K L University
(Koneru Lakshmaiah Education Foundation)
Estd. u/s 3 of UGC Act 1956
Green Fields, Vaddeswaram, (via) K.C. Works P.O. - 522 502, Guntur District, A.P.
Phones: 08645-246948, 246615 ' FAX: 08645-247249, 0866-2577902
Constituent College KLCE Accredited by NAAC with CGPA 3.76/4.00
Approved by A.I.C.T.E ± Accredited by N.B.A.± ISO 9001-2000 Certified

Report on

Faculty Development Program on Big Data Analytics

Sponsored by DST

(26th April 2016 to 6th May 2016)

Conducted by

Cloud computing Research Group

Department of Computer science and Engineering

Cloud computing Research group:

Cloud Computing Research Group in the Department of Computer Science and Engineering of K L University gives opportunities to faculty and students of the group for creating innovative products and e-services that has societal impact. The group celebrates freedom of thought, cultivates vision and encourages growth; it also inculcates human values and concern for the environment and society.

Focused areas of Research:

Cloud Computing; Big Data Technologies; Distributed Systems; Autonomic Computing; Green Computing; All interdisciplinary areas of computing.

DST- BDI Initiative:

Big Data, is data whose scale, diversity, and complexity require new architecture, techniques, algorithms, and analytics to manage it and extract value and hidden knowledge from it. In other words, big data is characterized by volume, variety (structured and unstructured data) velocity (high rate of changing) and veracity (uncertainty and incompleteness). In the Big Data research context, so called analytics over Big Data is playing a leading role. Analytics cover a wide family of problems mainly arising in the context of Database, Data Warehousing and Data Mining research. Analytics research is intended to develop complex procedures running over large-

scale, enormous in-size data repositories with the objective of extracting useful knowledge hidden in such repositories.

Some of the Science &Technology challenges that researchers across the globe and as well as in India facing are, related to data deluge pertaining to Astrophysics, Materials Science, Earth & atmospheric observations, Energy, Fundamental Science, Computational Biology, Bioinformatics & Medicine, Engineering & Technology, GIS and Remote Sensing, Cognitive science and Statistical data. These challenges require development of advanced algorithms, visualization techniques, data streaming methodologies and analytics. To tap the analytics momentum, India now needs to build a sustainable analytics eco-system that brings in a strong partnership across the industry players, government, and academia.

In this context, Govt.of India offers its Big Data Initiative Programme to promote and foster Big Data Science, Technology, applications and also to develop core technologies, tools and algorithms for wider applications.

Objectives of the Faculty Development program:

To expose the Faculty/ Research Scholars/ Students in emerging technologies in the areas of Big Data Analytics. This course provides practical foundation level training that enables immediate and effective participation in big data and other analytics projects.

Outcomes of the Faculty Development program:

Upon successful completion the training programme, participants should be able to participate and contribute as a Data Science Team Member on big data and other analytics projects by: a) Deploying the Data Analytics Lifecycle to address big data analytics projects.

b) Reframing a business challenge as an analytics challenge.

c) Applying appropriate analytic techniques and tools to analyze big data, create statistical models, and identify insights that can lead to actionable results.

d) Selecting appropriate data visualizations to clearly communicate analytic insights to business sponsors and analytic audiences.

