External Service Agreement Based on Scheduling of Spark Jobs

[1]G.Dhanalakshmi, [2]B M.Lakshmipriya, [3]S.Maheswari, [4]N.Nandhitha

[1]Associate Professor, [2,3,4]UG Scholar, Department of Information Technology,

Panimalar Institute of Techology

Abstract **-** The following Era of Sequencing (NGS-Next Generation Sequencing) innovation has brought about enormous sums of proteomics and genomics information. This data is of no use if it is not properly analyzed. ETL (Extraction, Transformation, Loading) is imperative in planning information analytics applications. ETL requires an appropriate understanding of the highlights of information. Information organization plays a key part in the understanding of information, representation of information, space required to store information, information I/O amid the preparation of information, middle of the road comes about of preparing, in-memory analysis of data, and overall time required to process data. Distinctive information mining and machine learning calculations require input information in particular sorts and designs.

Keywords – Next Generation Sequencing, External Service agreements, Fake job scam detection, Web Application.

I. INTRODUCTION

External Service Agreement Based Scheduling of Spark Jobs refers to a technique for scheduling jobs in a Spark cluster that takes into account the availability and capacity of external services. In this approach, the Spark cluster is integrated with external services such as databases, message queues, or web services, which are required by the Spark jobs for their execution.The scheduling of the Spark jobs is based on the availability and capacity of these external services. The scheduler takes into account the current load on the external services and schedules the Spark jobs accordingly, to avoid overloading the services and to ensure optimal performance.This approach is particularly useful for large-scale, complex Spark applications that require interaction with external services. By incorporating external service agreement-based scheduling, these applications can achieve better performance, scalability, and reliability.

Analysing data at a massive scale is becoming crucial due to the availability of huge data in various domains. A major information handling stage can be conveyed in nearby premises utilizing registering assets claimed by an organization. Furthermore, as cloud specialist organizations offer adaptable, versatile, and reasonable registering assets on-request, it is likewise becoming well known to convey a major information handling bunch in the cloud. Albeit the majority of the organizations of a major information registering bunch are either nearby, or on the cloud, numerous associations are likewise utilizing a crossover arrangement where both neighborhood and cloud assets are utilized together to frame the cluster1.These frameworks have turned into a very much embraced worldview for facilitating a huge number of computational specialist co-ops. Notwithstanding, it is trying to plan occupations in a group sent on half breed mists while guaranteeing the SLA boundaries like money related cost minimization, and cutoff time. Subsequently, a principal element of distributed computing, and foundation based administrations specifically, is the fuse of virtualization innovation.

II.RELATED WORK

[1] Authors are proposed that an innovative approach to image compression using machine learning and a pyramidal structure. The proposed technique achieves impressive results and has the potential to be applied in various applications, including medical imaging and remote sensing.

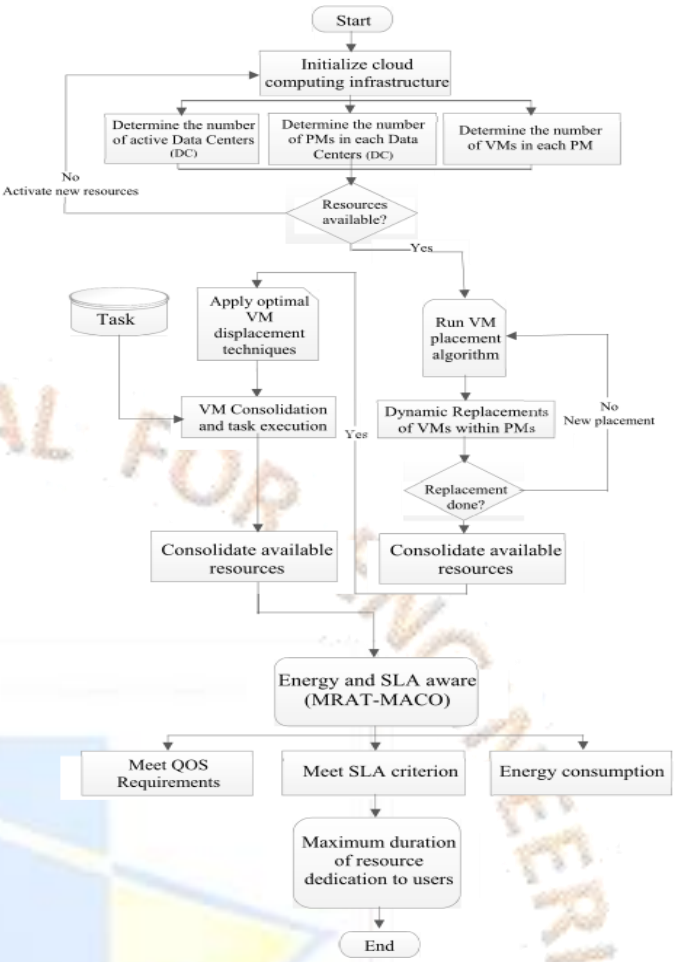
[2] The authors compared the performance of BPG with several other image compression techniques, including JPEG, JPEG2000, and WebP. They evaluate the performance in terms of compression ratio, PSNR (peak signal-to-noise ratio), and visual quality. The authors are introduced a detailed analysis of the performance of BPG for color image compression and highlights the importance of selecting appropriate compression settings to achieve optimal performance.

