



# SEARCH

IS THE NEW KILLER APP



## CHAPTER 1

---

# The Search Story is a Data Story

### **When we talk about search what are we actually talking about?**

It's easy to feel like the essence of search revolves around the search query. We type an idea, press a button, and then answers are culled from the ether, and delivered to us in digestible form. Our comfort with the search bar, constantly present in our everyday Internet lives, has spoiled us. Rarely do we consider the complexities triggered beneath the surface in the moment we hit "enter."

In reality, we developed search technologies, not primarily as a front-end service but to alleviate the challenges of the back end: our data. Search grew in direct response to data problems. The search story can't be told without the data story, and high-volume data would be nearly useless without search technology.

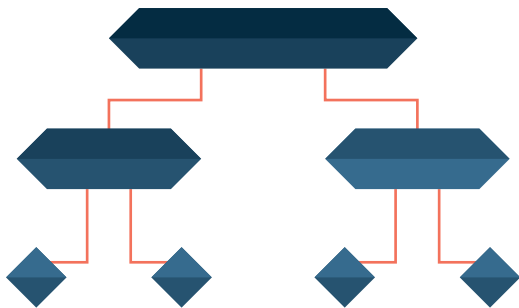
Data and search evolved together. Database advancements necessitated new search solutions, and search tech fundamentally shaped the way we conceptualize and interact with data.

---

## The Database Race: Early Database Structures

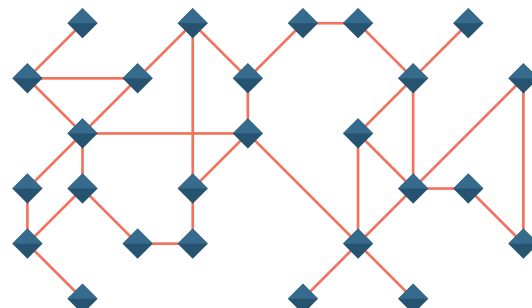
In the beginning, closed organizational networks were the primary data storage centers. Military, academic, and commercial enterprises collected data on custom-built storage platforms, requiring technical expertise to generate any value from them. Determining how to organize, model, and therefore read the data was a major challenge.

Two primary database structures prevailed through the 1970s:



### Hierarchical Database

The hierarchical system is a tree-like data structure with a single root. The data programmer must make low-level data calls using a navigational language to access stored data.



### Network Database

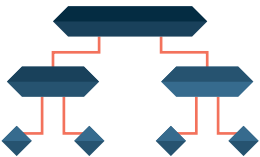
In the network database system, each data element “points” manually to other related data elements. To access network data, a programmer must be intimately familiar with the data structure and make low-level calls in the navigational language.

# The Relational Database Management System (RDBMS)

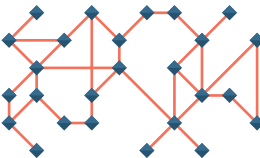
Data practitioners needed a standardized, seamless way to quickly access data. In 1970, an IBM research team developed a lasting answer: the Relational Database Management System (RDBMS). In this model, we started to see data stored in spreadsheet form. All data was tabular, organized into rows and columns.

Most importantly, the RDBMS was a major step toward democratizing data access. For the first time, users did not need advanced knowledge of a programming language or a custom algorithm to pull information out of a database. The RDBMS model laid the foundation for the forthcoming giant advancements in search technology.

Hierarchical Database



Network Database



Relational Database

	1	2	3	4	5	6	7
A	Blue Diamond	Red Diamond	Red Diamond	Green Diamond	Blue Diamond	Green Diamond	Red Diamond
B	Green Diamond	Green Diamond	Blue Diamond	Red Diamond	Blue Diamond	Green Diamond	Blue Diamond
C	Red Diamond	Blue Diamond	Green Diamond	Green Diamond	Blue Diamond	Red Diamond	Green Diamond
D	Red Diamond	Red Diamond	Blue Diamond	Green Diamond	Blue Diamond	Blue Diamond	Red Diamond
E	Blue Diamond	Green Diamond	Blue Diamond	Red Diamond	Red Diamond	Green Diamond	Green Diamond

## Structured Query Language (SQL)

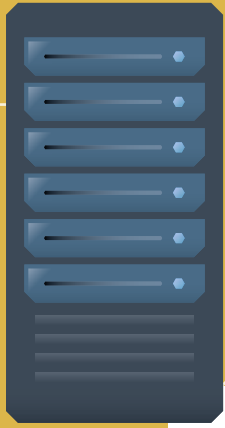
Along with this new relational theory of databases, SQL emerged to become an industry standard. Also introduced by the IBM team, SQL (read “sequel”) was one of the earliest programming languages to leverage the RDBMS, and it was widely adopted by 1986.

SQL would become instrumental in the development of search technology, remaining a major player in today’s storing and calling of data. Tech giants Microsoft, Oracle, IBM, and SAP each maintain their own extensions of SQL.

```
Cities
{
  CityID INT,
  CityName VARCHAR(200),
  State VARCHAR(200)
}
```



Although not yet accessible to the mass public, the advents of the relational database model and SQL language created blueprints for standardized storing and accessing of data. By 1990, the foundation was in place for organizations and consumers to start accessing data **like never before.**



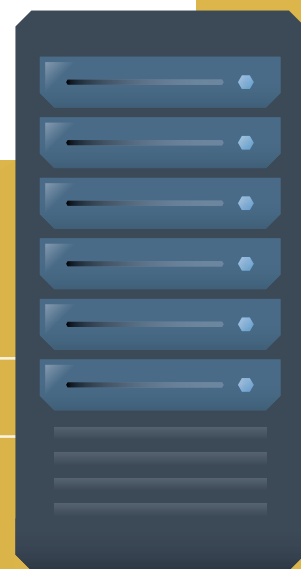
## CHAPTER 2

---

# Search for All

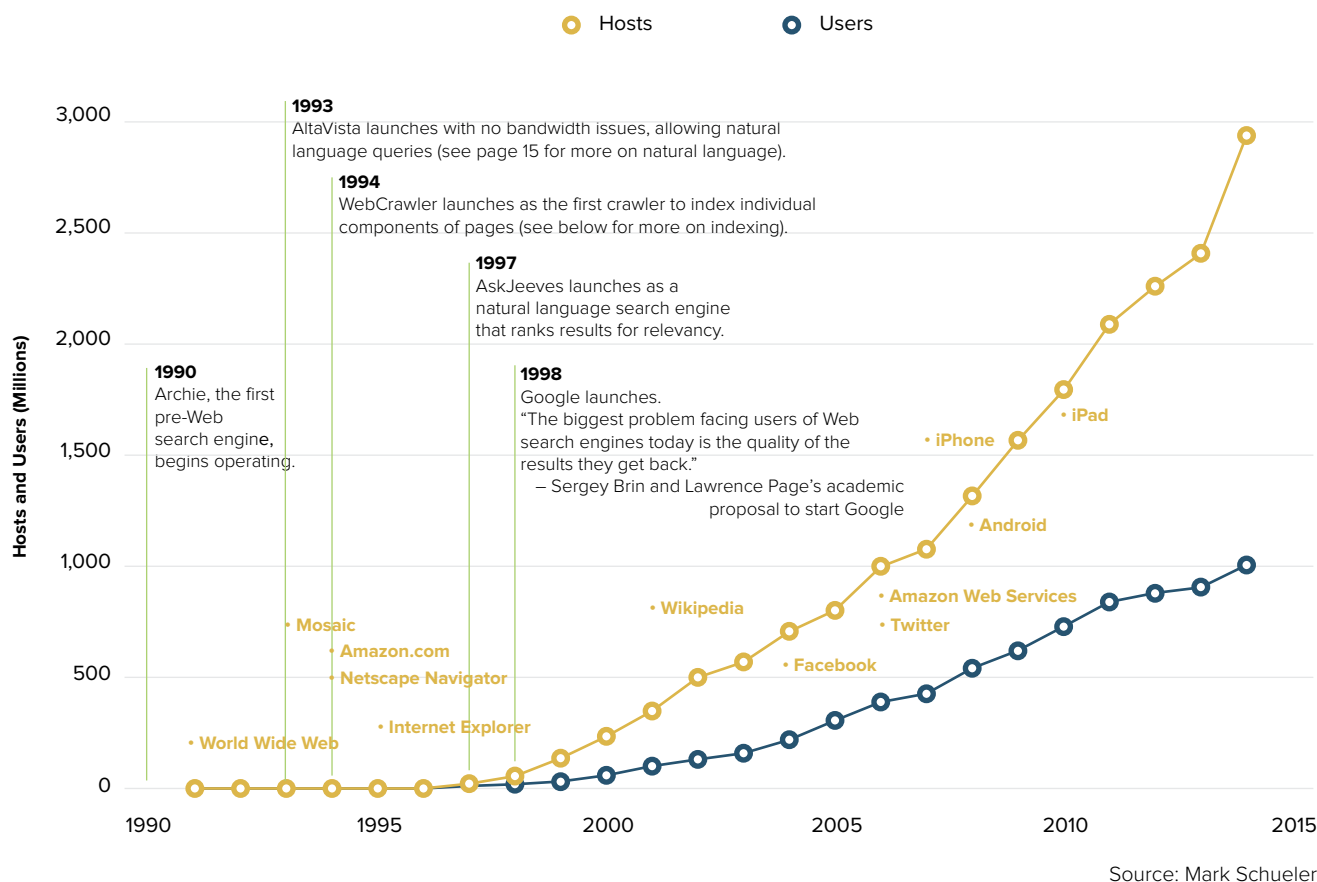
**Progress in search technology drives profound change in our fundamental relationship with the world's information.**

With relatively simple, standardized database management in place, data was ready for the big time. And, lucky for us all, the big time was coming. In the 1990s, the wide adoption of the Internet would be one of history's greatest technological changes, and search technology formed a crucial bridge between the Web-browsing public and the world's data stored away in the far corners of Web servers.



