

WORKSHOP ON BIG DATA.

Event Date : 28th and 29th September, 2019.

Venue: Computer Centre.

Event Speaker: Mr. Ismail Patel (Accion Labs, Spark Engineer).

A two day Workshop on “BIG DATA” was held on 28th and 29th September, 2019. The workshop had been organized by the Computer Engineering department in association with CSI for B.E students to provide them with an opportunity to get an in-depth knowledge about big data in the computer engineering field.

What is Big Data ?

"Big data" is a field that treats ways to analyse, systematically extract information from, or otherwise deal with data sets that are too large or complex to be dealt with by traditional data-processing application software. Data with many cases (rows) offer greater statistical power, while data with higher complexity (more attributes or columns) may lead to a higher false discovery rate. Big data challenges include capturing data, data storage, data analysis, search, sharing, transfer, visualization, querying, updating, information privacy and data source. Big data was originally associated with three key concepts: *volume*, *variety*, and *velocity*. When we handle big data, we may not sample but simply observe and track what happens. Therefore, big data often includes data with sizes that exceed the capacity of traditional usual software to process within an acceptable time and value

YARN stands for “*Yet Another Resource Negotiator* “. It was introduced in Hadoop 2.0 to remove the bottleneck on Job Tracker which was present in Hadoop 1.0. YARN was described as a “*Redesigned Resource Manager*” at the time of its launching, but it has now evolved to be known as large-scale distributed operating system used for Big Data processing. YARN architecture basically separates resource management layer from the processing layer. In Hadoop 1.0 version, the responsibility of Job tracker is split between the resource manager and application manager



The Workshop started with the ‘Qeerat’ and the felicitation of the speaker by Er. Mohammad Ahmed of Computer Department. After the felicitation Mr. Ismail Patel started with his introduction, then moving towards the topics of Big Data,



The topics that were covered during the two-day workshop were:

DAY 1

1. Big Data History, Problems and Solutions
2. Introduction to Map reduce 1-A card-based exercise.
3. MapReduce 2.0.
4. Hadoop- Architecture and working.

SESSION 1:

The commencement of the workshop was done with theoretical information about Big Data which was kind of a QA session where Mr. Ismail was explaining and asking questions to the students to understand the level of knowledge they have.

SESSION 2:

Word Count problem to many students can be boring, but instead of taking the theoretical approach, Mr. Ismail turned it to a fun activity by splitting students into a group of 5-6 and assigning them to arrange each word in the given passage in alphabetical order with number of occurrences marked. At the end of this activity I assure you no student will ever forget how mapper and reducer worked in a word count problem.

DAY 2

1. Hive DB and other services.
2. Hadoop Distributions- Hortonworks and Cloudera.
3. Spark introduction.
4. Kafka and Spark streaming- Twitter example with Kafka and Spark Streaming.

SESSION 1:

Conducted on Sunday was completely related to the project part. Starting with making tables into HDFS, inserting data and even uploading the dataset into our Cloudera System. Next was the implementation of many SQL queries into HIVE to analyze the dataset which we had uploaded earlier.

SESSION 2:

Coming to an end with session 2, we learned about SQOOP and Spark both of which were very informative and helpful for future studies and industrial approach.

Last but not the least the speaker had a question answer session again with the students to clear the doubts they had.

The workshop ended with the chance of discussion with the speaker about an internship in his company and finally by presenting a letter of appreciation that was given by Saiqa Khan and Ahlam Ansari to the speaker for conducting the workshop.



Er. Shabana Tadvi
(CSI In-Charge)

Er. Ahlam Ansari
(CSI In- Charge)

Dr Zainab Pirani
(CSI Moderator &
HOD Comp. Dept)