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
ode						
Nodes\\DEF_	Methodology_Tec	hnology				
Document						
Files\\U	C Berkeley School o	of Information 2	2014			
No		0.7240	29			
				1	JG	22/02/2019 16:23
opportunity to use	commerce and communice large scale, fine grained se that information to info	data to understand s	ystems, behavi	or and comme		ovation in technology makes it viable 22/02/2019 16:23
while the use of t really messy or wl structures amidst our frame of refer volume of data, b that's such a relat	e large scale, fine grained se that information to information to information to information to information to information to information in the result of the second series and the second series are the most importative term these days). In factorial second series in the second second series in the second second series in the second sec	data to understand sorm decisions and im and is often co-opteright questions or que aplex data points. Usuatop PC (e.g., larger than toncepts of big date, they can often be	ystems, behavion prove outcome d for other purperies to make — ally this revolve han a terabyte) ata don't actual applied to sma	or and comme s. 2 coses, I've und analysis that s around data and many of ti lly have much iller datasets a	JG lerstood "big d can help you f sets with a byt he tools aroun to do with it b s well. Natural	22/02/2019 16:23 lata" to be about analysis for data that ind patterns, anomalies, or new se size that seems fairly large relative to doing data are to help deal with a large eing "big" in this sense (especially sincoll language processing and lucene base
while the use of t really messy or wl structures amidst our frame of refer volume of data, b that's such a relat	e large scale, fine grained se that information to information to information to information to information to information to information in the result of the second seco	data to understand sorm decisions and im and is often co-opteright questions or que aplex data points. Usuatop PC (e.g., larger than toncepts of big date, they can often be	ystems, behavion prove outcome d for other purperies to make — ally this revolve han a terabyte) ata don't actual applied to sma	or and comme s. 2 coses, I've und analysis that s around data and many of ti lly have much iller datasets a en used with re	JG lerstood "big decan help you for sets with a bythe tools around to do with it best well. Natural elatively small	22/02/2019 16:23 lata" to be about analysis for data that ind patterns, anomalies, or new it is size that seems fairly large relative t d big data are to help deal with a large eing "big" in this sense (especially sinc I language processing and lucene base amounts of data.
while the use of treally messy or whistructures amidstour frame of data, bethat's such a relatisearch engines are	e large scale, fine grained se that information to information to information to information to information to information to information the received and the second and the second ence using files on a deskut to me the most importative term these days). In face good examples of big days	data to understand sorm decisions and important decisions and important decisions or questions or questions or questions. Usually, larger than toncepts of big diact, they can often be at a techniques and to	ystems, behavior prove outcome d for other purperies to make — ally this revolven nan a terabyte) ata don't actual applied to sma ols that are ofte	or and comme s. 2 coses, I've und analysis that s around data and many of ti lly have much iller datasets a en used with re	JG lerstood "big of can help you fisets with a bythe tools aroun to do with it bis well. Natural elatively small	22/02/2019 16:23 lata" to be about analysis for data that ind patterns, anomalies, or new se size that seems fairly large relative to dig data are to help deal with a large eing "big" in this sense (especially since I language processing and lucene base amounts of data. 22/02/2019 16:23
while the use of treally messy or whistructures amidstour frame of data, bethat's such a relatisearch engines are	e large scale, fine grained se that information to information to information to information to information to information to information the term is quite nebulous nere you don't know the rotherwise chaotic or comence using files on a deskut to me the most importaive term these days). In face good examples of big datulated the production of the collecting information information in the collecting in the collecting information in the collecting in the collec	data to understand sorm decisions and important decisions and important decisions or questions or questions or questions. Usually, larger than toncepts of big diact, they can often be at a techniques and to	ystems, behavior prove outcome d for other purperies to make — ally this revolven nan a terabyte) ata don't actual applied to sma ols that are ofte	or and comme s. 2 coses, I've und analysis that s around data and many of ti lly have much iller datasets a en used with re	JG lerstood "big of can help you fisets with a bythe tools aroun to do with it bis well. Natural elatively small	22/02/2019 16:23 lata" to be about analysis for data that ind patterns, anomalies, or new se size that seems fairly large relative to dig data are to help deal with a large eing "big" in this sense (especially since I language processing and lucene base amounts of data. 22/02/2019 16:23
