TEQIP Short Term Course on Big Data
August 07-11, 2017

Workshop Co-ordinator:
Dr. Veena Bansal
Department of Industrial and Management Engineering

Venue:
PBCEC Building, IIT Kanpur
TEQIP Short Term Course on Big Data
07-11 August, 2017
About

Knowledge Incubation for TEQIP, IIT Kanpur organized 5-day course on Big Data. This course aimed at introducing participants to the underlying technology, decision-making models, strategy and issues involved in a big data project.

One can become a part of Big Data project in following ways:

- Be a technical person and handle the technology part
- Work with the solution
- Lead and manage a big data initiative.

In this workshop the participants learnt about aforementioned topics in a systematic manner through lectures and some lab sessions on Hadoop, cognitive computing using Bluemix.

TOPICS DISCUSSED

- Introduction: What is big data
- Big Data: Modelling
- Tutorial: Hadoop installation and Examples
- Big Data: A Technical Perspective
- Descriptive, Prescriptive and Predictive analysis for big data
- Big Data Scenario: Energy
- Map Reduce
- DBMS-BDMS
- Scalability
- Social Media Analytics Marketing
- Algorithms/model: Classification
- MATLAB
- Data Modeling
- Data Interpretation Pitfalls
- Experiments with Big Data
SPEAKERS

1. Dr. Veena Bansal
   Workshop Organizer
   Associate Professor
   Department of Industrial and Management Engineering
   IIT Kanpur
   http://www.iitk.ac.in/ime/veena/

2. Dr. Faiz Hamid
   Assistant Professor
   Department of Industrial & Management Engineering
   IIT Kanpur
   http://iitk.ac.in/new/faiz-hamid

3. Dr. Devlina Chatterjee
   Assistant Professor
   Department of Industrial & Management Engineering
   IIT Kanpur
   https://www.iitk.ac.in/ime/devlina/index.html

4. Dr. Anoop Singh
   Associate Professor
   Department of Industrial & Management Engineering
   IIT Kanpur
   https://www.iitk.ac.in/ime/anoops/

5. Dr. Deepu Philip
   Associate Professor
   Department of Industrial & Management Engineering
   IIT Kanpur
   http://iitk.ac.in/new/deepu-philip
6. Dr. Shankar Prawesh  
   Assistant Professor  
   Department of Industrial & Management Engineering  
   IIT Kanpur  
   http://iitk.ac.in/new/shankar-prawesh

7. Dr. Medha Atre  
   Assistant Professor  
   Computer Science and Engineering  
   IIT Kanpur  
   https://www.cse.iitk.ac.in/users/atrem/

8. Dr. Purushottam Kar  
   Assistant Professor  
   Department of Computer Science & Engineering  
   IIT Kanpur  
   https://www.cse.iitk.ac.in/users/purushot/

9. Ms. Pallavi Kar  
   Application Engineer, Data Analytics  
   Mathworks  
   https://in.linkedin.com/in/pallavi-kar-2a591518
# Participating Institutes

<table>
<thead>
<tr>
<th>Institute</th>
<th>Number of Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Aligarh Muslim University</td>
<td>7</td>
</tr>
<tr>
<td>2. Birla Institute of Technology, Ranchi, Allahabad Campus</td>
<td>1</td>
</tr>
<tr>
<td>3. Guru Nanak Dev University, Amritsar</td>
<td>2</td>
</tr>
<tr>
<td>5. HBTU Kanpur</td>
<td>2</td>
</tr>
<tr>
<td>6. IFTM University, Moradabad</td>
<td>2</td>
</tr>
<tr>
<td>7. IIT Bombay</td>
<td>1</td>
</tr>
<tr>
<td>8. IIT Delhi</td>
<td>1</td>
</tr>
<tr>
<td>9. IIT Mandi</td>
<td>1</td>
</tr>
<tr>
<td>10. IIT Kharagpur</td>
<td>1</td>
</tr>
<tr>
<td>11. Institute of Engineering and Technology, Lucknow</td>
<td>1</td>
</tr>
<tr>
<td>12. JIS College of Engineering, Kalyani</td>
<td>1</td>
</tr>
<tr>
<td>13. KNIT Sultanpur</td>
<td>2</td>
</tr>
<tr>
<td>14. MJP Rohilkhand University, Bareilly</td>
<td>1</td>
</tr>
<tr>
<td>15. MMMUT Gorakhpur</td>
<td>1</td>
</tr>
<tr>
<td>16. MNNIT Allahabad</td>
<td>2</td>
</tr>
<tr>
<td>17. NIT Jalandhar</td>
<td>3</td>
</tr>
<tr>
<td>18. NIT Jamshedpur</td>
<td>2</td>
</tr>
<tr>
<td>19. Rajkiya engineering College, Kannauj</td>
<td>1</td>
</tr>
<tr>
<td>20. SDM College of Engineering &amp; Technology, Dharwad</td>
<td>4</td>
</tr>
<tr>
<td>21. UIET, M.D. University Rohtak</td>
<td>13</td>
</tr>
<tr>
<td>22. VBS Purvanchal University, Jaunpur</td>
<td>1</td>
</tr>
<tr>
<td>23. Rajasthan Technical University</td>
<td>1</td>
</tr>
<tr>
<td>24. JNTU, Hyderabad</td>
<td>1</td>
</tr>
<tr>
<td>25. NMAM institute of technology, Nitte</td>
<td>1</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>55</strong></td>
</tr>
</tbody>
</table>
STATEWISE PARTICIPATION

