2. OVERVIEW

2.1 The Big Data

- Big data can be applied to real-time fraud detection, complex competitive analysis, call centre optimization, consumer sentiment analysis, intelligent traffic management, and to manage smart power grids, to name only a few applications.

- Big data is characterized by three primary factors: volume (too much data to handle easily); velocity (the speed of data flowing in and out makes it difficult to analyze); and variety (the range and type of data sources are too great to assimilate).

- With the right analytics, big data can deliver richer insight since it draws from multiple sources and transactions to uncover hidden patterns and relationships

- There are four types of big data BI that really aid business:

    1. **Prescriptive** – This type of analysis reveals what actions should be taken. This is the most valuable kind of analysis and usually results in rules and recommendations for next steps.

    2. **Predictive** – An analysis of likely scenarios of what might happen. The deliverables are usually a predictive forecast.

    3. **Diagnostic** – A look at past performance to determine what happened and why. The result of the analysis is often an analytic dashboard.

    4. **Descriptive** – What is happening now based on incoming data. To mine the analytics, you typically use a real-time dashboard and/or email reports.

2.2 Big Data Analytics in Action:

Big data analytics can deliver big value to business, adding context to data that tells a more complete story. By reducing complex data sets to actionable intelligence, we can make more accurate business decisions.

1. **Prescriptive analytics:** is where big data analytics in general sheds light on a subject, prescriptive analytics gives you a laser-like focus to answer specific questions.

2. **Predictive analytics:** use big data to identify past patterns to predict the future.

3. **Diagnostic analytics**: are used for discovery or to determine why something happened.

4. **Descriptive analytics:** In other words, data mining are at the bottom of the big data value chain, but they can be valuable for uncovering patterns that offer insight

CoE-Data Analytics will focus on open source software and platforms as much as possible to stay nimble, foster efficiency and drive innovation.
It has to provide access to technologies in the following areas:

- Data collection Data storage
- Data curation and management
- Data processing
- Data exploration and visualization
- Unstructured data processing
- Data analytics and machine learning

2.3 Technologies Required to learn Big Data Analytics:

- **Hadoop Platform:** (Big Data Processing and analytics)

  → To Processing Big Data on Hadoop, understand how to use Hadoop to handle and arrange Big Data.

  → To teach how to perform analytical operations to gain insights from data processed through Hadoop.

- **Why is Big Data Analytics on Hadoop important?**

  → Businesses are now aware of the large volumes of data that they generate in their day to day transactions. They have also realized that this Big Data can provide very valuable insights once analysed.

  → The massive volume of Big Data and its unstructured format make it difficult to analyse Big Data. Hadoop brings the ability to cheaply process large amounts of data, regardless of structure.

  → Knowledge about Big Data Analytics on Hadoop will also prove to be a huge resume builder for Students who are aiming to work in the IT Industry

- **Different Hadoop Platforms:**

  → Apache, Horton work, Cloudera.

  → Tools Taught: Map-Reduce Programming, hive, Pig, NoSQL, Flume, Hue, HBase, Cassandra, Yarn.

- **Integration of Hadoop:**

  → With R: R-Hadoop, Python and Mahout

  → Level 1: Hadoop Platform: processing large volume of data using cluster techniques and convert unstructured, semi structure data into structure form.

  → Level2: Java, Python or R:

  → Level 3: ML

→ Level4: Visual Analysis tools: Tableau, Python with Pandas, tense flow, R with libraries

- **Paid Tools:**

  → Tableau, Microsoft azure, IBM Watson.

- **Associated Technologies:**

  → Cloud and IoT (Sensor data), Web App, Android App

3. EXECUTION PLAN

PHASE	DESCRIPTION	DURATION	TARGET STUDENTS
1	Start with Industry Experts (Importance of Data, use cases and Future Trends)	February 2019, 1st week	3rd Semester and 5th Semester
2	Deliver session on Hadoop Platform with all possible tools.	February 2019 - March 2019	3rd Semester and 5th Semester
3	Deliver sessions on Python Language Knowledge transfer to students through Project exhibition	April 2019 – May 2019	8th Semester
4	Deliver sessions on Machine Learning	August 2019 – October 2019	3rd Semester and 5th Semester
5	Assign projects to CoE-registered students and take up consultancy projects		3rd Semester and 5th Semester