



MASS  
TECHNOLOGY  
LEADERSHIP  
COUNCIL

## BIG DATA AND ANALYTICS

*A Major Market Opportunity for Massachusetts*



# Contents

<b>EXECUTIVE SUMMARY</b> .....	1
<b>BIG DATA AND ANALYTICS: EVERYONE IS TALKING</b> .....	2
<b>MASSACHUSETTS HAS DATA DNA</b> .....	4
<b>WHAT IS BIG DATA?</b> .....	6
The Data Tsunami of Volume, Velocity, and Variety .....	6
Responding to the Tsunami: Big Data Products and Services .....	7
<b>MASSACHUSETTS COMPANIES ARE INNOVATING ACROSS THE STACK</b> .....	8
The Technical Challenges of the Data Tsunami: Data Integration and Management .....	8
Generating Insights and Value: Innovation in Analytics, Visualization, and Applications .....	8
Helping Companies Harness Big Data: The Role of Service Providers .....	9
<b>NOTABLE MASSACHUSETTS COMPANIES: PAST AND PRESENT</b> .....	10
The Roots of Big Data in Massachusetts .....	10
Up-and-Coming Massachusetts Big Data Companies .....	12
<b>THE BIG DATA JOBS MARKET IN MASSACHUSETTS</b> .....	14
Already Six Percent of IT Jobs and Growing .....	14
Big Data Job Growth in the Massachusetts Technology Sector .....	14
Big Data Job Growth in the End Users .....	15
The Broader Impact on Massachusetts Employment .....	16
<b>SUMMARY: MASSACHUSETTS — A STATE WITH DATA DNA</b> .....	17
Looking Forward: MassTLC's Efforts to Support Massachusetts' Data DNA .....	17
<b>APPENDICES</b> .....	18
Massachusetts Database and Analytics Hall of Fame .....	18
Massachusetts Big Data and Analytics Company Roster.....	21
Database Sizes: International System of Units .....	23

## Mass Technology Leadership Council (MassTLC) Research Committee

**Co-Chairs:** Mike Kinkead, SVP, Urban Wallace, & former Chairman, MassTLC  
Steve O'Leary, Founder, Aeris Partners, & former Chairman, MassTLC

**Lead Contributors:** Bob Buday, Co-Founder, The Bloom Group  
Leo Hermacinski, Managing Director, Luneside Enterprises

**Committee Members:** Fran Grigsby, Founder, Next Level International & MassTLC Board Member  
Tom Hopcroft, President & CEO, MassTLC  
Juby Jacob, Director of Information Systems, Iron Mountain  
Dave Power, CEO, Power Strategy  
Bob Zurek, Vice President, Endeca Latitude Product Strategy

Thank you also to Kent Bennett, Nilanjana Bhowmik, Brenon Daly, Martin Fleming, Michael Goodman, Scott Jarr, Michael Mandel, Marilyn Matz, Mike Stonebraker, Angela Shen-Hsieh, and Evangelos Simoudis for their contributions to the material in this report.

# Executive Summary

The “Big Data” era has begun. Investors, technology entrepreneurs, giant companies of the technology industry, and the media and consulting sectors have seized upon the enormous Big Data opportunity.

It's not hype. We are drowning in digital data. As individuals, we use and generate data with our smart phones, GPS devices, laptops, Facebook, and Twitter, just to name a few examples. Machines are in on the act as well: Internet-enabled instrumentation and sensors capture huge and rapidly growing volumes of data. Retailers and consumer product companies analyze our Web surfing and our purchases to create better products, services, and customer experiences.

So where is the business opportunity? Businesses embrace Big Data to know more about their customers, sooner, generating large and rapid returns on their Big Data investments. Hospitals can deliver effective, personalized medicine to patients. Governments can mine enormous data resources to more effectively serve constituents. Energy companies can bring more renewable energy on-line using a highly instrumented smart grid, smart meters at the home, and increasingly accurate weather forecasts. Drug researchers use vast genomics databases to discover the next generation of cancer treatments. Big Data is already an important part of the \$64 billion database and data analytics market<sup>1</sup> and offers commercial opportunities of a comparable scale to enterprise software in the late 1980s, the Internet boom of the 1990s, and the social media explosion of today.

Massachusetts is a world leader in Big Data. Boston-area entrepreneurs have launched dozens of database and analytical technology startups over the last decade (some of which have sold recently for more than \$1 billion). The Massachusetts Technology Leadership Council (MassTLC) has identified nearly 100 companies in the Boston area currently engaged in Big Data technologies, plus some 20 startups still in stealth mode. Altogether, MassTLC estimates these Big Data vendors currently employ about 12,000 people in Massachusetts. Looking forward, the jobs will come not only from vendors but also from the end-user community. MassTLC estimates that there are currently an additional 58,000 people employed as data scientists and data-savvy managers in industries such as healthcare, financial services, life sciences, consumer products, and on-line media. Growth in both areas could add 50,000 jobs and bring the state total to as many as 120,000 Big Data jobs by 2018.

Massachusetts benefits from its “Data DNA.” This Data DNA encompasses decades of accumulated technical and managerial expertise in predecessor companies that created database, infrastructure, and application software, as well as communications and storage technologies. Today, MassTLC believes the fundamental elements are present in the local ecosystem to make Big Data the next major technology cluster in the state. Indeed, the essential elements of a vibrant cluster are here: we have technology expertise and entrepreneurial zeal in our nearly 100 companies, an experienced and active venture community, extensive and respected university research in both information technology and business, a deep well of management talent, and a driven yet collaborative culture supported, in part, by our industry networking associations. Equally important is the participation of our leading firms and institutions in such sectors as biotech, life sciences, healthcare, and financial services, all of which have pressing needs to incorporate big data and analytics into their operations.

In fact, **Massachusetts has the potential to become the premier location for Big Data expertise in the world.**

**Big Data is already an important part of the \$64 billion database and data analytics market<sup>1</sup> and offers commercial opportunities of a comparable scale to enterprise software in the late 1980s, the Internet boom of the 1990s, and the social media explosion of today.**

<sup>1</sup> Bank of America/Merrill Lynch analyst report, “Big Data Heralds Golden Era of Database Market”, November 22, 2010.



# BIG DATA AND ANALYTICS

## Everyone Is Talking

From IBM to McKinsey, *The Economist*, Wall Street, and Google, thought leaders are focused on the explosion of data collected, processed, and analyzed by whole new families of database and analytical software. Their message, delivered with remarkable consistency, is that Big Data and analytics software represent a transformational approach to a vast array of business and societal challenges—from identifying medical cures to creating a truly smart grid for electrical distribution to enhancing the efficiency of extended supply chains.

McKinsey's internal think-tank, the McKinsey Global Institute, published a major study in June 2011 on Big Data.<sup>2</sup> Its overarching conclusion: **Big Data is “a key basis of competition and growth.”** McKinsey predicts that successful companies will create a C-level data specialist role and that “data equity” will take its place alongside brand equity, financial equity, and human capital as a key business asset.

The normally hype-averse *Economist* called Big Data “the raw material for future revolutions...as usual, the reality of the digital age is outpacing fiction.”<sup>3</sup> The publication noted the extraordinary pace of data collection; the amount of data collected just last year would have filled 60,000 Libraries of Congress.

Venture funding for Big Data companies is at a fever pitch. The 451 Group reports venture capitalists invested \$157 million in the first eight months of 2011 in just two

categories of Big Data technologies (Hadoop and NoSQL). By mid-November, the total 2011 venture investment in these categories had grown to \$350 million—more than triple the \$95 million invested in all of 2010.<sup>4</sup>

Venture money is flowing because the opportunity is large and the financial returns are already proven. Four Big Data companies—Greenplum, Aster Data, Netezza, and Vertica—have sold for a combined \$2.5 billion. These four companies received total venture funding of \$245 million.<sup>5</sup>

There is a groundswell of interest and activity here in Massachusetts. People in the Big Data sector are participating in events ranging from formal programs, such as MassTLC's annual Big Data Summit, to “meet-ups” organized by the Boston Predictive Analytics Group, to informal attendee-generated discussions, such as a group that formed at MassTLC's 2011 unConference.

If further evidence of Massachusetts' strengths in Big Data is needed, look no further than IBM and its three-year-old “Smarter Planet” strategy based on Big Data and analytics. Smarter Planet builds upon important themes of “green” and energy efficiency. Underlying these themes are such broad concepts and Big Data applications as smart traffic, smart food, smart healthcare, smart government, smart water, and smart retail.<sup>6</sup> For IBM, “smart” refers to industry transformation by becoming “more instrumented, more interconnected, and more intelligent.”<sup>7</sup>

<sup>2</sup> McKinsey Global Institute report, “Big data: The next frontier for innovation, competition, and productivity,” published June 2011.

<sup>3</sup> *The Economist*, May 26, 2011, Schumpeter

<sup>4</sup> The 451 Group report, “Big Money for ‘Big Data.’”

<sup>5</sup> The four companies were Greenplum (acquired by EMC Corp.), Netezza, Vertica Systems, and Aster Data. Netezza and Vertica were based in Massachusetts.

<sup>6</sup> *Business Green*, January 09, 2009

<sup>7</sup> *Ibid*, *Business Green*

Through internal development and acquisition, IBM is building a range of Big Data products and services. Its acquisition strategy shows the importance of made-in-Massachusetts data and analytics expertise. In March 2005, IBM announced its purchase of Westborough-based Ascential Software for \$1.1 billion—a price that was four times Ascential's 2004 revenue and 73 times its earnings.<sup>8</sup> In 2007, IBM paid \$4.9 billion to acquire the Canadian business intelligence software company Cognos, whose U.S. headquarters were in Burlington, Massachusetts (Prior to the acquisition, Cognos purchased Applix, a Westborough-based analytics software company, for \$339 million). Most recently, in 2010, IBM paid \$1.7 billion for Netezza Corp., a

## These global heavyweights know the value of Massachusetts Big Data and analytics expertise.

Marlborough-based provider of high-performance data warehouse technology.<sup>9</sup>

IBM isn't the only global technology firm hunting for Big Data acquisitions in Massachusetts. In October 2011, Oracle announced a \$1.1 billion acquisition of Endeca Technologies, a Cambridge-based company that sells business intelligence software, to complement its November 2010 acquisition of Cambridge-based retail software vendor ATG. Also in 2010, Oracle paid \$685 million for Phase Forward, a company that manages pharmaceutical clinical data. These global heavyweights know the value of Massachusetts Big Data and analytics expertise.



