

Coding Summary By Code

Big Data Lay Definitions

22/02/2019 16:58

Aggregate	Classification	Coverage	Number Of Coding References	Reference Number	Coded By Initials	Modified On
-----------	----------------	----------	-----------------------------	------------------	-------------------	-------------

Node

Nodes\\DEF_Methodology_Technology

Document

Files\\UC Berkeley School of Information 2014

No	0.7240	29	1	XX	22/02/2019 16:23
----	--------	----	---	----	------------------

"Big Data" refers to a combination of an approach to informing decision making with analytical insight derived from data, and a set of enabling technologies that enable that insight to be economically derived from at times very large, diverse sources of data. Advances in sensing technologies, the digitization of commerce and communications, and the advent and growth in social media are a few of the trends which have created the opportunity to use large scale, fine grained data to understand systems, behavior and commerce; while innovation in technology makes it viable economically to use that information to inform decisions and improve outcomes.

2 XX 22/02/2019 16:23

While the use of the term is quite nebulous and is often co-opted for other purposes, I've understood "big data" to be about analysis for data that's really messy or where you don't know the right questions or queries to make — analysis that can help you find patterns, anomalies, or new structures amidst otherwise chaotic or complex data points. Usually this revolves around datasets with a byte size that seems fairly large relative to our frame of reference using files on a desktop PC (e.g., larger than a terabyte) and many of the tools around big data are to help deal with a large volume of data, but to me the most important concepts of big data don't actually have much to do with it being "big" in this sense (especially since that's such a relative term these days). In fact, they can often be applied to smaller datasets as well. Natural language processing and lucene based search engines are good examples of big data techniques and tools that are often used with relatively small amounts of data.

3 XX 22/02/2019 16:23

Big Data is the result of collecting information at its most granular level— it's what you get when you instrument a system and keep all of the data that your instrumentation is able to gather.

4 XX 22/02/2019 16:23

As computational efficiency continues to increase, "big data" will be less about the actual size of a particular dataset and more about the specific expertise needed to process it. With that in mind, "big data" will ultimately describe any dataset large enough to necessitate high-level programming skill and statistically defensible methodologies in order to transform the data asset into something of value.

5 XX 22/02/2019 16:23

You cannot give me too much data. I see big data as storytelling — whether it is through information graphics or other visual aids that explain it in a way that allows others to understand across sectors. I always push for the full scope of the data over averages and aggregations — and I like to go to the raw data because of the possibilities of things you can do with it.

6 XX 22/02/2019 16:23

Big data, which started as a technological innovation in distributed computing, is now a cultural movement by which we continue to discover how humanity interacts with the world — and each other — at large-scale.

7 XX 22/02/2019 16:23

Big data refers to the approach to data of "collect now, sort out later"...meaning you capture and store data on a very large volume of actions and transactions of different types, on a continuous basis, in order to make sense of it later. The low cost of storage and better methods of analysis mean that you generally don't need to have a specific purpose for the data in mind before you collect it.