While the use of treally messy or white the use of treally messy or which the contract of the	the large scale, fine grained se that information to information to information to information to information to information to information the result of the large state of the large s	data to understand sorm decisions and im and is often co-opter right questions or que uplex data points. Usua top PC (e.g., larger th ant concepts of big di act, they can often be ata techniques and to on at its most granula crease, "big data" will a mind, "big data" will	ystems, behavior prove outcome outcome outcome outcome of the purpose of the purp	or and commens. 2 coses, I've und analysis that s around data and many of tilly have much iller datasets aren used with refer to the actual size cribe any data.	JG lerstood "big decan help you fisets with a bythe tools aroun to do with it bes well. Natural elatively small JG nen you instruit JG of a particular set large enouges	22/02/2019 16:23 lata" to be about analysis for data that ind patterns, anomalies, or new the size that seems fairly large relative the dig data are to help deal with a large eing "big" in this sense (especially since I language processing and lucene base amounts of data. 22/02/2019 16:23 ment a system and keep all of the data 22/02/2019 16:23 r dataset and more about the specific gh to necessitate high-level programm
While the use of treally messy or white the use of treally messy or which the contract of the	e large scale, fine grained se that information to the result to me the most importaive term these days). In face good examples of big datult of collecting information is able to gather.	data to understand sorm decisions and im and is often co-opter right questions or que uplex data points. Usua top PC (e.g., larger th ant concepts of big di act, they can often be ata techniques and to on at its most granula crease, "big data" will a mind, "big data" will	ystems, behavior prove outcome outcome outcome outcome of the purpose of the purp	or and commens. 2 coses, I've und analysis that s around data and many of tilly have much iller datasets aren used with refer to the actual size cribe any data.	JG lerstood "big decan help you fisets with a bythe tools aroun to do with it bes well. Natural elatively small JG nen you instruit JG of a particular set large enouges	22/02/2019 16:23 lata" to be about analysis for data that ind patterns, anomalies, or new the size that seems fairly large relative the dig data are to help deal with a large eing "big" in this sense (especially since I language processing and lucene base amounts of data. 22/02/2019 16:23 ment a system and keep all of the data 22/02/2019 16:23 r dataset and more about the specific gh to necessitate high-level programm
While the use of the really messy or what structures amidstour frame of data, but that's such a relative search engines are sea	the large scale, fine grained se that information to information to information to information to information to information to design the term is quite nebulous here you don't know the rotherwise chaotic or comence using files on a desk at to me the most importaive term these days). In face good examples of big day and the term these days are the file to gather. The process it. With that in the process it. With that in the process it. With that in the process it.	data to understand sorm decisions and im sand is often co-opteright questions or que splex data points. Usua top PC (e.g., larger th ant concepts of big da act, they can often be ata techniques and to crease, "big data" will gies in order to transfi	ystems, behavior prove outcome d for other purperies to make — ally this revolve and a terabyte) at a don't actual applied to sma ols that are often ar level— it's will be less about a ultimately descorm the data as — whether it is sh for the full se	or and commens. 2 poses, I've und analysis that is around data and many of the second with reference and with reference and size or the actual size cribe any data is set into some in through information.	JG lerstood "big of can help you fisets with a bythe tools aroun to do with it bis well. Natural elatively small JG JG of a particular set large enoughing of value. JG mation graphi	22/02/2019 16:23 lata" to be about analysis for data that ind patterns, anomalies, or new se size that seems fairly large relative to dig data are to help deal with a large eing "big" in this sense (especially since I language processing and lucene base amounts of data. 22/02/2019 16:23 ment a system and keep all of the data 22/02/2019 16:23 r dataset and more about the specific gh to necessitate high-level programm

