

Smart E-Admission System with Feature Selection Technique using Hadoop

- ¹ Prof. Asmita R. Kamble, computer, SITS , Maharashtra India
² Ajinkya Paigude, computer, SITS , Maharashtra India
³ Sameer Shelar, computer, SITS , Maharashtra India
⁴ Shubham Shah, computer, SITS , Maharashtra India
⁵ Sumit Rahate, computer, SITS , Maharashtra India

ABSTRACT

Hadoop is very much suitable for high volume data and it also provide high speed access to the data of the application which we want to use. Using PSO Swarm Search and Feature Selection (FS) technique, which reduces the number of features, noisy, removes irrelevant and redundant data. As data coming from multiple university is vast in this includes storing it into database with proper manner and processing appropriate data in less time, so both this technology can make system smarter than present system. The objective of this paper is to design and implement a smart and efficient online admission system with more advance feature.

Keyword : - Hadoop, Feature Selection, Online Admission.

1. INTRODUCTION

Student admissions are essential part of any universities working because students are what keep a University alive. The student admission is one of the most prominent exercises within a university as no one can survive without students. A poor admissions system may lead to fewer students being admitted into a university because of mistakes or an overly slow response time.

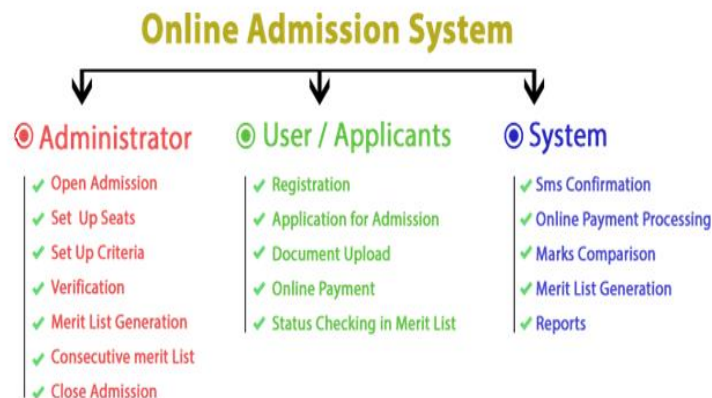
The increasing numbers of students looking for admission in the Universities are causing terrible pressure on the administrative body of the institutes to arrange and manage the admission process manually. It is not efficient now to handle the process accurately and in time. Hence, the need for online admission is necessary.

In case of a manual system, it is a time consuming process and involves immense manpower wherein the online admission system ensures precise and very fast computerized information processing. Maintaining backing up of data is also very easy using 'Online Admission System'. The aim of 'Online Admission System' is to automate the Academic Institute's admission structure and its related operation and functionality.

The objective of the initiative is to provide support to the administration and admission seeking candidates by providing a transparent, faster and easy way of keeping records and use them for reference and further proceeding. Online admission system present till date is not capable of handling large amount of data. Those system were time consuming and less efficient. In our system we are trying to overcome these drawbacks.

Hadoop is based on distributed computing having Hadoop Distributed File System (HDFS). Hadoop can be deployed on low cost hardware and highly fault-tolerant. So using Hadoop we can overcome those drawbacks as this technology can handle large amount of data and provide high speed access. Hadoop architecture is cluster based, which is consist of nodes i.e. Data node, name node, physically separate to each other, in ideal condition. The performance of Hadoop can be increased by proper assignment of the tasks in the default scheduler. In Hadoop a

program known as map-reduce is used to collect data according to query. As Hadoop is used for huge amount of data therefore scheduling can improve better performance in terms of storing and processing



The roles of each individual in the system are shown in above figure. The system has three tier access models. The jobs are divided among these three tiers. The jobs like registration to the system, apply for admission and document upload are done by student/applicant. Administrator is supposed to do the administrative jobs like set up criteria for seats, verification of documents, merit list generation etc.

2. MATERIAL

2.1. HADOOP:

As we know there is a lot of data surrounds us but we can't make that data beneficial to us. There is no capable tool that can make out analysis of information from this huge amount of data.