[3] The authors proposed a feature-aware compression framework that incorporates image features into the compression process. The framework uses a content-adaptive rate control technique that allocates more bits to important image regions, such as edges and textures, and fewer bits to regions with less important features.

Machine Learning plays a vital part in the Bioinformatics field. ML (Machine Learning) Calculations and Strategies are utilized for the Classification and Clustering of proteins information, sequencing information, genomics information, etc.A part of ML Calculations such as KNN (k Closest Neighbor), SVM (Bolster Vector

Machine), Calculated Relapse, Naïve Bayes, k-means, k-median, GLM

Fig.1 : Existing Model



(Generalized Direct Show), Choice Tree, and Irregular Timberland are accessible that perform Classification and Clustering assignments for Bioinformatics datasets.

III. PROPOSED SYSTEM

Detecting fake job postings can be a complex task that requires a combination of natural language processing and machine learning techniques. Here is a general approach to building a fake job detection system.

Platforms such as online job portals or social media for job advertisements are an exciting way of attracting potential candidates on which many enterprise companies are dependent on the hiring process. Fake jobs scam detection at an early stage can save a job seeker and make them only apply for legitimate companies. For this purpose, various machine learning techniques were utilized in this project. Specifically, supervised learning algorithms classifiers were used for scam detection. This project experimented with different algorithms

such as naïve Bayes, SVM, decision tree, random forest, and K-Nearest Neighbor. It is reported that the K-NN classifier gives a promising result for the value k=5 considering all the evaluating metrics. On the other hand,Random Forest is built based on 500 estimators on which the boosting is terminated.

The research understudy can be described as a three-tier approach starting with the dataset preprocessing, feature selection, and classifying by applying different machine learning models and evaluating them.

Let us look at the research that has already been done in this field of detecting fraudulent advertisements or detection of spam emails etc., over a period.

It is observed that many researchers have applied classification algorithms, including XGBOOST, SVM, etc., among which XGBOOST outperformed in many cases.

Considering this performance of XGBOOST as a parameter to be

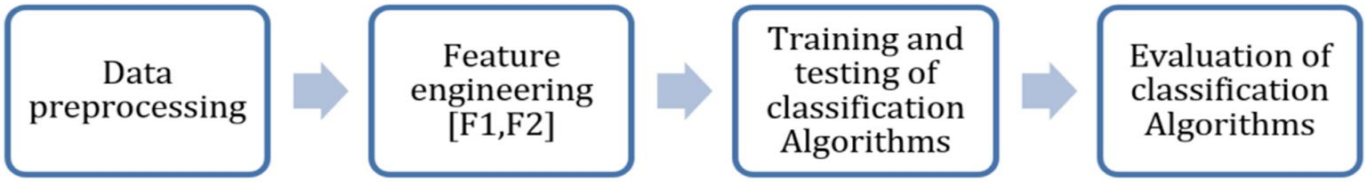
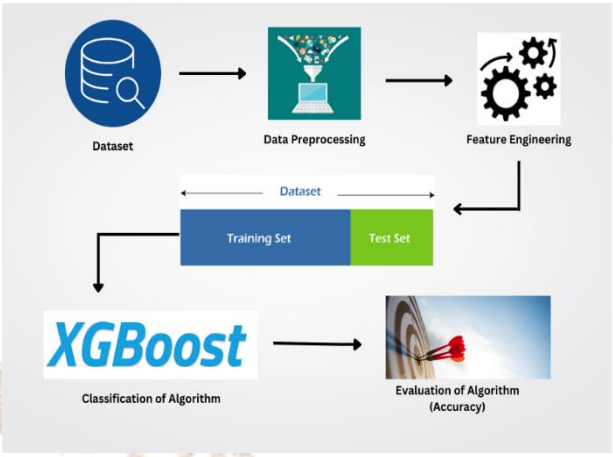


