International Journal of Psychosocial Rehabilitation, Vol. 24, Issue 02, 2020
ISSN: 1475-7192

Smart Online E-Learning Platform for Students using Educational Data Mining.

[1]G.Prudhvi, [2]P.Sri ramya

ABSTRACT-- *The focal point of this examination was to utilize Educational Data Mining (EDM) methods to direct a quantitative investigation of understudy's communication with an e-learning framework through teacher drove non-evaluated and reviewed courses. This activity is valuable for building up a rule for a progression of online short courses for them. A gathering of understudy's entrance conduct in an e-learning framework was dissected and they were assembled by their course get to log records. The outcome indicated that the distinction in the learning situations could change the online access conduct of an understudy gathering. Enormous Data Technology is utilized here for the unstructured information resembles recordings. The outcomes show that the understudies have a decent mechanical competency, have moderate competency in collaboration with learning substance, and absence of communication abilities with their learning network. A proposal to improve understudies' readiness in online cooperative learning is introduced.*

KEYWORDS-- *Educational Data Mining(EDM), instructor-led-non-graded.*

I. INTRODUCTION

On the off chance that you have taken eye to eye classes for your entire life, being somewhat uncertain toward the start is ordinary, regardless of whether you are technically knowledgeable. Be that as it may, taking an online course, instead of an up close and personal class, certainly has its advantages. Here are five favorable circumstances to contemplating on the web.

Considering on the web gives you greater adaptability. You can work and fit your work routine (and your side interests) around your coursework all the more effectively; considerably more so in the event that you are taking a nonconcurrent class: an online class where you don't need to sign in at a particular time for a live meeting however you can examine and collaborate with your teacher and your individual colleagues at your own pace through, for instance, the conversation discussion.

In a review led by The Learning House, 44% of online understudies detailed upgrades in their work remaining, for instance by getting an all day work inside a year of graduation, and 45% revealed a compensation increment. When you finish your online course, you will have increased more work understanding and learned new abilities that will assist you with progressing in your profession! By studying online, you choose your own learning

[1] *UG Scholar, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai.*

[2] *Associate Professor, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences , Chennai .*

Received: 02 Dec 2019 | Revised: 24 Jan 2020 | Accepted: 16 Feb 2020

environment that works best for your needs: be it your bedroom, your study, the café across the street, or your local gym, listening to your instructor's lecture podcast as you run on the treadmill. Isn't that awesome?

Taking an online course also means that you don't have to commute to class, which means less time spent on the bus and more study time sitting on your couch, the sound of a crackling fireplace in the background. You no longer have to worry about driving in the snowstorm and missing an important class!

Studying online means that you pay the tuition fee, possibly book supplies, an online application fee, and few other items. You don't, however, incur the costs of housing (which can range from $10,000 to $13,000 per year) furthermore, transportation, which means lower obligations and more investment funds.

Who says that being progressively self-restrained is a hindrance? The facts confirm that contemplating on the web requires increasingly self-inspiration and time-the board aptitudes, since you will invest a great deal of energy in your own without somebody genuinely near keep you concentrated on cutoff times. Take a gander at it along these lines: your online course won't just show you geography or verse, it will likewise assist you with turning out to be progressively self-roused, a quality that will make you hang out in the work environment and past. It will look incredible on your list of qualifications.

Let's be honest, when pondering what to examine, other than for intrigue and profession openings, where to contemplate is additionally an integral factor. This may confine the selection of subjects or courses to take. Considering on the web at your own accommodation permits you to no longer stress over class area while picking what to realize straightaway. By taking an online course, you can truly concentrate regarding the matter you are keen on and look over the assortment of online courses and projects.

An assortment of online courses and instructional exercises are accessible to families, bolster suppliers, life mentors and experts who wish to develop their insight or obtain certifications in supporting people with ASD and related conditions. Courses go from free initial projects to proficient improvement workshops to expense based online courses that can prompt specific affirmation. A portion of these online suppliers offer courses explicitly for people on the mental imbalance range.