Aggregate	Classification	Coverage	Number Of Coding References	Reference Number	Coded By Initials	Modified On
				8	XX	22/02/2019 16:23
[Big data is]	an opportunity to gain a more complex understanding of the relationships between different factors and to uncover previously undetected patterns in data by leveraging advances in the technical aspects of collecting, storing, and retrieving data along with innovative ideas and techniques for manipulating and analyzing data.					
				9	XX	22/02/2019 16:23
Big data is the broad name given to challenges and opportunities we have as data about every aspect of our lives becomes available. It's not just about data though; it also includes the people, processes, and analysis that turn data into meaning.						
				10	XX	22/02/2019 16:23
I prefer a flexible but functional definition of big data. Big data is when your business wants to use data to solve a problem, answer a question, produce a product, etc., but the standard, simple methods (maybe it's SQL, maybe it's k-means, maybe it's a single server with a cron job) break down on the size of the data set, causing time, effort, creativity, and money to be spent crafting a solution to the problem that leverages the data without simply sampling or tossing out records.						
				11	XX	22/02/2019 16:23
Historically, most decisions — political, military, business, and personal — have been made by brains [that] have unpredictable logic and operate on subjective experiential evidence. "Big data" represents a cultural shift in which more and more decisions are made by algorithms with transparent logic, operating on documented immutable evidence. I think "big" refers more to the pervasive nature of this change than to any particular amount of data.						
				12	XX	22/02/2019 16:23
Big data is data that even when efficiently compressed still contains 5-10 times more information (measured in entropy or predictive power, per unit of time) than what you are used to right now. It may require a different approach to extract value.						
				13	XX	22/02/2019 16:23
Big data is data that contains enough observations to demand unusual handling because of its sheer size, though what is unusual changes over time and varies from one discipline to another. Scientific computing is accustomed to pushing the envelope, constantly developing techniques to address relentless growth in dataset size, but many other disciplines are now just discovering the value — and hence the challenges — of working with data at the unwieldy end of the scale.						
				14	XX	22/02/2019 16:23
Big data has taken a beating in recent years, the accusation being that marketers and analysts have stretched and squeezed the term to cover a multitude of disparate problems, technologies, and products. Yet the core of big data remains what it has been for over a decade, framed by Doug Laney's 2001 three Vs, Volume, Velocity, and Variety, and indicating data challenges sufficient to justify non-routine computing resources and processing techniques.						
				15	XX	22/02/2019 16:23
Big data describes datasets that are so large, complex, or rapidly changing that they push the very limits of our analytical capability. It's a subjective term: What seems "big" today may seem modest in a few years when our analytic capacity has improved. While big data can be about anything, the most important kinds of big data — and perhaps the only ones worth the effort — are those that can have a big impact through what they tell us about society, public health, the economy, scientific research, or any number of other large-scale subjects.						
				16	XX	22/02/2019 16:23
Big data refers to using complex datasets to drive focus, direction, and decision making within a company or organization. This is done by deriving actionable insights from the analysis of your organization's data.						
				17	XX	22/02/2019 16:23
Big data is just the ability to gather information and query it in such a way that we are able to learn things about the world that were previously inaccessible to us.						
				18	XX	22/02/2019 16:23
Everything we know spits out data today — not just the devices we use for computing. We now get digital exhaust from our garage door openers to our coffee pots, and everything in between. At the same time, we have become a generation of people who demand instantaneous access to information — from what the weather is like in a country thousands of miles away to which store has better deals on toaster ovens. Big data is at the intersection of collecting, organizing, storing, and turning all of that raw data into truly meaningful information.						

Aggregate	Classification	Coverage	Number Of Coding References	Reference Number	Coded By Initials	Modified On
				19	XX	22/02/2019 16:23
At Aster Data, we originally used the term big data in our marketing to refer to analytical MPP databases like ours and to differentiate them from traditional data warehouse software. While both were capable of storing a “big” volume of data (which, in 2008, we defined as 10 TB or greater), “big data” systems were capable of performing complex analytics on top of that data — something that legacy data warehouse software could not do. Thus, our original definition was a system that (1) was capable of storing 10 TB of data or more and (2) was capable of executing advanced workloads, such as behavioral analytics or market basket analysis, on those large volumes of data. As time went on, diversity of data started to become more prevalent in these systems (particularly the need to mix structured and unstructured data), which led to more widespread adoption of the “3 Vs” (volume, velocity, and variety) as a definition for big data, which continues to this day.						
				20	XX	22/02/2019 16:23
In my view, big data is data that requires novel processing techniques to handle. Typically, big data requires massive parallelism in some fashion (storage and/or compute) to deal with volume and processing variety.						
				21	XX	22/02/2019 16:23
As our lives have moved from the physical to the digital world, our everyday tools like smartphones and ubiquitous Internet, create vast amounts of data. One of the best interpretations of the “big” in “big data” is expansive — whether you are a Fortune 500 company who just released an app that is creating a torrent of user data about every click and every activity of every user or a nonprofit who just launched a cellphone-based app to find the closest homeless shelters that are now spewing forth information about every search and every click, we all have data. Dealing with this so-called big data requires a massive shift in technologies for storing, processing, and managing data — but also presents tremendous opportunity for the social sector to gather and analyze information faster to address some of our world’s most pressing challenges.						
				22	XX	22/02/2019 16:23
I’m not fond of the phrase “big data” because it focuses on the volume of data, obscuring the far-reaching changes are making data essential to individuals and organizations in today’s world. But if I have to define it I’d say that “big data” is data that can’t be processed using standard databases because it is too big, too fast-moving, or too complex for traditional data processing tools.						
				23	XX	22/02/2019 16:23
The rising accessibility of platforms for the storage and analysis of large amounts of data (and the falling price per TB of doing so) has made it possible for a wide variety of organizations to store nearly all data in their purview — every log line, customer interaction, and event — unaggregated and for a significant period of time. The associated ethos of “store everything now and ask questions later” to me more than anything else characterizes how the world of computational systems looks under the lens of modern “big data” systems.						
				24	XX	22/02/2019 16:23
Big data originally described the practice in the consumer Internet industry of applying algorithms to increasingly large amounts of disparate data to solve problems that had suboptimal solutions with smaller datasets. Many features and signals can only be observed by collecting massive amounts of data (for example, the relationships across an entire social network), and would not be detected using smaller samples. Processing large datasets in this manner was often difficult, time consuming, and error prone before the advent of technologies like MapReduce and Hadoop, which ushered in a wave of related tools and applications now collectively called big data technologies.						
				25	XX	22/02/2019 16:23
Big data is when data grows to the point that the technology supporting the data has to change. It also encompasses a variety of topics relating to how disparate data can be combined, processed into insights, and/or reworked into smart products.						
				26	XX	22/02/2019 16:23
Big data used to mean data that a single machine was unable to handle. Now big data has become a buzzword to mean anything related to data analytics or visualization.						
				27	XX	22/02/2019 16:23
Big data is an umbrella term that means a lot of different things, but to me, it means the possibility of doing extraordinary things using modern machine learning techniques on digital data. Whether it is predicting illness, the weather, the spread of infectious diseases, or what you will buy next, it offers a world of possibilities for improving people’s lives.						
				28	XX	22/02/2019 16:23
For me, the technological definitions (like “too big to fit in an Excel spreadsheet” or “too big to hold in memory”) are important, but aren’t really the main point. Big data for me is data at a scale and scope that changes in some fundamental way (not just at the margins) the range of solutions that can be considered when people and organizations face a complex problem. Different solutions, not just ‘more, better.’						