<sup>8</sup> Ascential's GAAP-reported 2004 revenue was \$272 million and its net income was \$15 million. IBM's press release regarding its acquisition of Ascential can be found here: <http://www-03.ibm.com/press/us/en/pressrelease/7561.wss>

<sup>9</sup> IBM's press release regarding its acquisition of Netezza can be found here: <http://www-03.ibm.com/press/us/en/pressrelease/32514.wss>

# MASSACHUSETTS HAS DATA DNA

Massachusetts is ready to capitalize on Big Data. Massachusetts has Data DNA. The state enjoys a long history of creating sophisticated computing technology and software—including database and analytics technology—enabled by an impressive cluster of established companies, entrepreneurs, venture capital investors, leading academic researchers, and influential customers.

Historically, the firms behind many major IT innovations have clustered in certain geographic regions. Geographic centers of technological innovation become self-perpetuating, as their concentration of talent creates an ecosystem that attracts additional talent, companies, and investors. In the 1960s and 1970s, for example, Silicon Valley became the center for semiconductors and memory chips. By the 1980s, Boston and the Route 128 technology corridor became the hub for minicomputer firms, while Silicon Valley later became home to a breadth of PC hardware, components, and software startups. Microsoft and Amazon have made the Seattle area an important center of software activity. Most recently, Silicon Valley has become the epicenter for the Web's social media firms.

Massachusetts, as a technology region, has decades of innovation in database and analytical technologies. Cullinet Software, a database vendor often recognized as the industry's first packaged software provider, was the first software company to achieve a \$1 billion market capitalization.

**Currently, Massachusetts has nearly 100 market-shaping Big Data companies, with another 20 still in stealth mode.** These companies include established firms such as

Endeca (recently acquired by Oracle), and growing younger companies such as VoltDB, Tokutek, ParElastic, Paradigm4, NuoDB, and Hadapt. These younger companies are creating new database structures and analytical tools to address the unique challenges and opportunities posed by Big Data.

The local and national venture capital community recognizes Massachusetts' growing cluster of Big Data entrepreneurs, and several investors have already been handsomely rewarded. For example, Oracle's \$1.1 billion acquisition of Endeca delivered an exceptional return on investment for Venrock and Bessemer Venture Partners, which each maintain a major Cambridge presence. One year before the acquisition of Endeca, IBM acquired Netezza, a Big Data company that had been public for three years. The IPO and subsequent acquisition delivered strong returns to local venture firms Matrix Partners, Charles River Ventures, and Battery Ventures.

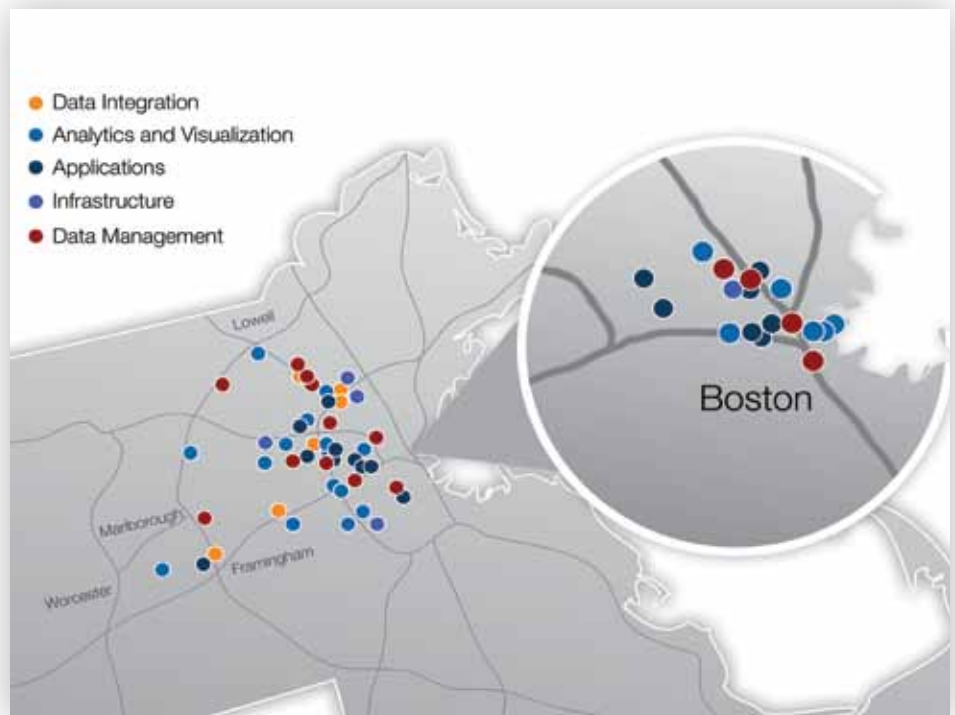


Figure 1: Massachusetts Big Data and Analytics Companies

**Although acquisitions of Massachusetts Big Data companies by firms headquartered elsewhere might suggest the state’s specialized expertise could migrate to other regions, it doesn’t appear to be playing out that way. In fact, IBM now employs an estimated 5,000 people in the state,** a large proportion of them in the data and analytics companies it has acquired over the last decade. Oracle has a similar presence in Massachusetts. After California and Colorado, Massachusetts has the third-largest number of Oracle employees of any state.

The region’s deep well of academic research in technology provides a strong foundation for Big Data startups. MIT has long been a center of thought leadership in new database and analytics technologies. MIT Professor Mike Stonebraker, a leading researcher and entrepreneur, has founded seven data companies on the basis of his research. Complementing MIT, Harvard, Northeastern, Tufts, Boston University, and other Massachusetts universities are world leaders in a range of disciplines, including biotech and life sciences, healthcare, and business management—all sectors at the forefront of applying Big Data and analytics and likely the first to reap the rewards.

In the end, those who solve previously unaddressable problems and create entirely new businesses will capture the value of Big Data. Massachusetts has leading companies and institutions in the sectors most likely to lead the adoption of Big Data. Native organizations, such as Partners Healthcare, Harvard Medical School, Tufts Medical School, Biogen/Idec, Boston Scientific, State Street, Fidelity, and Wellington Management are joined by others, like Novartis, Philips Healthcare, and Beijing Genomics, that chose to locate important pieces of business in the state. This physical proximity to our Big Data, analytics companies, and local academic wealth creates a crucible for real innovation in applying these powerful technologies to real-world

problems. This proximity underpins the well-proven “bump factor” that has contributed to Massachusetts innovation and business creation for over three centuries.

Big Data demands a pool of deep technical and analytical talent, which positions Massachusetts as a vital resource for regions lacking these skill sets. MassTLC estimates the State’s Big Data vendors currently employ about 12,000 people, approximately 6% of the overall number of jobs in Massachusetts’ IT sector.<sup>11</sup> MassTLC estimates there are currently an additional 58,000 people employed as data scientists and data managers in technology-driven industries such as healthcare, financial services, life sciences, consumer products, and on-line media. MassTLC expects an increase of Big Data vendor jobs as well as a need for highly skilled “practitioners” in companies and institutions that harness the power of Big Data. Based on recent growth trends, MassTLC projects that the State’s total Big Data employment could grow by 50,000 jobs for a total of 120,000 jobs by 2018 as organizations adopt and integrate Big Data into their operations.

Which region becomes the leader in Big Data is yet to be determined. We believe Massachusetts is positioned to become the world epicenter for Big Data. We have a strong foundation and we have all the necessary ingredients.

**Which region becomes the leader in Big Data is yet to be determined. We believe Massachusetts is positioned to become the world epicenter for Big Data.**

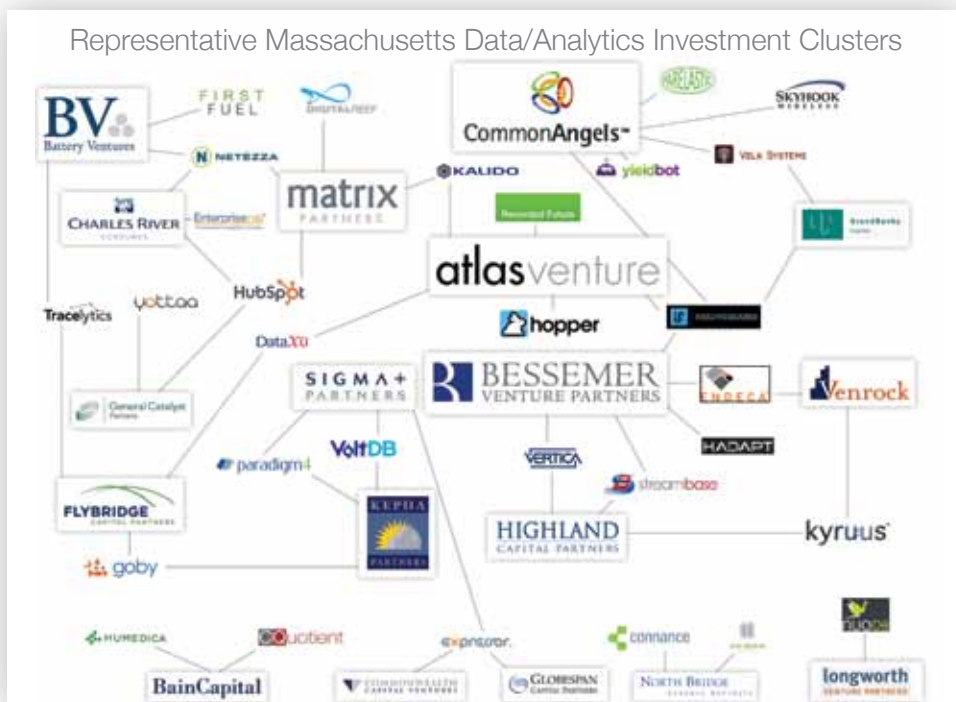


Figure 2: Massachusetts Big Data and Analytics Investment Clusters

<sup>10</sup> MassTLC estimate

<sup>11</sup> “The IT Industry: Hub of the Massachusetts Technology Economy.” University of Massachusetts Donahue Institute, November 2009



# WHAT IS BIG DATA?

## The Data Tsunami of Volume, Velocity, and Variety

McKinsey defines Big Data as “datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze.”<sup>12</sup> Thus, Big Data is a relative term. In the 1980s, the database management systems of IBM, Oracle, Cullinet, and Sybase could have been viewed as Big Data tools of that era. But they were not designed to handle the unforeseen explosion of data brought on by the Internet, mobile communications, and sensor networks.

