



Inspire...Educate...Transform.

CSE7322c

Engineering Big Data Applications with Hadoop and Spark Ecosystem

March 24th - 2018

Group Activity

– Case Study – Group Activity

- Pick one domain (i.e. Retail, Telecom etc.) And will work on the following.
- Identify different type and sources of data.
- Identify the engineering problems they may have solving.
- Classify the problem.
- How this can be solved using regular ML/Analytics tools.
- What is the type of data – Volume, Velocity and Variety aspect?
- Extend the same to big data analytics.

BIG DATA Application Examples



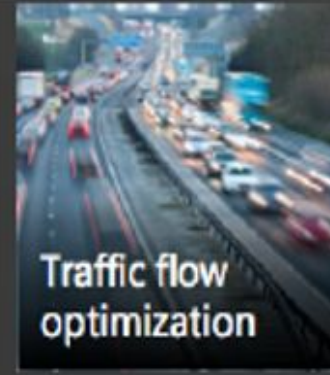
IT infrastructure
optimization



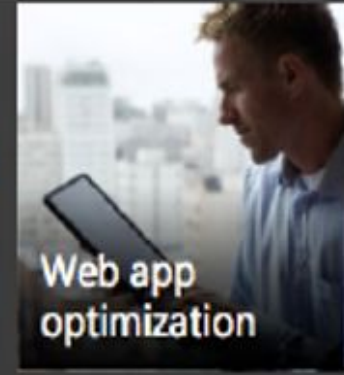
Legal
discovery



Social network
analysis



Traffic flow
optimization



Web app
optimization



Churn
analysis



Natural resource
exploration



Weather
forecasting



Healthcare
outcomes



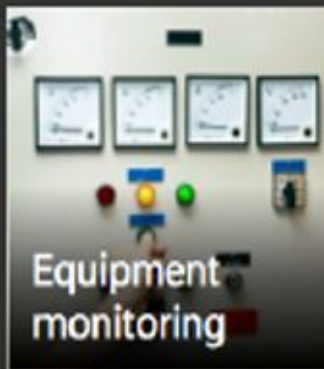
Fraud
detection



Life sciences
research



Advertising
analysis

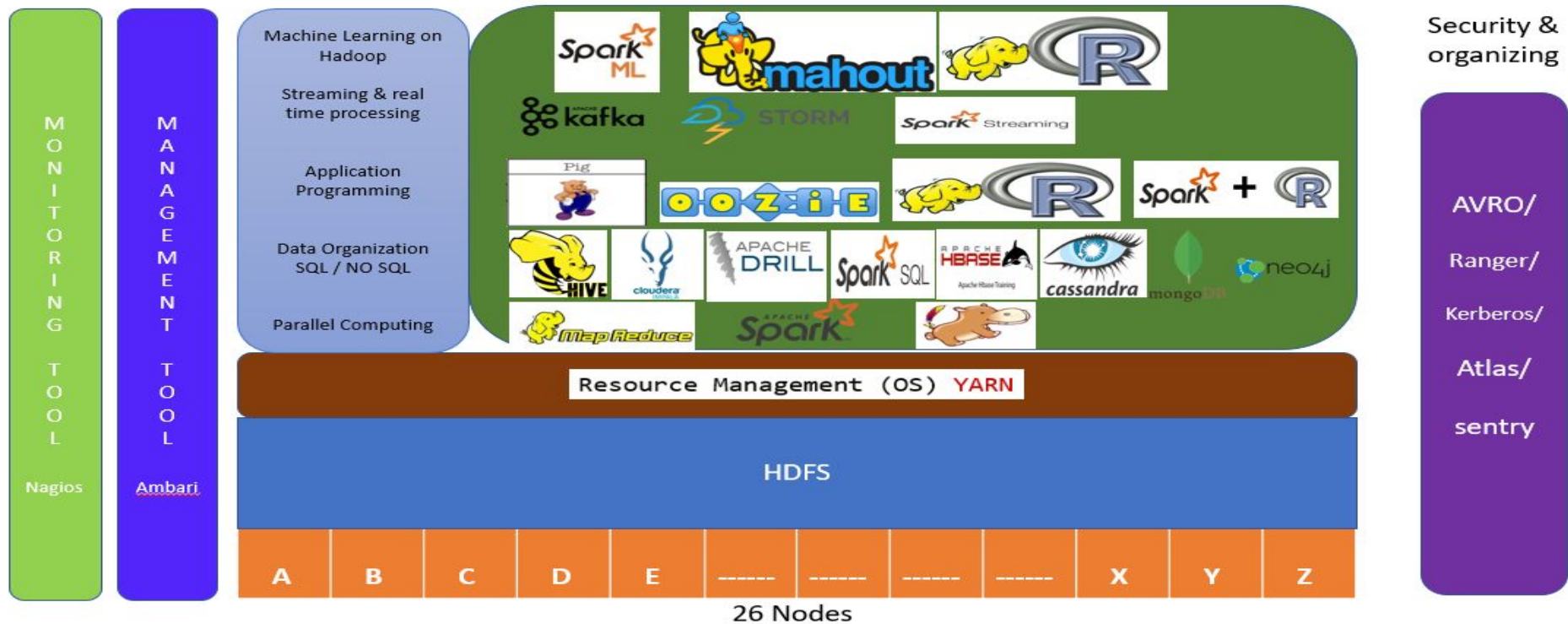


Equipment
monitoring



Smart meter
monitoring

Cluster Introduction



Every student will get access to the Real-time big data cluster environment for the next 6 months.