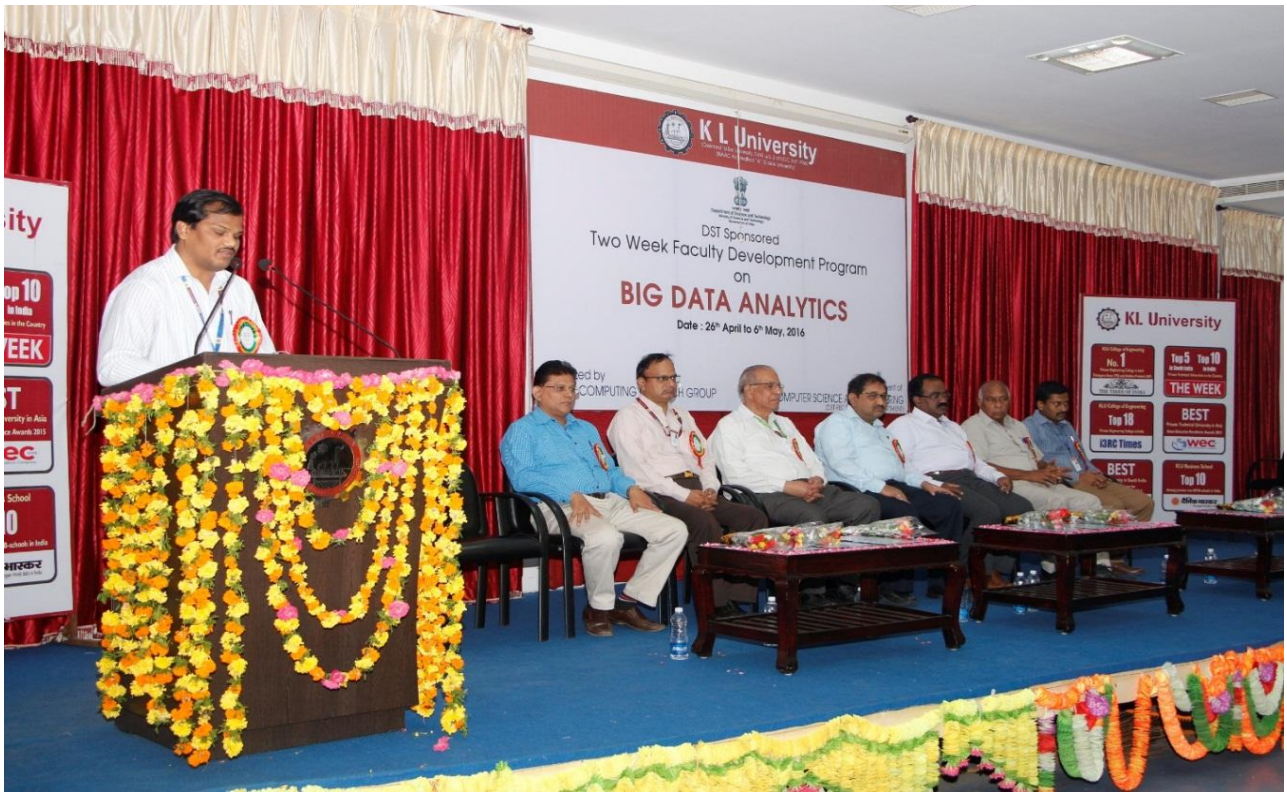
e) Using tools such as: R and RStudio, MapReduce/Hadoop.

Day 1: 26-04-2016

Morning session - Inauguration

The FDP is inaugurated on 26th May 2016 morning at 10AM at Peacock Hall, KL University. Dr.S.R.Rao, Advisor, DBT, Ministry of Science and Technology, Govt. of India, New Delhi is the chief guest for the Inaugural function.

Dr. K. Thirupathi Rao, convener of Bigdata FDP welcomed all the guests participants who came from various parts of India. Honorable Vice chancellor Dr LSS Reddy enlightened the audience regarding the emerging technology BigData. Honorable Chancellor Dr. RamaMoorty ignited all the participants by relating Bigdata analytics to the Power systems field. Chief guest Dr.S.R.Rao,Advisor, DBT, Ministry of Science and Technology explained the various research thrust areas in BigData analytics and BioInformatics, Registrar Dr K Rama Krishna and Principal KLUCE Dr. A Anand Kumar wished all the participants and organizers a very best. HOD CSE proposed the vote of thanks to the Inaugural session.



Dr. K Thirupathi Rao, Convenor Bigdata FDP welcoming all the participants in the inaugural session

Dignitaries Felicitating Dr.S.R.Rao, Advisor, DBT, Ministry of Science and Technology, Govt. of India