- UP: 46%
- Punjab: 11%
- Telangana: 4%
- Rajasthan: 4%
- Jharkhand: 4%
- Himachal Pradesh: 4%
- Delhi: 4%
- Maharashtra: 4%
- Haryana: 4%
- Karnataka: 8%
- West Bengal: 7%
# WORKSHOP SCHEDULE

## Day -1

<table>
<thead>
<tr>
<th>Date/Day</th>
<th>Slot</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monday 7 August</td>
<td>9:30-10:00</td>
<td>Registration &amp; networking</td>
</tr>
<tr>
<td></td>
<td>10:00-10:30</td>
<td>Inauguration</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Prof. C.S Upadhyay &amp; Dr. Veena Bansal</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>PBCEC</em></td>
</tr>
<tr>
<td></td>
<td>10:30 – 11:00</td>
<td>High Tea</td>
</tr>
<tr>
<td></td>
<td>11:00 – 12:45</td>
<td>Introduction to Big Data</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Dr. Veena Bansal</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>PBCEC</em></td>
</tr>
<tr>
<td></td>
<td>12:45 – 14:00</td>
<td>Lunch Break</td>
</tr>
<tr>
<td></td>
<td>14:00 – 15:30</td>
<td>Big Data: Modelling</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Dr. Veena Bansal</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>PBCEC</em></td>
</tr>
<tr>
<td></td>
<td>15:30 – 15:45</td>
<td>Tea Break</td>
</tr>
<tr>
<td></td>
<td>15:45 – 17:15</td>
<td>Tutorial: Hadoop installation and Examples</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Dr. Veena Bansal + Team</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>CC: L3</em></td>
</tr>
</tbody>
</table>

## Day -2

<table>
<thead>
<tr>
<th>Date/Day</th>
<th>Slot</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tuesday 8 August</td>
<td>9:30 – 11:00</td>
<td>Big Data: A technical perspective</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Dr. Deepu Philip</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>PBCEC</em></td>
</tr>
<tr>
<td></td>
<td>11:00 – 11:15</td>
<td>Tea Break</td>
</tr>
<tr>
<td></td>
<td>11:15 – 12:45</td>
<td>Descriptive, Prescriptive and Predictive analysis for big data</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Dr. Deepu Philip</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>PBCEC</em></td>
</tr>
<tr>
<td></td>
<td>12:45 – 14:00</td>
<td>Lunch Break</td>
</tr>
<tr>
<td></td>
<td>14:00 – 15:30</td>
<td>Your own Problem</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Dr. Veena Bansal</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>PBCEC</em></td>
</tr>
<tr>
<td></td>
<td>15:30 – 15:45</td>
<td>Tea Break</td>
</tr>
</tbody>
</table>
### Day -3

<table>
<thead>
<tr>
<th>Date/Day</th>
<th>Slot</th>
<th>Topic</th>
</tr>
</thead>
</table>
| Wednesday 9 August | 9:30 – 11:00 | Map Reduce  
*Dr. Veena Bansal*  
*PBCEC* |
|               | 11:00 – 11:15 | Tea Break                                  |
|               | 11:15 – 12:45 | DBMS-BDMS  
*Dr. Arnab Bhattacharya*  
*PBCEC* |
|               | 12:45 – 14:00 | Lunch Break                                |
|               | 14:00 – 15:30 | Lab  
*Dr. Veena Bansal & Team*  
*CC* |
|               | 15:30 – 15:45 | Tea Break                                  |
|               | 15:45 – 17:15 | Data Mining: Association Mining  
*Dr. Faiz Hamid*  
*PBCEC* |
|               | 19:30       | Workshop Dinner                            |