Fig.2 : Function Flow Diagram

validated, this research focuses on applying XGBOOST on the dataset and comparing their results .

Gather Data: Collect a large dataset of job postings that have been labeled as either real or fake. You can collect this data from job boards, social media, or other sources.

Fig.3 : System Architecture



Preprocessed Data: Clean and preprocess the data by removing irrelevant information such as HTML tags, stop words, and punctuations.

Extract Features: Extract relevant features from the job postings such as the job title, company name, location, salary range, experience requirements, and job description.

Build a Model: Train a machine learning model using the extracted features and the labeled data. You can use various algorithms such as decision trees, random forests, or deep learning techniques like neural networks to build a model.

Evaluate the Model: Evaluate the performance of the model using various evaluation metrics such as precision, recall, F1-score, and accuracy.

Deploy the Model: Once the model is trained and evaluated, you can deploy it as a web application or an API that can accept new job

postings and predict whether they are real or fake.

Here is a sample code that to train a Xgboost,SVM,NLP Classifier model on a dataset of job postings

PHP File :

```
<?php
require_once 'config/config.php';
require_once 'helpers/functions.php';
require_once 'helpers/auth.php';
require_once 'helpers/validation.php';
require_once
'helpers/sentimental/autoload.php';
// Autoload Core Libraries
spl_autoload_register(function($className) {
 require_once 'libraries/'. $className .'.php';
});
```

SQL File :

```
SELECT job_id, start_time, end_time, status,
 CASE
  WHEN TIMESTAMP_DIFF(end_time,
start_time, HOUR) <= 1 THEN 'MET'
  ELSE 'NOT MET'
 END AS sla
FROM job_schedule;
```

This is just a sample code, and we may need to tweak it based on your specific requirements and dataset. Additionally, it's important to note that building an accurate fake job detection system can be a challenging task that requires careful consideration of many factors.

IV. EXPERIMENTAL RESULTS

We have assessed the proposed calculation with respect to cost productivity, work cut-off time, and normal work fulfillment time. For these tests, we have utilized Evaluating Model (Genuine) as displayed for the VM valuing, which is like the Amazon AWS estimating plan for the cloud examples.

Step – 1 : Sign Up the Webapp

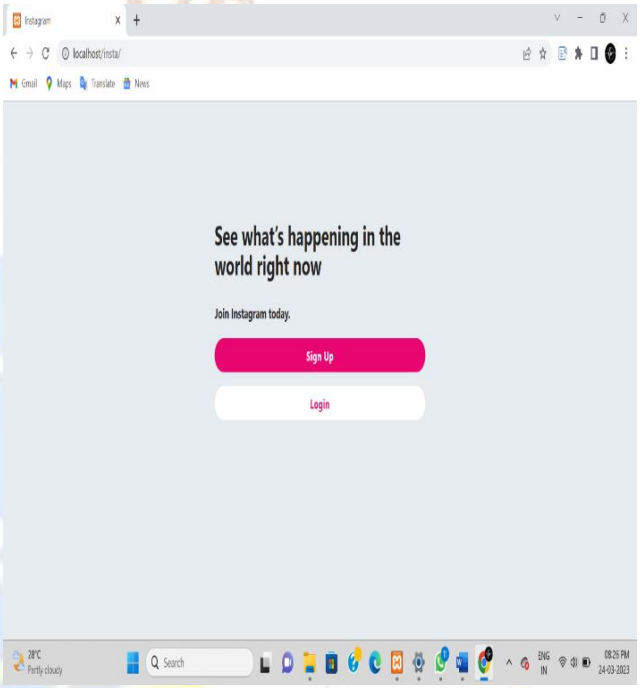


Fig.4 : Sign Up Page

Step – 2 : Enter the needed Credentials and then Submit.The Account will be created.