Day 1: 26-04-2016

Afternoon Session: R Language and its basic operations

Resource person: Resource: Raja Prabhakar Khaila, TimesPro

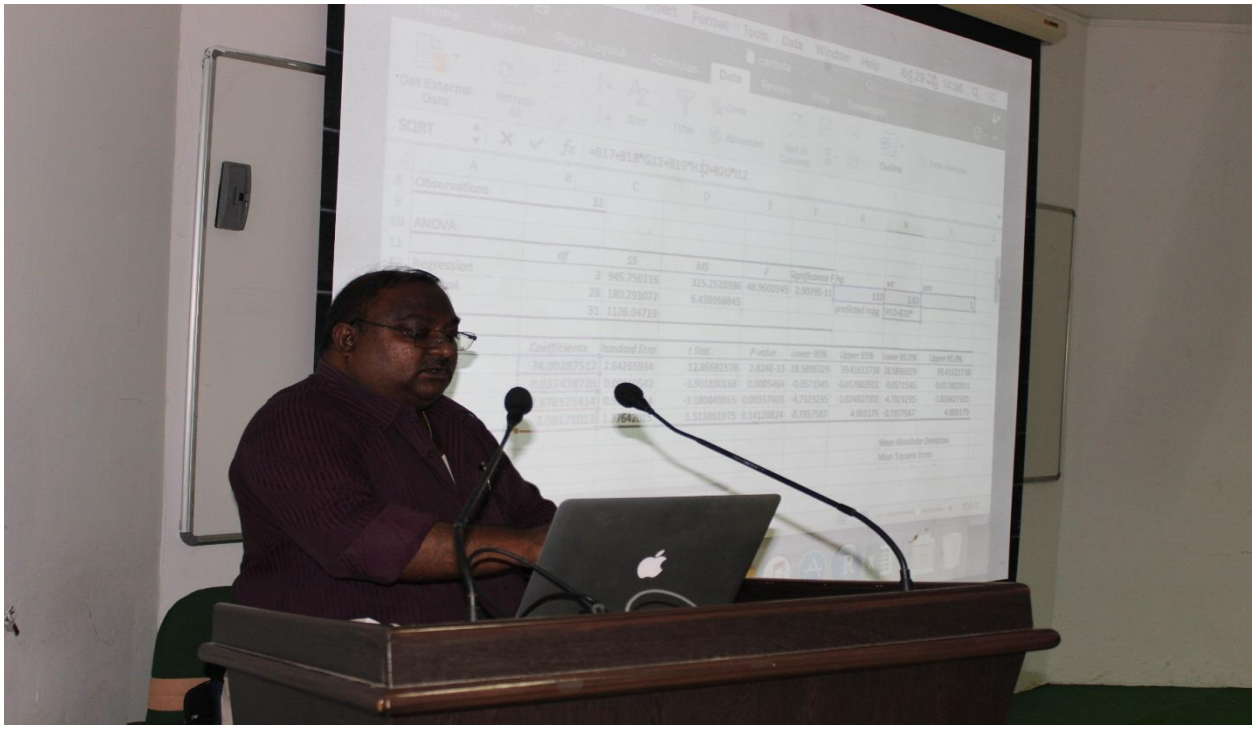
After the grand opening ceremony the audience moved to Rose hall where Mr. Raja Prabhakar Kaila, subject expert- R language from Times Pro introduced the Audience about R language.

He focused on the key features of R. Participants found it very useful. It is also explained about where and where not R can be applied too. R is an open source statistical analytical tool which is used very much in Big Data analytics.

It's a hand on session thereafter where all the participants installed R and R Studio

They have created various datasets and performed basic operations on R.

Mr Rajesh Prabhakar khaila Introducing R language and its features



R language operations demonstrated by Mr. Raja Prabhakar Khaila

DAY 2: 27-04-2016

Morning session: A session on Bigdata

Resource person : Dr. D. V Ramana, Hyderabad

The Bigdata FDP got huge response from both in-house and outside participants. So the venue of the workshop is moved to the Jasmine Hall with capacity of 150.

On Day two Subject experts on BigData analytics Dr. D V Ramana from Hyderabad has took two sessions.

Morning he introduced the Bigdata and storing and processing of Big data, Various tools used in Bigdata mining also were introduced to the audience. The participants found it very useful to know all such information.

DAY 2: 27-04-2016

Afternoon session: A session on Bigdata

Resource Person: Dr. D. V Ramana

Morning session everybody enlightened with various processing techniques in mining the information in Bigdata. Afternoon session Hadoop and Map reduce technologies are practically demonstrated by Dr. D. V. Ramana.

The participants have a hands on experience as all the facilities provided well in advance. They have practiced few Maping and reducing operations using Hadoop.