## Search Engines Make the Web Accessible

Search engines in the 1990s applied SQL's method of indexing and querying to Internet content. At first, the general public had no concept of just what information the World Wide Web made available. Search engines made exploring the Web possible, and the two have since evolved together in a push-and-pull relationship.



Innovations in data and search would tame the Internet's data chaos, bringing unprecedented information access into our homes and driving the rest of the tech industry to provide networks and hardware for it.

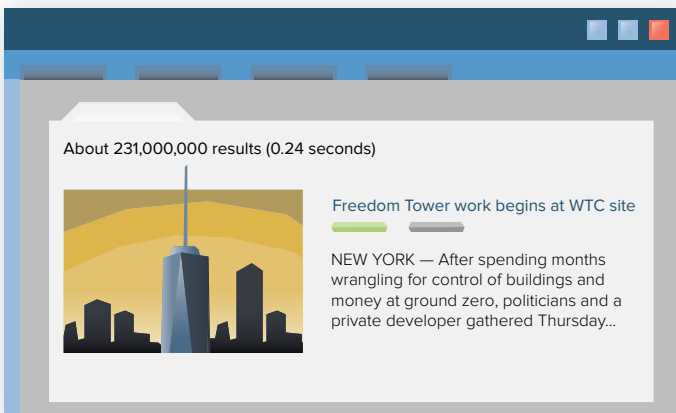
## Key Concept: Indexing

A search engine uses indexing to set virtual coordinates for every meaningful piece of information and then locates the information appropriate to the user's query. Without indexing, finding a piece of data would require manually scanning every inch of a database. Indexes work just like the indexes found in books: If you're looking for a specific topic, the index's listing will direct you straight to the page where that term is mentioned. But also like a book, indexing alone provides the user no indication of how relevant each result is to the search term. Search indexing's limitation is its disregard for the quality of results.

# The Genius of Google

## Google PageRank: Quality First

Responding to the major market need for results' relevance, Google launched with its patented PageRank in tow. A breakthrough for search, PageRank promotes sites with hyperlinks from many other credible pages. As a result, Google quickly became Internet users' default search engine.



## Sept. 11<sup>th</sup> Changed Everything, Even Search

It wasn't until Sept. 11, 2001, that the Google developers realized timely context was a crucial part of search. That day, querying "New York Twin Towers" would return static tourist information, rather than dynamic news updates.

**Google soon remedied the timeliness issue in 2002, with the launch of Universal Search.**

1999



1 Month

TODAY



1 Minute

## 50 million

In 1999, it took Google one month to crawl and build an index of about 50 million pages—a task now accomplished in less than one minute.



## 2013

A five-minute complete Google blackout resulted in a **40% dip** in global Internet traffic.





Providing value for **hundreds of millions of users** every day, search tech came a long way in just a decade; however, search was still centered around “finding what you need.” This search paradigm would shift quite a bit in the coming years.

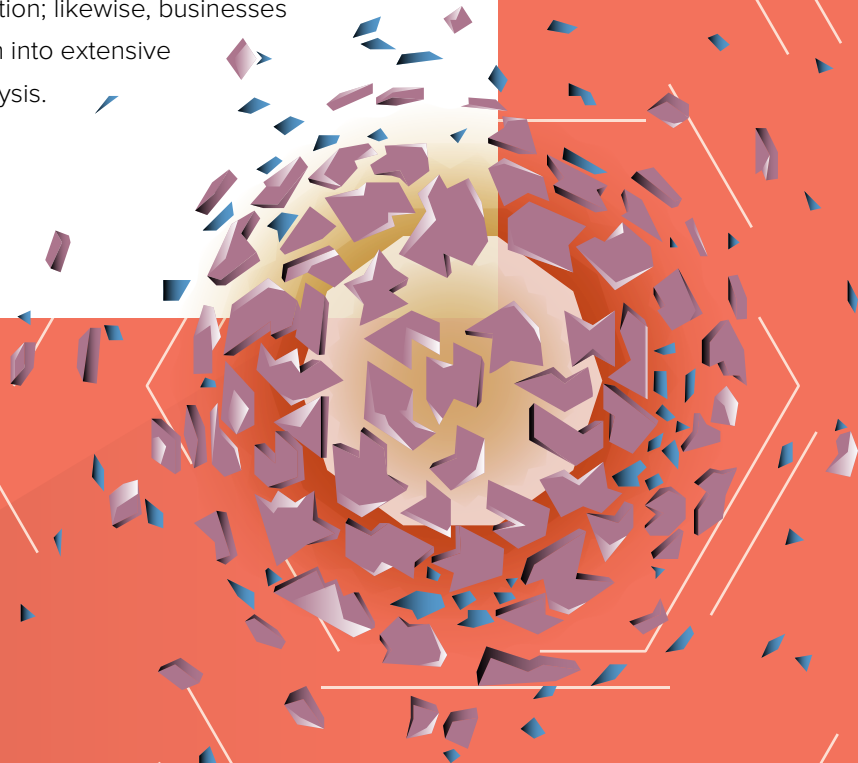
## CHAPTER 3

---

# Search Explodes as Data Explodes

**As data gets “big,” search tech reinvents itself to begin a golden age of information.**

The explosive data landscape made storage cheaper, easier, and more useful than ever before. Consumers started to discover value in the Internet’s wealth of information; likewise, businesses poured investments and research into extensive data acquisition, mining, and analysis.



# The Hunger for Bytes

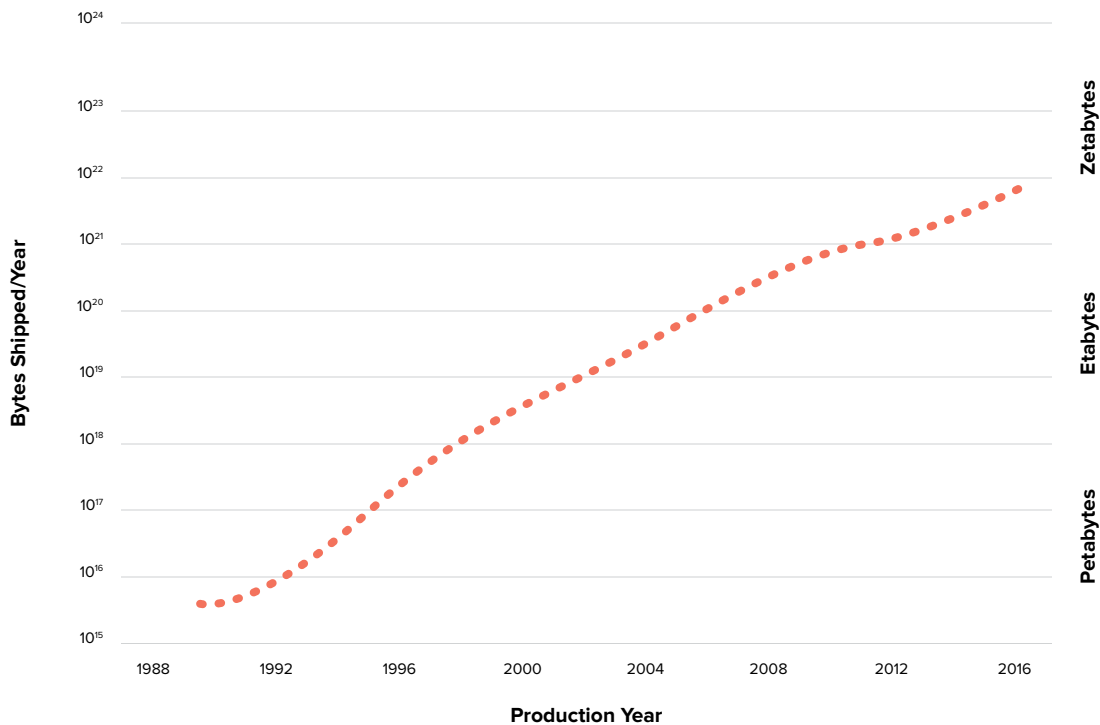
The demand for data storage has been on an exponential rise since the late 1990s. As data became more readable, the global appetite for collecting it soared.

