1.Characteristics of Big Data:

->VOLUME:

\*It refers to the large amount of data that needs to be stored.

\*The size varies from terabytes to petabytes.

\*Transactions,records,tables.

->VARIETY:

\*It refers to the fact that a variety or different types of data to be stored.

\*structured data like tables and files,

\*unstructured data like those of facebook,amazon etc.,

\*semi structured data like the additional information i.e meta data, say the names of files,title etc.

->VELOCITY:

\*It refers to the speed at which new data is generated and moves around.

\*Near-Time,Real-Time ...

\*for eg,the climatic condition changes.

2. Possible solutions to handle Big data:

->Scale Up:

\* It can be used to increase the configuration of a single system.

\* They maybe disk capacity, RAM, data transfer speed.

\* It is Complex, costly, and a time consuming process.

-> Scale Out:

\*It Uses multiple inexpensive (commodity)machines and distribute the load of storage or processing among those machines.

\*It is highly Economical and provides a quick way to implement as it works on distribution of load.

\*eg,It is feasible to use 20 machines with 100 GB storage instead of having a single system with 100 GB storage.

3. Differences between scaling up and scaling out:

|  |  |
| --- | --- |
| SCALE UP: | SCALE OUT: |
| \*Adds up the resources /increases the configuration of a single system. | \*It distributes the load among multiple machines. |
| \*It’s costly,less reliable compared to scale out. | \*It is highly reliable and efficient. |
| \* | \* |