II. LITERATURE SURVIVE

Description : In propose a new method for estimation in linear models. The 'lasso' minimizes the residual sum of squares subject to the sum of the absolute value of the coefficients being less than a constant. Because of the nature of this constraint it tends to produce some coefficients that are exactly 0 and hence gives interpretable models. Our simulation studies suggest that the lasso enjoys some of the favorable properties of both subset selection and ridge regression. It produces interpretable models like subset selection and exhibits the stability of ridge regression. There is also an interesting relationship with recent work in adaptive function estimation by Donoho and Johnstone. The lasso idea is quite general and can be applied in a variety of statistical models: extensions to generalized regression models and tree-based models are briefly described

Description : Convex empirical risk minimization is a basic tool in machine learning and statistics. We provide new algorithms and matching lower bounds for differentially private convex empirical risk minimization assuming only that each data point's contribution to the loss function is Lipschitz and that the domain of optimization is bounded. We provide a separate set of algorithms and matching lower bounds for the setting in which the loss functions are known to also be strongly convex. Our algorithms run in polynomial time, and in some cases even match the optimal non-private running time (as measured by oracle complexity). We give separate algorithms (and lower bounds) for (, 0)- and (, δ)-differential privacy; perhaps surprisingly, the techniques used for designing optimal algorithms in the two cases are completely different. Our lower bounds apply even to very simple, smooth function families, such as linear and quadratic functions. This implies that algorithms from previous work can be used to obtain optimal error rates, under the additional assumption that the contribution of each data point to the loss function is smooth. We show that simple approaches to smoothing arbitrary loss functions (in order to apply previous techniques) do not yield optimal error rates. In particular, optimal algorithms were not previously known for problems such as training support vector machines and the high-dimensional median.

Description : Approximation algorithms can sometimes provide efficient solutions when no efficient exact computation is known. In particular, approximations are often useful in a distributed setting where the inputs are held by different parties and may be extremely large. Furthermore, for some applications, the parties want to compute a function of their inputs securely without revealing more information than necessary. In this work, we study the question of simultaneously addressing the above efficiency and security concerns via what we call secure approximations. We start by extending standard definitions of secure (exact) computation to the setting of secure approximations. Our definitions guarantee that no additional information is revealed by the approximation beyond what follows from the output of the function being approximated. We then study the complexity of specific secure approximation problems. In particular, we obtain a sub linear communication protocol for securely approximating the Hamming distance and a polynomial-time protocol for securely approximating the permanent and related #P-hard problems.

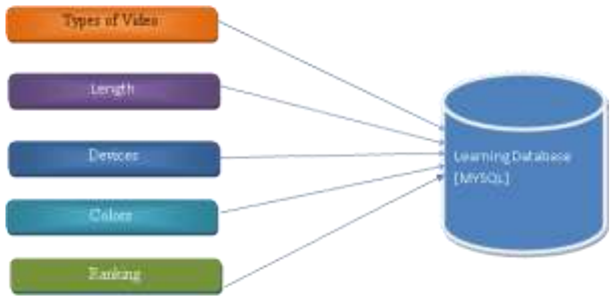
Description : This paper considers the problem of secure data aggregation (mainly summation) in a distributed setting, while ensuring differential privacy of the result. We study secure multiparty addition protocols using well known security schemes: Shamir's secret sharing, perturbation-based, and various encryptions. We supplement our study with our new enhanced encryption scheme EFT, which is efficient and fault tolerant. Differential privacy of the final result is achieved by either distributed Laplace or Geometric mechanism (respectively DLPA or DGPA), while approximated differential privacy is achieved by diluted mechanisms. Distributed random noise is generated collectively by all participants, which draw random variables from one of several distributions: Gamma, Gauss, Geometric, or their diluted versions. We introduce a new distributed privacy mechanism with noise drawn from the Laplace distribution, which achieves smaller redundant noise with efficiency. We compare complexity and security characteristics of the protocols with different differential privacy mechanisms and security schemes. More importantly, we implemented all protocols and present an experimental comparison on their performance and scalability in a real distributed environment. Based on the evaluations, we identify our security scheme and Laplace DLPA as the most efficient for secure distributed data aggregation with differential privacy