“The penalties for storing obsolete data are **less apparent** than the penalties for discarding potentially useful data.”

— I.A. Tjomsland,

Fourth IEEE Symposium on  
Mass Storage Systems, April 1980

Worldwide Hard Drive Capacity Shipments, 1999–2016



Source: Forbes

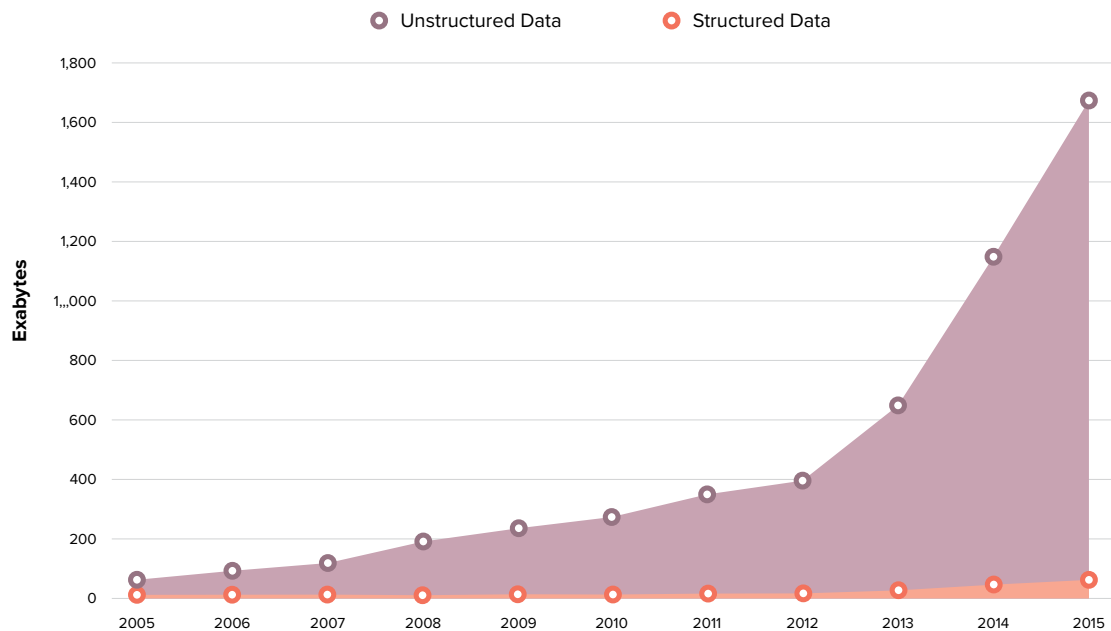
## A Sequel to SQL

A valuable evolution of the baseline SQL database language came in the form of NoSQL (short for “Not Only SQL”). It added scale and versatility to SQL, which had emphasized consistency over volume.

## Achievement Unlocked: Processing Unstructured Data

The relational database model was proving too rigid for large-scale data processing, as data wasn't always in the neatly organized, tabular structure needed for common relational search technologies.

The need to make sense of unstructured data became ever more urgent as it started to represent the overwhelming majority of collected data. *Unstructured data* is any stored information that is undetectable by software looking for a tabular structure. Items like the content of books, word processing documents, presentations, audio and video, analog data, images, metadata, and e-mails. Needless to say, these are items that can provide tremendous value to consumers and businesses alike.



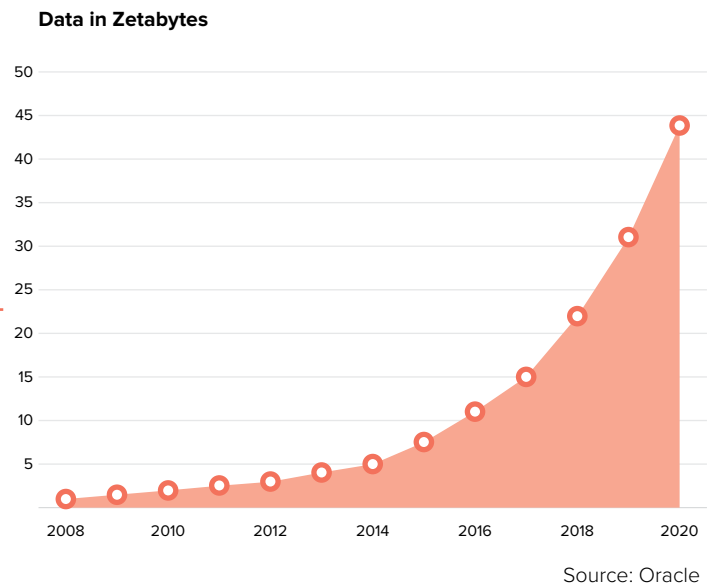
Source: IDC

Recognizing and reading unstructured data made the immense value of search an instrumental contributor to the information age. Now, machines could understand linguistic, auditory, and visual structures inherent in natural human communication.

## A Buzzword is Born

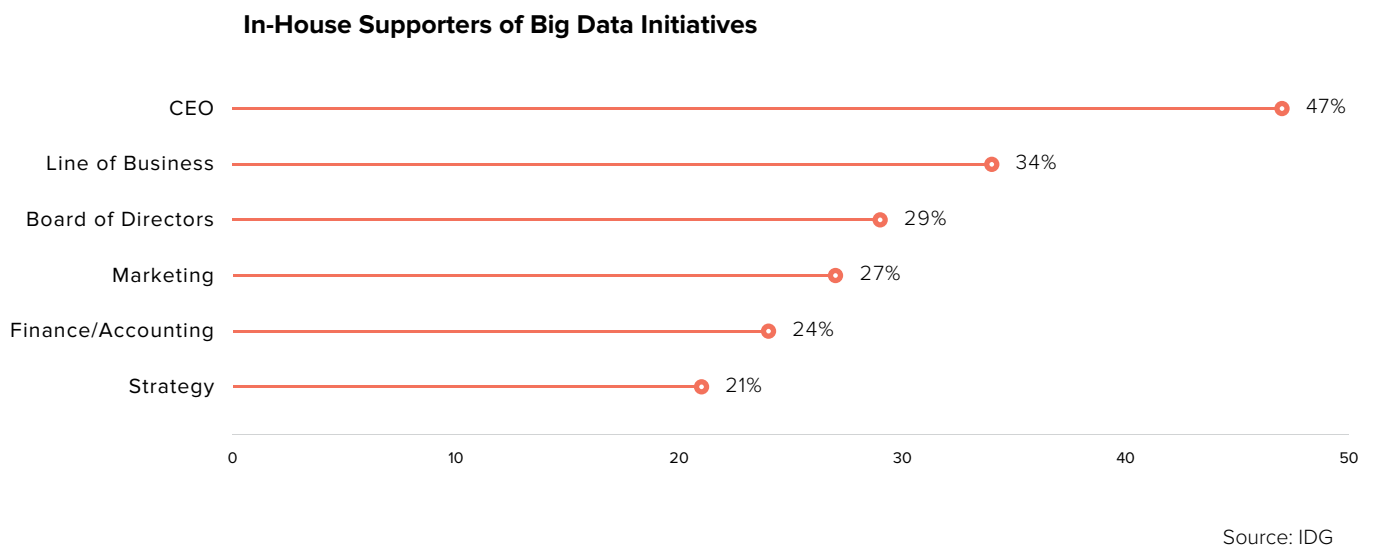
“Big Data” is a term we hear bandied about across all contemporary business sectors. Increasingly, data technology and insights are being put to use by entities both private and public. Where human cognition falls short of being able to synthesize massive amounts of information, machine analysis can quickly provide objective, actionable perspective.

● Data is growing at a **40% compound annual rate**, reaching nearly 40zb by 2020.