Dr. D V Ramana, Subject expert Bigdata addressing the audience

Dr. V Srikanth HOD CSE and Dr. V Krishna Reddy, Dr. B. Thirumala Rao felicitating Dr. D V Ramana

Day3 :28-04-16

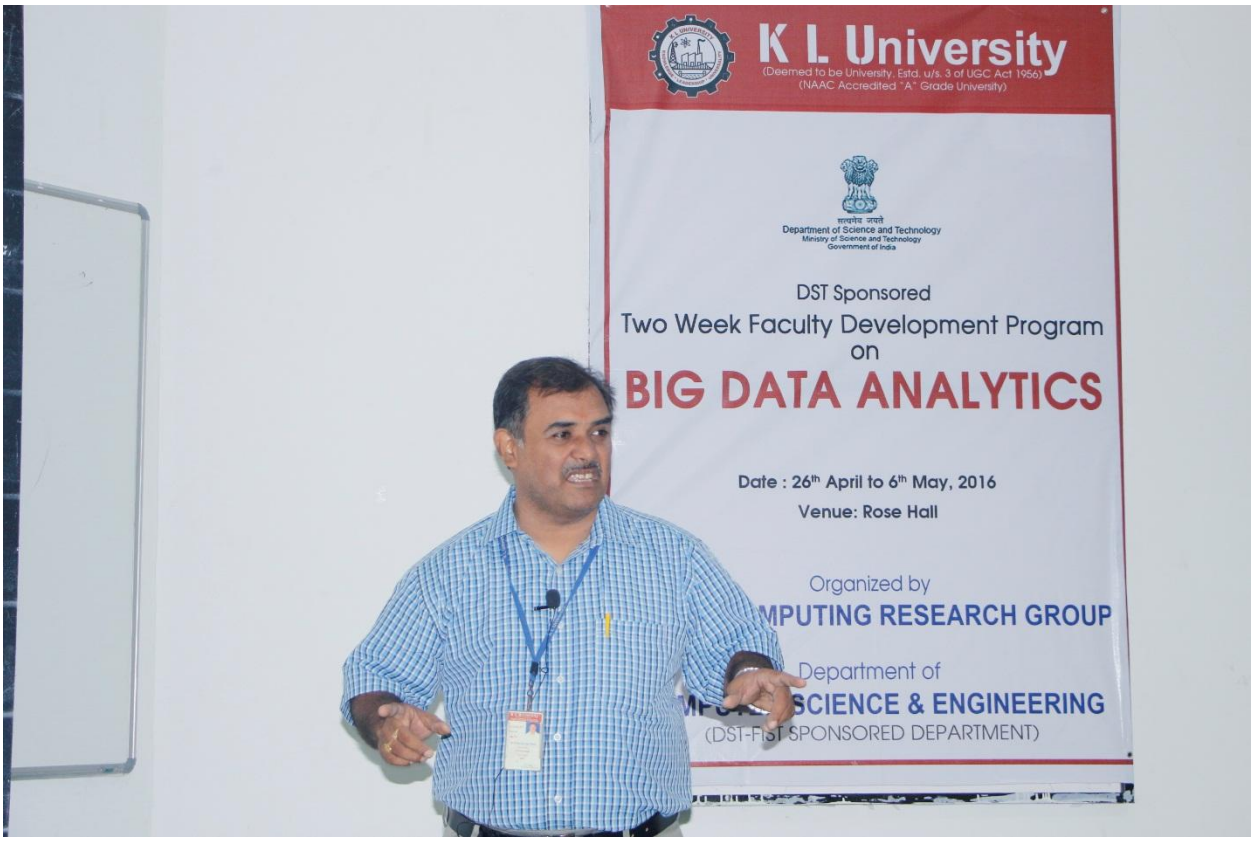
Morning session: A session on Knowledge Discovery in Bigdata

Resource Person: Dr. M.R. Narasinga Rao

Dr. M. R. Narsinga Rao an eminent professor from Department of CSE, KL University has addressed the addressed the audience on various Data mining issues.

He started with Model evaluation and Improvement, classification and accuracy in data mining. Various techniques are explained with suitable examples.Its very much interactive session as various queries from the participants posted to the speaker and all of those are addressed.

Post break various evaluation metrics, PCA, Ensemble methods, SVM techniques and applications of Big Data are discussed. It too went very live where the audience post lot many questions and got them clarified.



