THE MASSACHUSETTS

BIG DATA INDICATORS 2015

THE INNOVATION INSTITUTE
at the MassTech Collaborative

massbigdata

TABLE OF CONTENTS

ABOUT MASSTECH

The Massachusetts Technology Collaborative, (MassTech), is a public economic development agency working to support a vibrant, innovation-driven economy across Massachusetts. Through our three major divisions — the Innovation Institute, the Massachusetts eHealth Institute and the Massachusetts Broadband Institute — MassTech is fostering innovation and helping shape a vibrant economy. We develop meaningful collaborations across industry, academia and government which serve as powerful catalysts, helping turn good ideas into economic opportunity.

We accomplish this in three key ways: we work to foster the growth of dynamic, innovative businesses and industry clusters in the Commonwealth, by accelerating the creation and expansion of firms in technology-growth sectors; we work to accelerate the use and adoption of technology, by ensuring connectivity statewide and by promoting competitiveness; and we hope to harness the value of effective insight by supporting and funding impactful research initiatives.

About the Innovation Institute at MassTech

The Innovation Institute at MassTech was created in 2003 to improve conditions for growth in the innovation economy by enhancing industry competitiveness, promoting conditions which enable growth, and providing data and analysis to stakeholders in the Massachusetts innovation economy that promote understanding and informs policy development.

The Innovation Institute convenes with, and invests in academic, research, business, government and civic organizations who share the vision of enhancing the Commonwealth's innovation economy. Using an innovative, stakeholder-led process, we have been implementing a "cluster development" approach to economic development. Projects, initiatives and strategic investments in key industry clusters throughout all regions of the Commonwealth are creating conditions for continued economic growth. The Institute manages programs which focus on Advanced Manufacturing in the state, driving support for emerging sectors such as Big Data, Cybersecurity, and Robotics spurring programs which keep talented workers in the Commonwealth, whether through the Intern Partnership program or on entrepreneurship mentoring. Our mission is to strengthen the innovation economy in Massachusetts, for the purpose of generating more high-paying jobs, higher productivity, greater economic growth and improved social welfare.

Mass Big Data

The Innovation Institute at the Massachusetts Technology Collaborative directs the Mass Big Data efforts to expand, leverage, and deploy regional big data assets and resources, supporting strategic development of new opportunities that will boost the Commonwealth's comparative advantages, and addressing the ecosystem's unmet needs and barriers.

Launched in 2012, Mass Big Data was built on the extremely competitive strengths of the Commonwealth's big data ecosystem and continues to leverage the unique combination of focus on the advancement of data science including over 500 companies whose offerings, services and solutions are driven by data; the highest density talent pool of data science students anywhere in the country, and the computing infrastructure and development environments accessible and necessary to support the state's big data efforts.

The mission of Mass Big Data is to strengthen all dimensions of the Commonwealth's big data ecosystem. The program engages collaboratively with industry, academia, government, and nonprofit partners through a four-part strategic approach:

1. **Raise awareness** and interest around the region's big data assets, players, and successes;
2. **Talent & Workforce Development** — Develop and expand the essential, highly-skilled big data workforce across the state;
3. **Drive Cluster Growth through Novel Collaboration** — Accelerate regional innovation and company growth through novel collaborations while providing a forum for public and private sector engagement.
4. **Transforming Government and Civic Engagement through Data** — Advance public access to open government data sets and engage the Commonwealth and its residents in more data-driven activities and outcomes.

MASSACHUSETTS TECHNOLOGY COLLABORATIVE

massbigdata

INTRODUCTION

In 2014, MassTech, in conjunction with the Massachusetts Competitive Partnership (MACP), published ***The Mass Big Data Report: A Foundation for Global Leadership,*** a baseline analysis of the Mass Big Data ecosystem and its challenges, opportunities, and potential for growth. The research and reporting for the 2014 foundational report identified key indicators that shape the relative health and growth of the big data ecosystem in the Commonwealth. The 2014 Mass Big Data Report highlighted the prospects for growth in areas such as talent and workforce, ecosystem, and public data access, while identifying opportunities to promote and expand the Mass Big Data sector and enhance the Commonwealth's position as a global leader.