Aggregate	Classification	Coverage	Number Of Coding References	Reference Number	Coded By Initials	Modified On
			29	XX		22/02/2019 16:23

The term big data is really only useful if it describes a quantity of data that's so large that traditional approaches to data analysis are doomed to failure. That can mean that you're doing complex analytics on data that's too large to fit into memory or it can mean that you're dealing with a data storage system that doesn't offer the full functionality of a standard relational database. What's essential is that your old way of doing things doesn't apply anymore and can't just be scaled out.

Nodes\\DEF_Value

Document

Files\\UC Berkeley School of Information 2014

No	0.4032	17	1	XX	22/02/2019 13:40
----	--------	----	---	----	------------------

"Big Data" refers to a combination of an approach to informing decision making with analytical insight derived from data, and a set of enabling technologies that enable that insight to be economically derived from at times very large, diverse sources of data. Advances in sensing technologies, the digitization of commerce and communications, and the advent and growth in social media are a few of the trends which have created the opportunity to use large scale, fine grained data to understand systems, behavior and commerce; while innovation in technology makes it viable economically to use that information to inform decisions and improve outcomes.

2 XX 22/02/2019 13:48

As computational efficiency continues to increase, "big data" will be less about the actual size of a particular dataset and more about the specific expertise needed to process it. With that in mind, "big data" will ultimately describe any dataset large enough to necessitate high-level programming skill and statistically defensible methodologies in order to transform the data asset into something of value.

3 XX 22/02/2019 13:53

Big data refers to the approach to data of "collect now, sort out later"...meaning you capture and store data on a very large volume of actions and transactions of different types, on a continuous basis, in order to make sense of it later. The low cost of storage and better methods of analysis mean that you generally don't need to have a specific purpose for the data in mind before you collect it.

4 XX 22/02/2019 13:55

I prefer a flexible but functional definition of big data. Big data is when your business wants to use data to solve a problem, answer a question, produce a product, etc., but the standard, simple methods (maybe it's SQL, maybe it's k-means, maybe it's a single server with a cron job) break down on the size of the data set, causing time, effort, creativity, and money to be spent crafting a solution to the problem that leverages the data without simply sampling or tossing out records.

5 XX 22/02/2019 13:56

Big data is data that even when efficiently compressed still contains 5-10 times more information (measured in entropy or predictive power, per unit of time) than what you are used to right now. It may require a different approach to extract value.

6 XX 22/02/2019 14:02

Big data is data that contains enough observations to demand unusual handling because of its sheer size, though what is unusual changes over time and varies from one discipline to another. Scientific computing is accustomed to pushing the envelope, constantly developing techniques to address relentless growth in dataset size, but many other disciplines are now just discovering the value — and hence the challenges — of working with data at the unwieldy end of the scale.

7 XX 22/02/2019 14:03

Big data describes datasets that are so large, complex, or rapidly changing that they push the very limits of our analytical capability. It's a subjective term: What seems "big" today may seem modest in a few years when our analytic capacity has improved. While big data can be about anything, the most important kinds of big data — and perhaps the only ones worth the effort — are those that can have a big impact through what they tell us about society, public health, the economy, scientific research, or any number of other large-scale subjects.