Computer scientists are not alone in recognizing that the amount of data collected, processed, and stored in computers has skyrocketed over the last 10 years. The proliferation of computing devices—from personal computers in virtually every home and office to smart phones carried by hundreds of millions of people (including 90 million Americans this year<sup>13</sup>)—represent major sources of Big Data. The term “Big Data” encompasses customer data, healthcare data, sensor and machine data (often referenced as “the Internet of Things”), and the vast array of other data types including images, video, and mobile data, as well as huge volumes of unstructured data found in online social networks, blogs, and forums.

McKinsey estimates that in nearly every U.S. industry sector, companies with more than 1,000 employees each managed at least 200 terabytes of stored digital data in 2009. This represents twice the volume of data Wal-Mart stored 10 years earlier, when the giant retailer’s annual revenue was \$137 billion. In fact, **IBM claims 90% of today’s stored data was generated in just the last two years.**<sup>14</sup> From global giants to small local players, just managing this volume is problematic, let alone making smart, real-time decisions based on the data.

The challenge of Big Data is well-described by the three dimensions of volume, velocity, and variety:

■ **Volume**—the sheer amount of data that is created or captured by companies and other organizations.

According to market research company International Data

Corp., the world will store about 1,800 exabytes of digital data this year, a volume that is growing at a 60% compound annual growth rate (CAGR). (Note: One exabyte of data equals one trillion gigabytes.) MIT professor Michael Stonebraker, a co-founder of seven Big Data companies, refers to this crush of digitized information as a “data tsunami.”

■ **Velocity**—the speed with which companies must collect, process, analyze, and act upon information. Many industries now need databases that can operate at previously unfathomable speeds and require applications that can act with corresponding rapidity. In addition, Big Data “never sleeps,” as both the number of online businesses and consumers around the world make data collection, processing, and management a 24x7 task. Looking forward, the processing of velocity data in real-time will drive further innovation.

■ **Variety**—data types comprise (1) “structured” data (such as transactional data) that has long filled relational databases with their orderly rows and columns, and (2) “unstructured” data that might include social media postings, audio and video files, text files, and other complex data types. Popular social media sites represent a major source of unstructured data: Twitter, Facebook, and YouTube support hundreds of millions of users generating enormous amounts of valuable but unstructured data.

The companies behind the previous generation of database management and data mining tools (from the early 1980s to the early 2000s) optimized these tools for just two of the three “V’s”—velocity and volume—and at much lower levels than vendors address today. The universe in which they operated typically comprised high-level transactional data from sales, finance, and the administrative functions of business. Further, these vendors did not build products that could handle a wide variety of data types, because few

<sup>12</sup> McKinsey Global Institute report, “Big Data,” page 11.

<sup>13</sup> eMarketer.com

<sup>14</sup> IBM, “Bringing Big Data to the Enterprise.” <http://www-01.ibm.com/software/data/bigdata/>

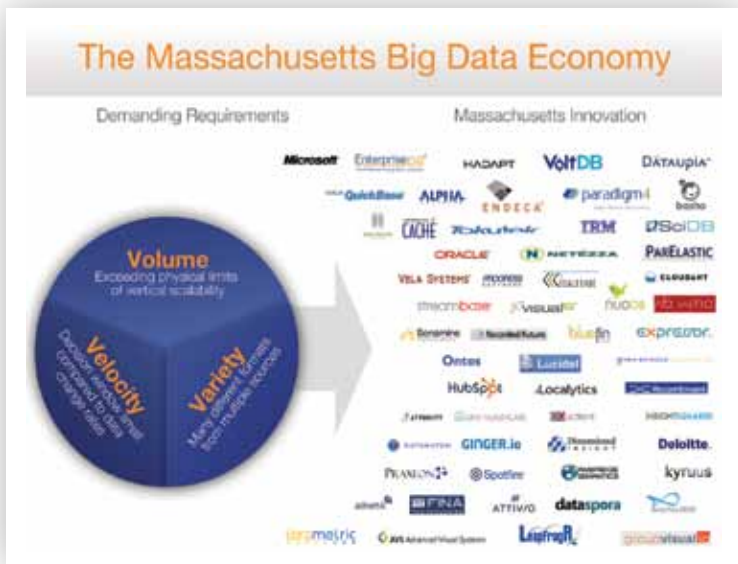


Figure 3: Big Data Drivers and Massachusetts Companies

organizations had any substantial amounts of digitized video, audio, or unstructured text.

Today, Massachusetts companies are aggressively solving the challenges of the “three Vs.” Figure 3 shows a sample of the companies and entrepreneurial activity here in the state.

## Responding to the Tsunami: Big Data Products and Services

Today, a rapidly expanding number of organizations need databases, data mining tools, and applications that can ingest, comb, and analyze vast quantities of data at unprecedented speeds.

The Internet companies were the first to be hit by the “data tsunami.” Their needs were so pressing that Google, Facebook, Yahoo!, eBay, and Twitter developed their own database infrastructures and technologies. The popular Hadoop Big Data application, now maintained as an Open Source project by the Apache Software Foundation, is the most prominent example. Yahoo! has been the largest contributor to the project and has launched Horton Works to commercialize its Hadoop implementation. Facebook is also a prominent Hadoop user. In July, Facebook reported that it stores some 30 petabytes of information on Hadoop, a volume equivalent to 3,000 Libraries of Congress.<sup>15</sup>

We see six areas of innovation in Big Data that we call the Big Data and Analytics Stack:

- **Infrastructure** to store and move large amounts of data<sup>16</sup>
- **Data integration tools** to ingest vast quantities of high-velocity data
- **Data management solutions** to manage the data
- **Analytics and visualization** that transform data analysis to business insights
- **Applications and solutions** that incorporate new database and analytical tools to address industry- and function-specific Big Data problems
- **Service providers**, including analysts, consultants, and system integrators, that help companies understand and implement new Big Data technologies

Massachusetts entrepreneurs, established companies, and their investors are lined up against all these opportunities (see Figure 4). In addition to the nearly 100 companies identified by MassTLC, we have found some 20 additional companies in stealth mode that will increase the state’s leadership in Big Data. Massachusetts leadership is well-understood: **IBM, Google, HP, Microsoft, Oracle, and Teradata came to Massachusetts to acquire local companies, because they recognized the know-how that exists here.**



Figure 4: The Big Data and Analytics Stack

<sup>15</sup> Paul Yang, Facebook Engineering Notes, July 27, 2011.

<sup>16</sup> Note that infrastructure is outside the scope of this document.



# MASSACHUSETTS COMPANIES ARE INNOVATING ACROSS THE STACK

The technical and business innovation spurred by Big Data runs the gamut from highly technical database structures, through astonishing data visualization and analytic tools, to service providers that help their clients implement these tools to benefit from Big Data.

## The Technical Challenges of the Data Tsunami: Data Integration and Management

The challenges of volume and velocity force a fundamental rethinking of “speed,” presenting two separate but related problems: a) handling the incoming data, and b) speedily processing that data for analysis. Massachusetts companies such as Expressor, Ab Initio, and Kalido have staked out positions in fast, efficient, parallel-based data integration. They perform the tasks of extracting, transforming, and loading large amounts of data into a data management system.

Handling data for ultra-fast and efficient transactions and analysis is the next challenge in the stack after data integration. Companies have taken a variety of technical approaches to give optimal performance to specific database problems. For example, VoltDB, based in Billerica, has built software delivering order-of-magnitude improvements in throughput to handle real-time feeds, machine-generated data, micro-transactions, and to serve content at super speeds. Other companies have focused on managing the data in anticipation of complex analytics. Paradigm4 provides an analytical data platform that quickly delivers versioned data and advanced analytic results from massive scientific, industrial, and commercial databases. Its technology enables scientists to conduct tasks such as analyzing the vast amounts of genomics sequencing data needed in drug discovery, and supports data scientists building new pricing and risk models.

The growing variety of data presents another important set of opportunities. Social media, texts, blog posts, Web browsing, and mobile devices are producing vast amounts of very valuable but unstructured data. Increasingly, an

organization’s most precious data are no longer in a neat and orderly structured form, but in 140-character tweets with arcane abbreviations and slang, in a doctor’s patient notes, or in a customer complaint form. Endeca and other Massachusetts companies are at the forefront of ingesting unstructured data, adding “traditional” data sets (that is, more structured information), and rapidly producing the analysis needed to make informed decisions.

Today, there are more than 30 Massachusetts companies involved in the data integration and data management layers of the stack—companies ranging from startups in stealth to well-established market leaders such as Vertica, Netezza, and Endeca. There is significant activity at all stages of development due to the technical expertise, investor interest, and academic underpinnings here in Massachusetts. We expect a continuing stream of innovation, new company formation, and rapid revenue and employment growth.

## Generating Insights and Value: Innovation in Analytics, Visualization, and Applications

**Although many see the greatest innovations in Big Data to be in the data integration and data management layers of the stack, more opportunities actually lie beyond: in software tools and analytic methods that will enable cutting-edge database technologies to address real-world problems.** Just as Big Data spans a wide variety of applications and a large number of industries, we see a similar breadth of companies active in building analytic and visualization tools, applications, and solutions. These tools enable organizations to make better business decisions based on the massive amounts of data available. Over 60 Massachusetts companies are solving these problems.

Innovation in analytics, applications, and visualization tools is critical for the widespread adoption of Big Data. In fact, at a Big Data session during MassTLC’s 2011 unConference, a wide cross-section of people from data management

companies said effective analytics and application tools were vital to their success. Fortunately, Massachusetts is well positioned for success in this area with more than 25 companies developing these tools.

Some of these companies offer powerful tools to build visually intuitive representations of large datasets. Examples include Advanced Visual Systems, which has tools to “convert data into actionable insight”<sup>17</sup> across a variety of business uses, and InsightSquared, enabling small and medium-sized businesses to become highly analytic and data driven in managing their sales activities. Endeca’s Latitude product (see Figure 5 below) builds clear and intuitive views of complex, large datasets. Other companies are targeting analytics and visualization in particular segments, such as Sonamine in digital gaming, VisualIQ and Yieldbot in online marketing and digital media, and Kyruus in healthcare delivery.

Massachusetts companies are also active in designing applications and solutions to particular problems. Not surprisingly, several firms are focused on healthcare—a data-rich and complicated area that is widely anticipated to be one of the sectors most impacted by Big Data. Ginger.io, GNS Healthcare, Praxeon, and LeapFrogRx are among the local companies bringing Big Data approaches to healthcare. Other companies such as Vela Systems, Interval Data Systems, and Recombinant are pursuing Big Data opportunities in the construction, energy management, and bio-informatics sectors.