In this update, ***Mass Big Data Indicators 2015***, we revisit the initial baseline indicators and provide new data and performance metrics in categories that can be useful in assessing the progress and impact of big data-related growth and performance in the Massachusetts innovation economy.

The most notable conclusion from this research is that in the last 24 months, enormous gains have been made in talent, the proliferation of research and university programs centered on big data, and in the number of new starts in the larger big data industry. There are now well over 500 companies in Massachusetts whose mission or products and services are data-driven or pure-play data offerings. In the last 24 months, 53 companies have started and over $2.4 billion in investor funding was raised by Mass Big Data companies. There has also been a significant uptick in the number of research centers throughout the Commonwealth focusing solely on computational science and big data applications. There are now 56 such research centers at universities and hospitals that have centered their research on advanced data science and analysis, compared to 10 just two years ago. Given these research assets, Massachusetts is a top patent producer, with 20% more patents than the average US state in select big data categories. Finally, the talent pipeline in data science and related concentrations continues to expand at a rapid pace. According to the most recently available data, approximately 6,170 students graduated in STEM fields driving the data science talent pipeline in 2013, a ten percent increase from the previous year.

To provide a view on current context and issues on the big data industry, a survey was conducted in the summer of 2015 of nearly 60 CEO and data executives at 'pure-play' and data-driven companies throughout the Commonwealth to identify key issues on workforce development, hiring, skills and talent, ecosystem collaboration and investment activity. Their sentiments on the future of the regional big data ecosystem — and the participants shaping it — are addressed in this update as well.

Mass Big Data Tech Trek,
MassMutual, Amherst, MA.

PHOTO CREDIT: JASON GIROUAR

EXECUTIVE SUMMARY

Technology State House Day 2015
Boston, MA

The Innovation Institute at MassTech has identified eight key indicators that summarize Massachusetts' competitive position in big data and the expansion of the big data ecosystem. This executive summary outlines these indicators and related findings. Each section of the full report following the executive summary details a particular aspect of the Big Data Ecosystem, based on the indicators outlined below. The report is designed to give a clear understanding of how the big data industry can spur economic growth in Massachusetts.

Continued Expansion of the Big Data Ecosystem:

Over 500 pure play big data and data-driven companies operate in Massachusetts today; 53 new big data companies founded in Massachusetts over the last two years.

There are 537 companies in the Mass Big Data ecosystem, including companies whose main products and/or services directly relate to big data; companies that are data-driven; and larger companies with one or more departments driven by big data. Most of these companies and business units develop software for data analysis, integration, visualization, business intelligence or other related work. They cover many different market sectors, with the top three industries being: Business Analytics (29%), AdTech & Marketing (14%), and Digital Health (11%) (Figure 2). The last two years have seen global brands growing their data-driven offices in the greater Boston area, including: Google, Facebook, Twitter, Amazon, Microsoft, and IBM.

Funding Ramp Up:

$2.4B reported investments in 102 big data companies over last two years.

134 big data and data-driven companies in Massachusetts received investment funding from the start of 2013 to August, 2015. Of these firms, 102 disclosed the amount of their investment funding, and the total investment into these companies hovers over $2.4B. LuminaCare Solutions, Wayfair, and HubSpot received the most investment, $617.8M, $366.8M, and $172.5M respectively since their first investment rounds.

LuminaCare Solutions uses predictive analytics to help physicians manage drug doses for their patients.[1] Wayfair is a data-driven eCommerce company that specializes in home décor. HubSpot is a software company that strives to streamline targeted marketing strategies.[2]

Pure play big data companies, Hadapt and Tamr, received $50M and $25.2M respectively. Hadapt provides an analytical platform for unstructured data and was acquired by Teradata in July 2014.[3] Tamr offers a data unification platform for companies who would like to gain insights from their own data.[4]

1 http://www.luminacaresolutions.com/

2 http://www.hubspot.com/

3 http://bostinno.streetwise.co/2014/07/22/teradata-hadapt-deal-hadapt-acquired-by-teradata-to-build-center-of-excellence-in-cambridge/

4 http://www.tamr.com/

Federal Grants Increase:

Close to 200 big data projects in Massachusetts have received new funds over the last two years.