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

JG

22/02/2019 16:23

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	Reference	Coded By	Modified On
	Cassillation	Coverage	Coding References	Number	Initials	
				8	JG	22/02/2019 16:23
undetected patteri		dvances in the techn	-	•		ctors and to uncover previously ving data along with innovative ideas a
				9	JG	22/02/2019 16:23
-	ad name given to challeng; it also includes the peop					lives becomes available. It's not just
				10	JG	22/02/2019 16:23
oroduce a product, down on the size o	, etc., but the standard, s	imple methods (may ne, effort, creativity,	be it's SQL, may	/be it's k-mear	ns, maybe it's a	olve a problem, answer a question, a single server with a cron job) break to the problem that leverages the data
				11	JG	22/02/2019 16:23
subjective experier	ntial evidence. "Big data"	represents a cultura	l shift in which	more and mor	e decisions are	nave unpredictable logic and operate on the made by algorithms with transparent is change than to any particular amoun
				12	JG	22/02/2019 16:23
-	it even when efficiently c you are used to right no	•			-	d in entropy or predictive power, per u
				13	JG	22/02/2019 16:23
and varies from on elentless growth i	e discipline to another. S n dataset size, but many	cientific computing i	s accustomed to	because of its pushing the	s sheer size, th envelope, cons	ough what is unusual changes over tim tantly developing techniques to addre
and varies from on	e discipline to another. S n dataset size, but many	cientific computing i	s accustomed to	because of its pushing the	s sheer size, th envelope, cons	22/02/2019 16:23 ough what is unusual changes over tim tantly developing techniques to address the challenges — of working with data 22/02/2019 16:23
and varies from on relentless growth i at the unwieldy en Big data has taken multitude of dispan aney's 2001 three	e discipline to another. S n dataset size, but many d of the scale. a beating in recent years rate problems, technolog vs, Volume, Velocity, an	cientific computing is other disciplines are , the accusation bein ies, and products. Ye	s accustomed to now just discover g that marketer at the core of big	g because of its o pushing the overing the value 14 rs and analysts g data remains	s sheer size, the envelope, cons e — and hence JG s have stretches s what it has be	ough what is unusual changes over tim tantly developing techniques to addre the challenges — of working with dat 22/02/2019 16:23 d and squeezed the term to cover a
and varies from on relentless growth i at the unwieldy en Big data has taken multitude of dispan aney's 2001 three	e discipline to another. S n dataset size, but many d of the scale. a beating in recent years rate problems, technolog vs, Volume, Velocity, an	cientific computing is other disciplines are , the accusation bein ies, and products. Ye	s accustomed to now just discover g that marketer at the core of big	g because of its o pushing the overing the value 14 rs and analysts g data remains	s sheer size, the envelope, cons e — and hence JG s have stretches s what it has be	ough what is unusual changes over tim tantly developing techniques to addre the challenges — of working with dat 22/02/2019 16:23 d and squeezed the term to cover a een for over a decade, framed by Doug
and varies from on elentless growth i at the unwieldy en Big data has taken multitude of dispar aney's 2001 three processing techniq Big data describes erm: What seems most important kir	e discipline to another. So not dataset size, but many dof the scale. a beating in recent years rate problems, technology by, Volume, Velocity, and ues. datasets that are so large "big" today may seem m	cientific computing is other disciplines are , the accusation bein ies, and products. Ye d Variety, and indicate, complex, or rapidly odest in a few years thaps the only ones v	g that marketed the core of bitting data challed when our analy worth the effort	pecause of its pushing the value of the valu	s sheer size, the envelope, conse — and hence JG s have stretches what it has been to justify non JG very limits of cast improved. We will have a	ough what is unusual changes over time tantly developing techniques to address the challenges — of working with date 22/02/2019 16:23 d and squeezed the term to cover a deen for over a decade, framed by Doug-routine computing resources and 22/02/2019 16:23 our analytical capability. It's a subjective
