

# **A Report on Faculty Development Programme on Computational Approach in Big Data Analytics Applications**

(16th Dec 2019 to 21st Dec 2019)

Organised by

**Department of Information Technology**

**Vasavi College of Engineering (Autonomous), Hyderabad**

in association with NIT, Warangal.

Under the scheme of Electronics & ICT Academy

(Sponsored by Ministry of Electronics and Information Technology (MeitY), GOI)

## **Objective of the Faculty Development Programme:**

The programme is designed to enhance the learning and research skills of the participants, to make them aware about modern analytical tools and methodologies. It provides an opportunity to acquire knowledge about current technological developments in relevant fields. The programme mainly focuses on Big Data Analytics and the different computational approaches with primary concern and motivation for application development.

## **Outcomes of the Faculty Development Programme:**

1. The participants would gain experience of how to work with data science, graph and social media analytics, as they may work with case studies related to them.
2. To gain mere knowledge between structured and unstructured databases.
3. To implement Map, Reduce Programming, R programming, unstructured data processing.
4. To acquire practical knowledge on different database tools like Spark, MongoDB, Cassandra, NoSQL Databases.

## **Contents of the programme:**

1. Introduction to Big Data Analytics
  - Big Data Analytics Life Cycle
  - Graph and Social Media Analytics
2. Overview of Hadoop
  - Map Reduce Programming
  - Data and Models Visualization
3. Data Processing using R
  - Unstructured Data Processing
  - Text Processing
  - Case Studies on Data Science
4. SPARK
  - Spark Features
  - Spark Components
  - Real time Analytics
5. NoSQL Databases.

## Day 1 (16-12-2019)- INAUGURAL SESSION

FDP Session began with inauguration at 10:00 A.M., venue was Project Lab in R-Block, Department of Information Technology, Vasavi College of Engineering.

Dignitaries of the program were Dr. S.V. Ramana, Principal, Vasavi College of Engineering, Prof. K. Ram Mohan Rao, Head, Department of Information Technology and the honourable chief guest Dr. P. Radha Krishna, Professor, Department of CSE, NIT Warangal has graced the inaugural session by lighting of the lamp.

Mrs. S. Aruna, Associate Professor, Department of IT, VCE addressed the gathering by welcoming guests and all the participants who came from various places.

Prof. K. Ram Mohan Rao, Head, Department of Information Technology, VCE, enlightened the audience by relating current trends of Big Data Analytics and its purposes in terms of real time applications. He also proposed the objective of this FDP program in concern to thrust areas of research as well as in relation with academics.

Principal Dr. S.V. Ramana, VCE congratulated all the participants and the organising team, Department of IT good luck.

Chief Guest of the day Dr. P. Radha Krishna, Professor, Dept. of CSE, NITW gave a primary motivation to all the participants regarding the FDP, role of EICT academy, encouraging recent trends in order to fulfil research-oriented industry-academics.

Vote of thanks of the inaugural was proposed by Mrs. B. Leelavathy, Asst. Prof, FDP Coordinator.



Prof. K. Ram Mohan Rao, HOD-IT, proposing objective of the FDP.

## Day 1 (16-12-2019 - MONDAY)

### Agenda for the day:

1. Introduction to Big Data
2. Big Data Architecture & Big Data Technologies
3. Big Data Analytics Life Cycle,
4. Research Trends in the Big Data Analytics

### Resource Person:

Dr. P. Radha Krishna is an eminent person working as a Professor in Department of Computer Science at NITW. His areas of research are Data Mining, Data Science, Databases and Workflow Systems.

### Session 1 (11:00 am – 1:15 pm)

First day the session was conducted by Prof P. Radha Krishna, who delivered a lecture on introduction to big data and its evolution. He also focused on different sources of data, how to work with data, what can be the future challenges, data fusion, Big Data Architecture and technologies in use.



Prof P. Radha Krishna delivering the lecture.

### Session 2 (2:00 pm - 4:00 pm)

Post Lunch, session began with Prof P. Radha Krishna in continuation to the morning session, he discussed about four Vs i.e., velocity, volume, variety and veracity with appropriate examples for each, also addressed three more Vs value, viscosity and virality. Later listed 16Vs for big data, also discussed about Google-BigTable, Facebook -Cassandra, Oracle-Era etc.

