

**Report**  
**On**  
**Cloudera Academic Partnership (CAP)**

Cloudera – VESIT Collaboration for “**Developing with Spark and Hadoop**” training was conducted from 26<sup>th</sup> September 2016 to 8<sup>th</sup> April 2017.

The schedule of the course is attached. The course span was 36 hrs including theory and lab session which was conducted by following faculty members

1. Dr. M. Vijayalakshmi (Department of Information Technology)
2. Mrs. Asha Bharambe (Department of Information Technology)
3. Mrs. Sangeeta Oswal (Department of MCA)
4. Mrs. Sujata Khedkar (Computer Engineering Department)

The course included topics on basics of Hadoop, its ecosystem, and components and MapReduce. The lab was conducted on the Virtual machine provided by Cloudera.

The course was attended by final year students of Computers, IT and Second MCA. The list of the students is attached.



A handwritten signature in blue ink, consisting of stylized initials and a long horizontal stroke extending to the right.

**Vivekanand Education Society's  
INSTITUTE OF TECHNOLOGY**

2016-2017

**List of Students**

Sr.No	Computers	IT	MCA
1	Himani Shah	Nikita Kandalkar	Sayali Chavan
2	Tanvi Punjabi	Ankita Tanawade	Prajakta Akerkar
3	Shikha Soneji	Vasant A. Pawar	Niveta Lavhate
4	Monish Narwani	Grishma Bhatt	Pooja Mulam
5	Divya Despande	Dinesh Rajpal	Renu Patil
6	Amala Despande	-	Akanksha Tembye
7	Deepika Khatri	-	Sneha Velankar
8	Aditya Wakade	-	Pravin Gawas
9	Aditya Suresh	-	Raju Gatla
10	Omkar Patinge	-	Adrian Viegas
11	Latika Wadhwa	-	Yatin
12	Sunny Ramrakhiani	-	Bharati Nadar
13	Kashyap Chetan Kotak	-	Sudarshan Shinde
14	Kevin Shah	-	Vinaya Poojari
15	Devendra Kamalapurkar	-	Snehal Johny
16	Varun Golani	-	Suraj Sreedharan
17	Sagar Bhojwani	-	Ajinkya Dalvi
18	Sunil Bhatia	-	Hitesh Patil
19	Neetal Chaudhari	-	M Prabhanath Nair
20	Apoorva Mahale	-	Nihaal Mehta
21	Payal Rajani	-	Sachith Poojary
22	Jayesh Mandhyani	-	-
23	Mrunmagee Mujumdar	-	-
24	Anant Gupta	-	-
25	Sharvari Jalit	-	-