and varies from on elentless growth i at the unwieldy en Big data has taken multitude of dispar aney's 2001 three processing techniq Big data describes erm: What seems most important kir	e discipline to another. So n dataset size, but many d of the scale. a beating in recent years rate problems, technolog vs, Volume, Velocity, an ues. datasets that are so large "big" today may seem mads of big data — and per	cientific computing is other disciplines are , the accusation bein ies, and products. Ye d Variety, and indicate, complex, or rapidly odest in a few years thaps the only ones v	g that marketed the core of bitting data challed when our analy worth the effort	pecause of its pushing the value of the valu	s sheer size, the envelope, conse — and hence JG s have stretches what it has been to justify non JG very limits of cast improved. We will have a	ough what is unusual changes over time tantly developing techniques to address the challenges — of working with date 22/02/2019 16:23 d and squeezed the term to cover a deen for over a decade, framed by Dough-routine computing resources and 22/02/2019 16:23 our analytical capability. It's a subjective while big data can be about anything, the
and varies from on elentless growth in the unwieldy en Big data has taken multitude of dispananey's 2001 three processing techniques and describes erm: What seems most important kir about society, pub	e discipline to another. So not dataset size, but many dof the scale. a beating in recent years rate problems, technology by, Volume, Velocity, and ues. datasets that are so large "big" today may seem minds of big data — and per lic health, the economy, so	cientific computing is other disciplines are other disciplines are , the accusation being ies, and products. Ye developed the disciplination of the complex, or rapidly odest in a few years shaps the only ones we scientific research, our drive focus, direction other disciplinations are disciplinations.	g that marketed the core of big ting data challed when our analy worth the effort or any number or on, and decision	pecause of its pushing the value of value	s sheer size, the envelope, consecuent of the envelope, consecuent of the envelope, consecuent of the envelope	ough what is unusual changes over time tantly developing techniques to address the challenges — of working with date 22/02/2019 16:23 d and squeezed the term to cover a deen for over a decade, framed by Dough-routine computing resources and 22/02/2019 16:23 our analytical capability. It's a subjective while big data can be about anything, the big impact through what they tell us
and varies from on relentless growth in the unwieldy en at the unwield en at t	e discipline to another. So no dataset size, but many dof the scale. a beating in recent years rate problems, technology by, Volume, Velocity, and ues. datasets that are so large "big" today may seem mads of big data — and per lic health, the economy, staining complex datasets to	cientific computing is other disciplines are other disciplines are , the accusation being ies, and products. Ye developed the disciplination of the complex, or rapidly odest in a few years shaps the only ones we scientific research, our drive focus, direction other disciplinations are disciplinations.	g that marketed the core of big ting data challed when our analy worth the effort or any number or on, and decision	pecause of its pushing the value of value	s sheer size, the envelope, consecuent of the envelope, consecuent of the envelope, consecuent of the envelope	ough what is unusual changes over time tantly developing techniques to address the challenges — of working with data 22/02/2019 16:23 d and squeezed the term to cover a seen for over a decade, framed by Dough-routine computing resources and 22/02/2019 16:23 bur analytical capability. It's a subjective while big data can be about anything, the big impact through what they tell us 22/02/2019 16:23
and varies from on relentless growth in the unwieldy end at the unwield end end end end end end end end end en	e discipline to another. So In dataset size, but many dof the scale. a beating in recent years rate problems, technology by, Volume, Velocity, and ues. datasets that are so large "big" today may seem mads of big data — and per lic health, the economy, so the sing complex datasets to from the analysis of your	cientific computing is other disciplines are other disciplines are , the accusation bein ies, and products. Ye d Variety, and indicate, complex, or rapidly odest in a few years thaps the only ones we scientific research, or organization's data	g that marketers the core of bitting data challers when our analy worth the effort any number or any number or an, and decision	they push the rate capacity had an alysts of data remains nges sufficient to they for they for they push the rate capacity had an alysts of they push the rate those to fother large-so they push th	s sheer size, the envelope, conse — and hence JG s have stretches what it has been to justify non JG very limits of conserved. We hat can have a cale subjects. JG n a company or	ough what is unusual changes over time tantly developing techniques to address the challenges — of working with date 22/02/2019 16:23 d and squeezed the term to cover a seen for over a decade, framed by Dough-routine computing resources and 22/02/2019 16:23 our analytical capability. It's a subjective while big data can be about anything, the big impact through what they tell us 22/02/2019 16:23 Torganization. This is done by deriving

