Zeal Education Society's

Zeal Institute of Business Administration, Computer Application & Research

Sr. No. 39, Narhe, Pune -411041, Phone No.:020-67206031, Website: www.zibacar.in

(Approved by A.I.C.T.E., Recognized by DTE (Govt. of Maharashtra) and

Affiliated to Savitribai Phule Pune University)

REPORT

NATIONAL WINTER TRAINING PROGRAM

ON

BIG DATA AND HADOOP

In Association with

IIT ROORKEE

Date: 17-24 December 2015

Time: 10:00 am - 5:00 pm

Report Prepared By:

Prof. Madhavi Shamkuwar, ZIBACAR

INDEX

ANNEXURE

1. EVENT DETAILS

Event Type	: Inter-College Faculty Development Programme
Description	: Workshop
Title	: **National Winter Training Program** on **Big Data and Hadoop**
In Association with	: IIT Roorkee
Speaker	: **Mr. Pankaj Singh** Sr. Research Engineer, Finland LabsTM
Venue	: **Computer Centre** Central Placement Cell Building, Zeal Education Society, Narhe, Pune-41
Date & Duration	: 17/12/2015- 24/12/15 (1Week workshop) 11:00am - 4:00pm
Reference	: Dr. Neha Sharma

2. CONCEPTION OF THE PROGRAMME

The development of Technical Quotient of the Faculty members of the institute is the need of the hour so as to have an edge over the competitors. For that reason, it is necessary to equip the faculty members on the latest technologies, so that they become capable of imparting knowledge on latest technologies to the students and can also use these technologies for research and problem solving. Such, a techno-driven environment when created will eventually contribute to the Technical development of the society. This was the start point of the idea to conduct a one week session on Big Data and Hadoop as its market is growing sharply.

(Annexure A- Workshop Proposal)

3. WHY BIGDATA & HADOOP TECHNOLOGY?

Importance of upgrading the knowledge:

Upgrading knowledge is not only an individual's requirement but also, requirement of an organization as it contributes to our Society's knowledge base. Such sessions enhance the quality of knowledge; provide employment opportunities and leads towards the professional excellence.

BigData:

Due to advancement in technology, lower costs of storage devices, extensive use of social networking sites, an enormous amount of data is generated. The amount is so large that even if it is just piled it will consume large amount of space. Traditional computing techniques are inefficient to process such a large amount of data; hence Big Data came into picture. It is a recent technology, which involves various tools, techniques and frameworks. Big data captures huge amount of data produced by different devices and applications and is characterized by 3V's i.e. Volume, Velocity and Variety. BigData is an infrastructure which provides more accurate analysis of huge volumes of structured and unstructured data in real-time and can protect data privacy and security.

Hadoop Technology:

Hadoop is a free, Java-based programming framework that supports the processing of large data sets in a distributed computing environment. Hadoop makes it possible to run applications on systems with thousands of nodes involving thousands of terabytes. Its distributed file system facilitates rapid data transfer rates among nodes and allows the system to continue operation uninterrupted in case of a node failure. This approach lowers the risk of catastrophic system failure, even if a significant number of nodes become inoperative.

4. OBJECTIVES

a) To implement a Hadoop Project
b) To learn to write Complex MapReduce programs
c) To perform Data Analytics using Pig and Hive
d) To understand Data Loading Techniques using SQOOP and Flume
e) To master the concepts of Hadoop Distributed File System and MapReduce Framework
f) To work on a Real Life Project on Big Data Analytics and gain Hands on Project Experience

5. RESOURCE PERSON

Mr. Pankaj Kumar Singh is Sr. Research and Development Engineer in Finland LabsTM. He is a Data Analyst and has got sound knowledge on Big Data, Cloud, Hadoop C Programming. On behalf of Finland LabsTM and Revert Technology Pvt. Ltd., he delivers technical corporate training to different organization as well as delivers seminars, training and workshop in association with "IIT-Roorkee" in India at different platforms. He has delivered more than 60 workshops on different types of advance robots (like gesture controlled) and more than 30 workshops on "Big – Data (Hadoop).