## Helping Companies Harness Big Data: The Role of Service Providers

IT services and research companies, as well as management consulting firms, are building new capabilities to help their clients harness Big Data. Massachusetts is rich with such firms. As they ramp up their Big Data practices, they too will provide significant local employment opportunities for people who can implement Big Data solutions, conduct research on the application of Big Data and Big Data vendors, and advise companies how to reorganize their businesses and the strategic thinking of their top managers. These companies are integral in assisting others harness Big Data, as described here:

■ **IT services companies:** Massachusetts is home to numerous IT services firms and branches of global IT services firms such as Accenture, Deloitte, PwC and

CSC. More than 3,500 such firms operated in the state in 2007, employing nearly 47,000 people, according to the US Census Bureau.<sup>18</sup>

■ **IT research companies:** Large research firms, such as International Data Corp. (IDC) and Forrester Research Inc., are based in Massachusetts. Their analysts enjoy local access to the over 100 Massachusetts firms that are leading the Big Data charge.

■ **Management consulting firms:** Massachusetts is headquarters to two of the world’s largest consulting firms: Boston Consulting Group and Bain & Co. In addition, the state has a number of influential mid-sized consulting firms including Parthenon Group and Monitor Group, and branch offices of big players such as McKinsey and Oliver Wyman. These firms will help their clients adopt Big Data into their strategies and operations.

These three sectors will be at the center of designing, implementing, and supporting real-world Big Data applications. The knowledge and expertise they build and disseminate to the broad business community will be an important catalyst in the widespread adoption of Big Data and analytics.



Figure 5: Endeca Latitude Visualization Example

<sup>17</sup> Advanced Visual Systems website

<sup>18</sup> US Census Bureau 2007 Economic Census



# NOTABLE MASSACHUSETTS COMPANIES—PAST AND PRESENT

Massachusetts brings a rich history of innovation and commercial success in database and analytical technologies. The following company profiles provide examples of the region's prominence.

## The Roots of Big Data in Massachusetts

### Cullinet Software

Among the early commercial success stories, Cullinet Software (originally known as Cullinane Corp.) stands out. According to Curt Monash's blog, *Software Memories*,<sup>19</sup> and a Wikipedia article on which he commented, John Cullinane and Larry English founded Cullinane Corp. in 1968 to develop packaged software for users of mainframe computers. This was a new idea at the time, because hardware vendors such as IBM, Unisys, Control Data, and Sperry Rand controlled the development of software systems.

After a series of successful, albeit minor products, Cullinane Corp. took over development rights to a database product it renamed IDMS that had been modified to operate on IBM mainframe systems. Cullinane made a highly successful decision in 1973 to focus on this product and built a strong management team to pursue the opportunity. On the market embrace of IDMS, the company (then known as Cullinet Database Systems, Inc.) went public in 1978 and was listed on the New York Stock Exchange in 1982. Cullinet was acquired by Computer Associates in 1989.

Cullinet established a number of industry firsts:<sup>20</sup>

- The first software company taken public by Hambrecht & Quist

- The first software company to achieve a \$1 billion market capitalization
- The first software company to advertise in the Super Bowl
- The first software company to be listed on the NYSE
- The first software company to be included in the S&P 500
- Greylock's first investment in a software company

Cullinet alumni represent a Who's Who of the industry including:

- John Cullinane – John has served as an angel investor in numerous Massachusetts startups and was instrumental in founding the Massachusetts Software Council, now MassTLC.
- Bob Goldman – After Cullinet, Bob went on to found AI Corp., which became Trinzic, a publicly traded provider of expert systems. Bob also founded Object Design, which merged to become Excelon and ultimately became part of Bedford-based Progress Software. Bob has also been active on a number of boards, including Net.Genesis, an early Web analytics company.
- Bob Weiler – Bob and John Landry founded a company called Distribution Management Systems (DMS) that was acquired by Cullinet. Bob went on to become a top executive at Lotus, CEO of Giga Information Group, CEO of Phase Forward, and is now a senior executive at Oracle, as well as a MassTLC trustee.
- John Landry – John, a co-founder of DMS, has also served as CTO of McCormick & Dodge and Lotus. He is an active angel investor in Massachusetts and a MassTLC trustee.

<sup>19</sup> Curt Monash, *Software Memories Blog*, May 27, 2008.

<sup>20</sup> Wikipedia

## Computer Corporation of America

Founded in 1965, Computer Corporation of America (CCA) developed high-performance mainframe databases including its flagship product, Model 204. CCA operated as an independent company until the late 1980s, when it was acquired by a Canadian insurer in a diversification move. Subsequently spun out, it again became an independent company, headquartered in Framingham and later Waltham, until its acquisition by Rocket Software, a Newton-headquartered company, in 2010.

Model 204 led the market for very large-scale, high-capacity, high-performance data management. CCA is credited with such inventions as hashing, bitmapped indexing technologies, the first online database management system (DBMS), the first enterprise-scale 4GL, and the first partitioned tables.<sup>21</sup> Model 204 continues to operate some of the most challenging database applications in the world.

## Interbase Software

Jim Starkey founded Interbase Software in 1984 in Groton, Mass., after a 10-year career with Digital Equipment Corp. The Interbase product, perhaps borrowing from concepts in DEC's relational database offering and work at CCA, was one of the first to offer multiversion concurrency control. Put simply, this feature enables a system to manage concurrent changes by many users. Other systems at the time used locking mechanisms to prevent multiple users from simultaneously accessing or changing a particular record. Interbase also offered a new rollback and recovery feature that enabled rapid recovery from faults. Other innovations included the first commercial implementations of heterogeneous networking, blobs, triggers, two-phase commit, and database events.<sup>22</sup>

The company achieved early success working with Apollo Computer in the engineering workstation market. Interbase was acquired by Ashton-Tate, which in turn was acquired by Borland. Today, the software is marketed by Embarcadero Technologies. A version exists as the Open Source Firebird project.<sup>23</sup>

Today, Starkey is a co-founder and CTO of NuoDB and works with co-founder and CEO Barry Morris. Morris is the former CEO of Iona Technologies (which was based in Dublin, Ireland, and Waltham) and StreamBase Systems.

## Applix/Netezza

Jit Saxena, a former Data General employee, founded Applix in 1984 to create UNIX desktop applications. As the desktop applications market moved to the PC, Applix repositioned itself as a provider of business performance management systems and business intelligence solutions, an early pioneer in the analytics market. Its TM1 product, a multidimensional online analytical processing (MOLAP) application, drew strong reviews and set the stage for the company's 2007 acquisition by Cognos, the leading business intelligence vendor. Cognos, which maintained its U.S. headquarters in Burlington, was subsequently acquired by IBM.

Saxena left Applix in 2000 to start a new company called Netezza. Backed again by Waltham-based Matrix Partners, Saxena and Netezza established the market for data-warehousing appliances, a combination hardware and software system designed to optimize management of information for analytics purposes. Netezza was acquired by IBM for \$1.7 billion, joining Applix, Cognos, Ascential Software, and a number of other Massachusetts companies in IBM's Big Data portfolio.

IBM now has an estimated 5,000 employees in Massachusetts including a strong cadre of database and business intelligence companies.

Saxena is now an angel investor and has joined the board of a new wave of Massachusetts companies including Waltham-based Actifio, which provides data management virtualization and storage solutions.

As one industry observer remarked, "the companies here come and go, but the people remain the same." Massachusetts benefits to a remarkable degree from this phenomenon. The above profiles represent just a few examples of the region's deep roots in this sector and clearly leave out many remarkable stories of pioneers like Peter Gyenes (Ascential Software), David Friend (Pilot Software and now Carbonite), and Joe Alsop of Progress Software, each of which have made major contributions.

Much like a family tree, IT innovations are passed down from one generation of technology founders to the next. The current generation of Massachusetts Big Data startups has inherited much from the prior generations of database technologies. This legacy forms our Data DNA.

<sup>21</sup> Xconomy Boston, March 29, 2010, Transaction Report.

<sup>22</sup> Nimbus DB Web site ([www.nimbusdb.com](http://www.nimbusdb.com))

<sup>23</sup> Firebird Web Site ([www.firebirdsql.org](http://www.firebirdsql.org))

# Up-and-Coming Massachusetts Big Data Companies

So which Massachusetts companies might become the leading vendors of this decade? Already there are many contenders, and new ones are constantly being formed. While we don't have a crystal ball, and we are not in the business of picking winners, we nonetheless believe that Big Data's future can be seen in several local companies that have had exciting starts out of the gate: Hadapt, Bluefin Labs, Kyruus, and two new Stonebraker companies, Paradigm4 and VoltDB.

## Hadapt

Hadapt has architected the first Big Data analytic platform to be natively integrated with Apache Hadoop™. Hadapt is expanding the Hadoop architecture with a more complete SQL interface, a patent-pending Adaptive Query Execution™ capability, and a hybrid storage engine to handle both structured and unstructured data in a single platform. It began offering early access to its flagship product in November 2011.

Justin Borgman, CEO, co-founded Hadapt in July 2010 with Daniel Abadi and Kamil Bajda-Pawlikowski. A native of Acton and a 2002 computer science graduate of the University of Massachusetts at Amherst and Yale MBA, Borgman spent the first six years of his professional career as a software developer at MIT Lincoln Laboratory and Raytheon. Abadi, Hadapt's chief scientist, received his PhD from MIT. His dissertation was about a technology (column-store database systems) that led to the launch of Vertica. Abadi is an assistant professor at Yale's computer science department. Bajda-Pawlikowski and Abadi developed the prototype for Hadapt's software at Yale.

Led by Norwest Venture Partners and Bessemer, the firm closed a Series A funding round of \$9.5 million in October 2011,<sup>24</sup> which followed an initial angel round of \$1 million in January 2011.

The company had 15<sup>25</sup> employees in October and has moved from New Haven, Conn., to Cambridge, Massachusetts.

## Paradigm4

Paradigm4 helps data scientists, analysts, and researchers solve some of their most challenging data management and analytical problems in industries from pharmaceutical,

biotechnology, and healthcare analytics to insurance, energy research, and finance. Just as the influx of physicists to Wall Street brought new tools and innovation to financial technologies, Paradigm4 brings the power of scientific data management and massively scalable complex analytics to commercial and industrial applications. Paradigm4 sponsors the development and use of SciDB, an open source, massively scalable, analytical platform that runs on both commodity hardware grids and in the cloud.

