



|  |  |  |
| --- | --- | --- |
|  | 4v of big data à®à¯à®à®¾à®© à®ªà® à®®à¯à®à®¿à®µà¯ | 11 vs of big data à®à¯à®à®¾à®© à®ªà® à®®à¯à®à®¿à®µà¯ |

|  |  |  |
| --- | --- | --- |
|  |  |  |

1.Large Volum

2.Multiple types of data

3.Rapaidly Growing

4.value correct meaning of Data

5.Veracity correcting quality of data( if it is not correct prediction may be wrong)

11.

1. **Vagueness:** The meaning of found data is often very unclear, regardless of how *much* data is available.
2. **Validity:** Rigor in analysis (e.g., [Target Shuffling](http://www.elderresearch.com/target-shuffling)) is essential for valid predictions.
3. **Valor:** In the face of big data, we must gamely tackle the big problems.
4. **Value:** Data science continues to provide ever-increasing value for users as more data becomes available and new techniques are developed.
5. **Vane:** Data science can point in the direction of correct decision making.
6. **Vanilla:** Even the simplest models, constructed with rigor, can provide value.
7. **Vantage:** Big data allows us a privileged view of complex systems.
8. **Variability:** Data science often models variable data sources. Models deployed into production can encounter especially wild data.
9. **Variety:** In data science, [we work with many data formats](https://www.mapr.com/blog/top-10-big-data-challenges-serious-look-10-big-data-vs) (flat files, relational databases, graph networks) and varying levels of data completeness.
10. **Varifocal:** Big data and data science together allow us to see both the forest *and* the trees.
11. **Varmint:** As big data gets bigger, so can software bugs!
12. **Varnish:** How end-users interact with our work matters, and polish counts.
13. **Vastness:** With the advent of the internet of things, the "bigness" of big data is accelerating.
14. **Vaticination:** Predictive analytics provides the ability to forecast. (Of course, these forecasts can be more or less accurate depending on rigor and the complexity of the problem. The future is pesky and never conforms to our March Madness brackets.)
15. **Vault:** With many data science applications based on large and often sensitive data sets, data security is increasingly important.
16. **Veer:** With the rise of [agile data science](http://www.elderresearch.com/consulting-services/agile-data-science), we should be able to navigate the customer's needs and change directions quickly when called upon.
17. **Veil:** Data science provides the capability to peer behind the curtain and examine the effects of latent variables in the data.
18. **Velocity:** Not only is the volume of data ever increasing, but the rate of data generation (from the internet of things, social media, etc.) is increasing as well.
19. **Venue:** [Data science work takes place in different locations and under different arrangements](https://www.mapr.com/blog/top-10-big-data-challenges-serious-look-10-big-data-vs): Locally, on customer workstations, and in the cloud.
20. **Veracity:** Reproducibility is essential for accurate analysis.
21. **Verdict:** As an increasing number of people are affected by models' decisions, Veracity and Validity become ever more important.
22. **Versed:** Data scientists often need to know a little about a great many things: mathematics, statistics, programming, databases, etc.
23. **Version Control:** You're using it, right?
24. **Vet:** Data science allows us to vet our assumptions, augmenting intuition with evidence.
25. **Vexed:** Some of the excitement around data science is based on its potential to shed light on large, complicated problems.
26. **Viability:** It is difficult to build robust models, and it's harder still to build systems that will be [viable in production](http://www.elderresearch.com/company/blog/making-analytics-work-in-production).
27. **Vibrant:** A thriving data science community is vital, and it provides insights, ideas, and support in all of our endeavors.
28. **Victual:** Big data — the food that fuels data science.
29. **Viral:** [How does data spread among other users and applications?](http://www.datasciencecentral.com/profiles/blogs/how-many-v-s-in-big-data-the-characteristics-that-define-big-data)
30. **Virtuosity:** If data scientists need to know a little about many things, we should also grow to know a lot about one thing.
31. **Viscosity:** Related to Velocity; how difficult is the data to work with?
32. **Visibility:** Data science provides visibility into complex big data problems.
33. **Visualization:** Often the *only* way customers interact with models.
34. **Vivify:** Data science has the potential to animate all manner of decision making and business processes, from advertising to fraud detection.
35. **Vocabulary:** Data science provides a vocabulary for addressing a variety of problems. Different modeling approaches tackle different problem domains, and different validation techniques harden these approaches in different applications.
36. **Vogue:** "Machine Learning" which becomes "Artificial Intelligence", which becomes...?
37. **Voice:** Data science provides the ability to speak with knowledge (though not *all* knowledge, of course) on a diverse range of topics.
38. **Volatility:** Especially in production systems, one has to prepare for data volatility. Data that should "never" be missing suddenly disappears, numbers suddenly contain characters!
39. **Volume:** More people use data-collecting devices as more devices become internet-enabled. The volume of data is [increasing at a staggering rate](http://www.infodocket.com/2014/04/16/how-large-is-the-digital-universe-how-fast-is-it-growing-2014-emc-digital-universe-study-now-available/).
40. **Voodoo:** Data science and big data aren't voodoo, but how can we convince potential customers of data science's value to deliver results with real-world impact?
41. **Voyage:** May we always keep learning as we tackle the problems that data science provides.
42. **Vulpine:** [Nate Silver would like you to be a fox, please](https://fivethirtyeight.com/features/what-the-fox-knows/).

Besides our own additions, this list of *V's* incorporates the lists of several articles written over the last few years:

**Structured Data**

* Your current data warehouse contains structured data and only structured data.
* It’s structured because when you placed it in your relational database system a structure was enforced on it, so we know where it is, what it means, and how it relates to other pieces of data in there.
* It may be text (a person’s name) or numerical (their age) but we know that the age value goes with a specific person, hence structured.

**Semi-Structured Data**

* The line between unstructured data and semi-structured is a little fuzzy.
* If the data has any organizational structure (a known schema) or carries a tag (like XML extensible markup language used for documents on the web) then it is somewhat easier to organize and analyze, and because it is more accessible for analysis may make it more valuable.

Example: Text ( XML, Email), Web Server logs and server patterns, sensor data

# Spatial Data

Spatial data refers to all types of data objects or elements that are present in a geographical space or horizon. It enables the global finding and locating of individuals  or devices.

Spatial data is used in geographical information systems (GIS) and other geolocation or positioning services. Spatial data consists of points, lines, polygons and other geographic and geometric data primitives, which can be mapped by location, stored with an object as metadata

|  |
| --- |
| Technology is Business enablers 🡺 using this technology they are going to make more $ |
|  |
| 1.Open soure 🡺  2.Reliability  3.Economic  4.Fault Tolerence 🡺 Rack awareness  5.Flexibe 🡺  6.Scalable |

|  |
| --- |
| 4.3 Minutes |
|  |

|  |  |
| --- | --- |
| Hadoop 1.X (Classic) Architecture |  |
|  | |
| SUBMITTED, ACCEPTED, RUNNING, FINISHED, FAILED, KILLED | |