Hadoop is based on distributed computing having HDFS file system (Hadoop Distributed File System). Hadoop is very much appropriate for high volume of data and it also gives the high speed access to the data of the application. Cluster based architecture of Hadoop, which consist of nodes (data node, name node), physically separate to each other, in ideal situation. The performance of hadoop can be improved by proper allotment of the tasks in the default scheduler. In hadoop a program known as map-reduce is used to gather data according to query. As hadoop is used for huge amount of data therefore scheduling in hadoop must be efficient for desirable performance.

Hadoop Architecture is based on HDFS, which is hadoop distributed file system. In which data is equally (ideally) divided into each node in the hadoop system. When a client want to fetch, add, modify or delete some data from Hadoop, then hadoop system gather the data from each node of our interest and do the knowledgeable actions of our interest.

Name Node is a special kind of master node, which is having the self-information or we can say that meta-data about the all data node there is address, free space, data stored, active & passive data node, task tracker, job tracker and many other configuration such as data replication.

Data Node is a type of slave node in the Hadoop, which is used to store the data and there is task tracker in data node which keep an eye on the on-going job on the data node and the jobs which are coming from name node.

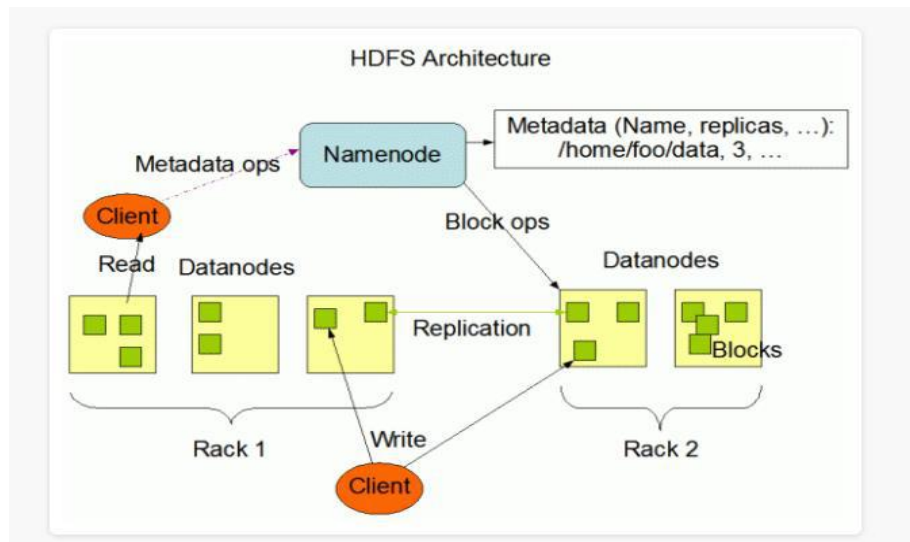


Figure. No.:3 HDFS Architecture

3. ARCTITECTURE

In following figure 5, the architecture of the system is given. Hadoop DB handles all the database related tasks. OLTP is used for securing the online transactions. OLTP itself is a secure protocol. Information analysis and processing takes care of processing of data. The data is submitted online via web page.