Paradigm4 was founded by renowned Boston database entrepreneur and MIT professor, Mike Stonebraker, along with Marilyn Matz who previously co-founded Massachusetts-based Cognex Corporation. They earned Series A funding from Sigma Partners and Kepha Partners. Their name, Paradigm4, comes from a monograph by Jim Gray on a new, fourth paradigm for research: data-driven discovery.

## VoltDB

This Billerica-based company created a relational database system for Big Data based on the design of an in-memory database by MIT professor Mike Stonebraker, Daniel Abadi (see Hadapt above), and Samuel Madden (an MIT computer science professor). Its technology was incubated within Vertica (a major Massachusetts Big Data company) and was spun out in 2009 before that firm's acquisition by HP.

VoltDB's technology is especially targeted at the "velocity" aspect of Big Data—helping organizations process, analyze, and make rapid decisions on huge volumes of transactional data. The firm is led by CEO Scott Jarr, who previously was Vice President of product management and marketing at online backup leader LiveVault Corp., before it was acquired by Iron Mountain. Mike Stonebraker is CTO. VoltDB has about 20 employees.

VoltDB released its database software in May 2010 as an open source product. In August 2010, it received a \$5 million "A round" investment from Sigma and Kepha. In October 2010, it released v1.2 as both an open source and a commercial product. The company has numerous customers, particularly in three sectors: capital markets, networking and communications, and digital advertising.

Jarr sees VoltDB's products becoming crucial to organizations that need to a) make real-time decisions based on huge volumes of transactional data, and b) conduct real-time analytics on such data.

<sup>24</sup> As reported by GigaOM

<sup>25</sup> Company-supplied data

## Bluefin Labs

Bluefin Labs is a social TV analytics company providing solutions to brand advertisers, advertising agencies, and TV networks. The company's technology and data enable clients, for the first time, to tap into data at scale that links people's social media commentary to the shows and commercials they watch on TV. Clients of Bluefin Labs include the TV networks CBS, ABC, Turner Broadcasting, Discovery Communications, MTV, and others. The company's agency and brand clients include: Publicis' Starcom MediaVest Group (SMG), WPP's Mediacom, IPG's Initiative, PepsiCo, Mars, and Humana.

The company was co-founded by CEO Deb Roy, an MIT Media Lab professor, and his PhD student, Michael Fleischman (now President and CTO). Based in Cambridge, they named their company after a Porter Square sushi bar where they dined. Since then, Bluefin Labs has raised more than \$8 million from Redpoint Ventures, Acacia Woods Partners, Lerer Ventures, and other investors.

Bluefin Labs is currently 40 employees and growing.<sup>26</sup>

## Kyruus

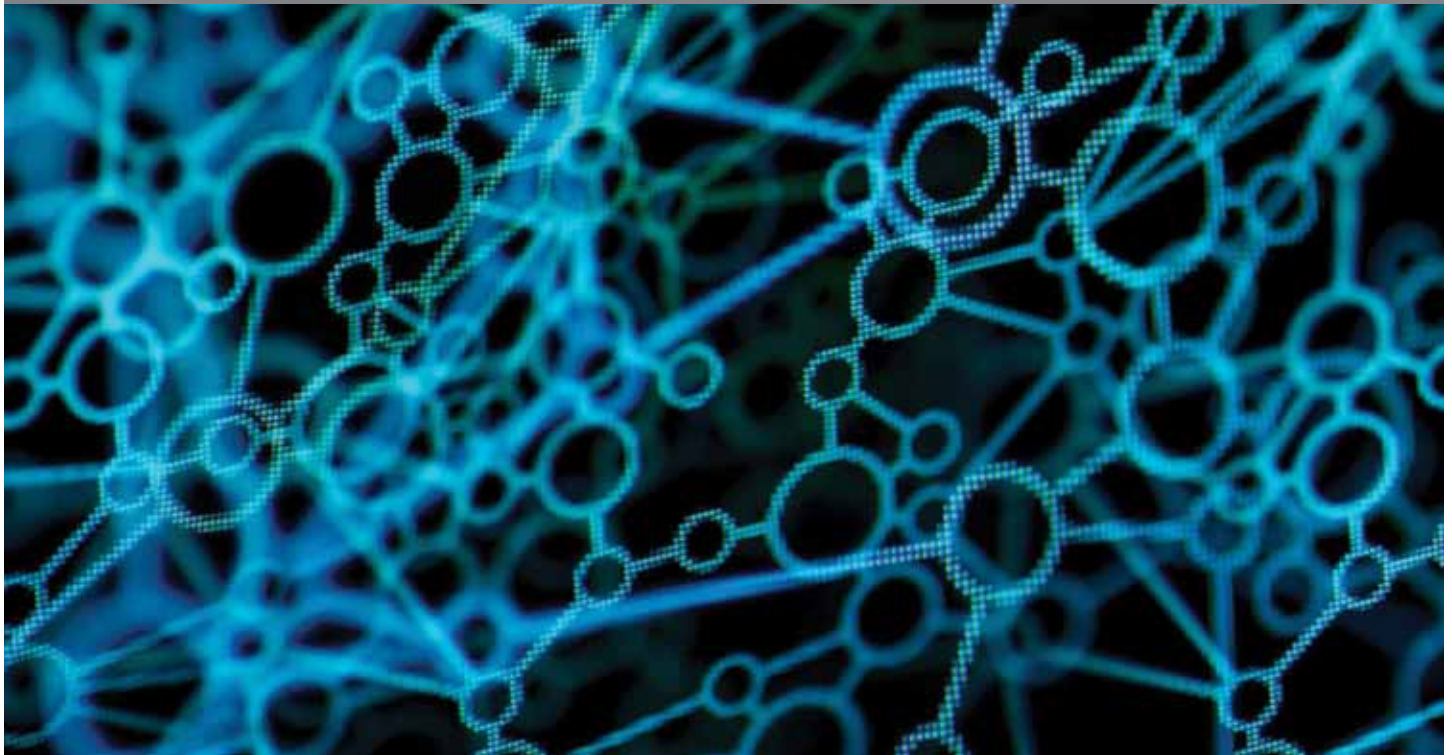
Kyruus takes a "Big Data" approach to helping organizations engage, manage, and optimize large networks of physicians involved in delivering high-quality health care and fundamentally improving our healthcare system.

Founded in 2010 to be the "ITA software of physician information," Kyruus aggregates, collates, and disambiguates millions of public and proprietary information sources for longitudinal analysis and data-driven decisions.

The Kyruus team comprises people from both healthcare- and non healthcare-related industries who are passionate about building compelling products that help their users generate actionable insights from complex data.

CEO and co-founder Dr. Graham Gardner received his medical and MBA degrees from Harvard. Gardner co-founded several companies (SynapDX and Generation Health) before launching Kyruus with co-founders Julie Yoo and Vinay Seth Mota.

Kyruus has raised \$8.5 million in funding to date from leading venture capital firms Highland Capital Partners and Venrock, as well as strategic investors such as Gerson Lehrman Group.



<sup>26</sup> Bluefin Labs-supplied information as of January 5, 2012.



# THE BIG DATA JOBS MARKET IN MASSACHUSETTS

## Already Six Percent of IT Jobs and Growing

Big Data already represents a meaningful component of the Massachusetts employment base. MassTLC estimates that the companies in the state that fit our categorization of Big Data—database, data mining tools, business analytics software, and vertical applications of such software—employ about 12,000 people (the vast majority of which are in-state). These 12,000 employees represent approximately 6% of the current number of IT jobs in Massachusetts. Of these 12,000 positions, about one-third are highly skilled IT jobs.<sup>27</sup> Of these 12,000 positions, about one-third are highly skilled IT jobs.

While Big Data is already a sizable employment sector today, we believe Massachusetts job growth in this area could outpace most other sectors in the coming years for three reasons:

- The growth rate of Big Data and analytics vendors and Massachusetts' significant share of these firms translates to meaningful local employment gains. Continued growth of the Big Data vendors that are already here in-state, and the launch and growth of the next round of Big Data contenders will make Massachusetts home to a vibrant job market that will attract people with the requisite skills.

- The “user multiplier effect:” The rapid adoption of Big Data and analytics applications software across other industries will cause significant job growth. A large number of Big Data jobs will be created outside of the Big Data technology vendors—that is, by the life sciences, healthcare services, retail, manufacturing, business services, and other sectors that purchase Big Data

technology to gain an information advantage for their businesses. McKinsey refers to these users as “data scientists” and “data-savvy managers.”

- The “indirect and induced effect:” These jobs are created “indirectly” in other firms as a result of the Big Data companies' purchases of items as supplies and professional services (e.g., lawyers, accountants, etc.), and is also “induced” by the consumer spending of each vendor's employees (e.g., restaurant employees, retail employees, auto repair, etc.) Many sectors of the Massachusetts economy will benefit from the success of the Big Data companies.

**A large number of  
Big Data jobs will  
be created outside  
of the Big Data  
technology vendors.**

## Big Data Job Growth in the Massachusetts Technology Sector

We already see signs of rapid growth in Big Data employment in Massachusetts. Consider two of the largest local Big Data firms, Netezza and Endeca, each of which commanded billion-dollar acquisition price tags since 2010:

- Netezza's revenue grew fourfold in four years, from \$54 million in 2006 to more than \$220 million in 2010 (when IBM acquired it for \$1.7 billion), a compound annual growth rate of 37%. At the time of sale, Netezza had 425 employees, of which 165 were in research and development.

- Endeca's revenue increased from \$108 million in 2007<sup>28</sup> to \$150 million in 2010, and the firm added more than 130 people between 2010 and 2011.<sup>29</sup> Between

<sup>27</sup> “The IT Industry: Hub of the Massachusetts Technology Economy.” University of Massachusetts Donahue Institute, November 2009.

<sup>28</sup> Endeca, 2007, SEC 10K filing

<sup>29</sup> Endeca, March 15, 2011 press release

2000 and 2010, the company's total employment grew eightfold, from 50 to 400.<sup>30</sup> By July 2011, Endeca had 450 employees, the majority of which are in Massachusetts.<sup>31</sup>

Many local firms are on strong growth tracks, and invariably revenue growth means employment growth (although, not all the jobs of Massachusetts Big Data companies will necessarily be located in the state).

**We see clear potential to more than double employment by Massachusetts Big Data vendors, adding more than 15,000 employees in-state by 2018.**<sup>32</sup>

One-third of these new jobs will be highly skilled R&D positions. We expect the Big Data companies, the broader business community, and local and state government will continue their efforts to make Massachusetts a top-tier location to work and live, and thereby keep the majority of these jobs here in Massachusetts.