III. MODULE DESCRIPTION

1. PREPROCESSING ONLINE LEARNING DATABASE
2. STORAGE
3. ANALYSE QUERY

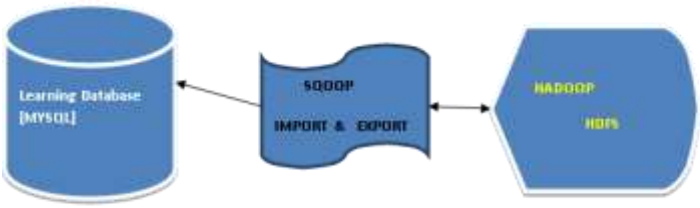
PREPROCESSING ONLINE LEARNING DATABASE

In this module, analyzing the data with different kinds of fields in Microsoft Excel then it converted into comma delimited format which is said to be csv (comma separator value) file and moved to MySQL backup through Database.



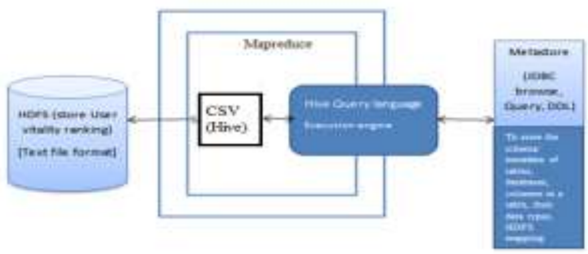
STORAGE

In this module we are getting every one of those reinforcement information which we have put away in MYSQL and bringing in each one of those information by utilization of sqoop orders to HDFS(Hadoop Distributed File System).now all the information are put away in HDFS were it is prepared to get handled by utilization of hive.



ANALYSE QUERY

In this module we are getting all those data from HDFS to HIVE by use of sqoop import command where hive is ready to analyze. Here in HIVE we can process only structured data to analyze. By extracting only the meaningful data and neglecting unclenched data we can analyze the data in more effective manner by use of hive.



Received: 02 Dec 2019 | Revised: 24 Jan 2020 | Accepted: 16 Feb 2020

IV. PROPOSED SYSTEM

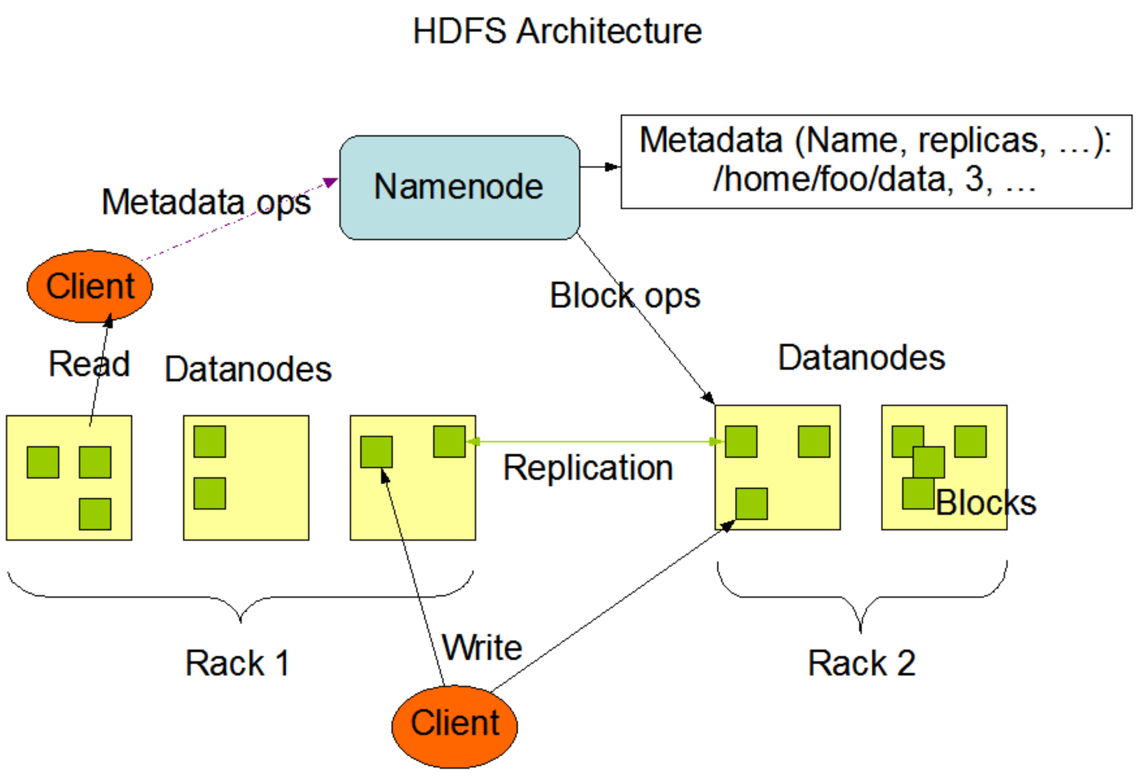
Proposed idea manages giving database by utilizing Hadoop apparatus we can dissect no impediment of information and straightforward add number of machines to the group and we get results with less time, high throughput and upkeep cost is less and we are utilizing segments and bucketing methods in Hadoop.