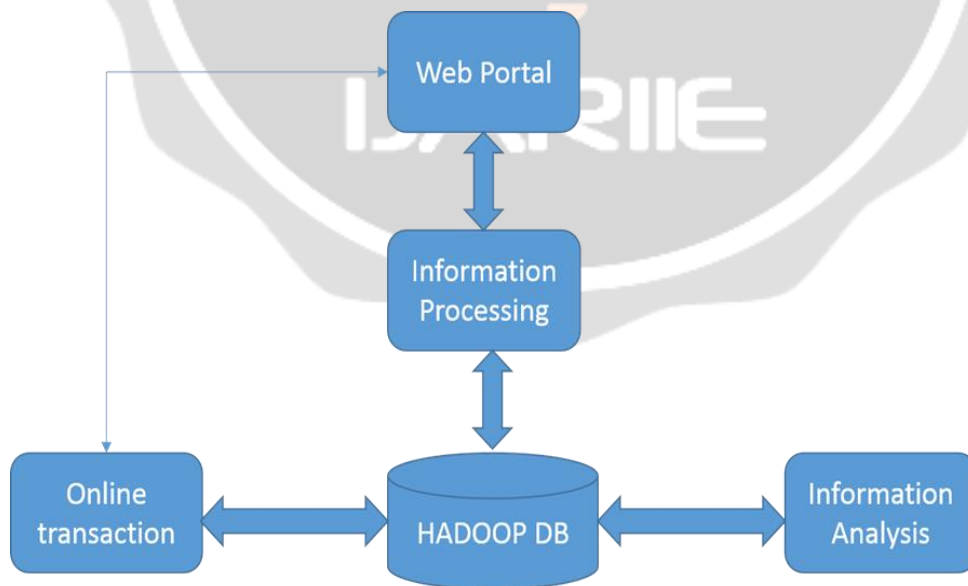
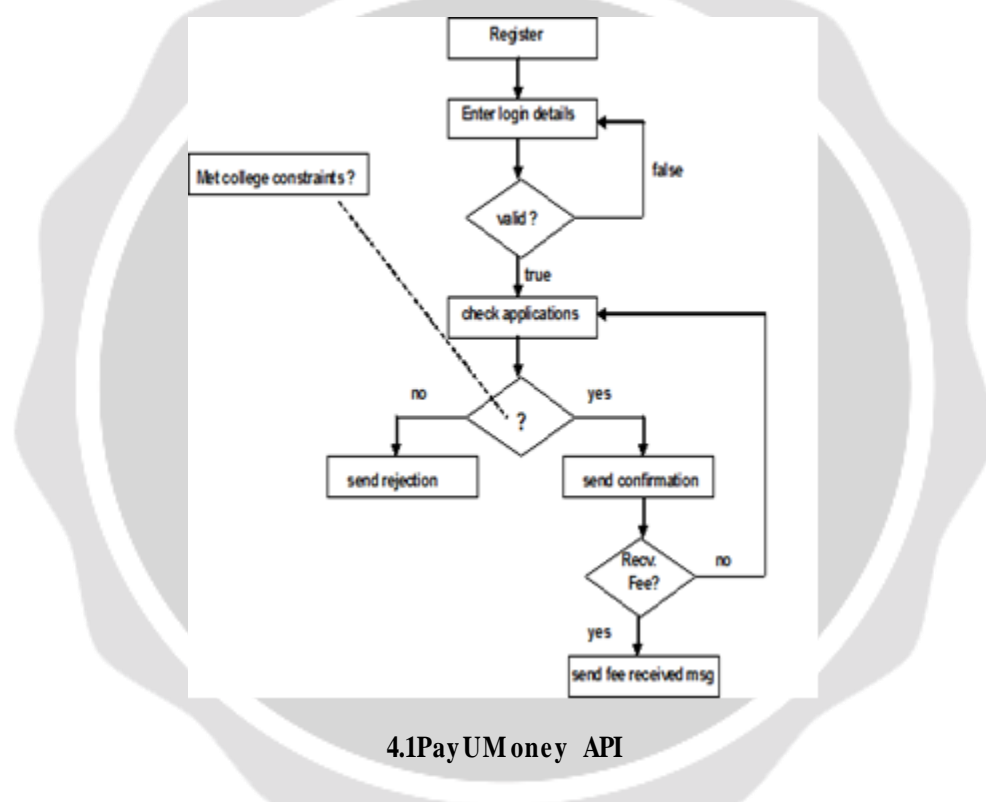


Figure. No. 5 System Architecture

4. IMPLEMENTATION

We are developing our project which is Web-Based and for its development we will be using JSP language. In our GUI first you have to provide the necessary information which is mentioned on the web page and then you will get authorized User-Id and Password from the system itself, You can login into the system using allocated pair of id-password and our system recognizes only the authorized users. Once you login into the system you have to upload the required documents and then the system validates the documents, You can proceed if and only if the uploaded documents match up the system requirements. All the documents and the candidate's details are stored in Hadoop Database. Later you have to pay the form fees online and for that we are using Dummy Gateway for Online-Transaction. After Successful transaction user will be notified through email. Here Our system is using two new functionalities first is online document uploading-validation and second is online form fee payment.



For online money transaction we are using PayUMoney API, this API provide us secure transaction over internet, In order to call API, we are passing authentication in the form of authorization header provided to PayUMoney API. To pass the header we are adding HTTP header with name as 'Authorization' and value as the value provided by us. Every merchant has a unique authorization.