 $\ \ \, \text{Everything we know spits out data today} - \text{not just the devices we use for computing.} \, \text{We now get digital exhaust from our garage door openers to} \\$ 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	IG	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 JG 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 JG 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 JG 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 JG 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 JG 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 JG 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 JG 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 JG 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 JG 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	JG	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	JG	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 JG 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 JG 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 JG 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 JG 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 JG 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 JG 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.

	Classification	Coverage	Number Of Coding References	Reference Number	Coded By Initials	Modified On	
				8	JG	22/02/2019 16:05	
rackers migrate a		iangulate information	in more ways,			ve, as digital sensors and beha known patterns in nature and s	
				9	JG	22/02/2019 16:05	
model the part of		about.Using their big	data, then, the			ey need to reconstruct, unders tates of the world, optimize th	
				10	JG	22/02/2019 16:12	
represents "a signa		in the business.) The	opportunity is t	o harness dat		rey opportunities. (Each source automate "harmonization" of	
				11	JG	22/02/2019 14:07	
						nt. Scientists and engineers have stand the value that can be cre	
				12	JG	22/02/2019 14:07	
our coffee pots, ar information — fro	nd everything in betweer	n. At the same time, w ke in a country thousa	e have become ands of miles aw	a generation ay to which st	of people who tore has better	xhaust from our garage door o demand instantaneous access deals on toaster ovens. Big da on.	to
				13	JG	22/02/2019 16:16	
		to the digital world, o		ols lika smartn	hones and which	quitous Internet, create vast ar	
data. One of the both that is creating a to find the closest ho called big data req	est interpretations of the orrent of user data abou meless shelters that are	t every click and every now spewing forth in echnologies for storin	expansive — w	hether you ar ry user or a no t every search nd managing o	e a Fortune 50 onprofit who ju and every clic data — but also	O company who just released a st launched a cellphone-based k, we all have data. Dealing wi o presents tremendous opport	an app app to th this so
data. One of the be that is creating a to find the closest ho called big data req	est interpretations of the orrent of user data abou meless shelters that are uires a massive shift in t	t every click and every now spewing forth in echnologies for storin	expansive — w	hether you ar ry user or a no t every search nd managing o	e a Fortune 50 onprofit who ju and every clic data — but also	O company who just released a st launched a cellphone-based k, we all have data. Dealing wi o presents tremendous opport	an app app to th this so
data. One of the bethat is creating a tofind the closest ho called big data reques the social sector to the rising accessible for a wide unaggregated and	est interpretations of the orrent of user data about meless shelters that are uires a massive shift in to o gather and analyze info ility of platforms for the evariety of organizations	t every click and every now spewing forth in echnologies for storin ormation faster to add storage and analysis of to store nearly all dat of time. The associated	expansive — wy activity of ever formation about g, processing, a ress some of out of large amount ta in their purvict d ethos of "store"	rhether you ar ry user or a no t every search nd managing our ur world's mos 14 as of data (and ew — every lo e everything n	e a Fortune 50 conprofit who ju cand every clic data — but also t pressing chal JG the falling pric g line, custome ow and ask qu	O company who just released a st launched a cellphone-based k, we all have data. Dealing with presents tremendous opport lenges. 22/02/2019 14:14 te per TB of doing so) has made interaction, and event — estions later" to me more than	an app app to th this so unity for
data. One of the bethat is creating a tefind the closest ho called big data reques the social sector to the rising accessible for a wide unaggregated and	est interpretations of the primary of user data about meless shelters that are uires a massive shift in to gather and analyze information of platforms for the evariety of organizations for a significant period of	t every click and every now spewing forth in echnologies for storin ormation faster to add storage and analysis of to store nearly all dat of time. The associated	expansive — wy activity of ever formation about g, processing, a ress some of out of large amount ta in their purvict d ethos of "store"	rhether you ar ry user or a no t every search nd managing our ur world's mos 14 as of data (and ew — every lo e everything n	e a Fortune 50 conprofit who ju cand every clic data — but also t pressing chal JG the falling pric g line, custome ow and ask qu	O company who just released a st launched a cellphone-based k, we all have data. Dealing with presents tremendous opport lenges. 22/02/2019 14:14 te per TB of doing so) has made interaction, and event — estions later" to me more than	an app app to th this so unity for
data. One of the bothat is creating a tofind the closest ho called big data require the social sector to the rising accessible for a wide unaggregated and else characterizes.	est interpretations of the prent of user data about meless shelters that are uires a massive shift in to gather and analyze information of platforms for the evariety of organizations for a significant period of how the world of computations.	t every click and every now spewing forth in- echnologies for storin ormation faster to add storage and analysis of to store nearly all da- of time. The associated stational systems look	expansive — wy activity of ever formation about g, processing, a ress some of out of large amount ta in their purvidents under the lense event data sets a	thether you ar ry user or a no t every search and managing our ur world's mos 14 as of data (and ew — every lo e everything no s of modern "b	e a Fortune 50 conprofit who ju cand every clic data — but also t pressing chal JG the falling pric g line, custome ow and ask qu big data" systen	O company who just released a st launched a cellphone-based k, we all have data. Dealing with presents tremendous opport lenges. 22/02/2019 14:14 the per TB of doing so) has made ar interaction, and event — estions later" to me more tharms.	an app app to th this so unity for e it

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

Big data enchants us with the promise of new insights. Let's not forget the knowledge hidden in the small data right before us.