## Friends in High Places

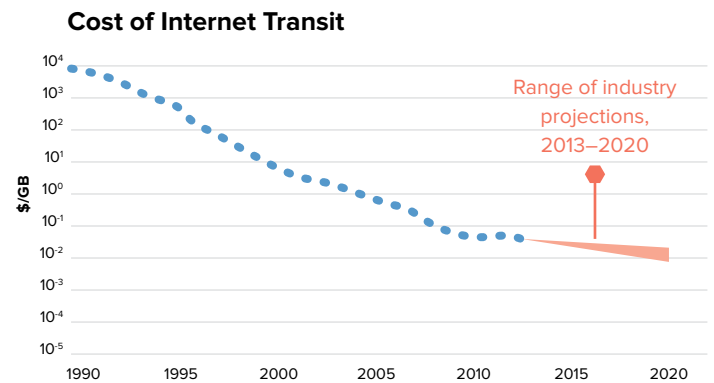
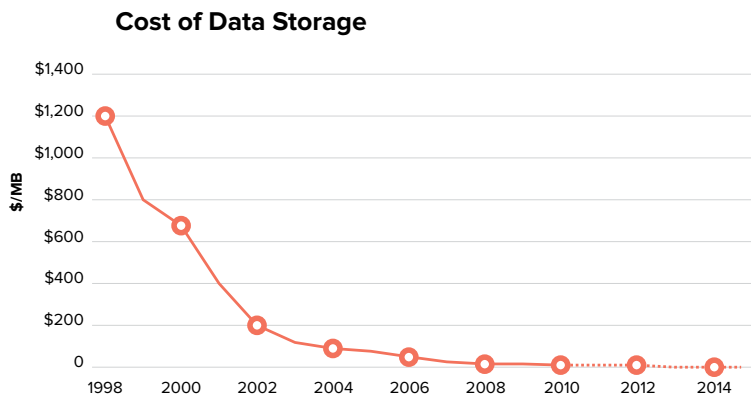
Big data's growth is being driven by corporate leaders looking to make the most informed decisions possible.



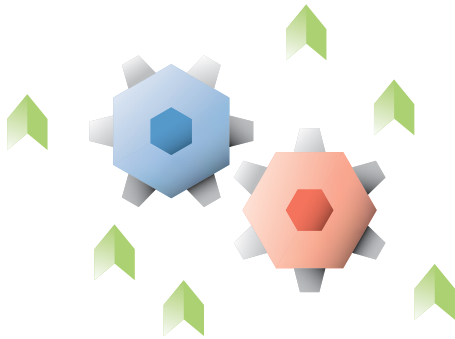
In 1996, digital storage became more cost-effective than paper for storing data, according to R.J.T. Morris and B.J. Truskowski in “The Evolution of Storage Systems,” *IBM Systems Journal*.

## Tech and Users Mature Together, Enabling Search’s Meteoric Rise

### Data Becomes Fast and Cheap



Front-end design advancements meant less technical expertise needed to access valuable data.



Digital literacy narrowed the access gap even further.

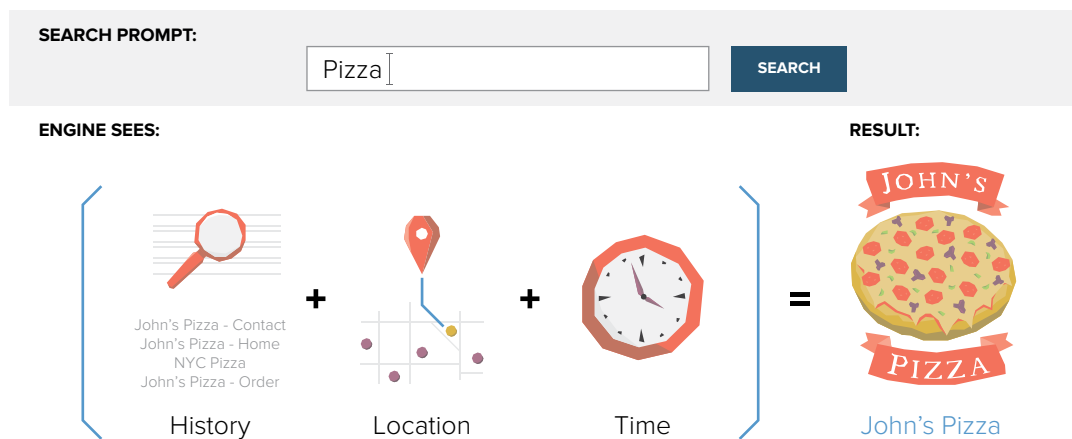


## Forethought vs. Post-Thought

Whereas data searching originated as a means for users to find things they needed, this paradigm has undergone a major shift. Now, search tech *predicts* a user's needs by leveraging two innovative features.

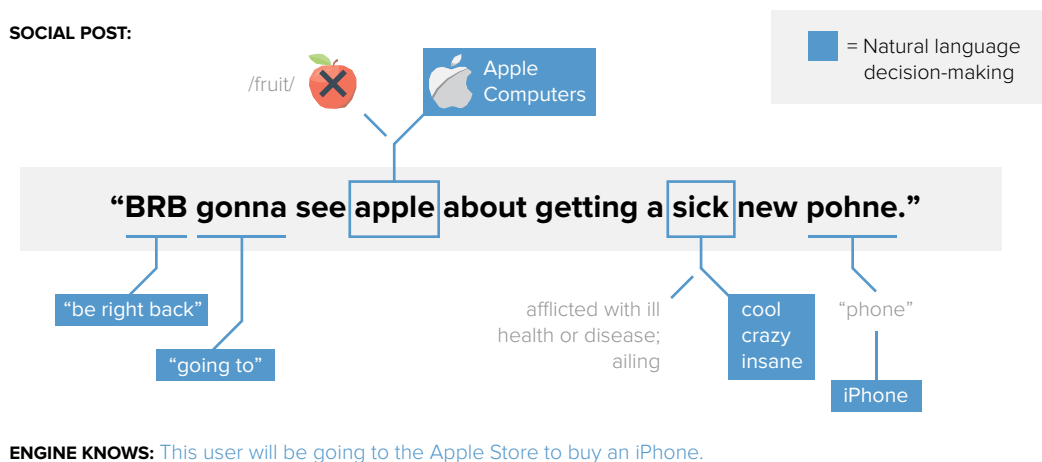
### Context Awareness

Use every piece of information available to determine the result most valuable to the user.



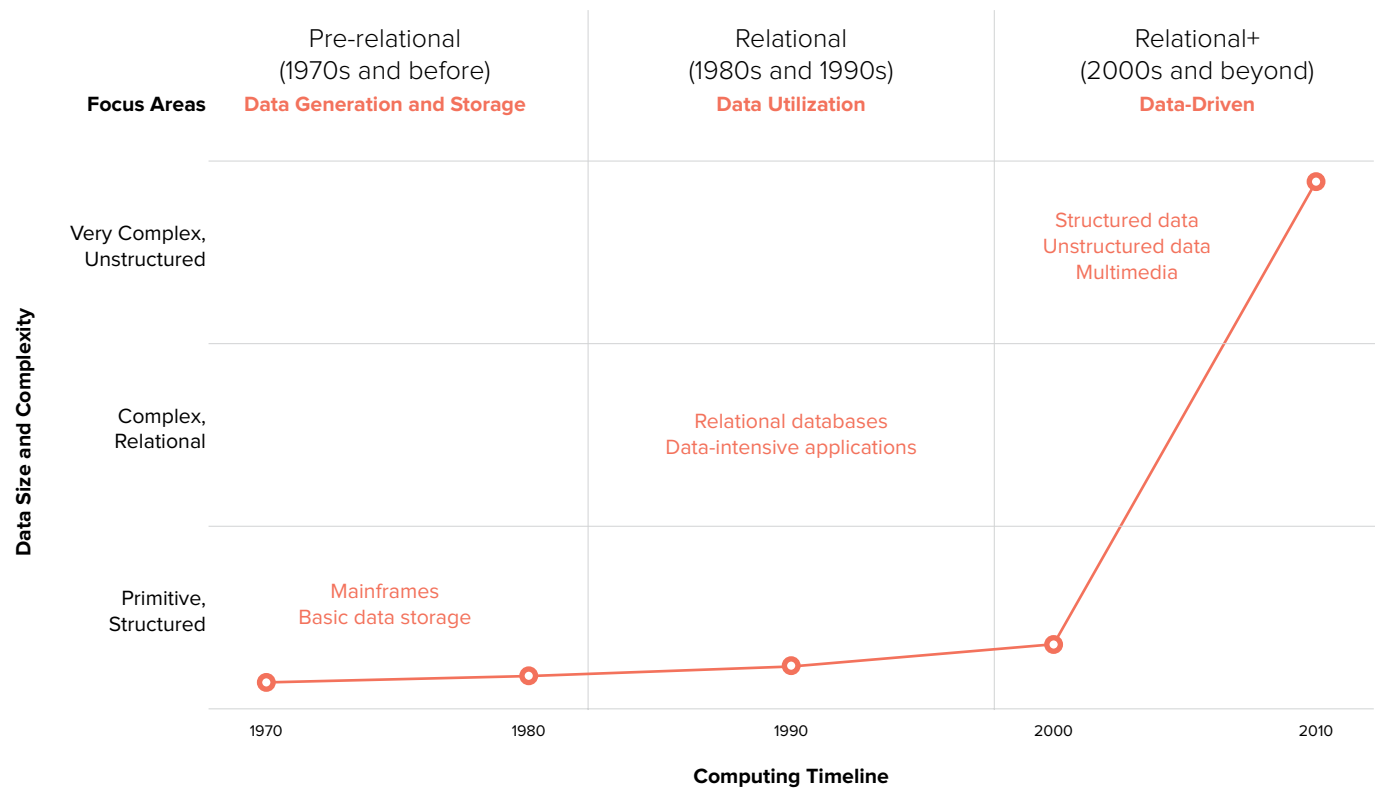
### Natural Language Processing

Pull from a deep linguistic database to understand the most probable meanings of data.