**Vivekanand Education Society's  
INSTITUTE OF TECHNOLOGY**

2016-2017

**“Developing with Spark and Hadoop”**

**ODD SEM**

**Schedule:** Timings: 8:10 to 10:20 (each day)

<b>Resource Person</b>	<b>Agenda</b>	<b>Task</b>	<b>Dates</b>
Dr. Mrs. M. Vijayalakshmi Mrs. Asha Bharambe Mrs. Sujata Khedkar Mrs. Sangeeta Oswal	Introduction Motivations for Hadoop Hadoop Ecosystem	Session 1	26/09/2016 (MV) (Monday)
Dr. Mrs. M. Vijayalakshmi Mrs. Asha Bharambe Mrs. Sujata khedkar Mrs. Sangeeta Oswal	Hadoop Architecture and HDFS	Session 2	27/08/2016 (MV) (Tuesday)
Dr. Mrs. M. Vijayalakshmi Mrs. Asha Bharambe Mrs. Sujata Khedkar Mrs. Sangeeta Oswal	Hadoop Architecture Continued(Yarn)	Session 3	28/09/2016 (AB) (Wednesday)
Dr. Mrs. M. Vijayalakshmi Mrs. Asha Bharambe Mrs. Sujata Khedkar Mrs. Sangeeta Oswal	Sqoop	Session 4	30/09/2016 (SO) (Friday)
Dr. Mrs. M. Vijayalakshmi Mrs. Asha Bharambe Mrs. Sujata Khedkar Mrs. Sangeeta Oswal	Introduction to Impala & Hive	Session 5	01/09/2016 (SKH) (Saturday)
Dr. Mrs. M. Vijayalakshmi Mrs. Asha Bharambe Mrs. Sujata Khedkar Mrs. Sangeeta Oswal	Modeling and Managing Data using Hive and Impala	Session 6	05/10/2016 (AB) (Wednesday)
Dr. Mrs. M. Vijayalakshmi Mrs. Asha Bharambe Mrs. Sujata Khedkar Mrs. Sangeeta Oswal	Data Format	Session 7	06/10//2016 (SKH) (Thursday)
Dr. Mrs. M. Vijayalakshmi Mrs. Asha Bharambe Mrs. Sujata Khedkar Mrs. Sangeeta Oswal	Data File Partitioning	Session 8	07/10/2016 (SO) (Friday)
Mid Term Break	Exam		

**Vivekanand Education Society's  
INSTITUTE OF TECHNOLOGY**

2016-2017

Resource Person	Agenda	Task	Dates
Dr. Mrs. M. Vijayalakshmi Mrs. Asha Bharambe Mrs. Sujata Khedkar Mrs. Sangeeta Oswal	Capturing Data with Apache Flume	Session 9	04/02/2017 (SKH) (Saturday)
Dr. Mrs. M. Vijayalakshmi Mrs. Asha Bharambe Mrs. Sujata khedkar Mrs. Sangeeta Oswal	Distributed Data Processing with Spark Spark Basics	Session 10	11/02/2017 (SO) (Saturday)
Dr. Mrs. M. Vijayalakshmi Mrs. Asha Bharambe Mrs. Sujata Khedkar Mrs. Sangeeta Oswal	Working with RDDs in Spark	Session 11	18/02/2017 (AB) (Saturday)
Dr. Mrs. M. Vijayalakshmi Mrs. Asha Bharambe Mrs. Sujata Khedkar Mrs. Sangeeta Oswal	Aggregating Data with Pair RDDs	Session 12	04/03/2017 (AB) (Saturday)
Dr. Mrs. M. Vijayalakshmi Mrs. Asha Bharambe Mrs. Sujata Khedkar Mrs. Sangeeta Oswal	Writing and Deploying Spark Applications	Session 13	11/03/2017 (SO) (Saturday)
Dr. Mrs. M. Vijayalakshmi Mrs. Asha Bharambe Mrs. Sujata Khedkar Mrs. Sangeeta Oswal	Parallel Data Processing in Spark	Session 14	18/03/2017 (SKH) (Saturday)
Dr. Mrs. M. Vijayalakshmi Mrs. Asha Bharambe Mrs. Sujata Khedkar Mrs. Sangeeta Oswal	Spark RDD Persistence	Session 15	25/03/2017 (AB) (Saturday)
Dr. Mrs. M. Vijayalakshmi Mrs. Asha Bharambe Mrs. Sujata Khedkar Mrs. Sangeeta Oswal	Common Pattern in spark Data Processing	Session 16	01/04/2017 (SO) (Saturday)
Dr. Mrs. M. Vijayalakshmi Mrs. Asha Bharambe Mrs. Sujata Khedkar Mrs. Sangeeta Oswal	Sprak SQL and Data Frame	Session 17	08/04/2017 (SKH) (Saturday)