Aggregate	Classification	Coverage	Number Of Coding References	Reference Number	Coded By Initials	Modified On
				8	XX	22/02/2019 16:05
What's "big" in big data isn't necessarily the size of the databases, it's the big number of data sources we have, as digital sensors and behavior trackers migrate across the world. As we triangulate information in more ways, we will discover hitherto unknown patterns in nature and society — and pattern-making is the wellspring of new art, science, and commerce.						
				9	XX	22/02/2019 16:05
To me, "big data" is the situation where an organization can (arguably) say that they have access to what they need to reconstruct, understand, and model the part of the world that they care about. Using their big data, then, they can (try to) predict future states of the world, optimize their processes, and otherwise be more effective and rational in their activities.						
				10	XX	22/02/2019 16:12
[Big data means] harnessing more sources of diverse data where "data variety" and "data velocity" are the key opportunities. (Each source represents "a signal" on what is happening in the business.) The opportunity is to harness data variety [and] automate "harmonization" of data sources to deliver fast-updating insights consumable by the line-of-business users.						
				11	XX	22/02/2019 14:07
While "big data" is often large in size relative to the available tool set, "big" actually refers to being important. Scientists and engineers have long known that data is valuable, but now the rest of the world, including those in control of purse strings, understand the value that can be created from data.						
				12	XX	22/02/2019 14:07
Everything we know spits out data today — not just the devices we use for computing. We now get digital exhaust from our garage door openers to our coffee pots, and everything in between. At the same time, we have become a generation of people who demand instantaneous access to information — from what the weather is like in a country thousands of miles away to which store has better deals on toaster ovens. Big data is at the intersection of collecting, organizing, storing, and turning all of that raw data into truly meaningful information.						
				13	XX	22/02/2019 16:16
As our lives have moved from the physical to the digital world, our everyday tools like smartphones and ubiquitous Internet, create vast amounts of data. One of the best interpretations of the "big" in "big data" is expansive — whether you are a Fortune 500 company who just released an app that is creating a torrent of user data about every click and every activity of every user or a nonprofit who just launched a cellphone-based app to find the closest homeless shelters that are now spewing forth information about every search and every click, we all have data. Dealing with this so-called big data requires a massive shift in technologies for storing, processing, and managing data — but also presents tremendous opportunity for the social sector to gather and analyze information faster to address some of our world's most pressing challenges.						
				14	XX	22/02/2019 14:14
The rising accessibility of platforms for the storage and analysis of large amounts of data (and the falling price per TB of doing so) has made it possible for a wide variety of organizations to store nearly all data in their purview — every log line, customer interaction, and event — unaggregated and for a significant period of time. The associated ethos of "store everything now and ask questions later" to me more than anything else characterizes how the world of computational systems looks under the lens of modern "big data" systems.						
				15	XX	22/02/2019 14:18
Big data is not all about volume, it is more about combining different data sets and to analyze it in real-time to get insights for your organization. Therefore, the right definition of big data should in fact be: mixed data.						
				16	XX	22/02/2019 14:22
The joke is that big data is data that breaks Excel, but we try not to be snooty about whether you measure your data in MBs or PBs. Data is more about your team and the results they can get.						
				17	XX	22/02/2019 16:23
Big data enchants us with the promise of new insights. Let's not forget the knowledge hidden in the small data right before us.						

