



COGNITIVE COMPUTING  
C O N S O R T I U M

# Big Data and Cognitive Computing

*By Hadley Reynolds*

See more at [cognitivecomputingconsortium.com](http://cognitivecomputingconsortium.com).

## A clear-eyed view

**D**iscussions of cognitive computing almost always include a reference to “big data.” Discussions of big data occasionally, but infrequently, reference “cognitive computing.” But are we truly confident that we know which is which and why that is so?

We felt that finding a way to describe these two trends in simple terms—and differentiate among them and define their relationship to each other—could help lower the level of hype and confusion in this active corner of the technology landscape. If we can achieve a new kind of clarity in this conversation, we can get on with the business of talking about cognitive computing in a much crisper and more intelligent manner than we’ve typically experienced to date.

It’s important first to pull apart the various levels that these terms operate on in our broader public conversation. To cut to the chase, I’m proposing that there are four important levels or “meanings” that these terms are operating on. We need to get better at understanding and differentiating these meanings. We need to be more accurate as we throw these terms around. The four levels are:

- 1) The mission or purpose of big data vs. that of cognitive computing
- 2) The foundation technologies of each
- 3) The functional description of what these trends and their technologies actually do for people
- 4) The symbolic level, where our public conversation has already transformed these terms into labels for various business strategies, worldviews, and hype campaigns.

In what follows, we will look at each of these levels in turn.

### Distinct missions

First, I want to suggest that big data and cognitive computing are highly distinct in their purpose or mission. To put it succinctly, the mission of big data is best understood as the next generation of the traditional IT function of storage and organization of machine-based enterprise information—now extended to include different types of data handled in new ways.

Cognitive computing, on the other hand, seeks the meaning in the data. Cognitive computing is best understood as an innovation in methodology for the field of analytics. Cognitive computing

seeks to break through the constraints of analytics based on backward facing numerical calculations and static presentations of results for human review. Instead, cognitive computing represents a unique form of computing combining analytics, problem solving, and communication with human decision makers. It uses big data if necessary to answer ambiguous questions and solve problems. But its key contributions go well beyond the charter of analytics as understood today. Cognitive computing looks within and across disparate data sets including rich media and text, identifies conflicting data, uncovers surprises, finds patterns, understands context, offers suggestions, requests clarifications, and provides ranked solution alternatives. Cognitive computing offers a new approach to uncover the potential in data – and capture value whether the data is big or small.

As its purpose, big data remodels the data center, the database, and the data warehouse to accommodate today's transformed digital environment. As its purpose, cognitive computing leverages a broad suite of evolving discovery, analysis, human interaction, and solution development technologies to offer a new kind of digital assistance that operates in near-human terms.

## Technology foundations

Beyond the issue of mission, we need to recognize that each trend—i.e. big data and cognitive computing—rests on a unique technology foundation. And we propose that the two foundations are related but also fundamentally different. So we have a “ground truth” based on distinct developments and innovations in the technology environment. For example, there is little dispute that big data is a phenomenon of the spread of digital technology across consumer, commercial, government, and scientific life (and most any other life you care to add). At the same time, cognitive computing has been associated with bringing computing machines to play in such challenges as bringing “human-like” insights to Jeopardy game-playing, making personal digital assistants intelligent, accelerating human genome analysis, and improving medical outcomes through diagnosis and treatment recommendations. All of these examples are based on the cognitive applications' ability to process “beyond-human” quantities of disparate data while analyzing and presenting suggestive, non-trivial, timely solutions.

Prefiguring the emergence of the big data trend, we all recognize that in the spread of the internet, global access to inexpensive content “publishing” both personal and professional, the adoption of online video and other rich media, the explosion of mobile devices, the overnight

rise of social and user-generated media, the proliferation of log files tracking all of this activity on a packet-by-packet basis, and on and on—the very nature of data has changed rapidly and is now irretrievably big.

So big is this big data, that in increasing numbers of applications, traditional means of creation, capture, organization, and storage of it threaten to break or become meaningless under the onslaught. As a result, innovative technologists are devising new approaches to try to keep pace. So Google, for example, faced the problem of how to manage the exponential growth of their web search indexes and came up with the idea of Map Reduce, an early approach to harnessing commodity hardware clusters to transform the level of efficiency of content processing and index creation. At roughly the same time Google was focusing on these distributed processing innovations, technologists at Yahoo!, seeking to supplement the reigning SQL database storage paradigms for performance and scalability reasons, invented non-SQL storage models designed to replace traditional DBMS parallel processing with distributed processing. The Hadoop Distributed File System they developed is now the most prominent model, supplemented by an ecosystem of multiple “non-SQL” software packages that extend the capabilities and connectibilities of the Hadoop storage core.