17

JG

22/02/2019 16:23

about your team and the results they can get.

Aggregate	Classification	Coverage	Number Of Coding References	Reference Number	Coded By Initials	Modified On	
Nodes\\DEF_	Variety						
Document							
Files\\U	C Berkeley School	of Information 2	2014				
No		0.4455	18				
				1	JG	22/02/2019 13:40	
really messy or wh	nere you don't know the	right questions or que	eries to make —	analysis that	can help you f	22/02/2019 15:43 lata" to be about analysis ind patterns, anomalies, o	r new
really messy or wh structures amidst our frame of refer volume of data, bu that's such a relati	nere you don't know the otherwise chaotic or con ence using files on a des ut to me the most impor	right questions or que nplex data points.Usu ktop PC (e.g., larger th tant concepts of big d act, they can often be	eries to make — ally this revolve nan a terabyte) ata don't actual applied to sma	ooses, I've und analysis that s around data and many of the ly have much	erstood "big of can help you for sets with a byth he tools aroun to do with it best well. Natural	data" to be about analysis ind patterns, anomalies, on the size that seems fairly lard big data are to help dealeing "big" in this sense (estinguage processing and	r new ge relative to with a large pecially since
really messy or wh structures amidst our frame of refer volume of data, bu that's such a relati	nere you don't know the otherwise chaotic or con ence using files on a desout to me the most importive term these days). In f	right questions or que nplex data points.Usu ktop PC (e.g., larger th tant concepts of big d act, they can often be	eries to make — ally this revolve nan a terabyte) ata don't actual applied to sma	ooses, I've und analysis that s around data and many of the ly have much	erstood "big of can help you for sets with a byth he tools aroun to do with it best well. Natural	data" to be about analysis ind patterns, anomalies, on the size that seems fairly lard big data are to help dealeing "big" in this sense (estinguage processing and	r new ge relative to with a large pecially since
really messy or wheatructures amidst our frame of refervolume of data, but that's such a relationsearch engines are big data refers to a transactions of dif	nere you don't know the otherwise chaotic or con rence using files on a desut to me the most importive term these days). In fee good examples of big dother the approach to data of the other these days.	right questions or que nplex data points.Usu ktop PC (e.g., larger th tant concepts of big d act, they can often be ata techniques and to "collect now, sort out uous basis, in order to	eries to make — ally this revolve nan a terabyte) ata don't actual applied to sma ols that are ofto later"meaning o make sense of	ooses, I've und analysis that s around datas and many of th ly have much ller datasets a en used with re 3 g you capture	erstood "big of can help you for sets with a byth ne tools aroun to do with it be swell. Natural elatively small JG and store data ow cost of store	lata" to be about analysis ind patterns, anomalies, o se size that seems fairly lar d big data are to help deal eing "big" in this sense (es I language processing and amounts of data.	r new ge relative to with a large pecially since lucene based
really messy or wh structures amidst our frame of refer volume of data, bu that's such a relati search engines are Big data refers to a transactions of dif	nere you don't know the otherwise chaotic or contence using files on a destut to me the most importive term these days). In fee good examples of big dother the approach to data of ferent types, on a contin	right questions or que nplex data points.Usu ktop PC (e.g., larger th tant concepts of big d act, they can often be ata techniques and to "collect now, sort out uous basis, in order to	eries to make — ally this revolve nan a terabyte) ata don't actual applied to sma ols that are ofto later"meaning o make sense of	ooses, I've und analysis that s around datas and many of th ly have much ller datasets a en used with re 3 g you capture	erstood "big of can help you for sets with a byth ne tools aroun to do with it be swell. Natural elatively small JG and store data ow cost of store	data" to be about analysis ind patterns, anomalies, o te size that seems fairly lard big data are to help dealeing "big" in this sense (es I language processing and amounts of data. 22/02/2019 13:53	r new ge relative to with a large pecially since lucene based
really messy or wheatructures amidst our frame of refervolume of data, but hat's such a relation search engines are by transactions of dift that you generally big data is the bro	nere you don't know the otherwise chaotic or contence using files on a destut to me the most importive term these days). In fee good examples of big dother the approach to data of ferent types, on a conting don't need to have a specific or content to day a specific or content	right questions or quenches data points. Usu ktop PC (e.g., larger than toncepts of big data, they can often be ata techniques and to collect now, sort out alous basis, in order the ecific purpose for the anges and opportunities	eries to make — ally this revolve han a terabyte) ata don't actual applied to sma ols that are ofte later"meaning o make sense of data in mind be s we have as da	ooses, I've und analysis that is around data; and many of the last	lerstood "big of can help you for sets with a byth ne tools around to do with it be so well. Natural elatively small and store data ow cost of store tit. JG JG JG JG JG JG JG JG JG J	data" to be about analysis ind patterns, anomalies, on the size that seems fairly lared big data are to help dealeing "big" in this sense (estanguage processing and amounts of data. 22/02/2019 13:53 a on a very large volume of tage and better methods of the size of th	r new ge relative to with a large pecially since lucene based actions and f analysis me