## Big Data Job Growth in the End Users

The number of new jobs in companies that become users of Big Data could be larger than the number of jobs in Big Data vendors. As executives across industries implore their organizations to master Big Data skills (a need that, if history repeats itself, will no doubt be sparked by Massachusetts IT research firms, business schools, and consulting firms), we will see many local companies with help-wanted ads for Big Data workers. These workers are the “user multiplier.”

A number of employment trend watchers have pointed to increased demand for workers with deep analytic skills—so-called “data scientists”—who help companies make sense of the digital data locked up in their computer systems. McKinsey estimates that U.S. organizations will create 290,000 to 340,000 new “Big Data” jobs by 2018. The consulting firm also predicts that more than half of those jobs (from 140,000 to 190,000) may go unfilled because of a shortage of skilled workers.

The ripple effect of Big Data on jobs will go beyond the new “data scientist” positions that companies will post. The line managers and executives who run these

companies—the brand managers; heads of marketing, procurement, production, supply chain, sales, finance; and other managers—will need high levels of analytic skills to be able to decipher and act on the data their internal data

scientists bring to them. McKinsey estimates that U.S. companies will need more than 4 million managers with such skills. We believe Massachusetts could generate a greater number of these jobs per capita than other states, given our cluster of Big Data companies, world-leading healthcare services, life science R&D, and financial companies located in the area.

**If Massachusetts gets just its historic share of these skilled analytic jobs, end-user organizations could add 3,700 new data scientist positions and 31,000 new data-savvy managerial jobs between now and 2018.**

Figure 6, below, shows how the job growth in the vendors and the end users builds on the current Big Data employment base, with potential to reach a total of 120,000 Big Data jobs in Massachusetts.

However, McKinsey's projections anticipate that *similar numbers of additional end-user jobs could go unfilled* due to a shortage of people with the right skills, leading to substantial further employment opportunities if universities and corporations rise to this training challenge.

The number of new jobs in companies that become users of Big Data could be larger than the number of jobs in Big Data vendors.

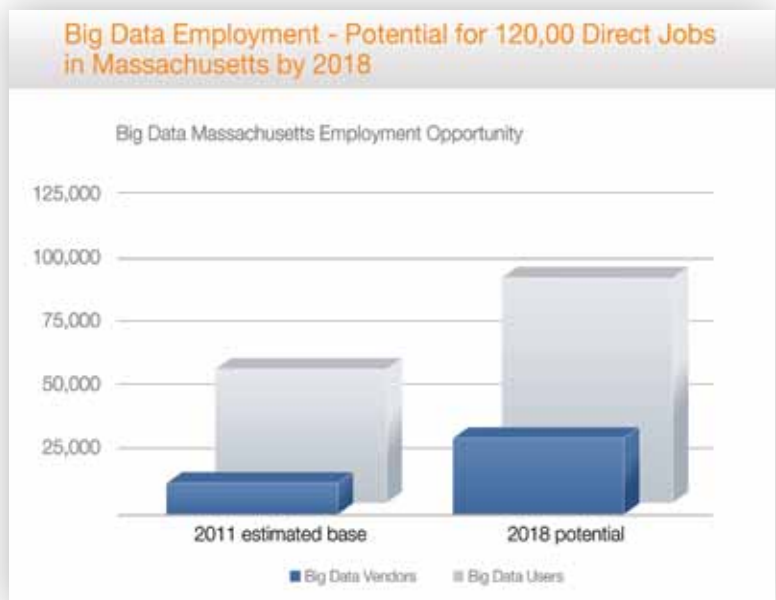


Figure 6: Impact on Massachusetts Employment

<sup>30</sup> Boston Globe, Oct. 24, 2011

<sup>31</sup> “Endeca gears up for likely IPO bid,” Boston Business Journal, July 1, 2011

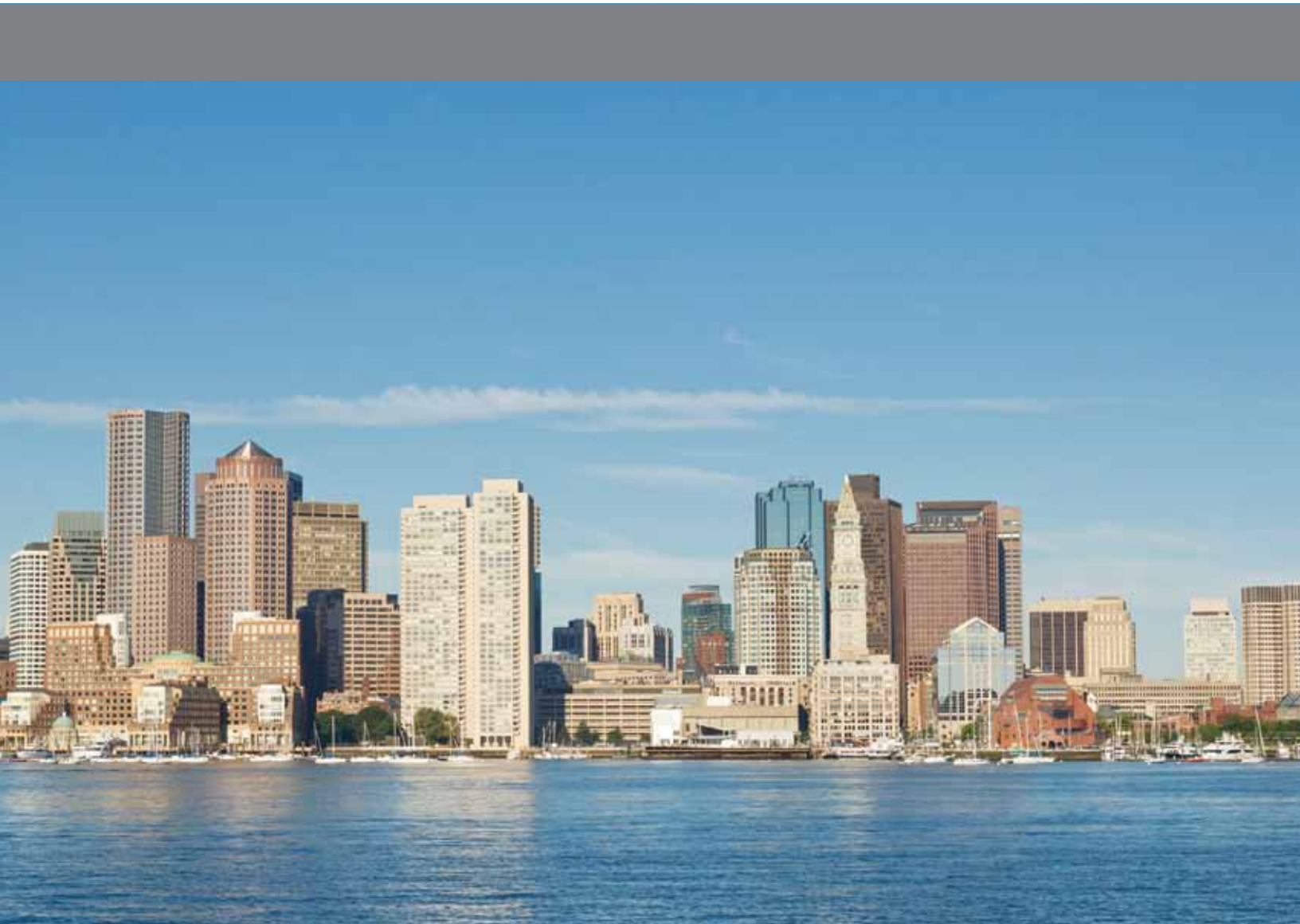
<sup>32</sup> MassTLC estimates

## The Broader Impact on Massachusetts Employment

Massachusetts will see additional jobs beyond Big Data vendors and users. As reported by the University of Massachusetts Donahue Institute in its 2009 study of the IT industry, every 100 IT jobs support an additional 163 jobs in the broader Massachusetts economy.<sup>33</sup> The 50,000 potential jobs added in Big Data vendors and users noted above will influence an additional 82,000 jobs needed to supply and support the vendors and users, as well as support their spending as consumers.

The predicted acute national skills shortage in managers and workers who can use Big Data tools creates the opportunity for some of these 82,000 jobs. Training, education, and data-based decision support are clear and present opportunities for Massachusetts companies, universities, and colleges that are already training executives around the world in such areas as business strategy, IT strategy, entrepreneurship, and other key business areas.

Massachusetts is poised for significant job growth from Big Data in its vendors, end users, and the broader economy.



<sup>33</sup> "The IT Industry: Hub of the Massachusetts Technology Economy," University of Massachusetts Donahue Institute, November 2009.



# Summary: Massachusetts — A State with Data DNA

Massachusetts is poised to remain a leader in the Big Data revolution given its historic and rich ecosystem of leading academic institutions; entrepreneurs spawning database, analytics, and specialized software companies to leverage Big Data; and the local investment community. We see seven factors critical to the Massachusetts cluster of world-leading Big Data activities:

- **The deep well of database and analytics**

**technical talent:** The state's legacy of decades of experience, and today's cutting-edge technical talent is found in more than 100 Big Data companies—from the dozens of startups that exist along Route 128, Interstate 495, and in Cambridge to the Massachusetts operations of the global technology giants.

- **The area's extensive university expertise and research:** Universities such as MIT, Harvard, and UMass are home to some of the world's premier technology professors and graduate students that are pushing Big Data forward in theory and in practice.

- **World-class companies and institutions in key markets:** These include life sciences, healthcare, and financial services.

- **A strong support network for entrepreneurs:** The Boston area is rich with serial entrepreneurs who have become mentors to the next generation of technology industry stars. Many of these mentors are also angel investors.

- **A VC community willing to make big bets on Big Data:** Venture capital firms such as Bessemer, Matrix Partners, Sigma Partners, Kepha Partners, and Highland Capital Partners have been key investors behind local Big Data success stories including Netezza and Endeca.

- **A broad supporting ecosystem for technology and its application:** Research houses, IT services, and management consulting firms, as well as organizations

such as MassTLC, the Boston Predictive Analytics Group, and others are active in bringing entrepreneurs, startups, established companies, investors, professors, and government together.

- **Government support:** Massachusetts government officials have long been active promoters of the region as a great place to start or move a business.

It is this cluster with all the ingredients of success that constitutes Massachusetts' Data DNA.