I review this bit of recent technology history to point out that the term “big data” is not simply a reference to the quantity of bytes we now generate, although of course intuitively it is that as well. But more importantly it also references a set of software resources, assets, and practices that have now built up a legacy of over a decade of development and are supporting many of the most critical compute applications on the planet.

So what can we say about cognitive computing’s technology foundation? The first observation is that it does not have to be involved with big data at all. While IBM’s Jeopardy-playing Watson ingested an impressive quantity of encyclopedias, history books, magazines, political broadsides, and previous Jeopardy questions and answers, this hardly constituted a big data application on a scale familiar to Google, the intelligence community, telecommunications carriers, etc. Watson was much more dependent on “big memory,” as it utilized recent innovations in “in-memory” processing approaches to discover, synthesize, and statistically analyze possible responses to those arcane Jeopardy questions in real time. The Watson Jeopardy application is much more usefully understood as a data science triumph than as a big data feat.

Quoting the definition of cognitive computing developed by the Cognitive Computing Consortium (2014), its use as a human problem solver is at the forefront:

"Cognitive computing makes a new class of problems computable. It addresses complex situations that are characterized by ambiguity and uncertainty; in other words it handles human kinds of problems. In these dynamic, information-rich, and shifting situations, data tends to change frequently, and it is often conflicting. The goals of users evolve as they learn more and redefine their objectives. To respond to the fluid nature of users' understanding of their problems, the cognitive computing system offers a synthesis not just of information sources but of influences, contexts, and insights."

The technology foundation of cognitive computing is not fundamentally about programming, processing, or storage paradigms, or about data flows and stream handling, but rather about the broad ranging data analysis technologies addressing discovery, disambiguation, contextual understanding, inference, recommendation, probabilistic reasoning, and human/machine communications. So instead of map reduce, hadoop, no-SQL, Pig, Hive, Spark, Sqoop, and other big data tools and technologies, cognitive computing relies on technologies like voice recognition, text-to-speech, language recognition, Natural Language Processing in its many forms, machine learning in its many forms, neural networks, Bayesian statistics and inferencing, support vector machines, many kinds of statistical analysis, voting algorithms, not to mention a heavy dependence on human interaction and visualization design. We can layer cognitive computing on a big data foundation, if it is available, in order to understand, infer or reason about the evidence the data contains.

## Symbiotic functions

While I have been focusing thus far on the distinctness of big data and cognitive computing in both the mission and the technology levels, I also want to highlight the valuable relationship between the two trends. The most important symbiosis between the two is that the availability of big data-scale quantities of data is tremendously helpful for the kinds of machine learning algorithms and methodologies on which cognitive computing depends for the accuracy and contextual appropriateness of its answers or solution recommendations. The flip side of this value of increased power of analysis for cognitive applications is of course the new kinds of analytic value these applications offer those who are trying to make some kind of sense out the petabytes, exabytes, or zettabytes of data that are collecting in their enterprise big data “lakes” or black holes, or other kinds of repositories.

Big data and cognitive computing will continue to be interrelated and will continue to be spoken about together as if they were all of a piece. Let us turn then to look at other distinctions as well as inter-relationships between the trends and the terms at the functional and symbolic levels.

On the level of function, confusion arises immediately from the many statements we can read which appear to wrap big data and cognitive computing into the same phenomenon.

Consider, for example, this marketing statement from a software vendor promoting cognitive computing: “The ... Cognitive Reasoning Platform is an artificial intelligence platform that allows organizations to capitalize on the power and potential of Big Data through advanced analytics and actionable insights that fundamentally inform organizations about the business, customers, and value chains in which they operate.” At the 50,000’ level, we could agree with this characterization of cognitive computing’s potential, but linking it up with the power of big data in the same sentence implies a symbiosis that does not in fact exist. It would be far more accurate to say that the power and potential of cognitive computing is one approach that organizations can consider when they face the challenge of teasing important business insights out of their many diverse sources of data, bigger and smaller and both together.

Another source of confusion about the respective function of these two trends arises from a phrase that has been used with increasing frequency over the past two years, the label “big data analytics.” While big data analytics often refers simply to traditional kinds of structured data analytics now applied to large data volumes, it also often points to the emergence of analytics

on unstructured or semi-structured data encountered in the new data lakes or hadoop clusters. Regardless of data format, the continuing priority of both the IT owners of these resources and the analysts attempting to get at them is to uncover actionable insights for the business. It is here that cognitive computing has entered the big data analytics picture, as a scanner and sense-maker for the diverse kinds of data in these stores.

The conversations around big data analytics often seem to ignore the reality that an application using cognitive computing approaches and technologies must be developed as a project distinct and largely independent of the “big data-ness” of the analytics environment.