really messy or wh structures amidst our frame of refer volume of data, but that's such a relati search engines are Big data refers to transactions of dif that you generally Big data is the bro	nere you don't know the otherwise chaotic or contence using files on a destut to me the most importive term these days). In fee good examples of big dother the approach to data of a ferent types, on a continuous don't need to have a spendad name given to challer	right questions or quenches data points. Usu ktop PC (e.g., larger than toncepts of big data, they can often be ata techniques and to collect now, sort out alous basis, in order the ecific purpose for the anges and opportunities	eries to make — ally this revolve han a terabyte) ata don't actual applied to sma ols that are ofte later"meaning o make sense of data in mind be s we have as da	ooses, I've und analysis that is around data; and many of the last	lerstood "big of can help you for sets with a byth ne tools around to do with it be so well. Natural elatively small and store data ow cost of store tit. JG JG JG JG JG JG JG JG JG J	data" to be about analysis ind patterns, anomalies, on the size that seems fairly lared big data are to help dealeing "big" in this sense (estanguage processing and amounts of data. 22/02/2019 13:53 If on a very large volume of the sage and better methods of the sage and th	r new ge relative to with a large pecially since lucene based actions and f analysis me
really messy or wheatructures amidst our frame of refervolume of data, but that's such a relation search engines are Big data refers to transactions of difthat you generally Big data is the broabout data though Big data has taken multitude of dispalaney's 2001 three	nere you don't know the otherwise chaotic or contence using files on a destut to me the most importive term these days). In fee good examples of big don't he approach to data of 'fferent types, on a conting don't need to have a spendad name given to challer h; it also includes the people as a beating in recent year arate problems, technologe Vs, Volume, Velocity, a	right questions or quenches data points. Usu ktop PC (e.g., larger than toncepts of big data, they can often be ata techniques and to collect now, sort out wous basis, in order to ecific purpose for the apple, processes, and a state accusation being gies, and products. Ye	eries to make — ally this revolve han a terabyte) ata don't actual applied to sma ols that are ofte later"meaning make sense of data in mind be s we have as da halysis that turr g that marketer at the core of big	analysis that a saround data: and many of the last and analysts analysts and analysts analysts and analysts and analysts and analysts and analysts analysts and analysts analysts and analysts analysts and analysts analysts analysts and analysts analysts and analysts analysts analysts analysts and analysts an	lerstood "big of can help you for sets with a byth ne tools around to do with it be so well. Natural elatively small and store data ow cost of store tit. JG	data" to be about analysis ind patterns, anomalies, o se size that seems fairly lar d big data are to help dealeing "big" in this sense (es I language processing and amounts of data. 22/02/2019 13:53 a on a very large volume of tage and better methods of 22/02/2019 15:57 I lives becomes available. I	r new ge relative to with a large pecially since lucene based f actions and f analysis me t's not just to cover a med by Doug
really messy or wh structures amidst our frame of refer volume of data, but that's such a relatisearch engines are Big data refers to transactions of difthat you generally Big data is the broabout data though	nere you don't know the otherwise chaotic or contence using files on a destut to me the most importive term these days). In fee good examples of big don't he approach to data of 'fferent types, on a conting don't need to have a spendad name given to challer h; it also includes the people as a beating in recent year arate problems, technologe Vs, Volume, Velocity, a	right questions or quenches data points. Usu ktop PC (e.g., larger than toncepts of big data, they can often be ata techniques and to collect now, sort out wous basis, in order to ecific purpose for the apple, processes, and a state accusation being gies, and products. Ye	eries to make — ally this revolve han a terabyte) ata don't actual applied to sma ols that are ofte later"meaning make sense of data in mind be s we have as da halysis that turr g that marketer at the core of big	analysis that a saround data: and many of the last and analysts analysts and analysts analysts and analysts and analysts and analysts and analysts analysts and analysts analysts and analysts analysts and analysts analysts analysts and analysts analysts and analysts analysts analysts analysts and analysts an	lerstood "big of can help you for sets with a byth ne tools around to do with it be so well. Natural elatively small and store data ow cost of store tit. JG	data" to be about analysis ind patterns, anomalies, on the size that seems fairly lard big data are to help deal eing "big" in this sense (estanguage processing and amounts of data. 22/02/2019 13:53 In on a very large volume of tage and better methods of tage and better methods of the second se	r new ge relative to with a large pecially since lucene based f actions and f analysis me t's not just to cover a med by Doug
