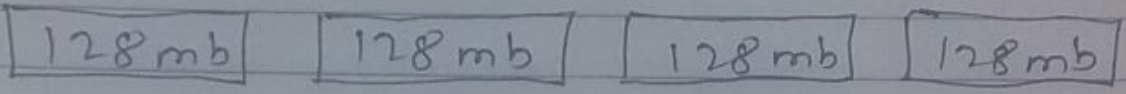
0209_BDA_18CE1009_ShivamKendre

Q.1]

2] Default blocks size in Hadoop is 2.x is 128 mb

So. a size of 614 mb can be divided into five blocks where first four will be of 128 mb and last blocks will be of 102 mb only.



since we are using default replication factor i.e 3 each blocks will be repeated thrice.

So we will have total of 15 blocks

12 blocks of size 128 mb

3 blocks of size 102 mb

0209 - BDA - 18CE1009 - Shivam Kendre

Q.1]

3]

NoSQL	RDBMS
1) it supports very simple query language	1) it supports a powerful query language
2) it has no fixed schema	2) it has a fixed schema
3) it is only eventually consistent	3) it follows acid properties
4) don't support transactions (support only simple transaction)	4) Relational database supports transactions (also complex transaction with join)
5) it is used to handle data coming in high velocity	5) it is used to handle data coming in low velocity
6) it's data arrives from many locations	6) it's data arrives from one or few location
7) it gives both read and write scalability	7) it gives only read scalability
8) it is deployed in horizontal fashion	8) it is deployed in vertical fashion

0209_BDA_18CE1009_Shivam Krndse

Q. 2]

1] 1) Data structure :-

A) Big data :-
it is unstructured in nature

B) Traditional :-
it has pre-defined structure.

2) Data relationship :-

A) Big data :-
unknown relationship in case of big data

B) Traditional :-
By default, stable and interrelationship

3) Data Analysis :-

A) Big data :-
in case of big data it is intermediate analysis

B) Traditional :-
After the complete build

4) Data location :-

A) Big data :-
it is physically highly distributed

B) Traditional :-
Centralized.

0209-18CE1009-BDA_Shivamkendre

Q. 2]

1] 5) Data Source :-

A) Big data :-
both inside and outside organization including traditional.

B) Traditional :-
it is mainly internal

6) Data reporting :-

A) Big data :-
Reporting in all possible direction across the data in real time mode.

B) Traditional :-
Mostly canned with limited and pre-defined interaction paths

7) Cost factor :-

A) Big data :-
Inexpensive commodity in cluster mode

B) Specialized high end hardware and software

8) CAP theorem :-

A) Big data :-
Availibility - Top priority

B) Consistency - top priority