Events          Online Learning          Onsite Education          **Membership**

Research & Resources          More Ways to Learn

# UP

DATA          BUSINESS

Upside Home          Trends in Analytics          Data & Business Leadership

IT & Enterprise Data Management          Practical Data Science

Share          Tweet          Share

# The 10 Vs of Big Data

Big data goes beyond volume, variety, and velocity alone. You
need to know these 10 characteristics and properties of big data

## Sign Up Now

**Explore by Topic**

Advanced Analytics

Artificial Intelligence
(AI) and Machine
Learning

Big Data

Blockchain

Career Development

to prepare for both the challenges and advantages of big data
initiatives.

By George Firican

February 8, 2017

The term *big data* started to show up sparingly in the early 1990s,
and its prevalence and importance increased exponentially as
years passed. Nowadays big data is often seen as integral to a
company's data strategy.

Big data has specific characteristics and properties that can help
you understand both the challenges and advantages of big data
initiatives.

You may have heard of the three Vs of big data, but I believe there
are seven additional important characteristics you need to know.
Conveniently, these properties each start with v as well, so let's
discuss the 10 Vs of big data.

### #1: Volume

Volume is probably the best known characteristic of big data; this
is no surprise, considering more than 90 percent of all today's
data was created in the past couple of years. The current amount
of data can actually be quite staggering. Here are some examples:

-- 300 hours of video are uploaded to YouTube every minute.

-- An estimated 1.1 trillion photos were taken in 2016, and that
number is projected to rise by 9 percent in 2017. As the same
photo usually has multiple instances stored across different
devices, photo or document sharing services as well as social
media services, the total number of photos stored is also
expected to grow from 3.9 trillion in 2016 to 4.7 trillion in 2017.

-- In 2016 estimated global mobile traffic amounted for 6.2
exabytes per month. That's 6.2 billion gigabytes.

### #2: Velocity

Velocity refers to the speed at which data is being generated,
produced, created, or refreshed.

Sure, it sounds impressive that Facebook's data warehouse stores upwards of 300 petabytes of data, but the velocity at which new data is created should be taken into account. Facebook claims 600 terabytes of incoming data per day.

Google alone processes on average more than "40,000 search queries every second," which roughly translates to more than 3.5 billion searches per day.

## #3: Variety

When it comes to big data, we don't only have to handle structured data but also semistructured and mostly unstructured data as well. As you can deduce from the above examples, most big data seems to be unstructured, but besides audio, image, video files, social media updates, and other text formats there are also log files, click data, machine and sensor data, etc.

## #4: Variability

Variability in big data's context refers to a few different things. One is the number of inconsistencies in the data. These need to be found by anomaly and outlier detection methods in order for any meaningful analytics to occur.

Big data is also variable because of the multitude of data dimensions resulting from multiple disparate data types and sources. Variability can also refer to the inconsistent speed at which big data is loaded into your database.

## #5: Veracity

This is one of the unfortunate characteristics of big data. As any or all of the above properties increase, the veracity (confidence or trust in the data) drops. This is similar to, but not the same as, validity or volatility (see below). Veracity refers more to the provenance or reliability of the data source, its context, and how meaningful it is to the analysis based on it.

For example, consider a data set of statistics on what people purchase at restaurants and these items' prices over the past five years. You might ask: Who created the source? What
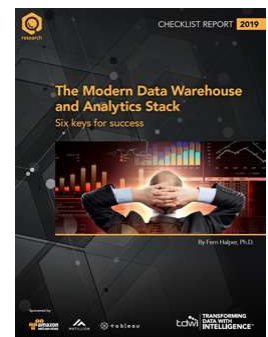
methodology did they follow in collecting the data? Were only certain cuisines or certain types of restaurants included? Did the data creators summarize the information? Has the information been edited or modified by anyone else?

Answers to these questions are necessary to determine the veracity of this information. Knowledge of the data's veracity in turn helps us better understand the risks associated with analysis and business decisions based on this particular data set.