## Looking Forward: MassTLC's Efforts to Support Massachusetts' Data DNA

MassTLC is a catalyst for growth and development of the Big Data and Analytics ecosystem in Massachusetts. The Council's unique role as an organization that spans the entire technology ecosystem places MassTLC at the heart of the Big Data tsunami. Below are some of the initiatives MassTLC is undertaking to foster the growth of this important sector:

- **Research** – Along with this report, MassTLC will continue to study the region's Big Data and Analytics community through surveys, focus groups, and on-going dialogue with leading big data thinkers.

- **Summit** – MassTLC's 2nd Annual Big Data Summit provides an opportunity for thought leaders to share ideas while the general community learns about opportunities in Big Data and Analytics.

- **Community** – MassTLC's Big Data "Cluster" community will continue the dialogue through a variety of Cluster meetings, events, and initiatives. MassTLC Clusters provide the venue to take a deep dive in a particular area, share knowledge, and build lasting relationships.

■ Bump – Integrating Big Data and Analytics across our other Clusters, Innovation unConference, Tech Tuesdays and other key Council initiatives, thereby fostering knowledge exchange with other segments will support the bump factor.

■ Recognition – Encouraging Big Data and Analytics companies to participate in the Mass Technology Leadership Awards.

■ Workforce – Helping to attract, retain, and develop the talent needed to grow Big Data and Analytics in Massachusetts through the MassTLC Education Foundation.

■ Leadership – Educate and advocate for policies that support the growth of big data and analytics technology development Massachusetts.



# APPENDICES

## Massachusetts Database and Analytics Hall of Fame

This honor roll highlights the Massachusetts individuals who have built substantial database/analytics enterprises and/or contributed to entrepreneurship in this area.

John Cullinane (Cullinet Software)

John Landry (Cullinet Software, Lotus)

Michael Stonebraker (MIT)

Terry Ragon (Intersystems)

Joe Alsop (Progress Software)

Peter Gyenes (Ascential Software)

Jit Saxena (Applix, Netezza)

Ray Ozzie (Iris Associates, Groove Networks)

Christopher Ahlberg (Spotfire, Recorded Future)

Steve Papa (Endeca Software)

### John Cullinane

John Cullinane founded Cullinane Corp., later known as Cullinet Software, a company most analysts view as the first successful software products company in the world. *Computerworld* recognized John as the “creator of the packaged software market.” He led a company that could claim many

industry firsts: first packaged application, first report writer software, first database to seriously compete with IBM on mainframes, and first pure software company to go public. He sold the company to Computer Associates in 1989.<sup>34</sup>

Following Cullinet, John became an active angel investor supporting a new generation of area companies. John invested in such companies as Segue Software and worked with the Massachusetts-based Common Angels. He also authored the book *The Entrepreneur’s Survival Guide*. John was a founder of the Massachusetts Software Council, which today is known as the Mass Technology Leadership Council (MassTLC).

### John Landry

John Landry needs no introduction to anyone involved in the Massachusetts technology community. According to his Web site, John is a “30+ year software veteran and innovator who has served as Chief Technology Officer/Senior VP Development at some great software leaders like Lotus Development, Dun & Bradstreet Software, Cullinet Software, and McCormack & Dodge and VP of Technology Strategy at IBM.”<sup>35</sup>

<sup>34</sup> *Computerworld* “Ten People in IT Who Mattered Most”

<sup>35</sup> Lead Dog Ventures Web Site ([www.leaddog.com](http://www.leaddog.com))

John has been a prolific angel investor and has served on the board of directors of more than 36 companies over the years. He is currently a director of Sonian, Inc., a provider of cloud-based archiving and search solutions. John is also a trustee of MassTLC.

## Michael Stonebraker

Professor and serial entrepreneur Mike Stonebraker has been a groundbreaking database researcher and a technology innovator for more than a quarter of a century. Through a series of academic prototypes and commercial startups, Stonebraker's research and products are central to many relational database systems on the market today. While a professor of Computer Science at University of California at Berkeley, Mike was the main architect of the Ingres relational DBMS and the object-relational DBMS, Postgres. Postgres was the first open source database, has been built-in to many prominent products, and remains in extensive use today.

After moving to MIT, Mike has done seminal work on streaming data, column stores, deep-Web search engines, and transaction processing. That research became the basis for four successful Massachusetts-based startups—StreamBase, Vertica (acquired by HP in 2011), Goby (acquired by TeleNav in 2011), and VoltDB. Most recently, Mike has led research and development for a new generation of databases that support the advanced analytical and data management needs for “big science” as well as for “big analytics” in commercial and industrial applications and has cofounded the open source SciDB project and Paradigm4 to bring that work to market.

In addition to cofounding seven startups, Mike has been awarded the prestigious ACM System Software Award, the first annual Innovation award by the ACM SIGMOD special interest group, the IEEE John Von Neumann award, and has been elected to the National Academy of Engineering and the American Academy of Arts and Sciences.

## Phillip T. (Terry) Ragon

Terry Ragon founded Cambridge-based InterSystems Corp. in 1978 and led the company to achieve current annual revenue of \$335 million and worldwide prominence in database, integration, analytics, and healthcare technology. InterSystems is known for its Cache and MUMPS database platforms. Prior to starting InterSystems, Terry cofounded a healthcare software company that he sold to GE Healthcare. In 2009, Terry and his wife, Susan, endowed the Ragon Institute of Massachusetts General Hospital, MIT, and

Harvard with a \$100 million gift to pursue vaccine treatment for AIDS, an effort that combines expertise in engineering, computational sciences, molecular biology, clinical research, and interdisciplinary research.

## Joseph Alsop

Joseph Alsop is former cofounder and chief executive officer of Progress Software Corp., a global supplier of application infrastructure software used to develop, deploy, integrate, and manage business applications. Alsop has led Progress since its founding to its position as a global software industry leader with annual revenue exceeding \$500 million.

Alsop has over 25 years of management and technical experience in the computer industry. He was founder and chief executive officer of Intercomp, Inc. (acquired by Logicon, Inc.) and later served as president and chief executive officer of Aristonics Corp. He is an angel investor in local Massachusetts database companies.

## Peter Gyenes

Peter Gyenes served as chairman and chief executive officer of Ascential Software Corp., a market leader in data integration software, and its predecessor companies VMark Software, Ardent Software, and Informix prior to their acquisitions by IBM. Peter also served on the board of directors of Applix, a business intelligence software company, and Netezza, a provider of data warehousing appliances, both of which are also owned by IBM. He currently serves on the boards of Lawson Software, Inc., a provider of enterprise applications software; Pegasystems, Inc., a provider of business process management software and services; and VistaPrint Limited, a global e-commerce provider of marketing services to small business. In addition, Peter served on the board of directors of BladeLogic, Inc., a data center automation company, from 2006 to 2008, and webMethods, Inc., an enterprise integration software business, from 2005 to 2007. Peter is an active mentor to software company CEOs and brings some 40 years of experience in technology, sales, marketing, and general management. He has also been an active participant in MassTLC and serves as Trustee Emeritus.

## Jit Saxena

Jit Saxena was the cofounder, chairman and CEO of Netezza. Previously, Jit was founder, chairman, and CEO of Applix Inc., a leading provider of analytical CRM software that he took public in 1994. While at Applix, Saxena guided the company to a leadership position in the CRM market and

achieved an eight-fold increase in market value subsequent to the IPO. Through his vision, the company was continually an innovator in new high growth markets. Prior to founding Applix, Saxena was a senior director at Data General where he was responsible for the development of the company's software products. At Data General, he held several other senior management positions and built a successful development organization that introduced industry-leading software in many categories. Saxena has an MS in Electrical Engineering from Michigan State University and an MBA from Boston University.

## Ray Ozzie

Ray received his bachelor's degree in computer science in 1979 from the University of Illinois at Urbana-Champaign, where he worked on the PLATO system and began his working career at Data General Corporation where he worked for Jonathan Sachs. After leaving Data General, Ozzie worked at Software Arts for Dan Bricklin and Bob Frankston, the creators of VisiCalc. Shortly thereafter, he was recruited by Sachs and Mitch Kapor to work for Lotus Development to develop what became Lotus Symphony. Ozzie left Lotus Development in 1984 and founded Iris Associates to create the product later sold by Lotus as Lotus Notes. Iris Associates was acquired by Lotus in 1994, and then Lotus itself was acquired by IBM in 1995.

Ozzie remained at IBM for several years before leaving to form Groove Networks. Groove was acquired by Microsoft in 2005, where Ozzie became one of three Chief Technical Officers. That year, he wrote a seven-page, 5,000-word internal memo, titled "The Internet Services Disruption:" "It's clear that if we fail to do so, our business as we know it is at risk ... We must respond quickly and decisively." On June 15, 2006, Ozzie took over the role of Chief Software Architect from Bill Gates.

Ozzie stepped down from his role at Microsoft on December 31, 2010. He is currently involved in new stealth-mode company called Cocomo, located in Boston and Seattle.

## Christopher Ahlberg

Chris Ahlberg is currently the CEO and co-founder of Recorded Future. Ahlberg is also an independent director of Hult International Business School, sits on the board of Swedish software corporation, Apptus AB, and holds a board seat at the hedge fund Fina Technologies.

Before cofounding Recorded Future, Ahlberg was the president of the Spotfire Division of Tibco, which he founded in 1996

and sold to Tibco in 2007. Spotfire was founded based on his research on information visualization at the University of Maryland. Ahlberg earned his doctorate from Chalmers University of Technology and worked as a visiting researcher at the University of Maryland. He has been granted two software patents and has multiple patents pending. Ahlberg was named among the World's Top 100 Young Innovators by MIT Technology Review.

## Steve Papa

Steve Papa serves as CEO of Endeca, which was acquired in October 2011 by Oracle. Papa founded Endeca in 1999 to use innovative technology to solve a long-standing class of business problems that stem from information overload. He has since led it to become one of the fastest-growing software companies in the world, with five consecutive years of more than 100% revenue growth. Prior to Endeca, he was an early team member of Inktomi, the pioneer in large-scale parallel Web search, and worked at Teradata, the world leader in large-scale structured data warehouses.

Papa holds a B.S. in Operations Research and Economics from Princeton University and an M.B.A. from Harvard Business School.