Dr. M. R. Narasinga Rao, Professor, Department of CSE addressing the audience

Day3 :28-04-16

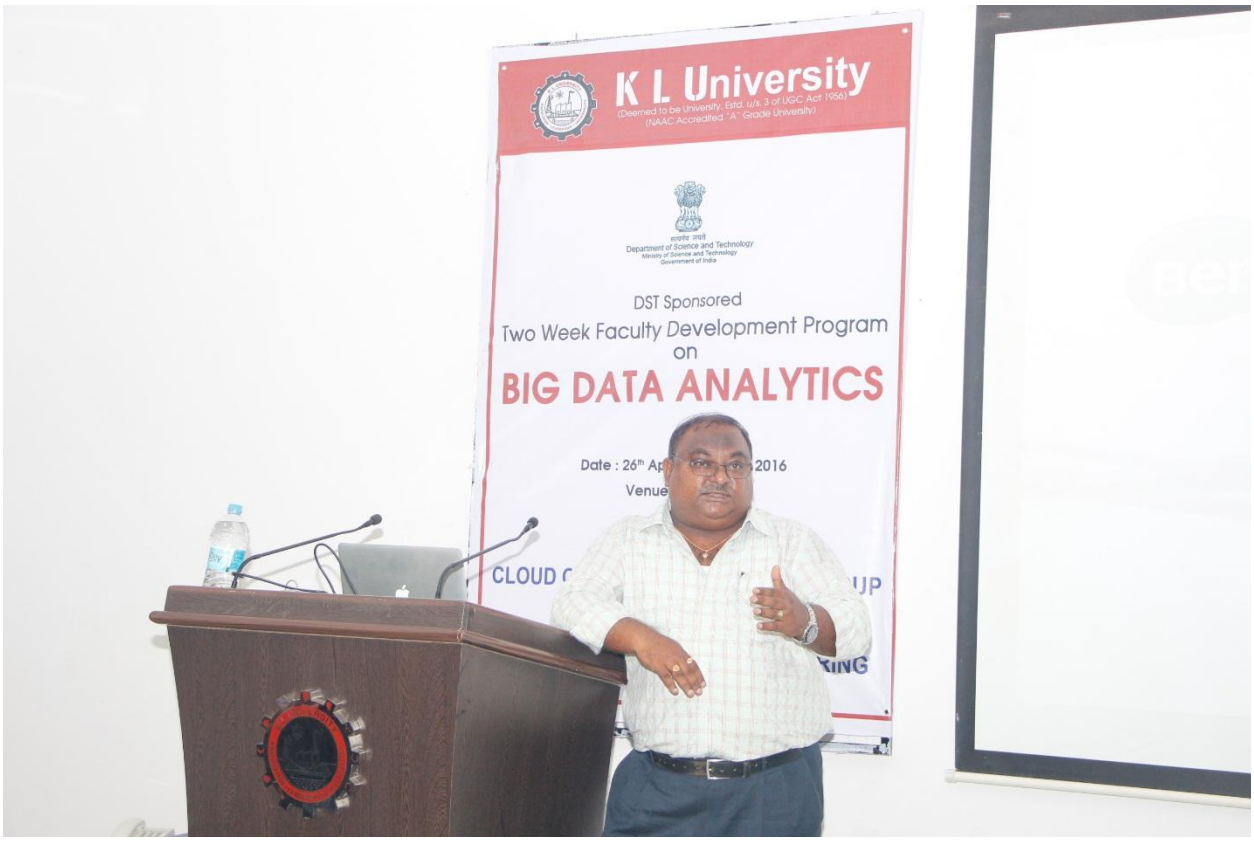
Afternoon session: R Language Plots, variants, testing techniques

Resource Person: Mr. Raja Prabhakar Khaila

After one and half day sessions on Bigdata and Mining techniques by Dr. DV Ramana and Dr MR Narasinga Rao, the afternoon session is on R by Mr. Khaila

On day 1 everybody installed R and R studio and done basic operations on R.More techniques are explained to the participants and they are practiced there itself by the participants.

Plots, Univariate, Bivariate data, Hypothesis testing, Chi square test, T test techniques are explained in detail.



Mr. Khaila discussed various testing techniques in R

Day-4 (29-4-2016)

Morning session: A session on NoSQL and column Row Databases

Resource person: Dr. Somayajulu, BoS chaiman, NIT Warangal

Dr. J V Somayajulu, known for his expertise in Databases, Mining, Warehouses and BigData technologies is visited KLU to an address on NoSQL databases.

With a help of a video he explained the evolution from SQL to NoSQL. Various categories of NoSQL databases are addressed. Row Databases and Column Databases are discussed. Google information storing patterns are also discussed in detail.

It is well demonstrated with a help of video to all the participants. It ends up with few queries from the audience.