## #6: Validity

Similar to veracity, validity refers to how accurate and correct the data is for its intended use. According to Forbes, an estimated 60 percent of a data scientist's time is spent cleansing their data before being able to do any analysis. The benefit from big data analytics is only as good as its underlying data, so you need to adopt good data governance practices to ensure consistent data quality, common definitions, and metadata.

## #7: Vulnerability

Big data brings new security concerns. After all, a data breach with big data is a big breach. Does anyone remember the infamous AshleyMadison hack in 2015?

Unfortunately there have been many big data breaches. Another example, as reported by CRN: in May 2016 "a hacker called Peace posted data on the dark web to sell, which allegedly included information on 167 million LinkedIn accounts and ... 360 million emails and passwords for MySpace users."

Information on many others can be found at Information is Beautiful.

## #8: Volatility

How old does your data need to be before it is considered irrelevant, historic, or not useful any longer? How long does data need to be kept for?

Before big data, organizations tended to store data indefinitely -- a few terabytes of data might not create high storage expenses; it

a few terabytes of data might not create high storage expenses, it could even be kept in the live database without causing performance issues. In a classical data setting, there not might even be data archival policies in place.

Due to the velocity and volume of big data, however, its volatility needs to be carefully considered. You now need to establish rules for data currency and availability as well as ensure rapid retrieval of information when required. Make sure these are clearly tied to your business needs and processes -- with big data the costs and complexity of a storage and retrieval process are magnified.

#9: Visualization

Another characteristic of big data is how challenging it is to visualize.

Current big data visualization tools face technical challenges due to limitations of in-memory technology and poor scalability, functionality, and response time. You can't rely on traditional graphs when trying to plot a billion data points, so you need different ways of representing data such as data clustering or using tree maps, sunbursts, parallel coordinates, circular network diagrams, or cone trees.

Combine this with the multitude of variables resulting from big data's variety and velocity and the complex relationships between them, and you can see that developing a meaningful visualization is not easy.

#10: Value

Last, but arguably the most important of all, is value. The other characteristics of big data are meaningless if you don't derive business value from the data.

Substantial value can be found in big data, including understanding your customers better, targeting them accordingly, optimizing processes, and improving machine or business performance. You need to understand the potential, along with the more challenging characteristics, before embarking on a big data strategy.

For Further Reading:

- Big Data and Analytics Spending Projected to Soar
- Adapting Data Governance to Big Data
- You Still Need a Model! Data Modeling for Big Data and NoSQL

## About the Author

**George Firican** is the director of data governance and business intelligence at the University of British Columbia. His innovative approach to data management received international recognition through award-winning program implementations in the data governance, data quality, and business intelligence fields. As a passionate advocate for the importance of data, he founded www.lightsondata.com, he is a frequent conference speaker, advises organizations about how to treat data as an asset, and shares practical takeaways on social media, industry sites, and in publications. He can be followed on Twitter (@georgefirican) or reached via email, or on LinkedIn.

### Related Articles

- How Process Mining Can Aid Your Digital Transformation
- Big Data, Infrastructure, and Performance
- 3 Fundamental Steps for Strong Big Data Security

### Trending Articles



**Data Stories: Creating Better Data Visualizations**

**CEO Perspective: How Marketers Can Get the Most from Analytics**

**Data Digest: Forecasts Begin for Skills, Tech, and Storage**

**AI's Impact in 2020: 3 Trends to Watch**

# TDWI MEMBERSHIP
## Accelerate Your Projects, and Your Career

TDWI Members have access to exclusive research reports, publications, communities and training.

### Individual, Student, & Team memberships available.

Membership Information

## FOLLOW TDWI

LinkedIn

Twitter

Facebook

YouTube

Instagram

## LINKS

Contact Us

Marketing Opportunities

Become an Instructor or Contributor

Become a TDWI Training Partner

Media Downloads

Author Submissions

Reprints & Permissions

Privacy Policy

Vendor News

Newsletters

Press Releases

Sitemap

TDWI Europe

## SUBSCRIBE TO OUR NEWSLETTER

Subscribe

Privacy Policy
© 2019 TDWI
All Rights Reserved

Become a Member

Become a Member