# Evolution of Big Data

We didn't arrive at big data overnight; it took decades for the foundation to form that would allow skyrocketing quantities of data to be useful. But as each supporting technology evolved—the relational database, SQL, the Worldwide Web, data storage, and networks—our power to process data amplified, converting our oceans of available data into meaningful information repositories.







## CHAPTER 4

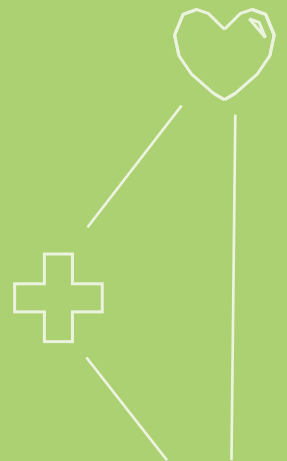
---

# Today, Search is Everywhere

**Apps you might not think of as “search” are, in fact, search apps at their core.**

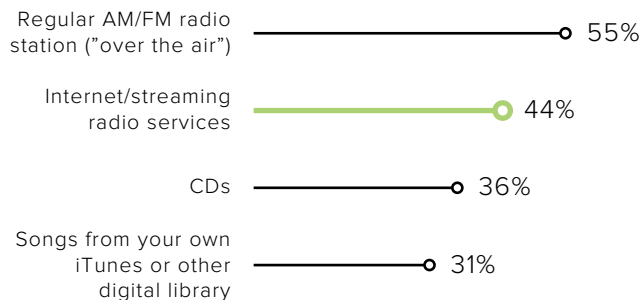
If you think your searching behaviors are limited to Google, ask yourself a few questions: Where did you find your last job? Or your last roommate? How do you find information about your computer? Do you listen to Pandora? Have you used a dating site recently?

The value of tools like Craigslist, Zillow, Amazon, streaming radio, and Match.com relies entirely on their ability to easily search and find relevant information. Music, clothing, dating, friends, jobs, furniture—we use these applications all the time, to find our most important things and to make our most important decisions.



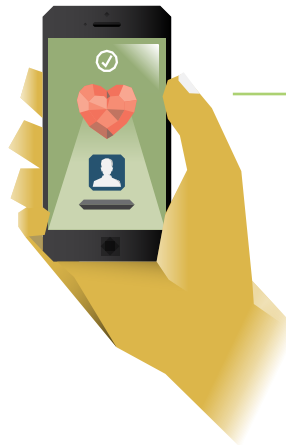
## Searching for Music

### Americans' Typical Music-Listening Sources



Source: Nielsen

## Searching for Love



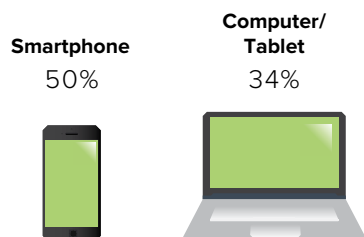
# 38%

of American adults who are "single and looking" have used online dating sites or mobile dating apps.

## Widespread Search Utilities

### Social/Local Shopping

Local searches lead to a high percentage of same-day store visits.



### Health

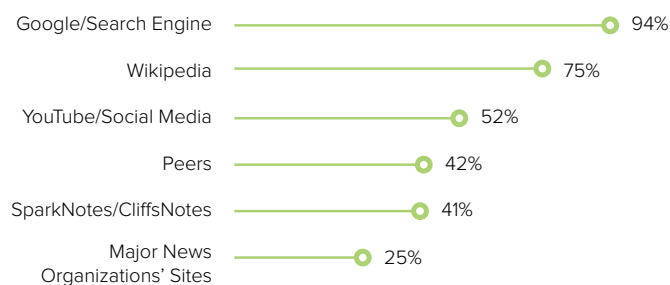
Medical professionals can receive insights on information like diagnoses, treatment results, and behavior-risk correlations from billions of past cases.



80% of health care data is unstructured.

### Education

The sources students are "very likely" to use in a typical research paper, according to middle and high school AP and MWP teachers:



### Live Event/Real-Time Conversations



## Search Sells

E-commerce pioneered big data search utilities to better sell to and understand its customers. Amazon, for example, co-opted the retail book market—and then the retail *everything* market—principally by innovating with large-scale data search technology.



### TV Binging: Brought to You by Search

To guarantee the success of its hit show *House of Cards*, Netflix famously leveraged a NoSQL platform, pinpointing exactly the type of content its users would enjoy by analyzing the viewing habits of its 33 million users.



The NSA reportedly crawls more than **850 billion** records of phone calls, emails, cell phone locations, and Internet chats.

With “Google-like” search capabilities, the United States government is able to flag and quickly respond to real-time suspicious behavioral patterns.

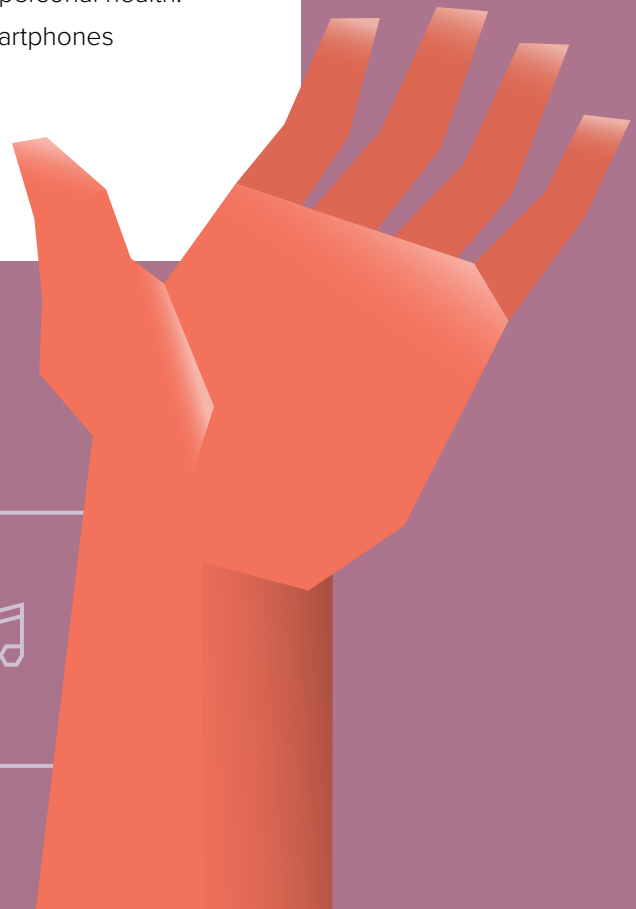
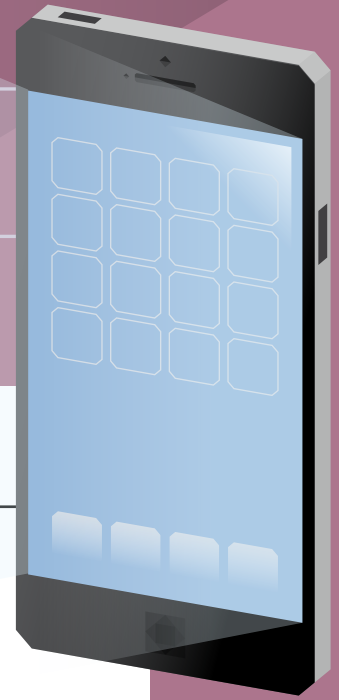
## CHAPTER 5