Cluster Configuration:

- 2 Namenodes with failover setup, 128 GB Ram, 8 Core processor, 500 GB HDD each.
- 24 Datanodes with 64 GB Ram, 8 Core processor, 500 GB HDD each.
- 1 MySQL Server on each node, with 64 GB Ram, 8 Core processor, 100GB HDD.
- 1 Login node (Load Balancer) with 64 GB Ram, 8 Core processor, 500 GB HDD each. Users will login from this and redirect to one of the datanodes in Round Robin method.
- 1 Ambari Node with 64 GB Ram, 8 Core processor, 500 GB HDD. This will give admin access to manage the Hadoop cluster.

Hadoop Eco System components / Software Stack:

- Hadoop Distribution: HDP 2.6
- HDFS, YARN, Mapreduce2, Tez, Hive, Hbase, Pig, Sqoop, Oozie, Zookeeper, Storm, Flume, Accumulo, Ambari Infra, Ambari Metrics, Atlas, Kafka, Knox, Log Search, Spark2, Zeppelin, Jupiter Notebook, Mahout, Slider.
- HUE
- MySQL

Login Details

- 1) Login Credentials have been sent to your registered email id's with INSOFE.
- 2) Login URL: <http://172.16.0.247> (Gateway IP)
- 3) UserName: 1234B31 (Enroll ID<BID>)
- 4) Password: *****
- 5) Once you login gateway will be routed to one of the node with in the cluster based on the resources availability.
- 6) Use the node address to which you login for file transfers etc.

Login details Email

From: **Chakradhar** <chakradhar.yaddala@insofe.edu.in>
Date: Fri, Nov 24, 2017 at 10:47 PM
Subject: INSOFE Hadoop Cluster Access
To: pranith.samanu@insofe.edu.in

Hi Praneeth,

Greetings from INSOFE!

You are receiving this mail as a confirmation of your subscription to the INSOFE Big Data Lab. Please find the below credentials to access the lab:

Note : This lab can be accessible only in INSOFE premises.

URL : <http://172.16.0.247>

Username : 1111B31

Password : 1234567890

Important Cluster Links:

Ambari: <http://ambari.insofe.edu.in:8080>

Job status in web: <http://b.insofe.edu.in:8088/cluster>

Job History: <http://b.insofe.edu.in:19888/jobhistory>

R Port: <http://j.insofe.edu.in:8787/>

Oozie: <http://b.insofe.edu.in:11000/oozie/?user.name=admin>

Grafana: <http://a.insofe.edu.in:3000/>

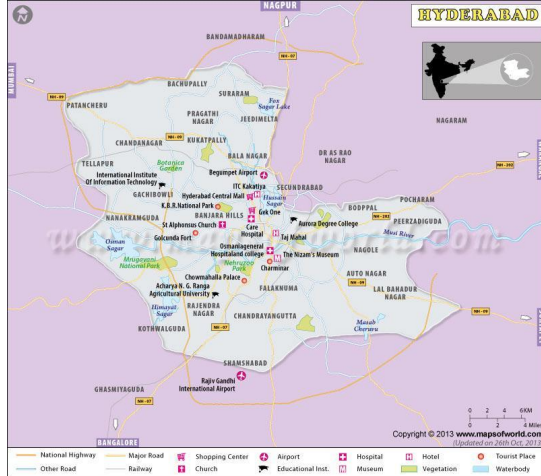
HBase: <http://b.insofe.edu.in:16010/master-status>

Spark History Server: <http://b.insofe.edu.in:18081/>

Zeppelin: <http://a.insofe.edu.in:9995/#/>

Jupyter port: `jupyter notebook --port 2BID --no-browser --ip=0.0.0.0`

Ex: Bnglr: 21524



HYDERABAD

Office and Classrooms

Plot 63/A, Floors 1&2, Road # 13, Film Nagar,
Jubilee Hills, Hyderabad - 500 033
+91-9701685511 (Individuals)
+91-9618483483 (Corporates)

Social Media

Web:

Facebook: <https://www.facebook.com/insofe>

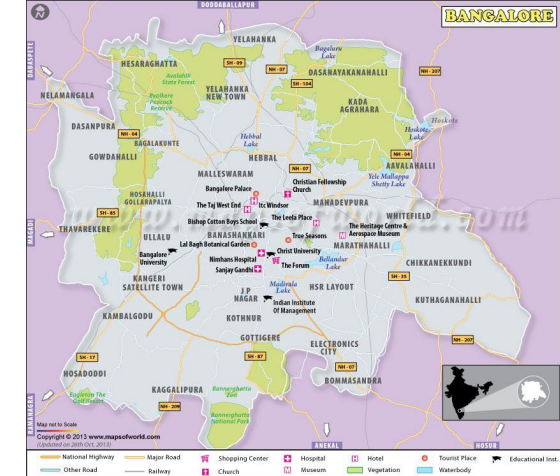
Twitter: <https://twitter.com/Insofeedu>

YouTube:

SlideShare:

LinkedIn:

This presentation may contain references to findings of various reports available in the public domain. INSOFI makes no representation as to their accuracy or that the organization subscribes to those findings.



BENGALURU

Office

Incubex, #728, Grace Platina, 4th Floor, CMH Road,
Indira Nagar, 1st Stage, Bengaluru – 560038
+91-9502334561 (Individuals)
+91-9502799088 (Corporates)

Classroom

KnowledgeHut Solutions Pvt. Ltd., Reliable Plaza,
Jakkasandra Main Road, Teacher's Colony, 14th Main
Road, Sector – 5, HSR Layout, Bengaluru - 560102