Dr. J Somayajulu delivering a lecture on NoSQL databases

Day-4 (29-4-2016)

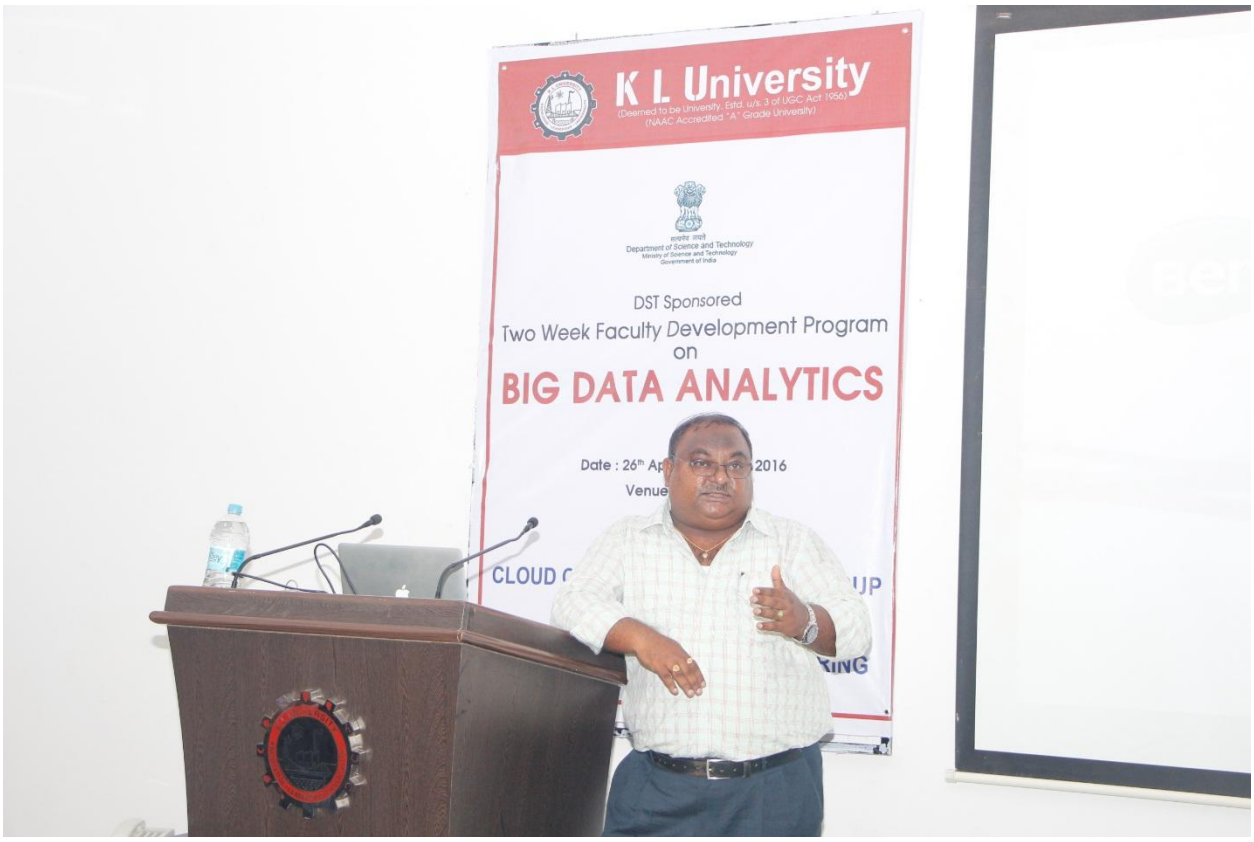
Afternoon session: A session on Excel operations in R

Resource person: Mr. Raja Prabhakar Khaila, Times Pro

After having a fruitful session on NoSQL databases in the forenoon session the afternoon session is on R by Mr. Khaila

In the previous session by Khaila Plots, Univariate, Bi variate data, Hypothesis testing, Chi square test, T test techniques are discussed.

Various excel operations, Importing and exporting datasets and Scatter plot, Box plot and histograms are discussed



Mr. Khaila Explaining Set operations plot and Histograms

Day 5: 30-4-16

Morning session: R Studio, Packages in R, Regression analysis, Linear models

Resource: Raja Prabhakar Khaila, Times Pro

Day 5 fully focused on advanced topics in R Language.

