1. Characteristics of bigdata:
2. VOLUME:

The name 'Big Data' itself is related to a size which is enormous. Size of data plays very crucial role in determining value out of data.Also, whether a particular data can actually be considered as a Big Data or not, is dependent upon volume of data. Hence, 'Volume' is one characteristic which needs to be considered while dealing with 'Big Data'.

Volume Refers to the vast amounts of data generated every second. We are not talking Terabytes but Zettabytes or Brontobytes.

If we take all the data generated in the world between the beginning of time and 2008, the same amount of data will soon be generated every minute. This makes most data sets too large to store and analyze using traditional database technology. New big data tools use distributed systems so that we can store and analyze data across databases that are dotted around anywhere in the world.

1. VELOCITY:

The term 'velocity' refers to the speed of generation of data. How fast the data is generated and processed to meet the demands, determines real potential in the data.Big Data Velocity deals with the speed at which data flows in from sources like business processes, application logs, networks and social media sites, sensors, mobile devices, etc. The flow of data is massive and continuous.

Velocity Refers to the speed at which new data is generated and the speed at which data moves around.

The Velocity is the speed at which the data is created, stored, analyzed and visualized.

The speed at which data is created currently is almost unimaginable.

1. VARIETY:

The next aspect of 'Big Data' is its variety.Variety refers to heterogeneous sources and the nature of data, both structured and unstructured.

During earlier days, spreadsheets and databases were the only sources of data considered by most of the applications. Now days, data in the form of emails, photos, videos, monitoring devices, PDFs, audio, etc. is also being considered in the analysis applications. This variety of unstructured data poses certain issues for storage, mining and analysing data.

Variety Refers to the different types of data we can now use. In the past we only focused on structured data that neatly fitted into tables or relational databases, such as financial data. In fact, 80% of the world’s data is unstructured (text, images, video, voice, etc.) With big data technology we can now analyze and bring together data of different types such as messages, social media conversations, photos, sensor data, video or voice recordings.

In the past, all data that was created was structured data, it neatly fitted in columns and rows but those days are over. Nowadays, 90% of the data that is generated by an organization is unstructured data. Data today comes in many different formats: structured data, semi-structured data, unstructured data and even complex structured data. The wide variety of data requires a different approach as well as different techniques to store all raw data.

There are many different types of data and each of those types of data require different types of analyses or different tools to use. Social media like Facebook posts

(IV)VERACITY:

Big Data Veracity refers to the biases, noise and abnormality in data. Is the data that is being stored, and mined meaningful to the problem being analyzed. Inderpal feel veracity in data analysis is the biggest challenge when compares to things like volume and velocity. In scoping out your big data strategy you need to have your team and partners work to help keep your data clean and processes to keep ‘dirty data’ from accumulating in your systems.

(V)VALUE:

The most important element of the big data is value.  Value that includes a large volume and variety of data that is easy to access and delivers quality analytics that enables informed decisions.  Providing a [fair market valuation](http://www.sagese.com/resources/their-market-value-or-fair-market-value) on used technology - one piece or an entire portfolio at a time.

2.POSSIBLE SOLUTION TO BIG DATA:

There are three possible solution to tackle bigdata

1. Scale up:The computing resources on a node via parallel processing & faster memory storage.
2. Scale out:The computing to distributed nodes in a cluster/node or at the edge
3. Scale down:The amoun of data processed or the resources needed to perform processing

3.Difference between scaling up and scaling out:

(i) Scaling up:

* Scale-up can solve a capacity problem without adding infrastructure elements such as network connectivity. However, it does require additional space, power, and cooling. Scaling up does not add controller capabilities to handle additional host activities. That means it doesn’t add costs for extra control functions either.

So the costs have not scaled at the same rate for the initial storage system plus storage devices – only additional devices have been added.

* Scaling up generally refers to purchasing and installing a more capable central control or piece of hardware. For example, when a project’s input/output demands start to push against the limits of an individual server, a scaling up approach would be to buy a more capable server with more processing capacity and RAM.

(ii) Scaling down:

* Scale-out storage usually requires additional storage (called nodes) to add capacity and performance. Or in the case of monolithic storage systems, it scales by adding more functional elements (usually controller cards).One difference between scaling out and just putting more storage systems on the floor is that scale-out storage continues to be represented as a single system.

There are several methods for accomplishing scale out, including clustered storage systems and grid storage. The definitions of these two types can also be confusing, and other factors add to the complexity (that’s a subject for another article), but the fundamental premise is that a scale-out solution is accessed as a single system.

* scaling out means linking together other lower-performance machines to collectively do the work of a much more advanced one. With these types of distributed setups, it's easy to handle a larger workload by running data through different system trajectories.