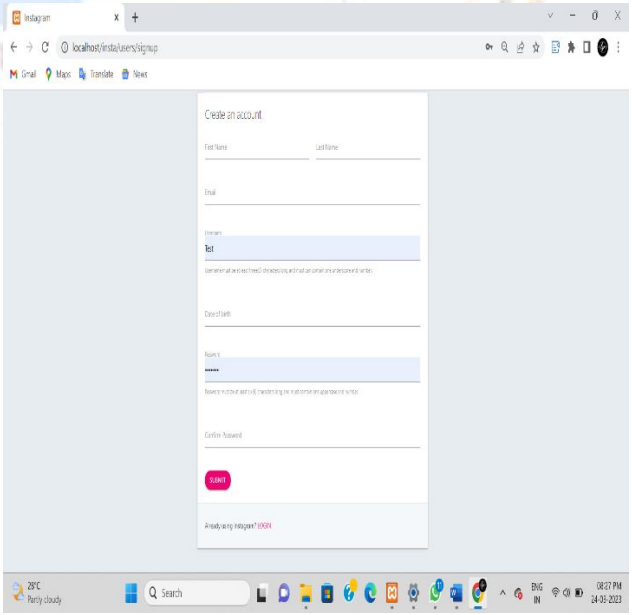


Fig.5 : Credential Site

Step – 3 : Login to the Webapp by entering your Username and Password

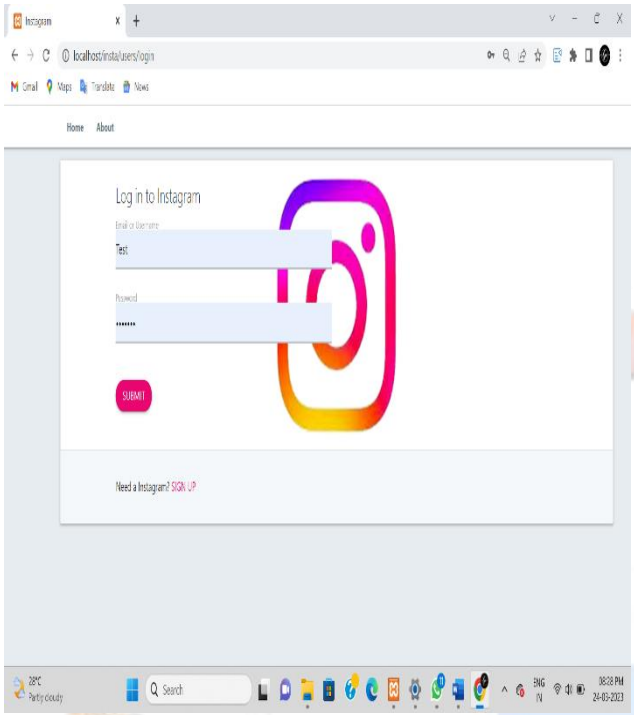


Fig.6 : Login Page

Step – 4 : The page will be displayed as shown below.

Here Im choosing public works department.