**Vivekanand Education Society's  
INSTITUTE OF TECHNOLOGY**

2016-2017

Sr. No.	Project Description	Dataset
1	<p>Politics &amp; Analytics</p> <p>-use Big Data Technology to analyze different parties performance using social media data</p> <ol style="list-style-type: none"> <li>1. Which candidate is being talked about the most?(top candidate of each party)</li> <li>2. What are the sentiments for the major parties in India?</li> <li>3. Where geographically are they being most talked about?</li> <li>4. Compare popularity and ranking in social media</li> <li>5. Observe the change in trends over time</li> </ol>	<p>-Twitter API</p> <p>- Tracking certain keywords and collecting data like user information, geo-location and the text</p>
2	<p>Negative reviews analysis and customer needs identification</p> <p>OR</p> <p>Identification of Fake Reviews</p> <ul style="list-style-type: none"> <li>•Spam Reviews</li> <li>•Reviews not helpful</li> </ul> <p>And Reviews Clustering and sentiment analysis.</p>	<p>Yelp dataset</p> <p><a href="https://www.yelp.com/dataset_challenge/dataset">https://www.yelp.com/dataset_challenge/dataset</a></p> <p>Algorithms</p> <p>Recommendation/</p> <p>Clustering:</p> <ul style="list-style-type: none"> <li>•SVM</li> <li>•Sentiment Analysis</li> <li>•Latent Dirichlet Allocation (LDA)</li> </ul>
3	<p>Book Recommender</p> <p>Algorithms:</p> <p>User based recommendation and Item based recommendation</p> <p>Enter a book you like and the application will analyze huge database of all users' information to recommend books for you as your next Suggested read.</p>	<p><a href="http://www2.informatik.uni-freiburg.de/~cziegler/BX/">http://www2.informatik.uni-freiburg.de/~cziegler/BX/</a></p> <p>It contains 278,858 users (anonymized but with demographic information) providing 1,149,780 ratings (explicit / implicit) about 271,379 books.</p> <p>Algorithms:</p> <p>Tools:</p> <p>→Java, Spark, Hadoop and Mahout for the analytics</p>
4	<p>Descriptive Analytics of Stock Data</p>	<p><a href="http://www.Nsebse.com">www.Nsebse.com</a>,</p>

**Vivekanand Education Society's  
INSTITUTE OF TECHNOLOGY**

2016-2017

	<ul style="list-style-type: none"><li>- Time plots of monthly price and returns</li><li>- High and Low value close</li></ul>	finance.yahoo.com
5	Bank / Financial Institute - Customer complaint analysis	<a href="https://catalog.data.gov/dataset/consumer-complaint-database">https://catalog.data.gov/dataset/consumer-complaint-database</a>
6	Web Server Log Analysis Find fraudulent behavior	Get dataset from internet
7	Twitter data analytics for movie recommendation	Get the data from twitter about different movies opinions
8	Hotel review analytics E.g. Taj Hotel Mumbai	TripAdvisor.com, Twitter data

**Big Data Projects**

# VESIT

## Cloudera Academic Training

for Developing with Spark and Hadoop Year : 2016-17

Department :- CMPN

List of student

Class:- D17A

Attendance Sheet

Sr.No.	Name	Date	Date	Date	Date	Date	Date	Date	Date	Date	Date	Date	Date	Date	Date
1	Himani Shah	26/09/2016	27/09/2016	28/09/2016	30/09/2016	01/10/2016	05/10/2016	07/10/2016	08/10/2016	04/11/16	11/11/16	18/11/16			
2	Tanvi Punjabi	P	Tanvi	Tanvi	Tanvi	Tanvi	Tanvi	Tanvi	Tanvi	A	A	A	Tanvi		
3	Shikha Soneji	P	Shikha	A	Shikha	Shikha	Shikha	Shikha	Shikha	Shikha	Shikha	Shikha	Shikha	Shikha	Shikha
4	Monish Narvani	P	Monish	Monish	Monish	Monish	Monish	Monish	Monish	Monish	Monish	Monish	Monish	Monish	Monish
5	Divya Despande	P	Divya	Divya	Divya	Divya	Divya	Divya	Divya	Divya	Divya	Divya	Divya	Divya	Divya
6	Amala Despande	P	Amala	Amala	Amala	Amala	Amala	Amala	Amala	Amala	Amala	Amala	Amala	Amala	Amala
7	Deepika Khatri	P	Deepika	Deepika	A	Deepika	A	Deepika	Deepika	Deepika	Deepika	Deepika	Deepika	Deepika	Deepika
8	Aditya Wakade	A	Aditya	Aditya	A	Aditya	Aditya	Aditya	Aditya	Aditya	Aditya	Aditya	Aditya	Aditya	Aditya