Aggregate	Classification	Coverage	Number Of Coding References	Reference Number	Coded By Initials	Modified On
Nodes\\DEF_Variety						
Document						
Files\\UC Berkeley School of Information 2014						
No		0.4455	18			
				1	XX	22/02/2019 13:40
"Big Data" refers to a combination of an approach to informing decision making with analytical insight derived from data, and a set of enabling technologies that enable that insight to be economically derived from at times very large, diverse sources of data. Advances in sensing technologies, the digitization of commerce and communications, and the advent and growth in social media are a few of the trends which have created the opportunity to use large scale, fine grained data to understand systems, behavior and commerce; while innovation in technology makes it viable economically to use that information to inform decisions and improve outcomes.						
				2	XX	22/02/2019 15:43
While the use of the term is quite nebulous and is often co-opted for other purposes, I've understood "big data" to be about analysis for data that's really messy or where you don't know the right questions or queries to make — analysis that can help you find patterns, anomalies, or new structures amidst otherwise chaotic or complex data points. Usually this revolves around datasets with a byte size that seems fairly large relative to our frame of reference using files on a desktop PC (e.g., larger than a terabyte) and many of the tools around big data are to help deal with a large volume of data, but to me the most important concepts of big data don't actually have much to do with it being "big" in this sense (especially since that's such a relative term these days). In fact, they can often be applied to smaller datasets as well. Natural language processing and lucene based search engines are good examples of big data techniques and tools that are often used with relatively small amounts of data.						
				3	XX	22/02/2019 13:53
Big data refers to the approach to data of "collect now, sort out later" ...meaning you capture and store data on a very large volume of actions and transactions of different types, on a continuous basis, in order to make sense of it later. The low cost of storage and better methods of analysis mean that you generally don't need to have a specific purpose for the data in mind before you collect it.						
				4	XX	22/02/2019 15:57
Big data is the broad name given to challenges and opportunities we have as data about every aspect of our lives becomes available. It's not just about data though; it also includes the people, processes, and analysis that turn data into meaning.						
				5	XX	22/02/2019 14:02
Big data has taken a beating in recent years, the accusation being that marketers and analysts have stretched and squeezed the term to cover a multitude of disparate problems, technologies, and products. Yet the core of big data remains what it has been for over a decade, framed by Doug Laney's 2001 three Vs, Volume, Velocity, and Variety, and indicating data challenges sufficient to justify non-routine computing resources and processing techniques.						
				6	XX	22/02/2019 14:03
Big data describes datasets that are so large, complex, or rapidly changing that they push the very limits of our analytical capability. It's a subjective term: What seems "big" today may seem modest in a few years when our analytic capacity has improved. While big data can be about anything, the most important kinds of big data — and perhaps the only ones worth the effort — are those that can have a big impact through what they tell us about society, public health, the economy, scientific research, or any number of other large-scale subjects.						
				7	XX	22/02/2019 14:04
What's "big" in big data isn't necessarily the size of the databases, it's the big number of data sources we have, as digital sensors and behavior trackers migrate across the world. As we triangulate information in more ways, we will discover hitherto unknown patterns in nature and society — and pattern-making is the wellspring of new art, science, and commerce.						
				8	XX	22/02/2019 14:05
Big data refers to using complex datasets to drive focus, direction, and decision making within a company or organization. This is done by deriving actionable insights from the analysis of your organization's data.						
				9	XX	22/02/2019 14:06
[Big data means] harnessing more sources of diverse data where "data variety" and "data velocity" are the key opportunities. (Each source represents "a signal" on what is happening in the business.) The opportunity is to harness data variety [and] automate "harmonization" of data sources to deliver fast-updating insights consumable by the line-of-business users.						

Aggregate	Classification	Coverage	Number Of Coding References	Reference Number	Coded By Initials	Modified On
				10	XX	22/02/2019 14:09
At Aster Data, we originally used the term big data in our marketing to refer to analytical MPP databases like ours and to differentiate them from traditional data warehouse software. While both were capable of storing a "big" volume of data (which, in 2008, we defined as 10 TB or greater), "big data" systems were capable of performing complex analytics on top of that data — something that legacy data warehouse software could not do. Thus, our original definition was a system that (1) was capable of storing 10 TB of data or more and (2) was capable of executing advanced workloads, such as behavioral analytics or market basket analysis, on those large volumes of data. As time went on, diversity of data started to become more prevalent in these systems (particularly the need to mix structured and unstructured data), which led to more widespread adoption of the "3 Vs" (volume, velocity, and variety) as a definition for big data, which continues to this day.						
				11	XX	22/02/2019 14:10
In my view, big data is data that requires novel processing techniques to handle. Typically, big data requires massive parallelism in some fashion (storage and/or compute) to deal with volume and processing variety.						
				12	XX	22/02/2019 14:14
I'm not fond of the phrase "big data" because it focuses on the volume of data, obscuring the far-reaching changes are making data essential to individuals and organizations in today's world. But if I have to define it I'd say that "big data" is data that can't be processed using standard databases because it is too big, too fast-moving, or too complex for traditional data processing tools.						
				13	XX	22/02/2019 14:15
The rising accessibility of platforms for the storage and analysis of large amounts of data (and the falling price per TB of doing so) has made it possible for a wide variety of organizations to store nearly all data in their purview — every log line, customer interaction, and event — unaggregated and for a significant period of time. The associated ethos of "store everything now and ask questions later" to me more than anything else characterizes how the world of computational systems looks under the lens of modern "big data" systems.						
				14	XX	22/02/2019 14:16
Big data originally described the practice in the consumer Internet industry of applying algorithms to increasingly large amounts of disparate data to solve problems that had suboptimal solutions with smaller datasets. Many features and signals can only be observed by collecting massive amounts of data (for example, the relationships across an entire social network), and would not be detected using smaller samples. Processing large datasets in this manner was often difficult, time consuming, and error prone before the advent of technologies like MapReduce and Hadoop, which ushered in a wave of related tools and applications now collectively called big data technologies.						
				15	XX	22/02/2019 14:16
Big data is when data grows to the point that the technology supporting the data has to change. It also encompasses a variety of topics relating to how disparate data can be combined, processed into insights, and/or reworked into smart products.						
				16	XX	22/02/2019 14:18
Big data is not all about volume, it is more about combining different data sets and to analyze it in real-time to get insights for your organization. Therefore, the right definition of big data should in fact be: mixed data.						
				17	XX	22/02/2019 14:19
Big data means data that cannot fit easily into a standard relational database.						
				18	XX	22/02/2019 14:20
I'm happy to repeat the definition I've heard used and think appropriately defines the over[all] subject. I believe it's Forrester's definition of Volume, Velocity, Variety, and Variability. A lot of different data coming fast and in different structures.						