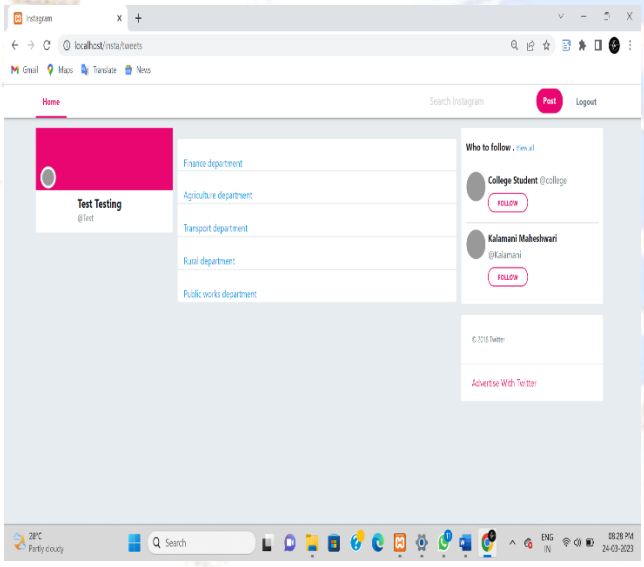


Fig.7 : Selecting the department regarding to the recommended profile

Step – 5 : Now I copy and pasted the job profile from an website

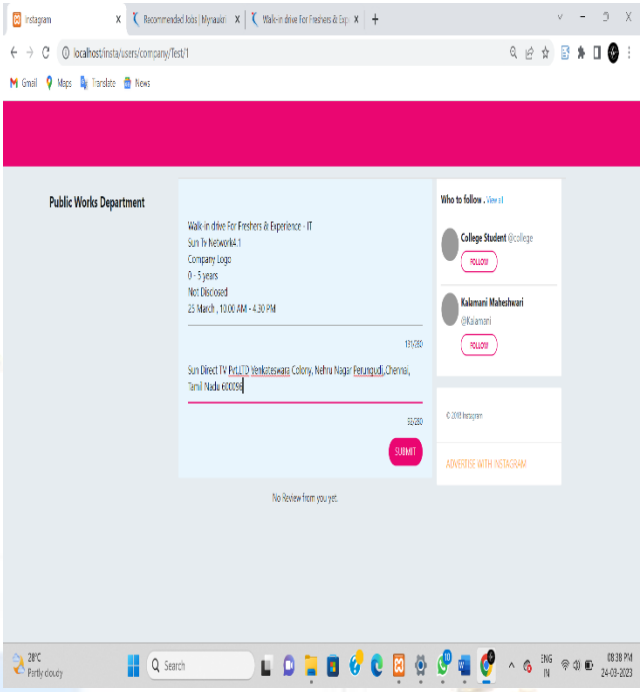


Fig.8 : Submiting the recommended profile in the review

Step – 6 : The Output is Displayed with result i.e., positive,negative and neutral and Star rating.

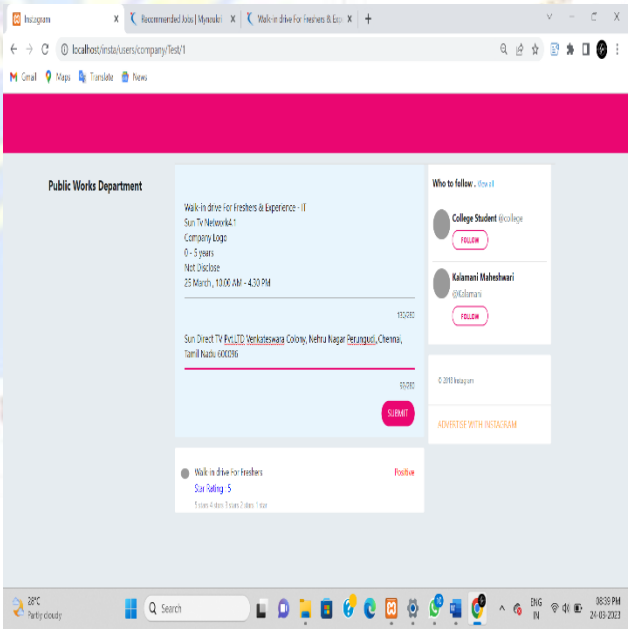


Fig.9 : Output is diaplayed

V.CONCLUSION

So, we concluded that these databases are very useful in handling large amount of data and are highly scalable and can handle semi structured and structured data in very efficient manner and also many other advantages over relational databases which makes them more useful and popular in future.

VI. REFRENCES

[1] M. Gashnikov, "Pyramidal Image Compression Based on Machine Learning," 2022 Ural-Siberian Conference on Biomedical Engineering, Radioelectronics and Information Technology (USBEREIT), Yekaterinburg, Russian Federation, 2022, pp. 240-243.

[2] B. Kovalenko and V. Lukin, "Analysis of color image compression by BPG coder," 2022 IEEE 3rd KhPI Week on Advanced Technology (KhPIWeek), Kharkiv, Ukraine, 2022, pp. 1-6.