Massachusetts research universities, institutions, and hospitals have received over $115M from the federal government in the form of prime grants to fund 194 different projects. A majority of these grants were funded by the Department of Health and Human Services, as well as the National Science Foundation, NASA, and NOAA, among others. The most highly funded projects work to combine disparate data sets for highly complex analysis. Other projects address connectomics (the field that bridges neuroscience and data analytics), STEM education in Massachusetts, sustainability, AIDS research, and climate science.

Data-driven Innovation:

The Mass Big Data Ecosystem includes 56 data-driven research centers, nine of which focus solely on computational science and big data applications, which drive innovation in Massachusetts.

More and more research centers within universities and hospitals, along with non-for-profit research institutes and federally funded research and development centers, are turning to data-driven directives and strategies to conduct research. The Mass Big Data ecosystem currently boasts 44 research centers at eight universities and five hospitals, in addition to 12 stand-alone research institutions across the Commonwealth.[5] Almost half of these centers use data-driven, analytical research methods to develop technologies, create computational tools, and provide insight into the fields of healthcare and life sciences. Other centers apply data-driven research to the fields of public policy, education, and cybersecurity.

5 See Appendix 5, Research Centers

The Big Data Community is Thriving:

52 active Meetup groups related to big data hold events, discussions, and meetings across the Commonwealth.

There are currently 52 active Meetup groups in Massachusetts that come together to discuss topics and issues related to big data, including machine learning, Hadoop, predictive analytics, data mining, programming education, and data's impact on social systems. The number of Meetup groups in the big data ecosystem has grown by 25 since 2013 (the total now includes 36 newly formed groups and 11 groups that have since become inactive). These Meetup groups have reached tens of thousands of data scientists, entrepreneurs, computer scientists, programmers, and those interested in data across the Commonwealth. In the past two years the total membership in Meetup groups and the number of Meetups has more than doubled.

Massachusetts is a Leader in Big Data Patents:

20 percent more patents than the United States as a whole in select Big Data-related categories, and a 46% increase in patents in big data-related categories from 2010 to 2014.

3,191 patents were granted to Massachusetts inventors in 2013 and 2014 combined in 23 technology classes that relate to the processing and use of data. The number of patents awarded in Massachusetts in these technology classes has grown by 46% from 2010 to 2014. The top patent-producing categories in the Commonwealth are finance and business, database management, and data transfer. Furthermore, in select big data categories Massachusetts produces at least 20% more patents than the average U.S. state.

Intensified Focus on Talent:

22 Massachusetts colleges and universities offer over 70 data science, analytics, informatics, and related undergraduate, graduate, and certificate programs; 10 of which are pure data science degree programs.

Massachusetts' colleges and universities are responding to the increasing need for skills in data science, data analytics, and informatics. As the amount of data continues to grow, so does the demand for employees who have the skills to work with this data. After 2013, the first data science graduate programs in the state began to emerge, including a Master in Computational Science and Engineering at Harvard; a PhD program in data science at WPI; and a Masters in Data Science at UMass Amherst through its new Center for Data Science. In total, five colleges and universities offer 10 programs in data science (with three additional data science concentrations).

In addition to these 10 programs in pure data science, there are 61 additional major and minor programs in disciplines that have a very strong connection to data science, including computational biology, business analytics, and health informatics, among others. About 40% of the programs have been founded in the past three years, an indication that academic institutions in Massachusetts are responding to the growing demand for skilled professionals to enter big data careers in a variety of verticals.

The Data Workforce is Here:

6,170 students in data-related STEM fields graduated from Massachusetts colleges and universities in 2013, a 10% increase from 2012.

6,170 students graduated from academic institutions in the Commonwealth in 2013 with degrees in data-related STEM fields driving data science, which include computer science, engineering, the life sciences, mathematics, operations research, and quantitative economics. These students have strong quantitative skills which can serve as a foundation for a career at a data-driven company or as a student in one of the Commonwealth's advanced data science degree programs. The STEM degree categories with the most growth from 2012 to 2013 are Medical Informatics, Management Information Systems and Services, and Management Sciences.