ADVANTAGES OF PROPOSED SYSTEM:

1. No data loss Problem
2. Efficient data Processing.

V. HDFS Architecture

Given below is the architecture of a HDFS. HDFS follows the distributed system architecture and it has the following elements..

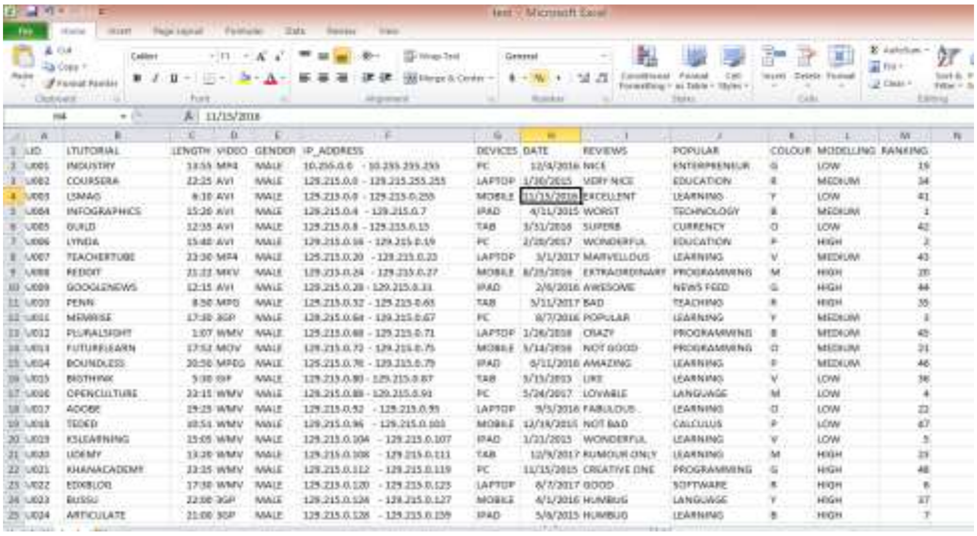


International Journal of Psychosocial Rehabilitation, Vol. 24, Issue 02, 2020
ISSN: 1475-7192

