

Executive Summary

On May 30, 2012, Governor Deval Patrick launched the Massachusetts Big Data Initiative, to leverage and expand the Commonwealth's position as a global leader in the rapidly growing big data sector. The Initiative, led by the Innovation Institute at the Massachusetts Technology Collaborative, has launched several pilot efforts to enhance and grow the region's vibrant and expanding Mass Big Data ecosystem, including strategic and collaborative partnership efforts with academia, industry and public sector organizations.

The purpose of the **2014 Mass Big Data Report** is to provide an assessment of the relative strengths and weaknesses of the Commonwealth in big data. The Mass Big Data Report is intended to highlight prospects for growth in areas such as talent and workforce, ecosystem, and public data access; and to identify opportunities to promote and expand the Mass Big Data sector, while enhancing the Commonwealth's position as a global leader. The 2014 Mass Big Data Report is intended to serve as a baseline assessment of the Massachusetts Big Data ecosystem and related economic factors. Subsequent updates to the report will track changes, trends, and metrics based on this foundational data.

Conducted by Nexus Associates and staff from the Innovation Institute at the Massachusetts Technology Collaborative, the study is based on a broad spectrum of sources, including interviews with 16 key industry stakeholders; the results of the first annual Mass Big Data Survey of over 60 Massachusetts big data companies; an analysis of publicly available federal, state, and university data; input from social media sources, including LinkedIn; and an extensive literature review.

Principal Findings

Close to 500 Companies Participate in the Massachusetts Big Data Ecosystem

The companies that make up the Mass Big Data ecosystem range from small start-ups with a handful of employees to large, well-established firms such as EMC, IBM, Akamai and Oracle. Mass Big Data companies are leaders in a wide variety of markets, including big data-enabled applications, data analysis tools, data management software, storage and other hardware, cloud services, and other supporting services. Many companies target a broad range of industries ("industry verticals"), including healthcare, life sciences, financial services, manufacturing, transportation, energy and utilities, telecommunications, e-commerce and retail trade, entertainment and media, social media, and marketing and advertising. There has been considerable acquisition activity among Mass Big Data companies in recent years as larger organizations seek to gain access to new technology or market share.

Research Centers Across the Commonwealth Differentiate the Mass Big Data Ecosystem

Massachusetts has a significant base of organizations with an interest in using big data to improve operations, to develop products, solutions, and services, and to inform decisions. Ten leading university and hospital-affiliated research centers across the Commonwealth provide an important foundation for advances in big data.

The Massachusetts Big Data Ecosystem

These centers are developing new technology platforms and analytical techniques, as well as using big data to address important research questions in healthcare, life sciences, communications, cyber security, transportation, energy, and other fields.

Nearly \$20 Million in Federal Grants Awarded for Big Data Initiatives in Massachusetts

From 2006 to 2013, Massachusetts organizations received close to \$20 million from the National Science Foundation, the National Institutes of Health and other federal agencies in support of research and educational activities related to big data. The federal Big Data Initiative has committed \$200 million in funding nationwide for 2012-2017.

Investment Funding in Mass Big Data Companies Topped \$2.5 Billion

More than 240 angel investor groups, venture capital firms, private equity firms, and strategic investors have invested more than \$2.5 billion in at least 123 Massachusetts-based big data related companies since 2000. In the three largest investments, Hubspot, Jumtapt, and Attvio have received \$130.5 million, \$101.5 million and 90.1 million, respectively.

Massachusetts Colleges and Universities Graduate Close to 5,600 Students Annually from 14 Data Science-related Programs

The Mass Big Data talent pipeline is robust and prepared to address the skills necessary in building the Mass Big Data ecosystem. Massachusetts offers a wide range of formal and informal educational opportunities for those interested in developing the skills identified as central to careers in big data. Massachusetts' colleges and universities graduate close to 5,600 students annually from 14 undergraduate and graduate data science-related programs, offering degrees in computer science and engineering, mathematics, statistics, physics, computational biology and other relevant fields. Hackathons, workshops, meet-ups and other industry-sponsored training are held on a regular basis on campuses across the Commonwealth. While most firms report these programs are generally well-aligned to the required skills, companies looking to fill positions report difficulty in recruiting sufficient numbers of qualified software engineers, data architects/engineers, and data scientists.