### Day -4

<table>
<thead>
<tr>
<th>Date/Day</th>
<th>Slot</th>
<th>Topic</th>
</tr>
</thead>
</table>
| Thursday 10 August | 9:30 – 11:00 | Lab  
*Dr. Veena Bansal & Team* |
|               | 11:00 – 11:15 | Tea Break                                  |
|               | 11:15 – 12:45 | Scalability  
*Dr. Medha Atre*  
*PBCEC* |
|               | 12:45 – 14:00 | Lunch Break                                |
|               | 14:00 – 15:30 | Social Media Analytics Marketing  
*Dr. Shankar Pravesh*  
*PBCEC* |
<p>|               | 15:30 – 15:45 | Tea Break                                  |</p>
<table>
<thead>
<tr>
<th>Date/Day</th>
<th>Slot</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Friday 11 August</td>
<td>9:30 – 11:00</td>
<td>MATLAB</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Pallavi Kar</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>PBCEC</em></td>
</tr>
<tr>
<td></td>
<td>11:00 – 11:15</td>
<td>Tea Break</td>
</tr>
<tr>
<td></td>
<td>11:15 – 12:45</td>
<td>Data Modeling</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Dr. Devlina Chatterjee</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>PBCEC</em></td>
</tr>
<tr>
<td></td>
<td>12:45 – 14:00</td>
<td>Lunch Break</td>
</tr>
<tr>
<td></td>
<td>14:00 – 15:30</td>
<td>Data Interpretation Pitfalls</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Dr. Nisheeth Srivastava</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>PBCEC</em></td>
</tr>
<tr>
<td></td>
<td>15:30 – 15:45</td>
<td>Tea Break</td>
</tr>
<tr>
<td></td>
<td>15:45 – 17:15</td>
<td>Experiments with Big Data</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Dr. Dhiraj Sati</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>PBCEC</em></td>
</tr>
<tr>
<td></td>
<td>17:15 – 17:30</td>
<td>Closing and Valedictory</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Dr. Veena Bansal</em></td>
</tr>
</tbody>
</table>
SUMMARY of FACULTY FEEDBACK

Workshop

1. Clarity of communication about workshop

2. Organization of the sessions
3. **Quality of Lectures**

- Excellent: 75%
- Good: 25%
- Ordinary: 0%

4. **Effectiveness of discussions**

- Excellent: 62%
- Good: 38%
- Ordinary: 0%
5. **Effectiveness of learning experience**

![Pie chart showing effectiveness of learning experience.](image)

- **Excellent**: 75%
- **Good**: 25%
- **Ordinary**: 0%

6. **Workshop duration**

![Pie chart showing workshop duration.](image)

- **Appropriate**: 62%
- **Short**: 38%
- **Long**: 0%
7. **Would you like to have more such sessions?**

8. **Would you like e-lectures by experts on special topics?**

9. **Suggest Specific topic that you would like additional expert lectures on**
   - Machine learning implementation modeling & problem formulation
   - Handsome session on closure and spark
   - Deep learning
   - RDBMS
   - MATLAB (Image Processing)
   - Hadoop
10. **Additional Suggestions**
   - Lab sessions must be properly planned and coordinated

**Teaching**

1. **Do you have additional support for teaching (tutors, graders, teaching, assistants, etc)?**

   ![Pie chart showing the percentage of respondents who have additional support for teaching.](chart1)
   - Yes: 33%
   - No: 67%

2. **Do you give class projects for UG classes?**

   ![Pie chart showing the percentage of respondents who give class projects for UG classes.](chart2)
   - Yes: 81%
   - No: 19%
3. Do you give class projects for PG classes?

Yes 100%

4. Do you have sufficient resources for laboratory courses?

Yes 83%

No 17%
5. *Is the library/journal/e-connection support adequate?*

- **Sufficient**: 83%
- **Inadequate**: 17%

6. *Would you like to have common (TEQIP) repository of course material?*

- **Definitely**: 87%
- **Maybe**: 13%
- **No**: 0%
7. Would you like to visit IITK to participate in and develop course material (existing or new)?

8. Would you like to participate in creation of the repository material (course file/lab Manuals/question bank etc.)?
9. How can IITK effectively help you prepare for teaching?

10. Which Subject do you teach?
- Evolutionary computing, NLP, Machine learning
- Web Technology, Data Structure
- DBMS, OS, Compiler
- Discrete Mathematics.