VI. OUTPUTS:

i.) **Data insertion**

ii.) **Tutorial details**



UID	LTUTORIAL	LENGTH	VIDEO	GENDER	IP_ADDRESS	DEVICES	DATE	REVIEWS	POPULAR	COLOUR	MODELLING	RANKING
U001	INDUSTRY	13:55	MP4	MALE	10.255.0.0 - 10.255.255.255	PC	12/9/2016	NICE	ENTERPRENEUR	G	LOW	19
U002	COURSERA	22:25	AVI	MALE	129.215.0.0 - 129.215.255.255	LAPTOP	1/30/2015	VERY NICE	EDUCATION	R	MEDIUM	34
U003	LSMAG	8:10	AVI	MALE	129.215.0.0 - 129.215.0.255	MOBILE	11/15/2018	EXCELLENT	LEARNING	Y	LOW	41
U004	INFOGRAPHICS	15:20	AVI	MALE	129.215.0.4 - 129.215.0.7	IPAD	4/11/2015	WORST	TECHNOLOGY	B	MEDIUM	1
U005	GUILD	12:35	AVI	MALE	129.215.0.8 - 129.215.0.15	TAB	3/31/2008	SUPERB	CURRENCY	O	LOW	42
U006	LYNDA	15:40	AVI	MALE	129.215.0.16 - 129.215.0.19	PC	2/20/2017	WONDERFUL	EDUCATION	P	HIGH	2
U007	TEACHERTUBE	23:30	MP4	MALE	129.215.0.20 - 129.215.0.23	LAPTOP	3/1/2017	MARVELLOUS	LEARNING	V	MEDIUM	43
U008	REDDIT	21:22	MKV	MALE	129.215.0.24 - 129.215.0.27	MOBILE	8/25/2008	EXTRAORDINARY	PROGRAMMING	M	HIGH	20
U009	GOOGLENEWS	12:15	AVI	MALE	129.215.0.28 - 129.215.0.31	IPAD	2/6/2016	AWESOME	NEWS FEED	G	HIGH	44
U010	PENN	8:50	MPG	MALE	129.215.0.32 - 129.215.0.63	TAB	5/11/2017	BAD	TEACHING	R	HIGH	35
U011	MEMRISE	17:30	3GP	MALE	129.215.0.64 - 129.215.0.67	PC	8/7/2016	POPULAR	LEARNING	Y	MEDIUM	3
U012	PLURALSIGHT	1:07	WMV	MALE	129.215.0.68 - 129.215.0.71	LAPTOP	1/26/2018	CRAZY	PROGRAMMING	B	MEDIUM	45
U013	FUTURELEARN	17:52	MOV	MALE	129.215.0.72 - 129.215.0.75	MOBILE	3/14/2008	NOT GOOD	PROGRAMMING	O	MEDIUM	21
U014	BOUNDLESS	20:50	MPEG	MALE	129.215.0.76 - 129.215.0.79	IPAD	6/11/2010	AMAZING	LEARNING	P	MEDIUM	46
U015	BIGTHINK	5:00	[illegible]	MALE	129.215.0.80 - 129.215.0.87	TAB	3/13/2013	LIKE	LEARNING	V	LOW	36
U016	OPENCULTURE	22:15	WMV	MALE	129.215.0.88 - 129.215.0.91	PC	5/24/2017	LOVABLE	LANGUAGE	M	LOW	4
U017	ADOBE	19:25	WMV	MALE	129.215.0.92 - 129.215.0.95	LAPTOP	9/3/2016	FABULOUS	LEARNING	O	LOW	22
U018	TEDED	10:54	WMV	MALE	129.215.0.96 - 129.215.0.103	MOBILE	12/19/2015	NOT BAD	CALCULUS	P	LOW	47
U019	KSLEARNING	15:05	WMV	MALE	129.215.0.104 - 129.215.0.107	IPAD	1/21/2013	WONDERFUL	LEARNING	V	LOW	5
U020	UDEMY	13:20	WMV	MALE	129.215.0.108 - 129.215.0.111	TAB	12/9/2017	RUMOUR ONLY	LEARNING	M	HIGH	23
U021	KHANACADEMY	23:35	WMV	MALE	129.215.0.112 - 129.215.0.119	PC	11/15/2015	CREATIVE ONE	PROGRAMMING	G	HIGH	48
U022	EDXBLOG	17:30	WMV	MALE	129.215.0.120 - 129.215.0.123	LAPTOP	8/7/2017	GOOD	SOFTWARE	R	HIGH	6
U023	BUSSU	22:00	3GP	MALE	129.215.0.124 - 129.215.0.127	MOBILE	4/1/2016	HUMBUG	LANGUAGE	Y	HIGH	37
U024	ARTICULATE	21:00	3GP	MALE	129.215.0.128 - 129.215.0.159	IPAD	5/9/2015	HUMBUG	LEARNING	B	HIGH	7

Enter password: training

Welcome to the MySQL monitor. Commands end with ; or \g.