***(Annexure B:** Profile of Mr. Pankaj Kumar Singh)*

Snapshot 1: Mr. Pankaj Singh

6. PREPARATION OF THE EVENT

Preparation of Faculty Development Programme had started immediately after identifying the inclination of faculties towards recent technologies. The approval for the FDP was sought almost two weeks in advance. The circular was circulated among the faculty members along with an E-Invitation to various institutes for their information regarding the FDP. The Auditorium and Computer centre was booked well in advance.

***(Annexure C**- Approval Note, Dated 03/12/2015)*

***(Annexure D** – e-Circular, Dated 03/12/2015)*

7. ITERNARY

Sr. No.	Timing	Activity
Day 1: 17/12/2015		
1	10.00 am to 10.05 am	Introduction of Resource Person- Mr. Pankaj Singh
2	10.05 am to 10.10 am	Welcome Address by Dr. Neha Sharma, Director
3	10.05 am to 10.10 am	Address by Prof. Jayesh Katkar, Ex-Director, ZES
4	10.05 am to 11.15 am	Session on Different Data Types
5	11.15 am to 11.30 am	**Tea Break**
6	11.30 am to 1.00 pm	Session on Introduction to Big data and Hadoop
7	1.00 pm to 2.00 pm	**Lunch Break**
8	2.00 pm to 3.30 pm	Session on Introduction to Hadoop
9	3.45 pm to 4.45 pm	Session on Applications to Big data and Hadoop
Day 2: 18/12/2015		
1	10.00 am to 01.00 pm	Technical session on VMWare
2	01.00 am to 02.00 pm	**Lunch Break**
3	2.00 pm to 04.00 pm	Technical session on Hadoop Architecture
Day 3: 19/12/2015		
1	10.00 am to 01.00 pm	Technical session on MapReduce: Data Flow Language
2	01.00 am to 02.00 pm	**Lunch Break**
3	2.00 pm to 04.00 pm	Technical session on Pig: Programming Model
Day 4: 21/12/2015		
1	10.00 am to 01.00 pm	Technical session on HIVE and its architecture
2	01.00 am to 02.00 pm	**Lunch Break**
3	2.00 pm to 04.00 pm	Technical session on Apache Hive data warehouse software
Day 5: 22/12/2015		
1	10.00 am to 01.00 pm	Technical session on HIVE and its architecture
2	01.00 am to 02.00 pm	**Lunch Break**
3	2.00 pm to 04.00 pm	Technical session on Apache Hive data warehouse software
Day 6: 23/12/2015		
1	10.00 am to 01.00 pm	Technical session on HBase and its architecture
2	01.00 am to 02.00 pm	**Lunch Break**
3	2.00 pm to 04.00 pm	Technical session on integration of HBase and MapReduce
Day 7: 24/12/2015		
1	10.00 am to 01.00 pm	Technical session on Sqoop and its architecture
2	01.00 am to 02.00 pm	**Lunch Break**
3	2.00 pm to 04.00 pm	Technical session on ZooKeeper
4	04.00 pm to 04.15 pm	Vote of Thanks by Prof. Archana Vechalekar

8. EXECUTIVE SUMMARY

All participants from various colleges across state and country arrived at Chhatrapati Shivaji Maharaj Auditorium by 9.30 am and the programme begun at 10.00 am as per schedule. One week National Winter Training Programme was formaly inaugurated in the presence of Prof. Jayesh Katkar, Executive Director. Dr. Neha Sharma Director, ZIBACAR gave the opening remarks for the session. She emphasized that the technical up-gradation is must to keep abreast with the Industry and Global World. Prof. Jayesh Katkar, Executive Director suggested to revolutionize the education process by imparting techno-friendly environment accommodating the needs, interest, current knowledge and learning styles of the E-learners. Prof. Archana Vechalekar, Asst. Prof., ZIBACAR introduced the resource person for the National Workshop, Mr. Pankaj Singh to the attendees.

Brief Information about the Workshop:

DAY-1: First session covered various types of data, introduction to Big Data and different components of Hadoop. Furthermore, application of BigData like Facebook, Amazon were discussed.

DAY-2: It focused on various VMware Player, installation and Horton Works Sand Box Introduction and Hadoop Architecture.