Department :- CMPN

List of student

Class:- D17B

Attendance Sheet

Sr.No.	Name	Date	Date	Date	Date	Date	Date	Date	Date	Date	Date	Date	Date	Date	Date
1	Aditya Suresh	26/09/2016	27/09/2016	28/09/2016	30/09/2016	01/10/2016	05/10/2016	07/10/2016	08/10/2016	4/11/16	11/11/16	18/11/16			
2	Omkar Paringe	P	Omkar	Omkar	Omkar	Omkar	Omkar	Omkar	Omkar	A	A	A	Omkar		
3	Larika Wadhwa	P	Larika	Larika	Larika	Larika	Larika	Larika	Larika	Larika	Larika	Larika	Larika	Larika	Larika
4	Sunny Ramrakhiani	P	Sunny	A	Sunny	Sunny	Sunny	Sunny	Sunny	A	A	A	Sunny		
5	Kashyap Chetan Kotak	P	Kashyap	Kashyap	Kashyap	Kashyap	Kashyap	Kashyap	Kashyap	Kashyap	Kashyap	Kashyap	Kashyap	Kashyap	Kashyap
6	Kevin Shah	P	Kevin	Kevin	Kevin	Kevin	Kevin	Kevin	Kevin	A	A	A	Kevin		
7	Devendra Kamalapurkar	P	Devendra	Devendra	Devendra	Devendra	Devendra	Devendra	Devendra	A	A	A	Devendra		
8	Vrjun Gohani	P	Vrjun	Vrjun	Vrjun	Vrjun	Vrjun	Vrjun	Vrjun	A	A	A	Vrjun		
9	Sagar Bhojwani	P	Sagar	Sagar	Sagar	Sagar	Sagar	Sagar	Sagar	A	A	A	Sagar		







**VESIT**  
**Cloudera Academic Training**  
**for Developing with Spark and Hadoop Year : 2016-17**

Department :- CMPN

List of student

Class:- D17C

Attendance Sheet

Sr.No.	Name	Date	Date	Date	Date	Date	Date	Date	Date	Date	Date			
1	Suniti Bharita	A	A	A	Sun	A						4	2	6
2	Neelal Chaudhari	A	A	A	A	A						2	0	2
3	Apoorva Mahale	A	A	A	A	A						8	8	16
4	Payal Rajani	A	A	A	A	A						7	6	13
5	Jayesh Mandhyani	A	A	A	A	A						7	6	13
6	Mrunalgece Mujumdar	A	A	A	A	A						7	2	9
7	Anant Gupta	A	A	A	A	A						6	0	6
8	Sharvari Jalit	A	A	A	A	A						6	0	6



# VESIT

Cloudera Academic Training

for Developing with Spark and Hadoop Year : 2016-17

Department :- MCA

List of student

Class:- Morning Shift

Attendance Sheet

Sr.N o.	Name	Date	Date	Date	Date	Date	Date	Date	Date	Date	Date	Date	Date
1	Bharati Nadar	A	A	A	A	A							
2	Sudarsshan Shinde	<del>absent</del>	<del>absent</del>	A	<del>absent</del>	A							
3	Vinaya Poojari	A	A	A	A	A							
4	Snehal Johny	A	A	A	A	A							
5	Suraj Sreedharan	A	A	A	A	A							
6	Ajinkya Dalvi	A	A	A	A	A							
7	Hitesh Patil	A	A	A	A	A							
8	M Prabhanath Nair	A	A	A	A	A							
9	Nihaal Mehta	A	A	A	A	A							
10	Sachith Poolary	A	A	A	A	A							