Massachusetts' Big Data Talent Density Among Highest in US

Massachusetts is a clear leader in per capita graduates from data science related programs as compared to other leading states, with a higher concentration of graduates in certain key degree programs, including biomathematics, bioinformatics, and computational biology.

Strength in Innovation: Data Integration Tools, Data Analysis Software, Data Management

Over two-thirds of the 485 companies researched develop big data applications for vertical industry markets, such as healthcare, life sciences and financial services. Nearly a third of the big data companies researched are in the data analysis software business.

5,250 Big Data Patents Granted in Massachusetts

Analysis of patent data provides insight into the technological strengths of organizations in Massachusetts. A total of 5,250 patents were granted to inventors in Massachusetts between 2008 and 2012 in 23 technology classes that relate to the processing and use of data.

Executive Summary

Overall Prospects for Growth:

Global Market for Big Data to Top \$48 Billion

The overall global big data market is expected to top \$48 billion by 2017, up from \$11.6 billion in 2012. While hardware and services are expected to continue to account for the lion share of revenue, the fastest growth is likely to be in big data-enabled applications.

Big Data Applications in Healthcare, Life Sciences and Financial Services Most Promising

The vast majority of respondents view applications in healthcare, life sciences and financial services as “very promising” or “extremely promising” in terms of their prospects for substantial growth in Massachusetts. Study respondents ranked “data integration tools” as the highest growth area within big data technology and market advancements, followed by data management and data analysis software.

Significant Demand for Mass Big Data Jobs Predicted Over Next 12 Months

Over 50 local big data related firms in the 2013 Mass Big Data Survey reported that they are seeking to fill almost 400 big data-related jobs in Massachusetts over the next 12 months. Considering this figure is drawn from just under 10% of the firms in the Mass Big Data regional ecosystem with possible job openings, the extrapolated figure for the region as a whole could be as high as 3,000-4,000 jobs, before any projection adjustments for additional sector growth.

Massachusetts Big Data

INDUSTRY

485 companies
across the Mass Big
Data Ecosystem

**\$2.5
billion
invested**

in 123 Mass Big Data-
related companies
since 2000

80 NEW

Big Data-related companies
launched since 2010

INNOVATION

\$20 million
in federal grants

awarded for big data
research in Massachusetts
since 2006

Since 2007 28 regional big
data meet-up groups held

368 meet-ups



Mass. inventors granted

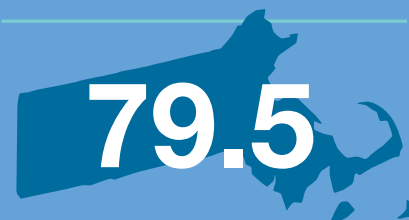
5,250 patents

in 23 Big Data technology
classifications since 2008

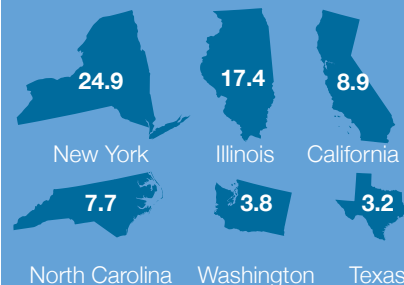
TALENT

5,600 students

from Mass. colleges and
universities graduate from
**14 data science-related
programs** annually



Massachusetts has the
highest per capita
Big Data- related graduation
concentration among leading
states (per 100,000 population)



Source: 2013 Mass Big Data Study



Six Key Mass Big Data Priorities Identified for Action

The Mass Big Data Study identifies six key areas around which the regional data science community is moving to organize and engage in order to realize the full potential of the Mass Big Data ecosystem. Based on industry input to the 2013 Mass Big Data Survey, these identified elements leverage the unique attributes of the Mass Big Data ecosystem and offer a roadmap to driving increased economic opportunity and public benefit.