DAY-3: The session was about giving practical exposure using MapReduce and Pig with the help of various examples. PIG is a data flow language, its key focus is to manage the flow of data from input source to output store. **MapReduce** is a programming model and an associated implementation for processing and generating large data sets with a parallel, distributed algorithm on a cluster. Further the trainer elaborated pig components, data model and execution modes.

DAY-4: The session provided a brief insight about differentiation between MapReduce and Pig. The data types, Modes of Execution, Execution Mechanism, Pig Commands etc were demonstrated with the help of live examples.

DAY-5: The session gave theoretical and practical explanation of HIVE along with its architecture, Query complier, Server, Hive QL and its Metastore. The Apache Hive data

warehouse software facilitates querying and managing large datasets residing in distributed storage. The session further explained Hbase characteristics, architecture, commands.

DAY-6: To have a detailed study of HBase and its architecture with its integration with MapReduce.

DAY-7 : The last session covered latest technology known as Sqoop, which is a tool designed for efficiently transferring bulk data between Apache Hadoop and structured datastores such as relational databases. The session discussed about ZooKeeper which is a centralized service for maintaining configuration information, naming, providing distributed synchronization, and providing group services.

9. PARTICIPANTS

Total 40 participants from various colleges across state and country attended the National Winter Training Programme.

External participants: 30

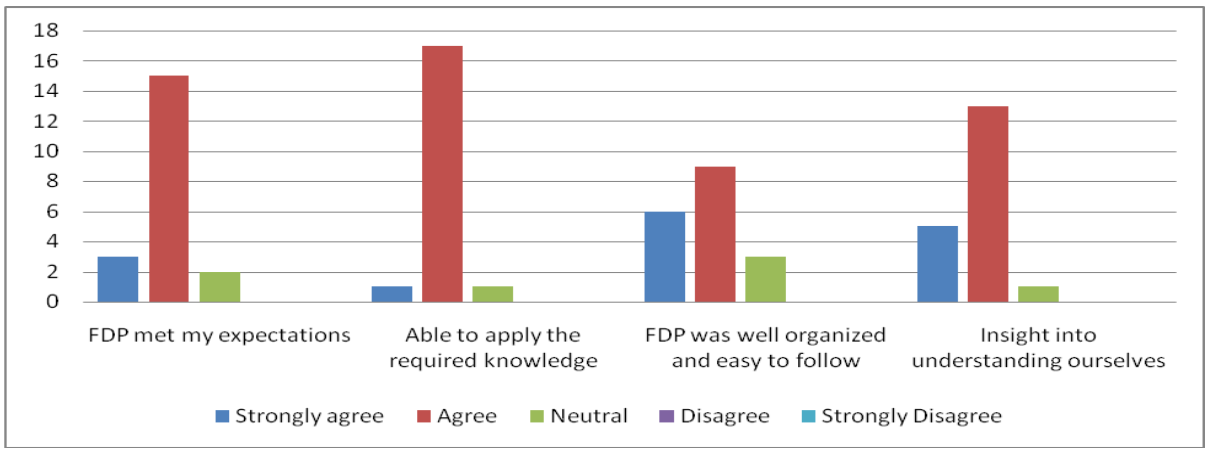
Internal Participants: 10

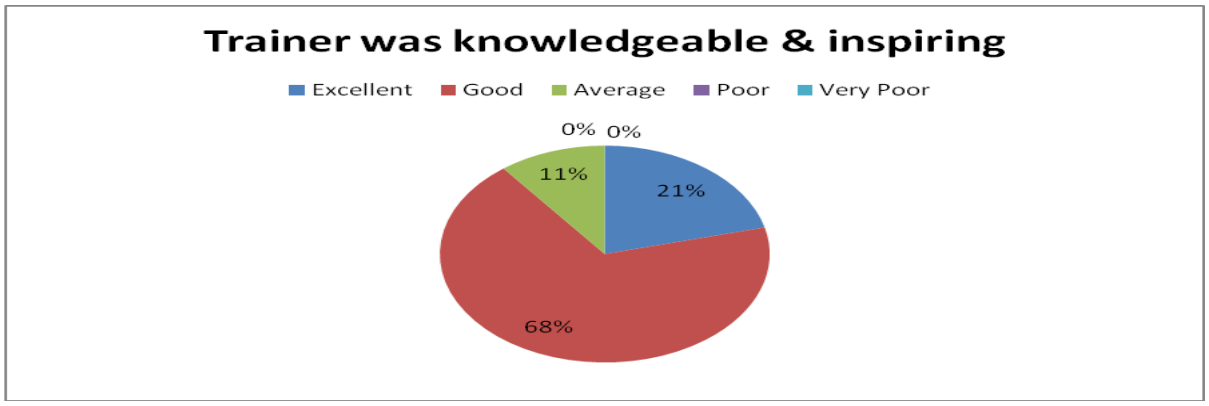
(Annexure E- Attendance Sheet)