11. What is average student to teacher ratio in your institute?
- 30:01
- 15:01
- 25:01

12. How TEQIP can improve your teaching?
- Course materials,
- Organizing more local workshops
- Please provide lab session after lunch
- By conducting workshops and short term courses on platforms
Research

1. **Would you like to visit an IIT for a short visit/internship/post-doctoral stint, if offered (via TEQIP)?**

2. **Would you like to share/use research infrastructure at IITK, if made available?**
3. Would you like to conduct collaborative research with IITK faculty?

4. Would you like lectures by experts (Indian and International) on niche research areas/topics?
5. **Do you want special-topic conferences?**

![Pie chart showing 71% Yes, 29% Maybe, and 0% No.]

6. **How can TEQIP help improve your research?**

- Practical oriented workshops
- Please provide image processing materials
- By providing research collaboration opportunities & research resources (like infrastructure, data, projects)
SUMMARY of STUDENT FEEDBACK

Workshop

1. Clarity of communication about workshop

2. Organization of the sessions
3. **Quality of Lectures**

![Quality of Lectures Pie Chart]

4. **Effectiveness of discussions**

![Effectiveness of Discussions Pie Chart]
5. **Effectiveness of learning experience**

![Pie chart showing effectiveness of learning experience with categories: Excellent 54%, Good 46%, Ordinary 0%]

6. **Workshop duration**

![Pie chart showing workshop duration with categories: Appropriate 100%, Long 0%, Short 0%]
7. **Would you like to have more such sessions?**

- **Definitely**: 92%
- **Maybe**: 8%
- **No**: 0%

8. **Would you like e-lectures by experts on special topics?**

- **Definitely**: 96%
- **Maybe**: 4%
- **No**: 0%

9. **Suggest specific topic that you would like additional expert lectures on**

- Machine learning with big data. Some practical sessions
- Some lab bases on SPSS Matlab on pattern recognition & classification of data set
- Data analysis, Matlab, SPSS, statistical inference
- Machine learning, neural networks
- Real life application through direct software implication like Matlab
- Image recognition, pattern matching, in machine learning, data collection through different source.
- Practical life model, few data can made effective & efficient
- Big Data and analysis internet of things
- Machine learning scope of computer sciences
- Lecture on the bases of real industry or plant
- Data science, machine learning, hadoop, spark, R
- More labs, social media, big data basics
- IOT (Internet of things)
- Regression models
- Hadoop, data analysis, SPSS
- Machine learning
- Big data analysis using statistical model
- Tools to extract data
- Big data uses in math
- Blockchain
- Programming, network, Network security

10. **Additional Suggestions**
- I think TEQIP workshops should held more after Android
- Provide paper study material
- Lab sessions need more attention
- To give opportunities for manufacturing, power plant etc. for experience purpose
- Would like to see more workshops that are even more detailed
- Duration of sessions is too large that it became too hectic to attend
- More importance given to practical work rather than lectures
- Provide more lab facility
- Just build a social blog in this group for future communication
- A very good & helpful initiative by TEQIP. Start more helpful courses which are job and research oriented.
- Please mention the target audience for the workshop before
- Workshop was quiet good. It helps a lot for us
- Lab sessions should have more time than theory at least for some specific applications
Learning

1. **Do you get enough class projects?**

   - Yes: 73%
   - No: 27%

2. **Is the learning adequate?**

   - Yes: 91%
   - No: 9%
3. Do you have sufficient resources for laboratory?

Yes 82%
No 18%

4. Is the library/journal support/e-connection adequate?

Sufficient 86%
Inadequate 14%
5. **Would you like have common (TEQIP) repository of course material?**

- **Definitely**: 73%
- **Maybe**: 27%
- **No**: 0%

6. **Would you like to visit IITK to attend specialized courses?**

- **Definitely**: 77%
- **Maybe**: 23%
- **No**: 0%
7. **Would you like MOOCs/e-resources based courses?**

8. **What is your area of specialization?**

- Back end web development
- Web designing, programming
- Android development
- Communication & signal processing
- Mathematics
- Machine learning
- Ordered random variable
- Explaining Big Data
- Manufacturing
- IOT & Big Data
- Image security
- Soft computing in mathematics
- Applications of soft computing in diseases diagnosis
- Social network analysis, graph mining, associate rule mining using big data

9. **How can TEQIP help improve your learning?**

- I am grateful to TEQIP for providing a learning platform to share our knowledge for free of cost really thankful to you.
- We got some ideas about related things to work which we are doing.
- Yes, it enhances my career
- Organization of different workshops and short term courses.
- By organizing more workshops in colleges if possible
- Organize webinars
- By introducing the current trends and linking it with basics
- By providing access to the laboratories here at IIT
- By including data analytics in this course
- It has provided my an outline of Big Data
- TEQIP give chances us to visit in Big institutes like IITs
- They gave me chance to attend these wonderful lectures
- By giving us the technical insights on more such technologies
- By providing technical knowledge
- By introducing various opportunities present in TEQIP that we don’t know about.