---

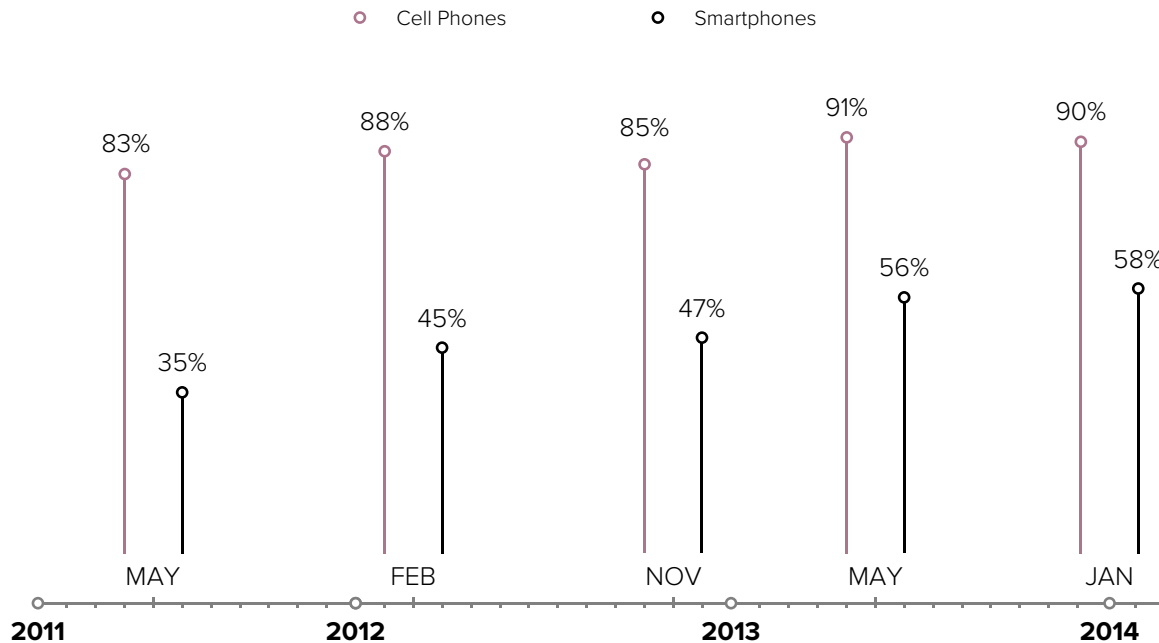
# Mobile Ubiquity Amplifies Search's Power

### **Is that a supercomputer in your pocket?**

It's been called the second digital revolution. Now a necessary accessory in nearly everyone's pocket or purse, the smartphone provides tools for everything from photography to personal health. In many ways, we can't imagine life without our smartphones because of everything we use them for.



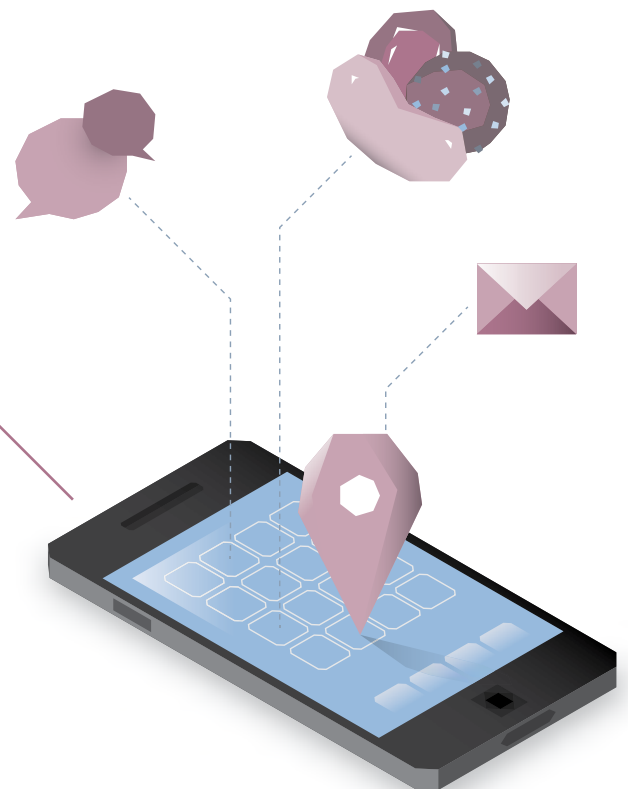
## Americans' Device Adoption



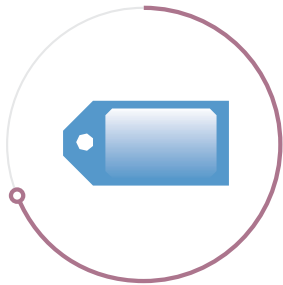
**We call, we text, we email, we read, we route, we film, we publish, we score-check, we Candy-Crush. All on one device. That is a ton of data we are accessing and submitting to our mobile devices on a daily basis.**

**29%** of cell phone owners describe their phones as “something they can’t imagine living without.”

The recent advent of virtual assistants across all devices isn’t merely a way for cell phone companies to attach a friendly personality to your technology. Insofar as its Natural Language Processing (NLP) abilities are useful for task-prompting, Siri, Cortana, and Google Now all make for amiable sidekicks. The big-picture value for these tools, however, is their unprecedented access to personal user data.



## Cortana Knows You Best: The Perks Users Say They Want From Virtual Assistants



**69%**

Take an instruction to find the best deals for you



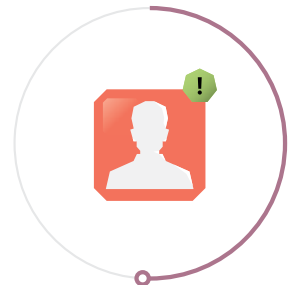
**65%**

Help manage a health condition by monitoring vitals



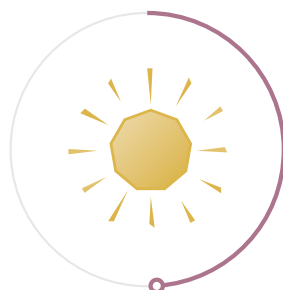
**59%**

Tell you to avoid walking down a particular road at night



**51%**

Alert you to a friend's presence nearby



**49%**

Tell you when to apply sunscreen based on your skin type and UV level



**47%**

Intervene when you're spending too much



**42%**

Support lifestyle changes through positive suggestions



**41%**

Provide small-talk updates before a work meeting

## Search and Machine Learning: From Awareness to Hyper-Awareness



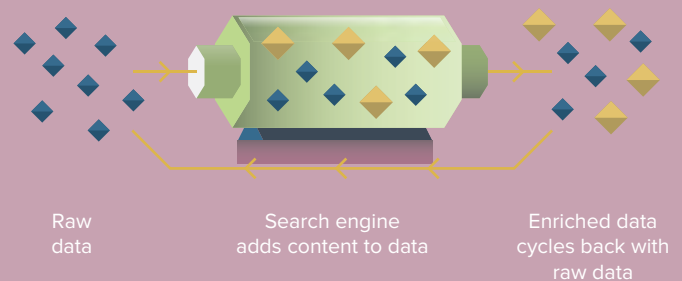
Indeed, virtual assistants represent a new era in data searching, equipping context awareness with a “global” lens—one that combines the knowledge of a smartphone user’s every action with the abundant information available via cloud servers. Combine hyper-awareness with perfect NLP, and you get something close to an ideal of artificial intelligence.

**This is why virtual assistants, born of search technologies, are such a critical piece of the tech story. They’re a launchpad for many of the not-so-far-off developments we’ll explore in the final chapter.**

## Key Concepts

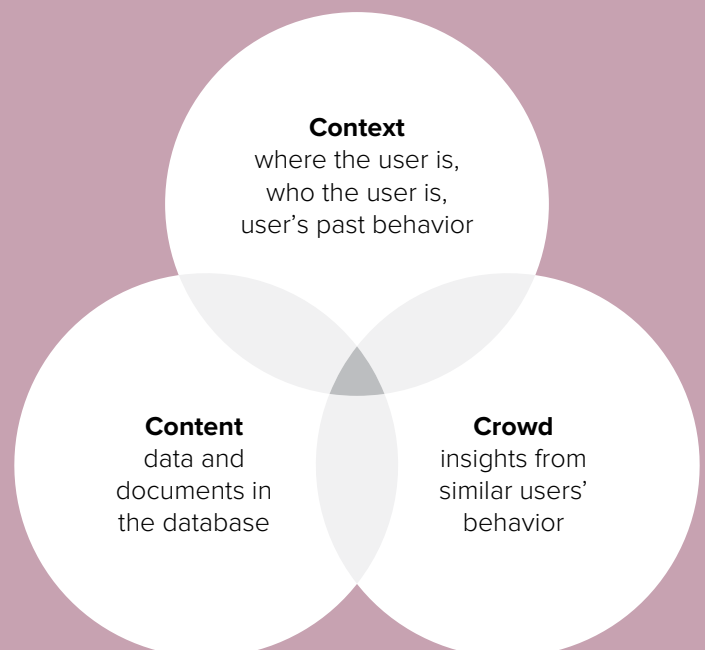
### Data Enrichment

The more data available to an engine, the more informed results it can return. So why leave data’s meaning to the humans? Search engines have gotten smarter by actively enriching data—essentially machine-indexing databases by affixing raw data with additional context, in order to make more sense of raw data.