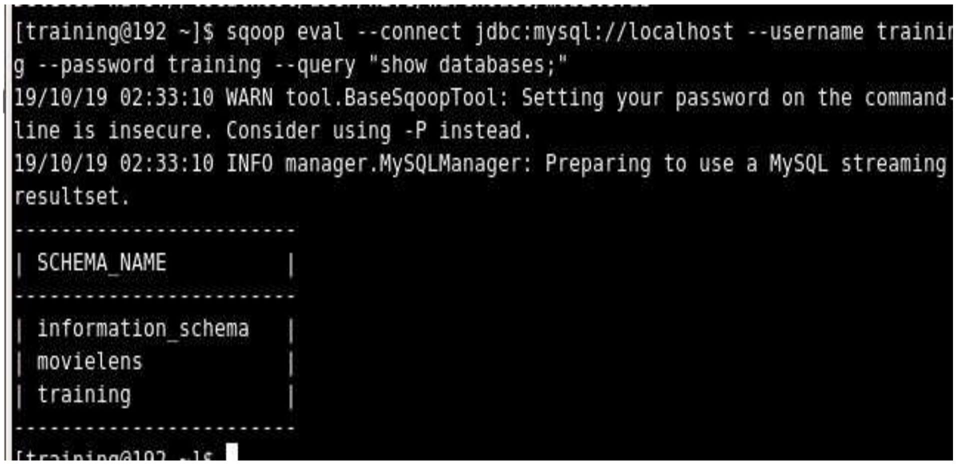
Your MySQL connection id is 19

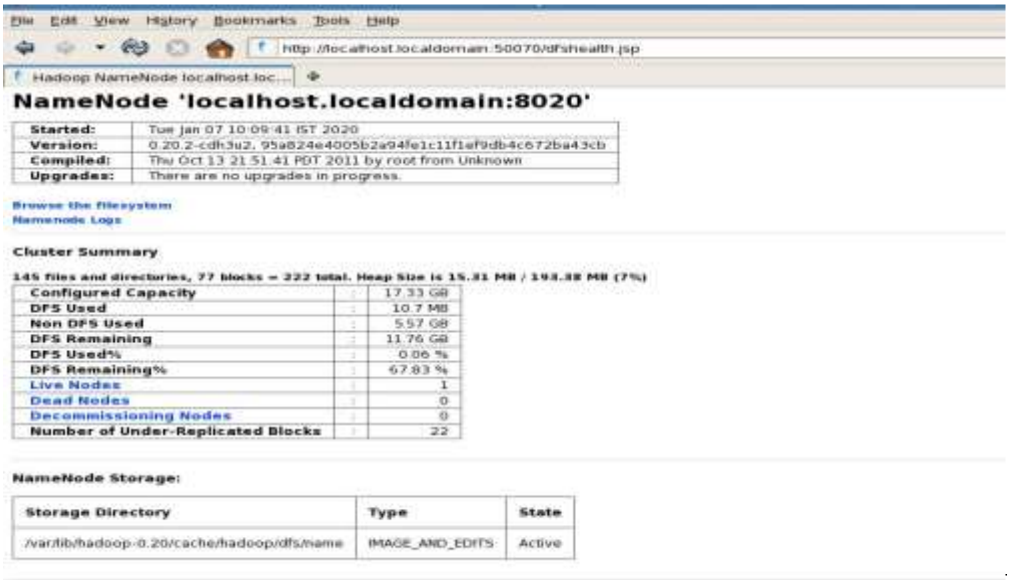
Server version: 5.0.77 Source distribution

Type 'help;' or '\h' for help. Type '\c' to clear the buffer.