**Research**

1. **Would you like to visit an IIT for a short visit/internship/post-doctoral stint, if offered (via TEQIP)?**

2. **Would you like to share/use research infrastructure at IITK, if made available?**
3. Would you like to conduct collaborative research with IITK faculty?

- Definitely: 68%
- Maybe: 32%
- No: 0%

4. Would you like lectures by experts (Indian and International) on niche research areas/topics?

- Definitely: 77%
- Maybe: 23%
- No: 0%
5. **Do you want special-topic conferences?**

![Pie chart showing the responses to the question.](image)

6. **How can TEQIP help improve your research?**

- By providing me with good study materials and getting in touch with IITK professors
- By stay connected with us & can tell about opportunities
- By giving more and more project
- By helping us in defining the problem statement in a specialized domain
- They gave me the chance to meet with wonderful people
- By sharing ideas
- Provision of paper study material
- By providing the infrastructure
- More practical learning courses
- TEQIP help in research area like experiment set up to provided
- By providing more research opportunities
- By providing a collaborative platform on which we can attract with IIT professors
- Provide more staff with good time propagation. Some time it came problematic to attend continuous session
- It provides rich set of technical knowledge to update my knowledge
- Permit to use some expensive software in our research work
- Actually by TEQIP short term courses we are growing our knowledge about different areas and there are something in these courses which is related to our work
- Through TEQIP we can meet experts and share our ideas and get knowledge from them to maintain quality of research in India.
In this workshop Participants learnt about various characteristics of big data. They were introduced to challenges involved in handling big data and various opportunities in this field. The lectures and lab sessions made participants familiar with following:

- Impact of big data
- Technology needed to solve big data problems
- How to develop big data strategy
- Why big data is important
- How to solve technical problems encountered in big data analysis

Some lectures with case studies were also delivered to show how big data is useful in energy, health care, marketing, social media etc.
Organizer’s Report

Report on TEQIP workshop on “Big Data” held from 7-11th Aug 2017 at PBCEC, VH, IIT Kanpur

Big Data term refers to data that is too big to fit into memory of one server, too unstructured to fit into row-column structure of an RDBMS and too continuously flowing to fit into a static data warehouse. Objective of a big data project is to convert data that has one or more of the characteristics of big data into a product or service to create value. There are three ways of becoming part of a big data project- one: be a technical person and handle the technology part, two: work with the solution and three: lead and manage a big data initiative. This course introduced participants to all three aspects of big data in a systematic manner. Speakers from academia as well as industry shared their knowledge and experience with the participants. We had total of 16 lecture sessions and 3 lab sessions. Lab sessions were conducted by M Tech students of Maths and Industrial & Management Engineering department students of IIT Kanpur.
Topics Discussed:
1. Introduction to Big data
2. Big Data Strategy
3. Big Data: Analysis & Interpretation
4. Xtreme Classification
5. SQL and NoSQL Databases
6. Scalability
7. Hadoop Stack & Map-Reduce
8. Data Mining
9. Social media analytics
10. Big Data in Energy Sector

Speakers:
1. Dr. Medha Atre, IIT Kanpur
2. Dr. Veena Bansal, IIT Kanpur
3. Dr. Arnab Bhattacharya, IIT Kanpur
4. Dr. Devlina Chatterjee, IIT Kanpur
5. Dr. Faiz Hamid, IIT Kanpur
6. Dr. Purushottam Kar, IIT Kanpur
7. Ms. Pallavi Kar, Matlab
8. Dr. Deepu Philip, IIT Kanpur
9. Dr. Shankar Prawesh, IIT Kanpur
10. Mr. Dhiraj Kumar Sati
11. Dr. Anoop Singh, IIT Kanpur
12. Dr. Nisheeth Srivastava, IIT Kanpur
Acknowledgements: The TEQIP staff members worked all through the course with a smile and took care of every aspect of the course flawlessly. Head, CC, IIT Kanpur gave some very useful suggestions and staff of CC helped in every possible way to make the labs a success. Participants were very