	Date	Date	Date	Date	Date	Date	Date	Date	Date	Date
	Sem VII	Sem VIII	Total							
1	0	0	0							
2	6	4	10							
3	4	0	4							
4	7	0	7							
5	7	0	7							
6	5	0	5							
7	0	0	0							
8	0	0	0							
9	6	0	6							
10	7	1	8							
	4	0	4							

**VESIT**  
 Cloudera Academic Training  
 for Developing with Spark and Hadoop Year : 2016-17

Department :- MCA

List of student

Class:- Afternoon Shift

Attendance Sheet

Sr.N	Name	Date	Date	Date	Date	Date	Date	Date	Date	Date	Date	Date	Date	Date
0.		26/09/2016	27/09/2016	28/09/2016	30/09/2016	01/10/2016	05/10/2016	07/10/2016	09/10/2016	11/10/2016	13/10/2016	15/10/2016	17/10/2016	19/10/2016
1	Sayali Chavan	<del>P</del>	A	A	<del>A</del>	A	A	A	A	A	A	A	A	A
2	Prajakta Akerkar	<del>P</del>	A	A	<del>A</del>	A	A	A	A	A	A	A	A	A
3	Niveta Lavhate	<del>P</del>	<del>Reshate</del>	<del>Reshate</del>	<del>Reshate</del>	<del>Reshate</del>	<del>Reshate</del>	<del>Reshate</del>	<del>Reshate</del>	<del>Reshate</del>	<del>Reshate</del>	<del>Reshate</del>	<del>Reshate</del>	<del>Reshate</del>
4	Pooja Mulam	<del>P</del>	<del>Red</del>	<del>Red</del>	<del>Red</del>	<del>Red</del>	<del>Red</del>	<del>Red</del>	<del>Red</del>	<del>Red</del>	<del>Red</del>	<del>Red</del>	<del>Red</del>	<del>Red</del>
5	Renu Patil	<del>P</del>	<del>Red</del>	A	<del>Red</del>	<del>Red</del>	<del>Red</del>	<del>Red</del>	<del>Red</del>	<del>Red</del>	<del>Red</del>	<del>Red</del>	<del>Red</del>	<del>Red</del>
6	Akanksha Tembye	<del>P</del>	<del>Red</del>	A	<del>Red</del>	<del>Red</del>	<del>Red</del>	<del>Red</del>	<del>Red</del>	<del>Red</del>	<del>Red</del>	<del>Red</del>	<del>Red</del>	<del>Red</del>
7	Sneha Velankar	<del>P</del>	A	A	A	A	A	A	A	A	A	A	A	A
8	Pravin Gawas	<del>P</del>	<del>Red</del>	<del>Red</del>	<del>Red</del>	<del>Red</del>	<del>Red</del>	<del>Red</del>	<del>Red</del>	<del>Red</del>	<del>Red</del>	<del>Red</del>	<del>Red</del>	<del>Red</del>
9	Raju Gadia	<del>P</del>	<del>Red</del>	<del>Red</del>	<del>Red</del>	<del>Red</del>	<del>Red</del>	<del>Red</del>	<del>Red</del>	<del>Red</del>	<del>Red</del>	<del>Red</del>	<del>Red</del>	<del>Red</del>
10	Adrian Viegas	<del>P</del>	<del>Red</del>	<del>Red</del>	<del>Red</del>	<del>Red</del>	<del>Red</del>	<del>Red</del>	<del>Red</del>	<del>Red</del>	<del>Red</del>	<del>Red</del>	<del>Red</del>	<del>Red</del>