really messy or wheatructures amidst our frame of refervolume of data, but that's such a relatisearch engines are being data refers to transactions of difthat you generally being data has taken multitude of dispataney's 2001 three processing technical big data describes term: What seems most important ki	nere you don't know the otherwise chaotic or contence using files on a destut to me the most importive term these days). In fee good examples of big don't he approach to data of 'fferent types, on a conting don't need to have a spood and name given to challer it also includes the pecture is a beating in recent year arate problems, technologe Vs, Volume, Velocity, a ques.	right questions or quenches data points. Usu ktop PC (e.g., larger than toncepts of big diact, they can often be ata techniques and to diact they can often be ata techniques and to diact they can often be ata techniques and to diact they can often be ata techniques and to diact they can often be at a techniques and to diact they can often be at a techniques for the accusation being gies, and products. Ye are they can often be at a technique and indicates, and products. Ye are they can often be at a technique at	eries to make — ally this revolve han a terabyte) ata don't actual applied to sma ols that are ofte later"meaning make sense of data in mind be s we have as da halysis that turr g that marketer at the core of big ting data challe of changing that when our analy worth the effort	analysis that is around data; and many of the later much iller datasets a en used with residue about every data into means and analysts analysts and analysts analysts and analysts analysts and analysts analysts and analysts and analysts analysts and analysts analysts analysts and analysts analysts and analysts analysts analysts analysts and analysts analysts analysts analysts and analysts a	erstood "big of can help you for sets with a byth ne tools around to do with it bigs well. Natural elatively small and store data ow cost of store tit. JG JG Aspect of our aning. JG Aspect of our aning.	data" to be about analysis ind patterns, anomalies, on the size that seems fairly lard big data are to help deal eing "big" in this sense (estanguage processing and amounts of data. 22/02/2019 13:53 If on a very large volume of tage and better methods of the size of th	r new ge relative to with a large pecially since lucene based f actions and f analysis me t's not just to cover a med by Doug rces and 's a subjective t anything, th

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 JG 22/02/2019 14:06

JG

22/02/2019 14:05

[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	JG	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 JG 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 JG 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 JG 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 JG 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 JG 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 JG 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 JG 22/02/2019 14:19

Big data means data that cannot fit easily into a standard relational database.

18 JG 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	Reference Number	Coded By Initials	Modified On	
			References				

Nodes\\DEF_Velocity

Document

Files\\UC Berkeley School of Information 2014

No 0.2922 11 1 JG 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 JG 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.

3 JG 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 JG 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 JG 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 JG 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 JG 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 JG 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	JG	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 JG 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 JG 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	JG	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 JG 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 JG 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 JG 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 JG 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 JG 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	JG	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.

8 JG 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.

9 JG 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.

10 JG 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 JG 22/02/2019 14:09

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 JG 22/02/2019 14:10

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.

13 JG 22/02/2019 14:11

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 JG 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.

15 JG 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.

16 JG 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.

•	Aggregate	Classification	Coverage	Number Of Coding References	Reference Number	Coded By Initials	Modified On
					17	JG	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 JG 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 JG 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 JG 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 JG 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 JG 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.