There are so many built in packages in R language for different purposes. Mr. Khaila explained how to make use of these packages and package dependencies and Regression analysis and linear models using R language

It's a full hand on session. Participants practiced Regression analysis and Linear models for knowledge discovery

Day-5: 30-4-2016

Afternoon session: A session on R language – Anova, Data prediction, tree pruning, Plots

Resource: Raja Prabhakar Khaila, Times Pro

Morning session focuses on Regression analysis and linear models. The post lunch session focused on Anova, Data prediction techniques and tree pruning techniques for data discovery. Various plot operations are discussed.

The advanced features in R language are practiced by the audience in Hall itself and the queries asked by the audience addressed by the resource person.

Audience practicing Data prediction techniques and Linear models in R

Day 6: 02-05-2016

Morning session: A session on Big Memory in Hadoop.

Resource person: Ramesh Ragala, VIT Chennai

R language has been practiced for 1 1/2 days continuously. On Day 6 an eminent researcher and expert in Hadoop and Big data techniques, Mr. Ramesh Ragala, Assistant professor, VIT Chennai addressed the gathering.

He addressed big memory in Hadoop in the morning session and discussed various operations in it. The environment gone very live when he related R language with Hadoop.

Day 6: 02-05-2016

Afternoon session: A session on High performance parallel computing with R

Resource person: Ramesh Ragala, VIT Chennai

Ramesh ragala in the morning explained using R with Hadoop. In the afternoon session Achieving high performance through parallel computing with R and Parallel computing Databases are discussed by the resource person.

Mr. Ramesh Ragala Presenting a lecture on R with Hadoop

Day 7: 03-05-2016

Morning session: A session on Big data analytics, Introduction and its importance

Resource person: Dr.R.B.V Subramanyam, Associate Dean Academics, NIT Warangal

In the morning session **Dr. R. B.V** Subramanyam explained about examples of Map reducing which enables all the researchers and the participants to understand Map and Reduce functions.

The elaboration by the Dr. Subramanyam is enlightened all the participants very much.

Dr. R.B.V Subramanyam addressing the audience about Matrix Multiplication in Map Reduce

Day 7: 03-05-2016

Afternoon session: A session on Big data Prediction and analytics

Resource person: Dr. R. B.V Subramanyam, Associate Dean Academics, NIT Warangal

Dr. Subramanyam outlined in detail about the Map and reduce functions. In the afternoon session he has detailed about concept of Big Data Prediction and Analytics.

The full day session found to be very cheerful to the audience

Dr. R.B.V Subramanyam interacting with the audience and clearing the doubts on Map Reduce

Day 8: 04-05-2016

Morning session: A session on Big Data Mining, Clustering

Resource person: Dr. D. Aruna Kumari, Professor ECMDepartment, KLUniversity

In the morning session **Dr. D. Aruna Kumari, Professor, ECM department detailed**about application of frequent item set mining in Big Data

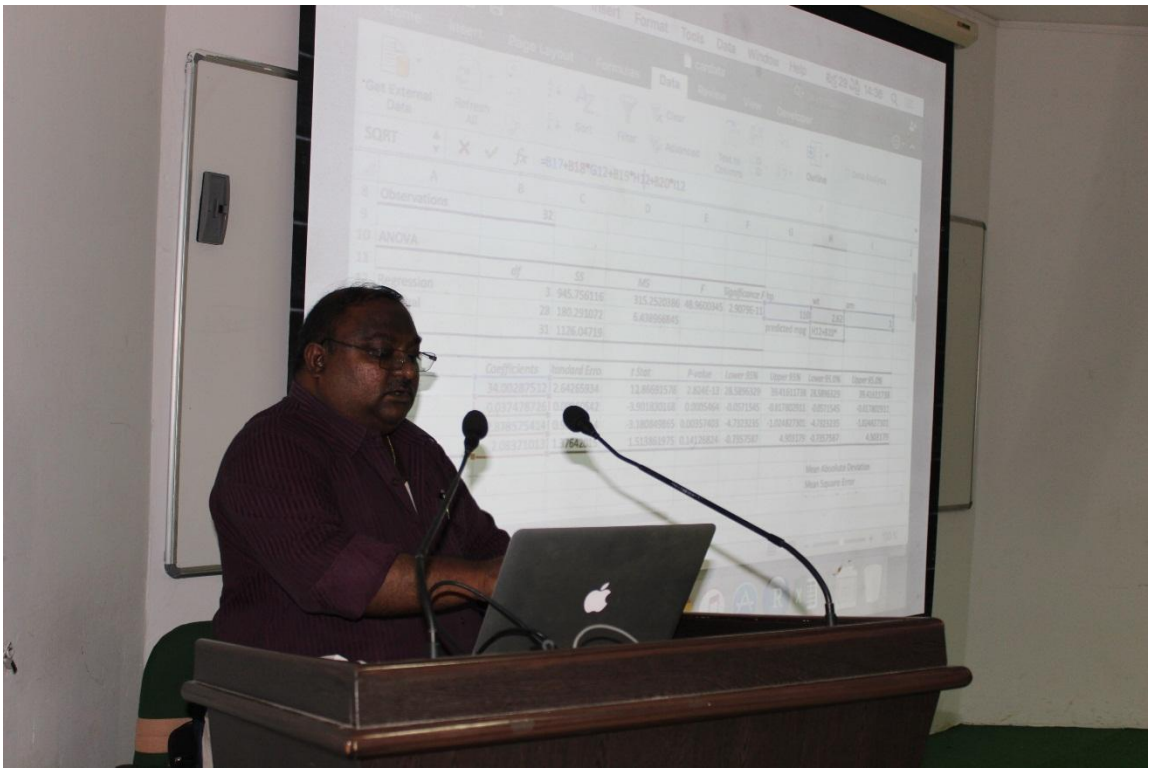
Dr. KVV Satyanarayana Cloud Computing Research Group Head and Dr. K. Thirupathi Rao, Convener felicitating Dr. A Aruna Kumar