The Mass Big Data Ecosystem 2015

INDUSTRY

537
COMPANIES driving the Mass Big Data Ecosystem

53 **COMPANIES LAUNCHED** since 2013

INNOVATION

200
New Big Data Projects received **FEDERAL INVESTMENT**

52 Active Big Data **MEETUP GROUPS**

TALENT

70+
DATA SCIENCE RELATED PROGRAMS at 22 Colleges & Universities

10 **NEW PROGRAMS** since 2013

$2.4B
INVESTED in Mass Big Data Companies since 2013

46%
INCREASE IN PATENTS in Big Data related categories since 2010

6,170
STEM GRADUATES in 2013

Source: Innovation Institute at MassTech

massbigdata.org

Massachusetts Green
High-Performance Computing Center,
Holyoke, MA

BIG DATA ECOSYSTEM

Mass Big Data Tech Trek,
Cambridge Innovation Center,
Cambridge, MA

COMPANIES

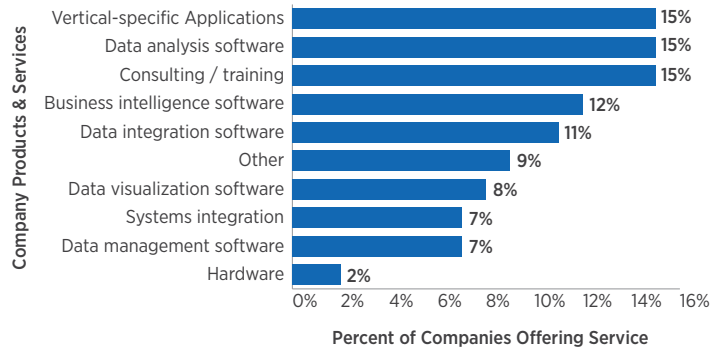
There are 537 companies in the Mass Big Data ecosystem, most of which offer software as their primary product.[6] One third of the companies produce either data analysis software or applications geared toward specific verticals (Figure 1).[7] These companies serve a variety of verticals, tailoring their products to one or more specific industries. From both the products they offer and specific verticals to which they cater, companies can be categorized by industry. The top three industries for big data companies are Business Analytics (29%), AdTech & Marketing (14%), and Digital Health (11%) (Figure 2).

Burning Glass' platform analyzes labor market data to better inform the work of HR teams and college career centers

Companies are considered to be in the analytics industry if they analyze data as their primary business, whereas they are considered pure play big data companies if they offer products and/or services that are specific to big data. More specifically, business analytics companies analyze their clients' data, or provide software that does this analysis, to drive revenue and growth for their clients. One example of business analytics is Burning Glass' platform which analyzes labor market data to better inform the work of HR teams and college career centers.[8] Another example is the software provided by Machine Metrics, which collects and analyzes manufacturing data to increase efficiency in the workplace.[9] Some industries represented in the "Other" category in Figure 2 include weather, fantasy sports, and education.

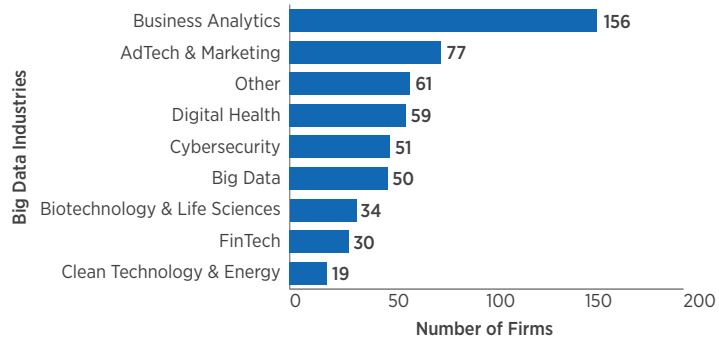
6 See Appendix 1, list of Mass Big Data Companies
7 Results of the MassTech 2015 Mass Big Data Industry Survey
8 http://burning-glass.com/
9 http://www.machinemetrics.com/