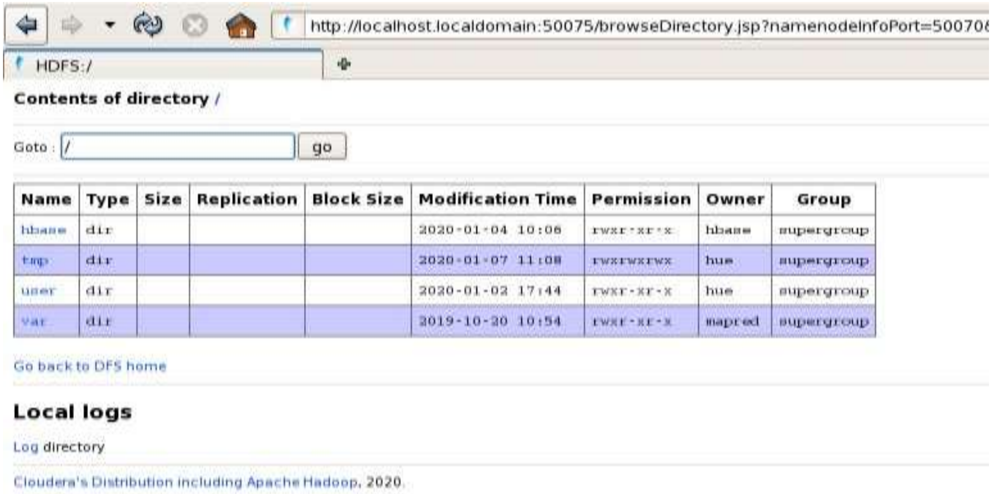
```
[training@192 ~]$ mysql -u training -p movielens  < ~/Desktop/Driving/test1
Enter password:
[training@192 ~]$ mysql -u training -p
Enter password:
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 8
Server version: 5.0.77 Source distribution

Type 'help;' or '\h' for help. Type '\c' to clear the buffer.

mysql> show databases;
+--------------------+
| Database           |
+--------------------+
| information_schema |
| movielens          |
| training           |
+--------------------+
3 rows in set (0.03 sec)

mysql>
```

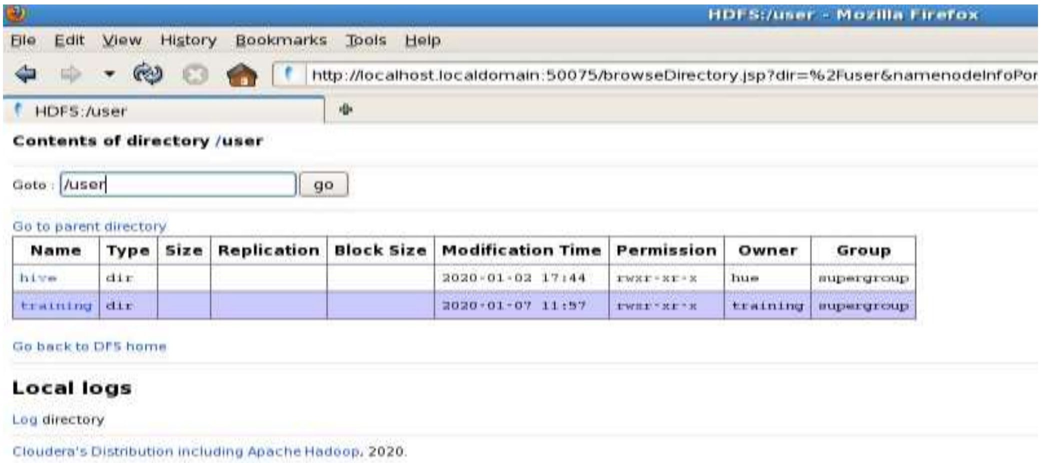
International Journal of Psychosocial Rehabilitation, Vol. 24, Issue 02, 2020
ISSN: 1475-7192