## Massachusetts Big Data and Analytics Company Roster

Company	Category	Location	URL
Ab Initio	Data integration	Lexington	<a href="http://www.abinitio.com">www.abinitio.com</a>
Activate Networks	Analytics and visualization	Newton	<a href="http://www.activate networks.net">www.activate networks.net</a>
Adnetik	Applications	Boston	<a href="http://www.adnetick.com">www.adnetick.com</a>
Advanced Visual Systems	Analytics and visualization	Waltham	<a href="http://www.av.s.com">www.av.s.com</a>
Akamai	Infrastructure	Cambridge	<a href="http://www.akamai.com">www.akamai.com</a>
Akiban Technologies, Inc.	Data management	Boston	<a href="http://www.akiban.com">www.akiban.com</a>
Alpha Software	Data management	Burlington	<a href="http://www.alphasoftware.com">www.alphasoftware.com</a>
Attivio	Analytics and visualization	Auburndale	<a href="http://www.attivio.com">www.attivio.com</a>
Attunity	Data integration	Burlington	<a href="http://www.attunity.com">www.attunity.com</a>
Basho	Data management	Cambridge	<a href="http://www.basho.com">www.basho.com</a>
Beijing Genomics Institute	Applications	Cambridge	<a href="http://en.genomics.cn">http://en.genomics.cn</a>
Biff Labs (Stealth)	Applications	Cambridge	<a href="http://www.biff labs.com">www.biff labs.com</a>
Bluefin Labs	Analytics and visualization	Cambridge	<a href="http://www.bluefin labs.com">www.bluefin labs.com</a>
Cambridge Semantics	Data management	Boston	<a href="http://www.cambridgesemantics.com">www.cambridgesemantics.com</a>
Cloudant	Data management	Boston	<a href="http://www.cloudant.com">www.cloudant.com</a>
Connance	Applications	Waltham	<a href="http://www.connance.com">www.connance.com</a>
CQuotient	Analytics and visualization	Cambridge	<a href="http://www.cquotient.com">www.cquotient.com</a>
Crimson Hexagon	Analytics and visualization	Boston	<a href="http://www.crimsonhexagon.com">www.crimsonhexagon.com</a>
Dataspora	Analytics and visualization	Cambridge	<a href="http://www.dataspora.com">www.dataspora.com</a>
Dataupia	Data management	Cambridge	<a href="http://www.dataupia.com">www.dataupia.com</a>
Datawatch Corp.	Analytics and visualization	Chelmsford	<a href="http://www.datawatch.com">www.datawatch.com</a>
Dataxu	Analytics and visualization	Boston	<a href="http://www.dataxu.com">www.dataxu.com</a>
Deloitte (Oco)	Analytics and visualization	Waltham	<a href="http://www.oco-inc.com">www.oco-inc.com</a>
Digital Reef, Inc.	Analytics and visualization	Boxborough	<a href="http://www.digitalreefinc.com">www.digitalreefinc.com</a>
Echonest (The)	Applications	Somerville	<a href="http://the.echonest.com">the.echonest.com</a>
Dimensional Insight	Analytics and visualization	Burlington	<a href="http://www.dimins.com">www.dimins.com</a>
EMC (Greenplum)	Data integration	Hopkinton	<a href="http://www.greenplum.com">www.greenplum.com</a>
Enterprise DB	Data management	Westford	<a href="http://www.enterprisedb.com">www.enterprisedb.com</a>
Expressor Software	Data integration	Burlington	<a href="http://www.expressor-software.com">www.expressor-software.com</a>
Fina Technologies	Analytics and visualization	Cambridge	<a href="http://www.finatechnologies.com">www.finatechnologies.com</a>
FirstFuel Software	Applications	Waltham	<a href="http://www.firstfuel.com">www.firstfuel.com</a>
GenomeQuest, Inc.	Applications	Westborough	<a href="http://www.genomequest.com">www.genomequest.com</a>
Ginger.io	Applications	Cambridge	<a href="http://ginger.io">http://ginger.io</a>
GNS Healthcare	Applications	Cambridge	<a href="http://www.gnshealthcare.com">www.gnshealthcare.com</a>
Great Minds Interactive (Terametric)	Analytics and visualization	Cambridge	<a href="http://www.terametric.com">www.terametric.com</a>

Company	Category	Location	URL
Group Visual IO	Analytics and visualization	Cambridge	<a href="http://groupvisual.io">http://groupvisual.io</a>
Hadapt	Data management	Cambridge	<a href="http://www.hadapt.com">www.hadapt.com</a>
Harte-Hanks (Trillium Software)	Data integration	Billerica	<a href="http://www.harte-hanks.com">www.harte-hanks.com</a>
Hewlett Packard (Vertica)	Data management	Billerica	<a href="http://www.vertica.com">www.vertica.com</a>
Hopper (Stealth)	Applications	Cambridge	
Hubspot	Applications	Cambridge	<a href="http://www.hubspot.com">www.hubspot.com</a>
Humedica, Inc.	Applications	Boston	<a href="http://www.humedica.com">www.humedica.com</a>
IBM (Applix)	Analytics and visualization	Westborough	<a href="http://www.ibm.com/software/analytics/cognos/">www.ibm.com/software/analytics/cognos/</a>
IBM (Ascential)	Data integration	Westborough	<a href="http://www.ibm.com/software/data/integration">www.ibm.com/software/data/integration</a>
IBM (Netezza)	Data management	Marlborough	<a href="http://www.netezza.com">www.netezza.com</a>
IBM (Unica)	Analytics and visualization	Waltham	<a href="http://www.unica.com">www.unica.com</a>
Icosystem	Analytics and visualization	Cambridge	<a href="http://www.icosystem.com">www.icosystem.com</a>
Insight Squared	Analytics and visualization	Cambridge	<a href="http://www.insightsquared.com">www.insightsquared.com</a>
InStream Media, Inc.	Applications	Wellesley	<a href="http://www.instreamglobal.com">www.instreamglobal.com</a>
Intersystems	Data management	Cambridge	<a href="http://www.intersystems.com">www.intersystems.com</a>
Interval Data Systems, Inc.	Applications	Waltham	<a href="http://www.intdatasys.com">www.intdatasys.com</a>
Intuit (Turning Mill Software)	Data management	Waltham	<a href="http://www.quickbase.intuit.com">www.quickbase.intuit.com</a>
Ironside Group	Analytics and visualization	Lexington	<a href="http://www.ironsidegroup.com">www.ironsidegroup.com</a>
Jana (FKA Txteagle)	Applications	Boston	<a href="http://jana.com">http://jana.com</a>
Kalido	Infrastructure	Burlington	<a href="http://www.kalido.com">www.kalido.com</a>
KNODE, Inc.	Applications	Boston	<a href="http://www.knodeinc.com">www.knodeinc.com</a>
Kyrus Mobile	Applications	Lincoln	<a href="http://www.kyrusmobile.com">www.kyrusmobile.com</a>
Kyruus	Applications	Boston	<a href="http://www.kyruus.com">www.kyruus.com</a>
LeapFrogRX	Applications	Waltham	<a href="http://www.leapfrogrx.com">www.leapfrogrx.com</a>
Localytics	Analytics and visualization	Cambridge	<a href="http://www.localytics.com">www.localytics.com</a>
Lucidel	Analytics and visualization	Needham	<a href="http://www.lucidel.com">www.lucidel.com</a>
Mathworks, Inc.	Analytics and visualization	Natick	<a href="http://www.mathworks.com">www.mathworks.com</a>
Mobile Intelligence Solutions (Umbra Systems)	Analytics and visualization	Concord	<a href="http://umbrasystems.com">http://umbrasystems.com</a>
Nexaweb	Infrastructure	Burlington	<a href="http://www.nexaweb.com">www.nexaweb.com</a>
NuoDB	Data management	Cambridge	<a href="http://www.nimbusdb.com">www.nimbusdb.com</a>
ObjectRiver, Inc.	Data integration	Wayland	<a href="http://www.objectriver.net">www.objectriver.net</a>
OpenLink Software	Infrastructure	Burlington	<a href="http://www.openlinksw.com">www.openlinksw.com</a>
Oracle (Endeca)	Data management	Cambridge	<a href="http://www.endeca.com">www.endeca.com</a>
Oracle (Sleepy Cat Software)	Data management	Lincoln	<a href="http://www.oracle.com">www.oracle.com</a>
Paradigm4	Data management	Waltham	<a href="http://www.paradigm4.com">www.paradigm4.com</a>
ParElastic	Data management	Boston	<a href="http://www.parelastic.com">www.parelastic.com</a>
Perkin Elmer (Geospiza)	Applications	Waltham	<a href="http://www.geospiza.com">www.geospiza.com</a>

Company	Category	Location	URL
Praxeon, Inc.	Applications	Jamaica Plain	www.praxeon.com
Progress Software	Data management	Bedford	www.progress.com
Quant5 (Stealth)	Analytics and visualization	Cambridge	http://quant5.com
Recombinant	Applications	Newton	www.recomdata.com
Recorded Future	Analytics and visualization	Cambridge	www.recordedfuture.com
Rocket Software	Data management	Auburndale	www.rocketsoftware.com
RockTech	Analytics and visualization	Cambridge	www.rocktech.com
ScaleBase	Data management	Newton	www.scalebase.com
SciDB	Data management	Cambridge	www.scidb.org
Seven Bridges Genomics	Applications	Cambridge	www.sbgenomics.com
Skyhook Wireless	Applications	Boston	www.skyhookwireless.com
Sonamine	Analytics and visualization	Waltham	www.sonamine.com
Sonian	Infrastructure	Newton	www.sonian.com
StreamBase Systems, Inc.	Data management	Lexington	www.streambase.com
TeleNav (Goby Technologies)	Analytics and visualization	Boston	www.goby.com
Teradata (xkoto)	Data management	Waltham	www.xkoto.com
TIBCO (Spotfire)	Analytics and visualization	Somerville	www.spotfire.tibco.com
TNS (Compete)	Analytics and visualization	Boston	www.competeinc.com
Tokutek	Data management	Lexington	www.tokutek.com
Vela Systems	Applications	Burlington	www.velasystems.com
Verisk (AIR Worldwide)	Analytics and visualization	Boston	www.air-worldwide.com
ViaScience	Analytics and visualization	Cambridge	www.viascience.com
VisualIQ	Analytics and visualization	Needham	www.visualiq.com
VoltDB	Data management	Billerica	www.voltdb.com
Yieldbot	Analytics and visualization	Maynard	www.yieldbot.com
Zettapoint, Inc.	Infrastructure	Concord	www.zettapoint.com



MASS  
TECHNOLOGY  
LEADERSHIP  
COUNCIL

[www.masstlc.org](http://www.masstlc.org)