FIGURE 1: MASS BIG DATA FIRMS BY PRODUCTS & SERVICES OFFERED



Source: Innovation Institute at MassTech

FIGURE 2: MASS BIG DATA FIRMS BY INDUSTRY

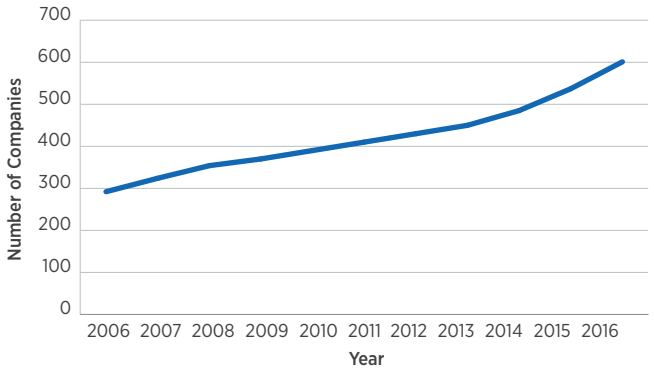


Source: Innovation Institute at MassTech

The regional big data ecosystem continues to expand at an impressive rate: the number of companies in the big data ecosystem has more than doubled in the past ten years (Figure 3). Some of the companies, started before the term "big data" was coined, have since been identified as data-driven if data is an integral component of their business platform. The companies in the ecosystem are diverse in terms of both age and employment (Figures 4 & 5). At 65% of the companies, over half of the employees work in big data, proving that the companies in this ecosystem are committed to gaining insight from big data and that data is an essential part of the companies' missions.[10]

Bridj applies a data-driven approach to transportation in the cities of Boston, Kansas City, and Washington, D.C., making it a unique and competitive company. 80 percent of its employees work directly with big data, a comparatively high percentage across the Mass Big Data ecosystem.

FIGURE 3: MASS BIG DATA ECOSYSTEM EXPANSION CONTINUUM



Source: Innovation Institute at MassTech

10 MassTech 2015 Mass Big Data Industry Survey

With a strong network of entrepreneurs and venture capitalists, a sizeable number of new companies begin in Boston each year. Since 2013, 53 big data companies, across a wide range of verticals, have been founded in the Commonwealth (Figure 6). Some examples follow.

Tamr, founded in 2013 by Andy Palmer, Mike Stonebraker, and Ihab Ilyas, all with expertise in the field of database engineering, allows companies to integrate disparate data sets in a time and cost efficient manner.[11] Tamr's clients — including Toyota, Thomson Reuters, and General Electric — speak to the company's relevance and leadership in multiple industries. In September 2015, Tamr unleashed a Beta version of its new software, Tamr Catalog, which categorizes data sets, visualizes relationships between data sets, and provides metadata related to each data set. Their software is targeted to all employees who work with data, in order to maximize the time they spend analyzing their company's data and ensure that they are aware of the potential uses and implications of the data sets available to them.[12]

Cazena was founded in 2014 by Prat Moghe and Jit Saxena to transform the way companies approach big data processing.[13] Launched in 2015, Cazena's big data as a service platform gives companies access to enterprise big data through the cloud. Cazena's groundbreaking technology is able to overcome the challenges that data processing in the cloud presents and has created a "workload intelligence" component which optimizes performance. Furthermore, the service saves companies both labor and capital resources that are spent determining how to store and process their own big data.[14]

Bridj, founded in 2014 and a leader in the mass transportation industry, continually develops shuttle routes based on real-time and historical transportation demand data.[15] Bridj applies a data-driven approach to transportation in the cities of Boston, Kansas City, and Washington, D.C.,

11 http://www.tamr.com

12 http://www.tamr.com/tamr-opens-public-beta-program-for-its-free-enterprise-meta-data-catalog-tool/

13 http://www.cazena.com

14 https://www.cazena.com/blog/announcing-cazena-big-data-service

15 http://www.bridj.com

making it a unique and competitive company. 80 percent of its employees work directly with big data, a comparatively high percentage across the Mass Big Data ecosystem. Bridj plans to add five more developers to its engineering team in the near future.[16]

Evervest, was founded in 2014 by two MIT alums to create transparency in an industry bogged down by data. EverVest provides risk analysis on wind energy projects by applying a machine learning algorithim to wind measurement data.[17] The results from this analysis predict future wind measurements at a specific location, crucial information for investors in wind energy. The software also provides a more detailed look into a potential wind energy investment by analyzing credit risk data and providing portfolio monitoring and analytical reporting features.[18] The company's software sits at the intersection of CleanTech and FinTech, making it a unique company in the regional big data ecosystem.