Yatin A yatin yatin yatin A A yatin yatin A A A A

**VESIT**  
 Cloudera Academic Training  
 for Developing with Spark and Hadoop Year : 2016-17

Department :- MCA

List of student

Class:- Afternoon Shift

Attendance Sheet

Sr.N o.	Name	Date	Date	Date	Date	Date	Date	Date	Date	Date Sem vii	Date Sem viii	Date Total
1	Sayali Chavan	A	A	A	A	A				2	0	2
2	Prajakta Akerkar	A	A	A	A	A				2	0	2
3	Niveta Lavhate	<del>A</del>	A	A	A	<del>A</del>				6	5	11
4	Pooja Mulam	<del>A</del>	A	A	<del>A</del>	<del>A</del>				8	6	14
5	Renu Patil	A	A	A	A	A				7	0	7
6	Akanksha Tembye	A	A	A	A	A				7	0	7
7	Sneha Velankar	A	A	A	A	A				0	1	01
8	Pravin Gawas	<del>A</del>	<del>A</del>	<del>A</del>	<del>A</del>	<del>A</del>				7	7	14
9	Raju Gatta	<del>A</del>	<del>A</del>	<del>A</del>	<del>A</del>	<del>A</del>				7	3	10
10	Adrian Viegas	<del>A</del>	<del>A</del>	<del>A</del>	<del>A</del>	<del>A</del>				7	8	15
11	Yatin	A	A	A	A	A				4	0	5

**VESIT**  
**Cloudera Academic Training**  
**for Developing with Spark and Hadoop Year : 2016-17**

Department :- INFY

List of student

Class:- D20

Attendance Sheet

Sr.N o.	Name	Date	Date	Date	Date	Date	Date	Date	Date	Date	Date	Date	Date	Date	Date
		26/09/2016	27/09/2016	28/09/2016	30/09/2016	01/10/2016	05/10/2016	07/10/2016	08/10/2016	11/10/2016	11/11/17	10/12/17			
1	Nikita Kandalkar	P <i>Nikita</i>	P <i>Nikita</i>	A <i>Nikita</i>	A <i>Nikita</i>	A <i>Nikita</i>	A	A	A	A	A	A	A	A	A
2	Ankita Tanawade	P	P	P	A	A	A	A	A	A	A	A	A	A	A
3	Vasant A. Pawar	P	P	P	<del>AB</del>	<del>AB</del>	<del>AB</del>	<del>AB</del>	<del>AB</del>	<del>AB</del>	<del>AB</del>	<del>AB</del>	<del>AB</del>	<del>AB</del>	<del>AB</del>
4	Grishma Bhatt	P <i>GB</i>	P <i>GB</i>	P <i>GB</i>	<del>AB</del>	<del>AB</del>	<del>AB</del>	<del>AB</del>	<del>AB</del>	<del>AB</del>	<del>AB</del>	<del>AB</del>	<del>AB</del>	<del>AB</del>	<del>AB</del>
5	Dinesh Rajpal	P	A	A	A	A	A	A	A	A	A	A	A	A	A
6															
7															

*5/11/17*

**VESIT**  
**Cloudera Academic Training**  
**for Developing with Spark and Hadoop Year : 2016-17**

Department :- INFT

List of student

Class:- D20

Attendance Sheet

Sr.N o.	Name	Date	Date	Date	Date	Date	Date	Date	Date	Date	Date	Date
1	Nikita Kandalkar	A	A	Nikita	A	Nikita						
2	Ankita Tanawade	A	A	A	A	A						
3	Vasant A. Pawar	<del>VR</del>	<del>VR</del>	<del>VR</del>	<del>A</del>	<del>VR</del>						
4	Grishma Bhatt	A	<del>VR</del>	<del>VR</del>	<del>A</del>	<del>VR</del>						
5	Dinesh Rajpal	A	<del>VR</del>	<del>A</del>	<del>A</del>	A						
6												
7												

Date Sem VII      Date Sem VIII      Date Today  
 6                      5                      11  
 3                      0                      3  
 7                      5                      12  
 6                      6                      12  
 01                      3                      4