Aggregate	Classification	Coverage	Number Of Coding References	Reference Number	Coded By Initials	Modified On
-----------	----------------	----------	-----------------------------	------------------	-------------------	-------------

Nodes\\DEF_Velocity

Document

Files\\UC Berkeley School of Information 2014

No	0.2922	11
----	--------	----

1	XX	22/02/2019 15:48
---	----	------------------

Big data refers to the approach to data of “collect now, sort out later”...meaning you capture and store data on a very large volume of actions and transactions of different types, on a continuous basis, in order to make sense of it later. The low cost of storage and better methods of analysis mean that you generally don’t need to have a specific purpose for the data in mind before you collect it.

2	XX	22/02/2019 14:02
---	----	------------------

Big data has taken a beating in recent years, the accusation being that marketers and analysts have stretched and squeezed the term to cover multitude of disparate problems, technologies, and products. Yet the core of big data remains what it has been for over a decade, framed by Doug Laney’s 2001 three Vs, Volume, Velocity, and Variety, and indicating data challenges sufficient to justify non-routine computing resources and processing techniques.

3	XX	22/02/2019 14:03
---	----	------------------

Big data describes datasets that are so large, complex, or rapidly changing that they push the very limits of our analytical capability. It’s a subjective term: What seems “big” today may seem modest in a few years when our analytic capacity has improved. While big data can be about anything, the most important kinds of big data — and perhaps the only ones worth the effort — are those that can have a big impact through what they tell us about society, public health, the economy, scientific research, or any number of other large-scale subjects.

4	XX	22/02/2019 14:06
---	----	------------------

[Big data means] harnessing more sources of diverse data where “data variety” and “data velocity” are the key opportunities. (Each source represents “a signal” on what is happening in the business.) The opportunity is to harness data variety [and] automate “harmonization” of data sources to deliver fast-updating insights consumable by the line-of-business users.

5	XX	22/02/2019 14:07
---	----	------------------

Everything we know spits out data today — not just the devices we use for computing. We now get digital exhaust from our garage door openers to our coffee pots, and everything in between. At the same time, we have become a generation of people who demand instantaneous access to information — from what the weather is like in a country thousands of miles away to which store has better deals on toaster ovens. Big data is at the intersection of collecting, organizing, storing, and turning all of that raw data into truly meaningful information.

6	XX	22/02/2019 14:09
---	----	------------------

At Aster Data, we originally used the term big data in our marketing to refer to analytical MPP databases like ours and to differentiate them from traditional data warehouse software. While both were capable of storing a “big” volume of data (which, in 2008, we defined as 10 TB or greater), “big data” systems were capable of performing complex analytics on top of that data — something that legacy data warehouse software could not do. Thus, our original definition was a system that (1) was capable of storing 10 TB of data or more and (2) was capable of executing advanced workloads, such as behavioral analytics or market basket analysis, on those large volumes of data. As time went on, diversity of data started to become more prevalent in these systems (particularly the need to mix structured and unstructured data), which led to more widespread adoption of the “3 Vs” (volume, velocity, and variety) as a definition for big data, which continues to this day.

7	XX	22/02/2019 14:12
---	----	------------------

As our lives have moved from the physical to the digital world, our everyday tools like smartphones and ubiquitous Internet, create vast amounts of data. One of the best interpretations of the “big” in “big data” is expansive — whether you are a Fortune 500 company who just released an app that is creating a torrent of user data about every click and every activity of every user or a nonprofit who just launched a cellphone-based app to find the closest homeless shelters that are now spewing forth information about every search and every click, we all have data. Dealing with this so-called big data requires a massive shift in technologies for storing, processing, and managing data — but also presents tremendous opportunity for the social sector to gather and analyze information faster to address some of our world’s most pressing challenges.

8	XX	22/02/2019 14:14
---	----	------------------

There is certainly a colorful variety of definitions for the term big data out there. To me it means working with data at a large scale and velocity.