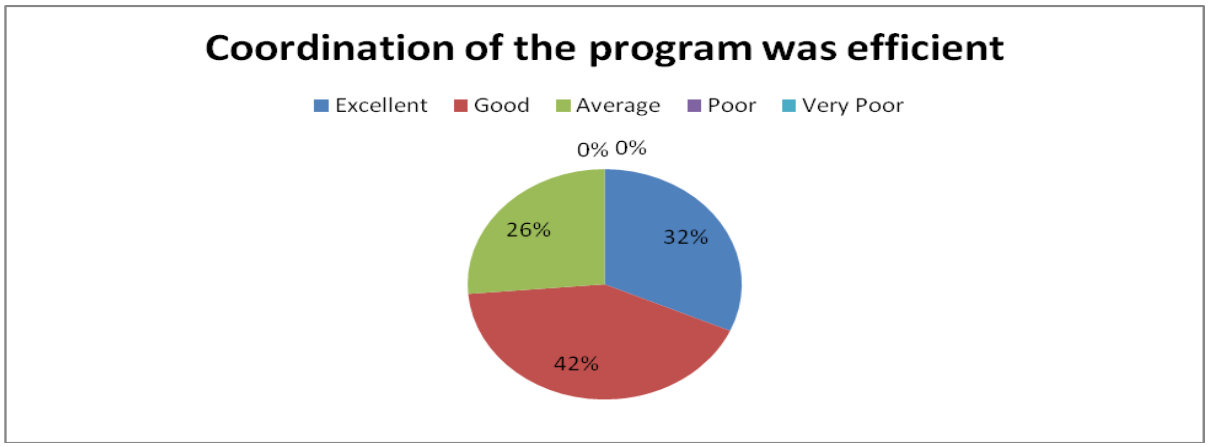
10. FEEDBACK OF THE WORKSHOP

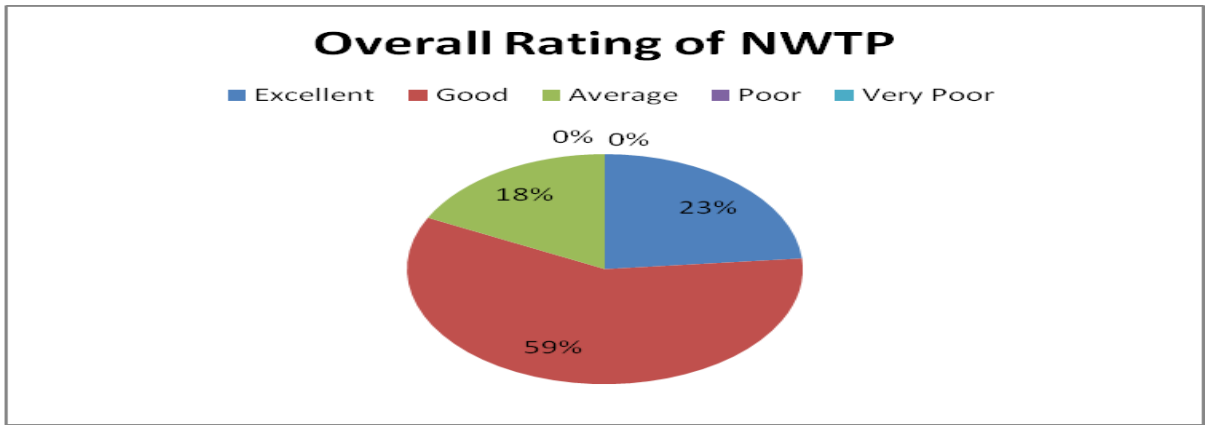
To confirm the impact of the workshop, a feedback session was organized at the end of session. The feedback was gathered with the help of exclusive form designed for the same purpose.