### Signal Processing

Signal processing is the tech that goes into weighing the various sources of information to calculate the most relevant results.







## CHAPTER 6

# The Future of Search

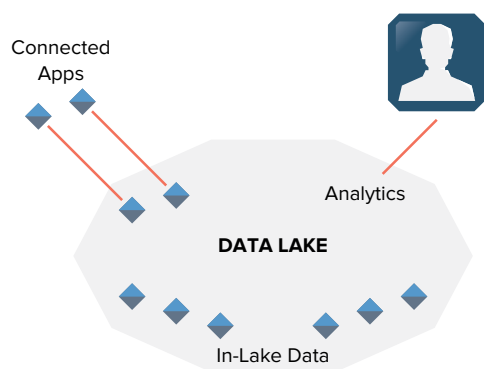
Search technology will have a major role in some of the biggest stories of the 21<sup>st</sup> century.

### **Let's Address These From Mildest to Most-Likely-to-Freak-You-Out:**

- A new data paradigm
- Artificial intelligence
- Real-time, omniscient virtual assistants
- Singularity



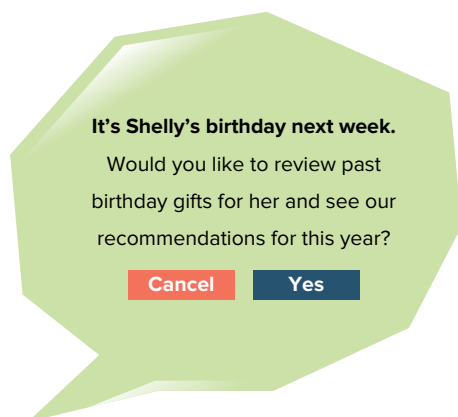
# A New Data Paradigm



## Metaphor Metamorphosis: From the File Cabinet to the Data Lake

When we first started dealing with digital data, we were accustomed to using the file cabinet to store our information in real life. Thus, the digital filing metaphor was an analog translation to the digital world: a tree of files and folders.

Now that search is king, search engines' ability to pick up on files' multidimensional attributes makes manual data organization into simple buckets and themes (files and folders) completely obsolete.



## The Searcher Trumps the Things Being Searched

As search increases its hold on holistic and context-aware data apps like virtual assistants, familiarity with human search habits will become second-nature to the search engine process. Proactive, unsolicited recommendations will become the norm, as stored data takes a backseat to user habits.

## 2

# Artificial Intelligence

## Higher Machine Learning

We previously discussed machine hyper-awareness as a precursor to artificial intelligence. Now consider the immense quantities of valuable data being added to the cloud every second of the day.

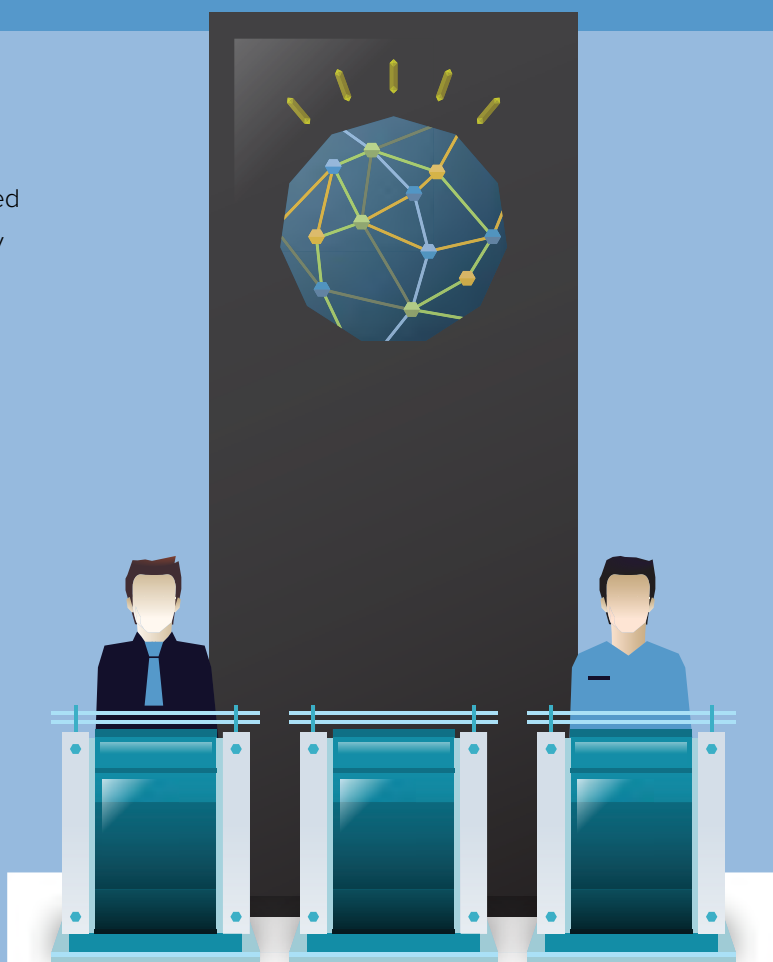
We are not far-removed from the existence of hyper-intelligent machines that leverage search technologies to cull, process, and deliver real-time answers—pulling context both from the user and infinite reservoirs of cloud data.

## You Know My Methods, Watson

Millions of households watched as IBM Watson, an artificially intelligent machine, famously bested super-champions Ken Jennings and Brad Rutter on the televised game show *Jeopardy!* Watson's intuition is based largely on tech developed by search engines.

- Using natural language processing to understand questions
- Querying an enriched, 15-terabyte database using the prompt's keywords
- Processing and comparing top answers with the context given

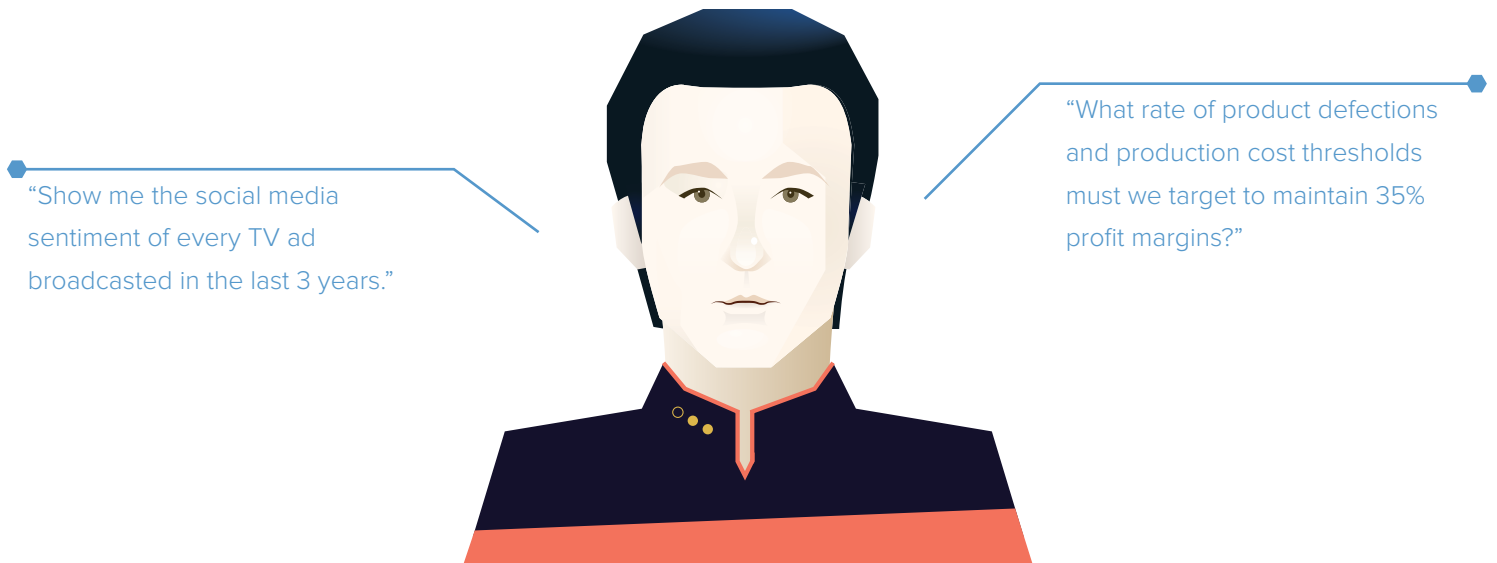
**Watson has begun beta testing for business analytics, and before long it will be a public tool for consumers.**



### 3

## Real-Time, Omniscient Virtual Assistants

*Star Trek* fans will recall the way crew members talked to Computer—the all-knowing, articulate, and sometimes-sassy software aboard the ship. We'll likely interact with virtual assistants in similar ways. The future of Siri, Cortana, and related technology is developing quick, accurate answers to your questions.

