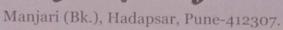


Pune District Education Association's College Of Engineering





Accredited by NAAC

DSBDAL	Assignment No:15
*	Aim: - Hoite a case study to process data
	driven for digital marketing or health
	care system with hadoop ecosystem compo-
	nents as shown:
•	HDFS: Hadoop distributed file system.
	YARM; Let onother resource negotiator map-
	Reduce programming based data processing
	spark: In memory data processing ma,
	HIGE query based processing or data
	Sewices.
*	theory: - Apache hadoop is an open source
	Framework intended to make interaction
•	With big data easier.
	- Hadoop Ecosystem is a platform or a
	suite which provides various sources to
	solve the big data problems.
	Mac et al de la lipiait de la cil
	-HDFS:- Hadoop distributed file system.
	HOFS is the primary or major components
	of hadoup ecosystem and is responsible for storing large data sets of structured
	or unstructured data across various
	hodes and there by maintaining the
	metadata in the foom or log files.
	109

- 1) Name Node
- 2) pata Mode

YARN:Yet another resource negotiator, as the name implies YARN is the one. Who helps to manages the resources across the clusters. Inshort it performs the scheduling & resource allocation for the hadoop system.

-consists of three major components:-

- D Resource manager
- 2) Hodes manager
- 3) Application manager

Map Reduce -

By making the use of distributed & parallel algorithms mapreduce make it positive to carry over the processing logic & helps to larite applications which transform big data sets into a manageable one.

mapreduce makes the use of the tho

- 1) map Reduce ()
- 2) Reduce ()

Apache spust :-



Pune District Education Association's College Of Engineering



Manjari (Bk.), Hadapsar, Pune-412307.

Accredited by NAAC

It is platform that handles all the processing interactive or iterative real time processing graph conversion in and visualization.

spoot is best suited for real-time data whereas Hadoop is best suited for structured data or both processing.

pIG was basically developed by yahoo which works on a big latin language, which is queen based language Similar to sal.

PIG does not work of executing the command & in the background all the activities of map.

Pune District Education Association

HIVE -

Hith the help of sol methodology & interface HTUE performs reading fulliting of large data sets.

J+ is highly scalable as it allows realtime processing & batch processing both.

processing easier.

* conclusion:-I studied the case study to process data driven for digital marketing or health care systems with the hadoop ecosystem components.