol. 23, no. 1, pp. 2943, Feb. 2015.
- [2] X. Wu, X. Zhu, G.-Q.Wu, and W. Ding, "Data mining with big data," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 1, pp. 97107, Jan. 2014.
- [3] N. Bizanis and F. A. Kuipers, "SDN and virtualization solutions for the Internet of Things: A survey," *IEEE Access*, vol. 4, pp. 55915606, May 2016.
- [4] J. Zhou et al., "Securing outsourced data in the multi-authority cloud with fine-grained access control and efficient attribute

revocation," Comput. J., vol. 60, no. 8, pp. 12101222, Aug. 2017.

[5] L.Wu, Y. Zhang, K.-K. R. Choo, and D. He, "Efficient and secure identitybasedencryption scheme with equality test in cloud computing," FutureGener. Comput. Syst., vol. 73, pp. 2231, Aug. 2017.

[6] H. He, R. Li, X. Dong, and Z. Zhang, "Secure, efficient and fine-graineddata access control mechanism for P2P storage cloud," IEEE Trans. CloudComput., vol. 2, no. 4, pp. 471484, Oct./Dec. 2014.

[7] K. Zakir Hussain, M. Durairaj and G. Rabia Jahani Farzana ,2012 Application of Data Mining Techniques for Analyzing Violent Criminal Behavior by Simulation Model

AUTHORS



MURALIKRISHNA B has received his B.Tech in Computer Science and Engineering and M.Tech degree in Computer Science and Engineering from JNTU, Hyderabad in 2006 and JNTU, Anantapur in 2014 respectively. He is dedicated to teaching field from the last 12 years. He has guided 20 P.G and 20(batch) U.G students. His research areas included AI, Data Mining and Computer

Architecture, Cloud Computing. At present he is working as Assistant Professor in PBR Visvodaya Institute Of Technology And Science, Kavali, Andhra Pradesh, India.



GOTTIPATI SUMANTH, pursuing B.Tech in Computer Science and Engineering (CSE) from PBR Visvodaya Institute of Technology and Science, Kavali, Nellore (dt),Andhra Pradesh.



POTHUGANTI JATISH, pursuing B.Tech in Computer Science and Engineering (CSE) from PBR Visvodaya Institute of Technology and Science, Kavali, Nellore (dt),Andhra Pradesh.



LINGAREDDY RAVI CHANDRAREDDY, pursuing B.Tech in Computer Science and Engineer-ing (CSE) from PBR Visvod-aya Institute of Technology and Science, Kavali, Nellore (dt),Andhra Pradesh.



THANNEERU AVINASH, pursuing B.Tech in Computer Science and Engineering (CSE) from PBR Visvodaya Institute of Technology and Science, Kavali, Nellore (dt),Andhra Pradesh.