Parameters	**Strongly agree**	**Agree**	**Neutral**	**Disagree**	**Strongly Disagree**
FDP met my expectations	3	15	2	0	0
Able to apply the required knowledge	1	17	1	0	0
FDP was well organized and easy to follow	6	9	3	0	0
Insight into understanding ourselves	5	13	1	0	0


18
16
14
12
10
8
6
4
2
0
FDP met my expectations
Able to apply the required knowledge
FDP was well organized and easy to follow
Insight into understanding ourselves
Strongly agree
Agree
Neutral
Disagree
Strongly Disagree


Trainer was knowledgeable & inspiring
Excellent
Good
Average
Poor
Very Poor
0%
0%
11%
21%
68%


Coordination of the program was efficient
Excellent
Good
Average
Poor
Very Poor
0%
0%
26%
32%
42%


Overall Rating of NWTP
Excellent
Good
Average
Poor
Very Poor
0%
0%
18%
23%
59%

Interpretation: The feedback analysis illustrates that majority of the participants have had enriching experiences at the Workshop on "Big Data and Hadoop".

Annexure F- (*Duly Filled Feedback Forms)*

11. ECONOMIC IMPACT

The registration fee for the Workshop was Rs. 3000/- per participant from Outside and Rs. 1000/- for in-house faculty members. The fee included training, certification, and Event registration at IIT Roorkee, a free Big Data & Hadoop Kit, Lunch and Snacks to each participant. The in-house faculty registration fee was sponsored by the Institute.

Annexure G- (*Registration form)*

12. CONTRIBUTION OF THE MANAGEMENT

Management was very supportive during the planning and overall conduct of the Workshop on "Big Data and Hadoop".

13. IMPACT OF THE WORKSHOP

Participants realized that Big Data and Hadoop is the cornerstone of business technology agenda. This workshop helped the participants to get acquainted with a must-know technologies like Hadoop, Pig, Hive, Sqoop, MapReduce, HDFS etc and equipped them to stay ahead in the game.

14. MEDIA COVERAGE

Zeal Education Society hosting training prog for teachers

Pune: Zeal Education Society, on the occasion of completion of 20 years of its establishment, is conducting a seven-day National Winter Training Programme (NWTP) on 'Big Data and Hadoop' to hone the teaching and technical skills of faculty members of computer application and information technology.

The NWTP is being held in association with EDC-IIT Roorkee from December 17 to 24 at Zeal Education Society's Narhe campus.

The NWTP aims to provide a common platform to academia, industry practitioners and participants to take knowledge and upgrade themselves on various topics including Big Data Technology, Hadoop Architecture and Pig and Hive.

The event was inaugurated by Jayesh Katkar, Executive Director Zeal Education Society. Dr Neha Sharma, Director Zeal Institute of Business Administration, Computer Application and Research (ZIBACAR) appreciated the presence of participants. *sr*

15. PHOTOGRAPHS / CERTIFICATES

The workshop was captured well with the help of photographs

(Annexure H*- Photographs)*

(Annexure I*- Certificates)*

ANNEXURE - H

Snapshot 1: Registration of Participants for NWTP

Snapshot 2: Left to Right: Dr. Neha Sharma, Director-ZIBACAR, Prof. Jayesh Katkar, Executive-Director-ZES, Mr. Pankaj Singh, Sr. Research Engg in Finland, Dr. Anand G. Jumle

Snapshot 3: Felicitation of MR. Pankaj Singh by Prof. Jayesh Katkar

Snapshot 4: Dr. Neha Sharma, Director-ZIBACAR addressing the Participants

Snapshot 5: Prof. Jayesh Katkar, Ex-Director , ZES addressing the Participants

Snapshot 6: Mr. Pankaj Singh conducting Training session in Computer Lab

Snapshot 7: Mr. Pankaj Singh conducting Training session in Computer Lab

Snapshot 8: Mr. Pankaj Singh conducting Training session in Computer Lab

Snapshot 9: Dr. Neha Sharma, Mr. Pankaj Singh with Participants

Snapshot 10:Mr. Pankaj Singh with NWTP Participants