We are using this api for two specific reason first is payment API , This API is used to post a payment request on PayUmoney's server , for this we need some mandatory field like Unique merchant key provided by PayUmoney, transaction_id, payment amount, product_information , first_name, email, phone number, Success URL where PayUmoney will redirect the customer after successful payment, failure url, hash checksum , service_provider

Second is getpayment response API, this API require Authorization Header, merchant Key is Mandatory, from date , todate , count are optional parameter but atleast one parameter should be provided



5. ALGORITHM

5.1 Template matching

Template matching is a technique in digital image processing for finding small parts of an image which match a template image. It can be used as a way to detect edges in images

A basic method of template matching uses a convolution mask (template), tailored to a specific feature of the search image, which we want to detect. This technique can be easily performed on grey images or edge images. The convolution output will be highest at places where the image structure matches the mask structure, where large image values get multiplied by large mask values.

This method is normally implemented by first picking out a part of the search image to use as a template: We will call the search image $S(x, y)$, where (x, y) represent the coordinates of each pixel in the search image. We will call the template $T(x_t, y_t)$, where (x_t, y_t) represent the coordinates of each pixel in the template. We then simply move the center (or the origin) of the template $T(x_t, y_t)$ over each (x, y) point in the search image and calculate the sum of products between the coefficients in $S(x, y)$ and $T(x_t, y_t)$ over the whole area spanned by the template. As all possible positions of the template with respect to the search image are considered, the position with the highest score is the best position. This method is sometimes referred to as 'Linear Spatial Filtering' and the template is called a filter mask.

$$\text{Diff}(x_s, y_s, x_t, y_t) = |I_s(x_s, y_s) - I_t(x_t, y_t)|$$

6. RESULTS AND TABLES

The system is expected to give following results after successful execution:

- 1) Student Admission System gives the admission to the student.
- 2) Maintains student personal, academic and fee related data.
- 3) Marks details & verifications of filled forms for eligibility checking.
- 4) Generate Student's Personal Detail Report.
- 5) Generate Student's Fee Deposition Status Report.
- 6) It Stores Merit list provided by Universities.
- 7) The time required for whole admission process is reduced by huge extent.
- 8) Efficient data processing by removing redundant data.
- 9) Preparation of merit list for each category and each course.
- 10) Sending call letters to online users on e-mail/Mobile.

Sr. No	Present System	Proposed System
1	Older system was not on Hadoop	The system will be in Hadoop
2	OLTP (Online Transactions were not present)	Online transaction were present
3	Manual Scheduling was taking place	Automatic Scheduling was taking place
4	Reduce efforts of the university	Reduces efforts of the university more than previous system.
5	Results will be displayed in more time.	Results will be displayed in less time.
6	Commission system is still present	Commission system will be eliminated

7. CONCLUSIONS

Big Data grows continually with fresh data are being generated at all times; hence it requires an incremental computation approach which is capable of monitoring large scale of data dynamically. Since managing data from multiple universities is time consuming process. So, lightweight incremental algorithms should be used which are capable of achieving robustness, high accuracy and minimum pre-processing latency. In this paper, we designed a system which handles and process tremendous amount of data at the same time giving accurate and efficient results.

8. ACKNOWLEDGEMENT

We would like to thank our project guide Prof. Asmita R. Kamble and all the staff members of our department for giving us valuable time and contributing their experience towards this project.

9. REFERENCES

- 1) Quinlan, J.R., C4.5: Programs for Machine Learning. Morgan Kauffmann Publishers, 1993
- 2) Ping-Feng Pai, Tai-Chi Chen, "Rough set theory with discriminant analysis in analysing electricity loads", *Expert Systems with Applications* 36 (2009), pp.8799–8806
- 3) Mohamed Medhat Gaber, Arkady Zaslavsky, Shonali Krishnaswamy, "Mining data streams: a review", *ACM SIGMOD Record*, Volume 34 Issue 2, June 2005, pp.18-26
- 4) Wei Fan, Albert Bifet, "Mining Big Data: Current Status, and Forecast to the Future", *SIGKDD Explorations*, Volume 14, Issue 2, pp.1-5
- 5) Suman Arora," Survey Paper on Scheduling in Hadoop", ISSN: 2277 128X.
- 6) Job Scheduling for Multi-User Map Reduce Clusters, Matei Zaharia, hruba Borthakur, Joydeep Sen Sarma, Khaled Elmeleegy, Scott Shenker, Ion Stoica; Electrical Engineering and Computer Sciences, University of California at Berkeley.