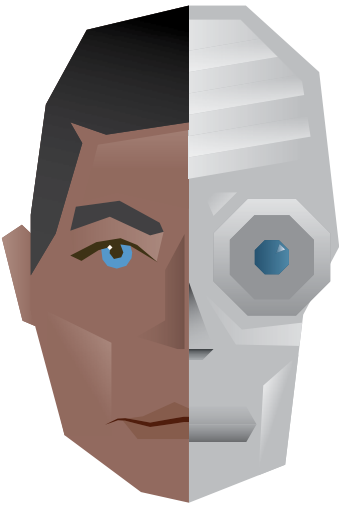


“A string of recent acquisitions in the field of robotics and machine intelligence, along with the recent hiring of [leading singularity theorist] Ray Kurzweil as a director of engineering, shows that Google is by no means done with machine learning: It is clear that the company is just getting started.”

– **Andrew Sheehy, Generator Research**

## 4

# Singularity

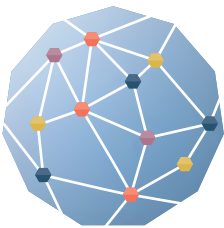


## The End of Search as Merely an Appendage

We possess technology that can make sense of uncured, chaotic, unstructured data. Billions of searches are performed per day—each one a data point to help search engines grow and learn. An enormous portion of the world's information is being consolidated on cloud servers. Before long, machines will be able to understand even the most casual human speech, and the distinct separation between human and device will blur. The emerging wearable technology industry will have a profound impact on our interactions with tech as a whole.

All of these conditions contribute to a realistic vision of singularity: the merging of man and machine. Our search habits won't so much be an extension of our lives as an extension of our physical beings. We'll communicate seamlessly and naturally with technology, depending on it to return valuable insights and manage our lives for us.

### Early Signs of Machine Development



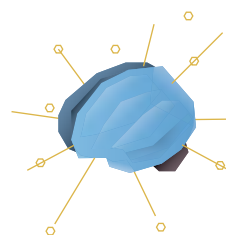
Netflix is building a “neural” network to connect data to each other in the cloud.



A mind-controlled body apparatus helped a paralyzed man walk and kick a soccer ball at the World Cup.



Google built a network able to identify videos of cats.



IBM simulated 4.5% of the human brain, with 147,456 processors working in tandem to imitate 1 billion neurons and 10 trillion synapses.



# How Far Do We Want it to Go

Our level of comfort (or discomfort) with our data being collected will be one of the only major limitations on search's influence in our lives.

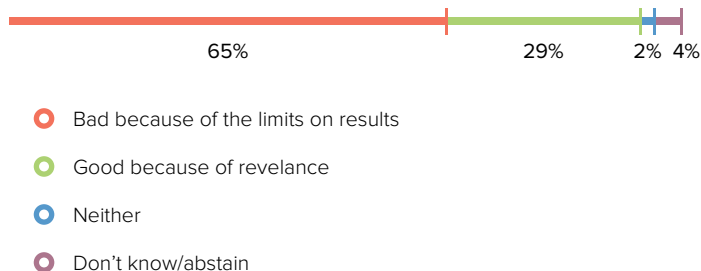
## Public Sentiments About Personal Data Being Used by Search Engines



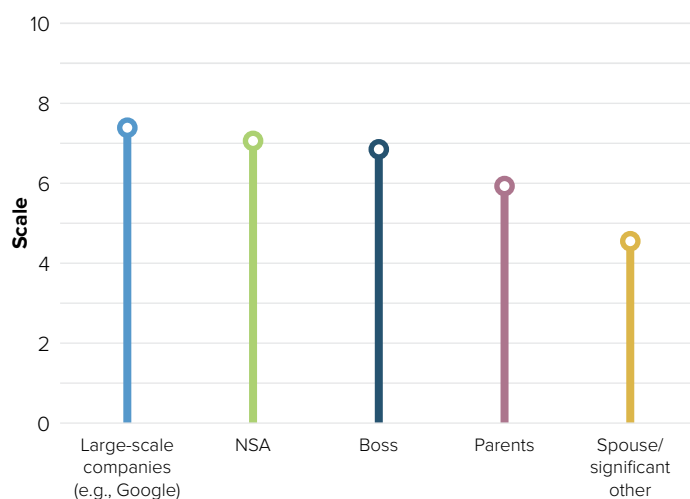
Interestingly, a majority of those same people also did not perceive an improvement in quality.

Search engines may, in fact, be the *least* trusted of all data gatherers, causing more outrage than even the NSA.

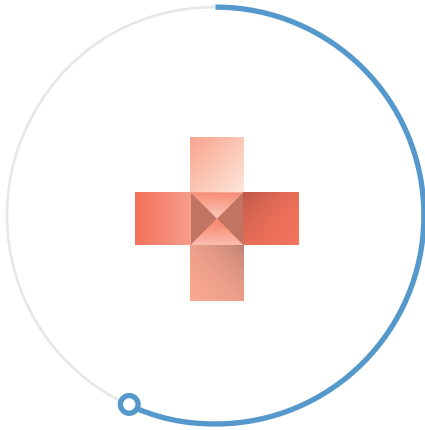
## Public Sentiments About Quality of Personalized Search Results



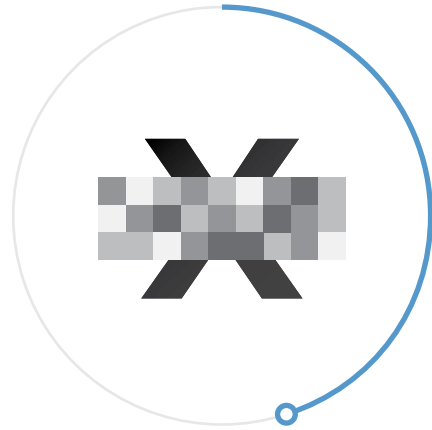
## Public Outrage About Data Collecting, by Organization



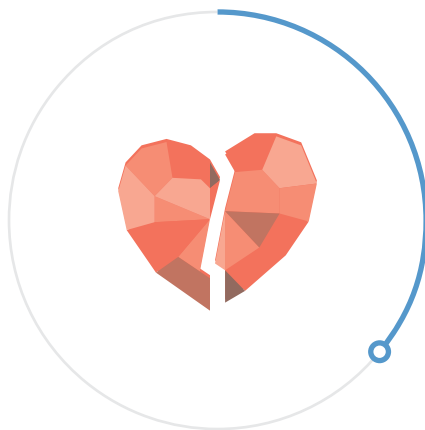
## What Data is Most Sensitive to Us?



**57%** Medical



**45%** Pornography



**36%** Information on  
old relationships



**15%** Pirated media

## Skeptical About Singularity

If devices or implants fed most people information,  
it would be a change for the:



If brain implants were used for memory or mental  
capacity, it would be:



*Note: These charts omit neutral responses and therefore do not add up to 100%.*



Search is at the center  
of humanity's next great  
evolutions.



This has been a Lucidworks production.

Lucidworks builds enterprise search solutions for some of the world's largest brands. Fusion, Lucidworks' advanced search platform, provides the enterprise-grade capabilities needed to design, develop and deploy intelligent search apps—at any scale. Companies across all industries, from consumer retail and health care to insurance and financial services, rely on Lucidworks every day to power their consumer-facing and enterprise search apps. Lucidworks' investors include Shasta Ventures, Granite Ventures, Walden International, and In-Q-Tel.

**Learn more at [lucidworks.com](http://lucidworks.com).**

**Sources, in order of first use:** World Wide Web Consortium (W3C), University of Oslo, Mark Scheuler, Stanford University, Pew Research Center, International Data Group, Oracle, A.T. Kearney, International Data Corporation, *Forbes*, IBM, Nielsen, edUi Conference, *The Intercept*, Google, *CNNMoney*, Mindshare, *Wired*, NPR, *Scientific American*, *InformationWeek*

Share this Ebook with a colleague or friend.