Aggregate	Classification	Coverage	Number Of Coding References	Reference Number	Coded By Initials	Modified On
				9	XX	22/02/2019 14:14
I'm not fond of the phrase "big data" because it focuses on the volume of data, obscuring the far-reaching changes are making data essential to individuals and organizations in today's world. But if I have to define it I'd say that "big data" is data that can't be processed using standard databases because it is too big, too fast-moving, or too complex for traditional data processing tools.						
				10	XX	22/02/2019 14:18
Big data is not all about volume, it is more about combining different data sets and to analyze it in real-time to get insights for your organization. Therefore, the right definition of big data should in fact be: mixed data.						
				11	XX	22/02/2019 14:20
I'm happy to repeat the definition I've heard used and think appropriately defines the over[all] subject. I believe it's Forrester's definition of Volume, Velocity, Variety, and Variability. A lot of different data coming fast and in different structures.						

Nodes\\DEF_Volume

Document

Files\\UC Berkeley School of Information 2014

No	0.5629	22	1	XX	22/02/2019 13:40
"Big Data" refers to a combination of an approach to informing decision making with analytical insight derived from data, and a set of enabling technologies that enable that insight to be economically derived from at times very large, diverse sources of data. Advances in sensing technologies, the digitization of commerce and communications, and the advent and growth in social media are a few of the trends which have created the opportunity to use large scale, fine grained data to understand systems, behavior and commerce; while innovation in technology makes it viable economically to use that information to inform decisions and improve outcomes.					
2			2	XX	22/02/2019 13:48
As computational efficiency continues to increase, "big data" will be less about the actual size of a particular dataset and more about the specific expertise needed to process it. With that in mind, "big data" will ultimately describe any dataset large enough to necessitate high-level programming skill and statistically defensible methodologies in order to transform the data asset into something of value.					
3			3	XX	22/02/2019 13:50
Big data, which started as a technological innovation in distributed computing, is now a cultural movement by which we continue to discover how humanity interacts with the world — and each other — at large-scale.					
4			4	XX	22/02/2019 13:52
Big data refers to the approach to data of "collect now, sort out later"...meaning you capture and store data on a very large volume of actions and transactions of different types, on a continuous basis, in order to make sense of it later. The low cost of storage and better methods of analysis mean that you generally don't need to have a specific purpose for the data in mind before you collect it.					
5			5	XX	22/02/2019 13:55
I prefer a flexible but functional definition of big data. Big data is when your business wants to use data to solve a problem, answer a question, produce a product, etc., but the standard, simple methods (maybe it's SQL, maybe it's k-means, maybe it's a single server with a cron job) break down on the size of the data set, causing time, effort, creativity, and money to be spent crafting a solution to the problem that leverages the data without simply sampling or tossing out records.					
6			6	XX	22/02/2019 13:56
Big data is data that even when efficiently compressed still contains 5-10 times more information (measured in entropy or predictive power, per unit of time) than what you are used to right now. It may require a different approach to extract value.					

Aggregate	Classification	Coverage	Number Of Coding References	Reference Number	Coded By Initials	Modified On
				7	XX	22/02/2019 14:02
				8	XX	22/02/2019 14:02
				9	XX	22/02/2019 14:03
				10	XX	22/02/2019 14:09
				11	XX	22/02/2019 14:09
				12	XX	22/02/2019 14:10
				13	XX	22/02/2019 14:11
				14	XX	22/02/2019 14:14
				15	XX	22/02/2019 14:14
				16	XX	22/02/2019 14:14

Big data is data that contains enough observations to demand unusual handling because of its sheer size, though what is unusual changes over time and varies from one discipline to another. Scientific computing is accustomed to pushing the envelope, constantly developing techniques to address relentless growth in dataset size, but many other disciplines are now just discovering the value — and hence the challenges — of working with data at the unwieldy end of the scale.

Big data has taken a beating in recent years, the accusation being that marketers and analysts have stretched and squeezed the term to cover a multitude of disparate problems, technologies, and products. Yet the core of big data remains what it has been for over a decade, framed by Doug Laney's 2001 three Vs, Volume, Velocity, and Variety, and indicating data challenges sufficient to justify non-routine computing resources and processing techniques.

Big data describes datasets that are so large, complex, or rapidly changing that they push the very limits of our analytical capability. It's a subjective term: What seems "big" today may seem modest in a few years when our analytic capacity has improved. While big data can be about anything, the most important kinds of big data — and perhaps the only ones worth the effort — are those that can have a big impact through what they tell us about society, public health, the economy, scientific research, or any number of other large-scale subjects.