Day 8: 04-05-2016

Afternoon session: A session on OLTP and OLAP

Resource person: Dr. P Sai Kiran, Professor, Department of CSE

Dr. P. Sai Kiran explained the audience about concept of OLTP, OLAP and RTAP Introduction. Post his**session a practical session on Apriori Algorithm is done in supervision of Mr. Raja Prabhakar Khaila.** The advanced features in Apriori algorithm were practiced by the audience in Hall itself and the queries asked by the audience addressed by the resource person



Audience practicing Apriori algorithms Techniques

Day-9: 05-5-2016

Afternoon session : A session on text mining -Introduction

Resource person : Raja Prabhakar Khaila, Times Pro

Morning session focuses on text mining and its introduction concepts . A Practical session on Analysis of text miningwere practiced by the audience in Hall itself and the queries asked by the audience addressed by the resource person.

Afternoon session: A session on Cluster Analysis

Resource: Raja Prabhakar Khaila, Times Pro

Afternoon session focuses on concepts of Cluster Analysis and also about concepts that are related to K-Means. Mr. Khaila outlined in detail about the concepts of Hierarchical clustering.

Audience practicing hierarchical clustering programs

Day-10: 06-5-2016

Morning session: A session on Hadoop Commands

Resource: Ramesh Ragala, VIT University

The Morning session focuses on Hadoop Commands. A practical session on Hadoop Installation and commands are practiced by the audience in Hall itself and the queries asked by the audience addressed by the resource person.

Afternoon session: A session on Program using Hadoop, Map reduce classes

Afternoon session focuses on concepts of programming using Hadoop and also about Map reducing classes. Mr. Ramesh has given questions about Hadoop Classes and asked audience to practice.This short sessions is followed by the felicitation.

Mr. Ramesh Ragala delivered enlightened speech on Mapper and reducer

Felicitation Ceremony:

It has been a 10 days long journey in the FDP program, but the participants felt that it's a very short duration.

Dr. K. Rama Krishna, Registrar is the chief guest for the felicitation ceremony. Dr. A Anand Kumar, Principal KLUCE, Dr. V. Srikanth, HOD CSE, Convener of the FDP Dr. K. Thirupathi Rao graced the dais.

The guests and participants congratulated the Organizer of the workshop Dr. K. Thirupathi Rao, HOD CSE Dr. V Srikanth and Research group head Dr. K.V.V. Satyanarayana and other organizers for making this FDP program a grand success.

The participants from outstations expressed their experience at KLU.

Everybody thanked that the hospitality is wonderful and the Session management is excellent and prominent speakers made this FDP very effective.

Participants of the FDP

All together it is a wonderful journey of two weeks which enlighten all the participants with full of knowledge.

Live practice sessions make this FDP program very handy. The participants expressed their wish to visit KLU for such future programs.

The registrar and the Principal congratulated all the participants organizers of the FDP especially the Cloud computing research group of CSE Department.

The End