Another consideration generating confusion around the functions of the two trends arises from statements that conveniently ignore the relative maturity of each. Big data has well over a decade of development and is already seeing the adoption of its second generation of tools and techniques. Cognitive computing, on the other hand, is in its earliest stages, with very few products even ready for the market and many more promises in the air than tangible results on the ground. When statements are put forward linking the two trends like peas in a pod, they rarely pause to consider that fifteen years from now, when cognitive computing arrives at a similar stage of maturity as today’s big data, the only certain thing is that we will be talking about both of these trends in very different, probably unrecognizable terms.

## Maturing of the marketplace

We turn now to how the terms “big data” and “cognitive computing” are working on the symbolic level. We need to recognize first that when terms or phrases become buzz labels for entire new phases in the development of the computing business and the information economy, they take on something of a life of their own. This new life in the symbolic atmosphere has proven over the years to turn buzz terms into potent sources of confusion and obfuscation rather than guideposts toward understanding. This is the situation today with big data and cognitive computing.

For example, simply by looking at the impact of the phenomenon of big data on the marketplace, we see a remarkable difference between the attitude and language of companies who actually run their businesses on big data and those companies who are using these terms but are primarily worried about being left behind.

Google, Facebook, Yahoo, Amazon—these are the firms who first encountered big data and learned how to harness it into web advertising, social media, digital publishing, and online commerce and web services, respectively. But these firms rarely make a big noise about the term “big data.” They view data as being simply raw material for business propositions that drive well beyond the narrow view of the technologies that underpin them. (An exception, of course, is Amazon’s cloud services business, which offers its expertise and resources directly to others as a service.) These firms don’t view themselves as being in the computing business, despite the fact that they have invented the state-of-the-art tools and practices required to cope intelligently with massive volumes of flowing data.

On the other hand, the legacy enterprise software vendors who dominated the computing markets in the recent past have been trumpeting about “big data” and how they can save their corporate customers from its perils while enabling them to reap giant gains from its opportunities. IBM, Oracle, HP, Microsoft, SAP, EMC—the list goes on. All these firms missed the industry turn to cloud computing, software-as-a-service, and big data, and now are re-engineering their product lines and marketing approaches around the new terminology.

At this stage of the maturing of the big data marketplace, most analysts are predicting a slowing of innovation, a period of consolidation, and a converging of big data offerings on a broadly comparable suite of products and services from a smaller number of vendors. In this environment, all the legacy vendors need to differentiate their big data value. This is where the emerging ideas and terminology around cognitive computing come into play.

IBM deserves special mention in this discussion of the symbolic importance of the cognitive computing term, since the company has been years ahead of competitors in uniting their R&D in cognitive technologies with their event marketing prowess (Watson playing Jeopardy) and with a serious, long-term, multi-layered business and investment initiative. They deserve recognition for proposing and successfully establishing the term in the first place. That said, IBM recognizes the value of cognitive computing in the short term in giving them symbolic differentiation from their competitors in the trenches of the big data, cloud, and analytics marketing wars.

We have left the consumer out of this discussion altogether, but we can’t close without noting that cognitive computing is now operating, for better and for worse, in your smartphone



homescreen, and Apple and Google and Microsoft and Amazon are betting that their nascent digital assistants will be taking over a greater and greater part of the operating interfaces of these devices. IBM and Apple have announced alliances around these emerging capabilities, and we will certainly see another round of consumer-driven IT arrive in the near term, with smarter “cognitive” applications delivered in mobile form factors available at work as well as in civilian life. And leveraging big data to boot!

In closing, we hope that we have offered a series of perspectives on big data and cognitive computing that provide food for thought, a degree of clarity about similarities and differences, and a framework for looking at these fascinating and emerging technology trends with confidence and a level of immunity from industry hype.

## Author biography

### Hadley Reynolds

Co-founder of Cognitive Computing Consortium, Principal Analyst at NextEra Research

In 35 years in the software industry, Hadley has held a variety of product development and industry analyst roles. Hadley is a principal analyst at NextEra Research that offers a lens on emergent intelligent systems deployed to augment human decision making. Prior to NextEra Research, he most recently was associated with International Data Corporation as a senior analyst developing the practice on Search and Digital Marketplace technologies. Prior to IDC, he founded and acted as VP and Director of the Centre for Search Innovation at FAST/Microsoft.

For over a decade, he headed the research practice at Delphi Group, with a focus on knowledge management, search, content & collaboration management, and business process automation. Prior to his work at Delphi, he held product management, marketing, and strategy executive roles at Project Software & Development, Inc., a leader in enterprise application software for project management.