Received: 02 Dec 2019 | Revised: 24 Jan 2020 | Accepted: 16 Feb 2020





Now and again the explanations for something achievement can be gazing you directly in the face. For Hadoop, the greatest helper in the market is basic: Before Hadoop, information stockpiling was costly. Hadoop, nonetheless, lets you store as much information as you need in whatever structure you need, basically by adding more servers to a Hadoop group. Each new server (which can be ware x86 machines with generally little sticker prices) includes more stockpiling and all the more handling capacity to the general group. This makes information stockpiling with Hadoop far less expensive than earlier strategies for information stockpiling.

VII. APPLICATION

Facebook using Hadoop:

At Facebook, Hadoop has for the most part been utilized related to hive for utmost and examination of expansive educational records. The vast majority of this examination happens in isolated bunch occupations and the supplement has been on developing throughput and proficiency. These outstanding loads ordinarily read and make a lot out of information from plate dynamically. In that limit, there have been fewer supplements on making

Hadoop performant for sporadic access remaining occupations waiting be finished by giving low slowness access to HDFS. Or on the other hand perhaps, we have utilized a mix of immense gatherings of MySQL databases and securing measurements made utilizing memcached. All around, results from Hadoop are moved into MySQL or memcached for use by the web level

Twitter using Hadoop:

Twitter has far reaching data collecting and getting ready necessities, and from this time forward we have attempted to execute a great deal of front line data taking care of and work process structures inside Hadoop. In particular, we store a huge segment of our data LZO stuffed, considering the way that the LZO weight winds up striking a remarkably not too awful congruity between weight degree and speed for use in Hadoop. Hadoop occupations are everything seen as IO-bound, and average weight figuring resembles gzip or bzip2 are so computationally focused that vocations promptly advanced toward finding the opportunity to be CPU-bound. LZO strangely was worked for speed, so you get 4-5x weight degree while leaving the CPU open to accomplish affirmed work. For more talk of LZO, complete with execution associations we finished at some point back.

VIII. CONCLUSION

In this paper, we presented a study on Online learning which can handle huge amount of datasets of videos and documents developed for the students. We are using Hadoop to Store and analyze the datasets of the Online learning data and the student access behavior in Hadoop ecosystem using Big Data Technology.

IX. RESULT

In this paper we have discussed about how the software is going to work from above screenshot we have an idea about inserting the datasets and creating the tables according to the size. Later we will upload the videos according to the syllabus and set of time it will play automatically.

REFERENCES

1. Breiman, L. (1993) Better subset selection using the non-negative garotte. Technical Report. University of California, BerkeleChen, S. and Donoho, D. (1994) Basis pursuit. In 28th Asilomar Conf. Signals, Systems Computers, Asiloma
2. Efron, B. and Tibshirani, R. (1993) An Introduction to the Bootstrap. London: Chapman and Hall.
3. Lawson, C. and Hansen, R. (1974) Solving Least Squares Problems. Englewood Cliffs: Prentice Hall
4. LeBlanc, M. and Tibshirani, R. (1994) Monotone shrinkage of trees. Technical Report. University of Toronto, Toront
5. Hastie, T. and Tibshirani, R. (1990) Generalized Additive Models. New York: Chapman and Hall
6. Zhang, P. (1993) Model selection via multifold cv. Ann. Statist., 21, 299-311.
7. A. Beimel, H. Brenner, S. P. Kasiviswanathan, and K. Nissim. Bounds on the sample complexity for private learning and private data release. Machine learning, 94(3), 2014.

8. S. Boyd and L. Vandenberghe. Convex Optimization. Cambridge University Press, New York, NY, USA, 2004.
9. J. C. Duchi, M. I. Jordan, and M. J. Wainwright. Local privacy and statistical minimax rates. In IEEE Symp. on Foundations of Computer Science (FOCS), 2013.
Received: 02 Dec 2019 | Revised: 24 Jan 2020 | Accepted: 16 Feb 2020