### Session 3 (4:15pm- 5:15pm)

This session began by Prof. P. Radha Krishna by delivering a lecture on big data analytics specifically spatial big data. Few real time big data applications like US health care, Europe public sector administration, Global personal location data, US retail and manufacturing, current technological innovations and scope of research trends in the market were discussed.



## Day 2 (17-12-2019 - TUESDAY)

### Agenda for the day:

1. Big Data Batch Processing Technologies/Methods
2. Hadoop-Map Reduce Programming
3. Big Data Batch Processing Algorithms
4. Graph Data Analysis

### Resource Person:

Dr. R.B.V. Subramanyam, Professor, Head of the Department, Department of Computer Science at NITW. His areas of research are Data Mining - Association Rule Mining, Temporal Data Mining, Sequential Pattern Mining, Distributed Data Mining, Graph Databases, Fuzzy Data Mining.

### Session 1 (9:30 am – 11:00am)

Forenoon session began by Prof Subramanyam, who delivered a lecture on distributed systems and their design issues, batch processing technologies for big data and different methods that can be applied on big data to perform batch processing.

### Session 2 (11:15 am – 1:15 pm)

Session focused on how Hadoop takes call of distribution of data that is load balancing aspects, different algorithms in use. Hadoop-Map Reduce Programming with a case study in practice to map program to reduce program using local and global knowledge aspects how to map was explained.



### Session 3 (2:00 pm - 4:00 pm)

Post Lunch, session began with in continuation to the morning session by Prof. Subramanyam, he discussed about Hadoop- mapper, reducer, matrix multiplication on the case study as well a practice problem was also given on the same exercise to sort the names of the available dataset in dictionary format. Participants were enthusiastic in solving the given exercise.

### Session 4 (4:15pm- 5:15pm)

This session began by delivering a lecture on few batch processing algorithms, graph data analytics with real time examples and applications that are in use.

Session was concluded by felicitating the guest of honour Prof. R.B.V. Subramanyam with a memento by Prof. Venkat Krishna, Dept of IT,VCE as a token of gratitude.

## Day 3 (18-12-2019 - WEDNESDAY)

### Agenda for the day:

1. Data Visualization: Hands-on sessions
2. Visualization Analytics: Hands-on sessions
3. Data Analytics using R: Hands-on sessions

### Resource Person: (Coordinator of this FDP)

Dr. T. Ramakrishnudu currently working as an Assistant Professor in Department of Computer Science at NITW. His areas of research are Data Mining-Association rule mining, Distributed data mining., Big Data Analytics, Web Mining, Text Mining, Urban Computing, Social Media Analysis, Database Systems, Web Technologies.

### Session 1 (9:30 am – 11:00am)

Forenoon session began by introduction to data visualization using R. It was a hands-on session for which participants worked on R-tool. The basics of the R CRAN, R-GUI & RStudio were explained.

### Session 2 (11:15 am – 1:15 pm)

In continuation to the forenoon session, Dr. T. Ramakrishndu, a hands-on session was conducted on Unstructured Data Analytics, Text processing and Analytics like vector representation, word to vector, stop words, stemming, parsing and tokenizing, how can regular expression used to validate were demonstrated.

### Session 3 (2:00 pm - 4:00 pm)

Post Lunch, session began in continuation to forenoon session by mainly focusing on bag of words, document frequency, inverse document frequency, Image analytics were demonstrated. Participants were deeply involved in the hands-on practice session.

### Session 4 (4:15pm- 5:15pm)

This session began Dr. Ramakrishnudu, by focusing on Image analytics were demonstrated. Participants were deeply involved in the hands-on practice session.



Session was concluded by felicitating the guest of honour Dr. T. Ramakrishnudu with a memento by Prof. Venkat Krishna, Dept of IT as a token of gratitude followed by a group photograph.

## Day 4 (19-12-2019 - THURSDAY)

### Agenda for the day:

1. Text Mining
2. Social Media Analytics

### Resource Person:

Dr. P. Radha Krishna is currently working as a Professor in Department of Computer Science at NITW. His areas of research are Data Mining, Data Science, Databases and Workflow Systems.