[3] L. Huang and T. Suzuki,"WeightedWavelet-Based Spectral-Spatial Transforms For CFA-Sampled Raw Camera Image Compression Considering Image Features," ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Singapore, Singapore, 2022, pp. 1850-1854.

[4] N. Shyamala and S. Geetha, "Fusion model of Modified Wavelet Transform and Neural Network for medical image compression," 2022 3rd International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2022, pp. 1676-1682.

[5] B. A. Lungisani, C. K. Lebekwe, A. M. Zungeru and A. Yahya, "Image Compression Techniques in Wireless Sensor Networks: A Survey and Comparison," in IEEE Access, vol. 10, pp. 82511-82530, 2022

[6] V. Makarichev, G. Proskura, O. Rubel, V. V. Lukin, B. Vozel and K. Chehdi, "Lossy Compression of Three-Channel Remote Sensing Images with “Color” Component Downscaling," IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium, Kuala Lumpur, Malaysia, 2022, pp. 2215-2218.

[7] M. Lu, P. Guo, H. Shi, C. Cao and Z. Ma, "Transformer-based Image Compression," 2022 Data Compression Conference (DCC), Snowbird, UT, USA, 2022, pp. 469-469.

[8] P. Bacchus, R. Fraisse, A. Roumy and C. Guillemot, "Quasi Lossless Satellite Image Compression," IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium, Kuala Lumpur, Malaysia, 2022, pp. 1532-1535.

[9] R. Yuzkiv and M. Gashnikov, "Modification of Machine Learning Algorithms for Embedding in Image Compression Methods," 2022 VIII International Conference on Information Technology and Nanotechnology (ITNT), Samara, Russian Federation, 2022, pp. 1-4.

[10] V. O. Makarichev, V. V. Lukin, I. V. Brysina and B. Vozel, "Spatial Complexity Reduction in Remote Sensing Image Compression by Atomic Functions," in IEEE Geoscience and Remote Sensing Letters, vol. 19, pp. 1-5, 2022, Art no. 6517305.

[11] A. J. Ahmed, M. M. Hamdi, A. S. Mustafa and S. A. Rashid, "WSNApplication Based on Image Compression Using AHAAR Wavelet Transform," 2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), Ankara, Turkey, 2022, pp. 1-4.

[12] M. Simka, J. Kufa and L. Polak, "Picture Quality of 360° Images Compressed by Emerging Compression Algorithms," 2022 32nd International Conference Radioelektronika (RADIOELEKTRONIKA),Kosice,Slovakia, 2022, pp. 1-4.

[13] G. Spasova and I. Boychev, "A Method of Color Images Compression," 2021 International Conference on Biomedical Innovations and Applications (BIA), Varna, Bulgaria, 2022, pp. 111-114.

[14] Y. Chen and F. Yuan, "Lossless Compression Method of Color Image Based on Improved Particle Swarm Optimization," 2022 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC), Dalian, China, 2022, pp. 22-26.

[15] A. Maksimov and M. Gashnikov, "Generalization of Machine Learning-Based Compression Method to Hyperspectral Images," 2022 VIII International Conference on Information Technology and Nanotechnology (ITNT), Samara, Russian Federation, 2022, pp. 1-4.

[16] G. Dhanalakshmi, Victo Sudha George,Security threats and approaches in E-Health cloud architecture system with big data strategy using cryptographic algorithms,Materials Today:Proceedings,Volume 62, Part 7,2022,Pages 4752-4757,ISSN 2214-7853,https://doi.org/10.1016/j.matpr.2022.03.254.

[17] Dhanalakshmii, G., and G. V. George. "An Enhanced Data Integrity for the E-Health Cloud System using a Secure Hashing Cryptographic Algorithm with a Password Based Key Derivation Function2 (KDF2)." Int J Eng Trends Technol 70 (2022): 290-7.

[18] Jaya Sree, R. V., et al. "A Comprehensive Survey of Approaches Used For Detecting Events in Twitter." International Journal of Applied Environmental Sciences 11.1 (2016): 259-266.

[19] Dhanalakshmi, G., et al. "Explosion detection and drainage monitoring system by Automation System." International Journal of Innovative research in computer and communication engineering 6.2 (2018).