1. Strengthening Opportunities for Data Science Education and Training

Over two-thirds of survey respondents identified a need for new and refreshed data science programs in Massachusetts, citing growing demand for a big data workforce with training in computer science and mathematics/statistics, as well as familiarity with specific industry verticals. Other respondents suggested that, in parallel with creating new degree programs, courses in computer science and mathematics/statistics should also be integrated into a broader range of other existing degree programs to support multiple paths to the mix of interdisciplinary skills sought by employers. Suggestions supported an emphasis on developing bachelors' degree programs at public universities and matching professional certification courses with industry needs to extend the training for workers already in the labor force.

2. Increasing Regional Talent Retention and Industry Recruiting Success

Respondents highlighted that in an industry driven by talent resources, securing talent is a top industry priority, especially in the current context of high global demand for skilled data professionals. With an existing world-class talent pipeline in the region, industry growth can be enhanced by improving access and engagement between recent and rising graduates and local firms. Collaborative projects, hackathons and internships were cited as critical to engaging and expanding a local community of practitioners.

3. Expanded Access to Public Data

Respondents identified significant value in the availability of state and local public data sets in formats readily accessible by researchers, application developers, and others to create practical applications targeted at specific issues related to the delivery of public services and the quality of life in the Commonwealth. Strong initial efforts in health records, transportation, and education data should be expanded, regularized, and supported with improved public access to the data. Additional efforts to engage the developer community around the use of this data, through meet-ups, hackathons, and other events, were cited as critical to strengthening the Mass Big Data ecosystem.

4. Increased Awareness

Respondents felt that the Commonwealth should strengthen promotional efforts to raise regional, national, and international awareness of the strengths, assets, and ongoing leadership of top performers in the Mass Big Data ecosystem. Successful efforts would support increases in the attraction and retention of individual talent as well as companies. A broad-based campaign would increase buzz about Mass Big Data through websites, social media, and other press to highlight the innovative uses of big data around the region, the important role played by data scientists in industry verticals, and the success of big data related entrepreneurs in Massachusetts.

5. Mass Big Data Ecosystem Expansion

According to study respondents, additional Mass Big Data initiatives should accelerate regional innovation and company growth by supporting novel collaboration to stimulate partnerships and opportunities to enhance the unique innovation environment in the Commonwealth. Increasing cross-sector collaboration among university researchers, enterprise system suppliers, and software developers will improve use of existing strong regional expertise and assets. Strengthening ties among the major computer science research centers and celebrating student-led innovation and competition should increase opportunities for collaborative development of new technologies and products. Supporting the formation of new partnerships between big data firms and top Massachusetts industry verticals enables companies to open new big data markets and exploit opportunities in particular industry verticals.

6. Federal Grants

Study participants consistently recommended that Massachusetts big data companies and academic departments should actively seek out and apply for federal grants where appropriate and collaborations should be explored as early in the process as possible. Greater awareness of federal grants allows researchers to more effectively put together competitive proposals.

Big Data Defined

“Big Data” describes a range of data, data types, and tools to address the rapidly increasing amount of data that organizations around the globe are handling.¹ The amount of data collected, stored and processed by this diverse spectrum of organizations has grown exponentially. This has been driven, in part, by an explosion in the amount of data sourced from web-based transactions, social media and sensors. IDC projects that the digital universe will reach 40 zettabytes (ZB) by 2020, an amount that exceeds previous forecasts by 5 ZBs, resulting in a 50-fold growth from the beginning of 2010.²

There are a variety of ways for organizations to use big data to create value. Data can be used to develop a better understanding of customers and to tailor products and services for narrowly defined segments. Organizations can use data to monitor performance of key functions, identifying factors contributing to observed variances and highlighting needed remedial actions or new ways to optimize systems. Some use data to predict behavior or forecast events, and as a result, take appropriate action. Data can assist in helping to meet regulatory compliance or legal discovery requirements. Finally, organizations can use data as the building blocks for new products and services found across all industries.

¹ http://www.ssc.upenn.edu/~fdiebold/papers/paper112/Diebold_Big_Data.pdf

² IDC, The 2011 Digital Universe study, “Extracting Value from Chaos”, sponsored by EMC Corp.