### Session 1 (9:30 am – 11:00 am)

Forenoon session was conducted by Prof P. Radha Krishna, who delivered a lecture on structured and unstructured data, statistical analysis, different pre-processing techniques for information retrieval, content analysis.

### Session 2 (11:15pm – 1:15 pm)

Session began with a lecture by Prof. P. Radha Krishna on challenges of text mining, various insights to be focused in context to text mining.

### Session 3 (2:00 pm - 4:00 pm)

Post Lunch, session began with Prof P. Radha Krishna in continuation to the morning session, he discussed about social media analytics related topics like social network, graph analytics, social network features.

### Session 4 (4:15pm- 5:15pm)

This session began by Prof. P. Radha Krishna by focusing on different applications of social network analytics like community clustering, spectral clustering, etc., were discussed with examples.



Session was concluded by felicitating Prof. P. Radha Krishna with a token of gratitude by Prof Venkat Krishna, department of IT, VCE.



## Day 5 (20-12-2019 - FRIDAY)

### Agenda for the day:

1. Big Data Storage Technologies
2. NoSQL Databases: Hands-on sessions
3. MongoDB: Hands-on sessions

### Resource Person:

Mr. Vivek Unnikrishnan is Working at JP Morgan Chase as a Software Analyst. He is currently working with 7+ projects using current real time technologies like AI, MongoDB, Cassandra, NLP techniques etc.



### Session 1 (9:30 am – 11:00 am)

Mr. Vivek has delivered a lecture on introduction to big data storage technologies, architecture of big data, pros and cons of big data storage.

### Session 2 (11:15pm – 1:15 pm)

In continuation to the forenoon session Mr. Vivek gave a hands-on demonstration on NOSQL database. Installation of software's like MongoDB, Cassandra were also demonstrated to the participants.

### Session 3 (2:00 pm - 4:00 pm) & Session 4 (4:15pm- 5:15pm)

Post Lunch, session was a hands-on session where Mr. Vivek explained how to deal with bigdata using MongoDB and Cassandra. It was an interactive session along with hand on practise by the participants. Session ended with felicitation to Mr. Vivek Unnikrishnan a memento as a token of appreciation from department of IT, VCE.

## Day 6 (21-12-2019 - SATURDAY)

### Agenda for the day:

1. Stream data Processing: Hands-on sessions
2. Spark: Installation, setup, Architecture: Hands-on sessions
3. Big Data Analytics with Spark: Hands-on sessions
4. Real time Analytics: Hands-on sessions
5. Valedictory Function

### Resource Person:

Mr. B. Sachin Govind is Working as Tech Lead, Data Engineering at the start up yellow messenger, having 4 years of experience in Data Engineering.

### Session 1 (9:30 am – 11:00 am)& Session 2 (11:15pm – 1:15 pm)

Forenoon session of the last day of the fdp began with introduction to stream data processing followed by spark installation, setup, architecture of YARN, HADOOP, also about google Collab.

### Session 3 (2:00 pm - 4:00 pm)

Post Lunch, session was a hands-on session where Mr. Sachin explained how to deal with bigdata analytics using spark. Some real time analytics use case scenarios were also demonstrated by Mr. Sachin to the participants. It was an excellent hand-on session experienced by participants.

### Session 4 (4:15pm- 5:15pm)

The Programme was concluded by Valedictory function, the dignitaries who graced the occasion were Dr. S.V. Ramana, Principal, Vasavi College of Engineering, Prof. K. Ram Mohan Rao, Head, Department of Information Technology, VCE and the chief guest Mr. B. Sachin Govind, Tech Lead, Yellow Messenger.

Program was initiated by Prof. K. Ram Mohan Rao, who addressed the gathering by congratulating all the participants for attending the workshop. Later text books distribution to the participants was done by Dr. S.V. Ramana, followed by the feed back of the participants regarding the 6-day Faculty Development Programme. Finally vote of thanks was proposed by FDP coordinator Ms. B.Leelavathy, to all the dignitaries, participants, supporting teaching and non-teaching staff of Department of IT, VCE.



Session was concluded with felicitating Mr. Sachin Govind with a memento as token of appreciation followed by a group photograph with all the participants.