At Aster Data, we originally used the term big data in our marketing to refer to analytical MPP databases like ours and to differentiate them from traditional data warehouse software. While both were capable of storing a "big" volume of data (which, in 2008, we defined as 10 TB or greater), "big data" systems were capable of performing complex analytics on top of that data — something that legacy data warehouse software could not do. Thus, our original definition was a system that (1) was capable of storing 10 TB of data or more and (2) was capable of executing advanced workloads, such as behavioral analytics or market basket analysis, on those large volumes of data. As time went on, diversity of data started to become more prevalent in these systems (particularly the need to mix structured and unstructured data), which led to more widespread adoption of the "3 Vs" (volume, velocity, and variety) as a definition for big data, which continues to this day.

In my view, big data is data that requires novel processing techniques to handle. Typically, big data requires massive parallelism in some fashion (storage and/or compute) to deal with volume and processing variety.

The best definition I saw is, "Data is big when data size becomes part of the problem." However, this refers to the size only. Now the buzzword "big data" refers to the new data-driven paradigm of business, science and technology, where the huge data size and scope enables better and new services, products, and platforms. #BigData also generates a lot of hype and will probably be replaced by a new buzzword, like "Internet of Things," but "big data"-enabled services companies, like Google, Facebook, Amazon, location services, personalized/precision medicine, and many more will remain and prosper.

As our lives have moved from the physical to the digital world, our everyday tools like smartphones and ubiquitous Internet, create vast amounts of data. One of the best interpretations of the "big" in "big data" is expansive — whether you are a Fortune 500 company who just released an app that is creating a torrent of user data about every click and every activity of every user or a nonprofit who just launched a cellphone-based app to find the closest homeless shelters that are now spewing forth information about every search and every click, we all have data. Dealing with this so-called big data requires a massive shift in technologies for storing, processing, and managing data — but also presents tremendous opportunity for the social sector to gather and analyze information faster to address some of our world's most pressing challenges.

There is certainly a colorful variety of definitions for the term big data out there. To me it means working with data at a large scale and velocity.

I'm not fond of the phrase "big data" because it focuses on the volume of data, obscuring the far-reaching changes are making data essential to individuals and organizations in today's world. But if I have to define it I'd say that "big data" is data that can't be processed using standard databases because it is too big, too fast-moving, or too complex for traditional data processing tools.

The rising accessibility of platforms for the storage and analysis of large amounts of data (and the falling price per TB of doing so) has made it possible for a wide variety of organizations to store nearly all data in their purview — every log line, customer interaction, and event — unaggregated and for a significant period of time. The associated ethos of "store everything now and ask questions later" to me more than anything else characterizes how the world of computational systems looks under the lens of modern "big data" systems.

Aggregate	Classification	Coverage	Number Of Coding References	Reference Number	Coded By Initials	Modified On
				17	XX	22/02/2019 14:16
Big data originally described the practice in the consumer Internet industry of applying algorithms to increasingly large amounts of disparate data to solve problems that had suboptimal solutions with smaller datasets. Many features and signals can only be observed by collecting massive amounts of data (for example, the relationships across an entire social network), and would not be detected using smaller samples. Processing large datasets in this manner was often difficult, time consuming, and error prone before the advent of technologies like MapReduce and Hadoop, which ushered in a wave of related tools and applications now collectively called big data technologies.						
				18	XX	22/02/2019 14:16
Big data is when data grows to the point that the technology supporting the data has to change. It also encompasses a variety of topics relating to how disparate data can be combined, processed into insights, and/or reworked into smart products.						
				19	XX	22/02/2019 14:18
Big data is not all about volume, it is more about combining different data sets and to analyze it in real-time to get insights for your organization. Therefore, the right definition of big data should in fact be: mixed data.						
				20	XX	22/02/2019 14:20
I'm happy to repeat the definition I've heard used and think appropriately defines the over[all] subject. I believe it's Forrester's definition of Volume, Velocity, Variety, and Variability. A lot of different data coming fast and in different structures.						
				21	XX	22/02/2019 16:20
For me, the technological definitions (like "too big to fit in an Excel spreadsheet" or "too big to hold in memory") are important, but aren't really the main point. Big data for me is data at a scale and scope that changes in some fundamental way (not just at the margins) the range of solutions that can be considered when people and organizations face a complex problem. Different solutions, not just 'more, better.'						
				22	XX	22/02/2019 14:21
The term big data is really only useful if it describes a quantity of data that's so large that traditional approaches to data analysis are doomed to failure. That can mean that you're doing complex analytics on data that's too large to fit into memory or it can mean that you're dealing with a data storage system that doesn't offer the full functionality of a standard relational database. What's essential is that your old way of doing things doesn't apply anymore and can't just be scaled out.						