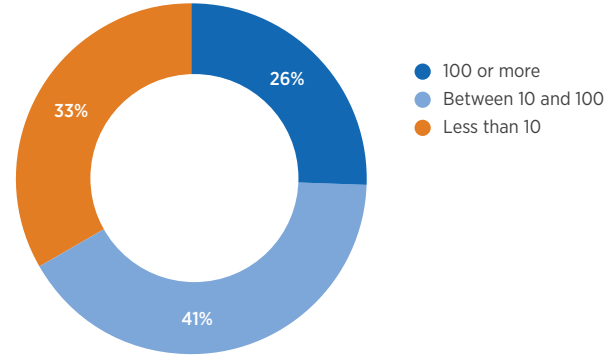
FIGURE 4: MASS BIG DATA FIRM LONGEVITY

Company Age	Number of Companies	Percent of Companies
1-3 years	54	10%
4-10 years	266	50%
11-20 years	135	25%
21 or more	79	15%

*Data unavailable for four companies.

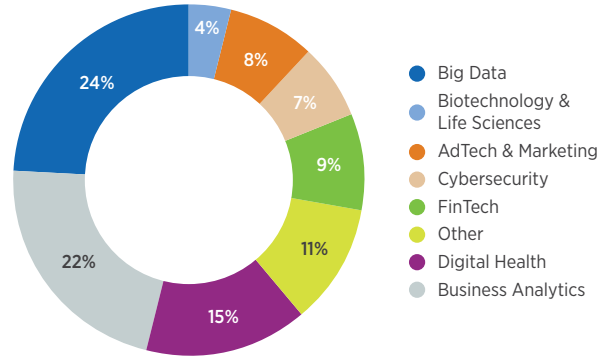
Source: Innovation Institute at MassTech

FIGURE 5: EMPLOYEES WORKING DIRECTLY IN BIG DATA AT MASS BIG DATA FIRMS



Source: MassTech 2015 Mass Big Data Industry Survey

FIGURE 6: NEW MASS BIG DATA FIRMS SINCE 2013 BY INDUSTRY



Source: Innovation Institute at MassTech

16 MassTech 2015 Mass Big Data Industry Survey

17 http://www.evervest.co

18 http://bostinno.streetwise.co/2015/02/03/cardinal-wind-wind-power-investments-risk-analysis-software/

INVESTMENT

Investment funding serves as a metric to evaluate how much confidence financial experts have in the big data ecosystem. There are several rounds of investment funding, starting with seed funding and then moving to series A, B, and C. Additionally, there are two types of investors: angel investors, individuals who often invest in addition to working another job, and venture capital firms, which are dedicated to researching and investing in startups. 134 Massachusetts big data companies received investment funding since 2013, a quarter of which are pure play big data and machine learning companies.[19] 102 of these companies disclosed the amount of investment they have received, totaling $2.4 billion.[20] Figures 7 and 8 below show the top investments in data-driven and pure play big data companies respectively.

FIGURE 7: TOP INVESTMENTS IN DATA-DRIVEN FIRMS

LuminaCare Solutions	$618M
Wayfair	$367M
HubSpot	$173M
Rapid7	$111M
Nexage	$108M

Source: Pitchbook.com

FIGURE 8: TOP INVESTMENTS IN PURE-PLAY BIG DATA FIRMS

Hadapt	$50M
Tamr	$25M
Cazena	$20M
Quantopian	$15M
Deep Information Sciences	$8M

Source: Pitchbook.com

19 http://pitchbook.com/
20 http://pitchbook.com/

Some examples of companies that have sustained investor interest and growth:

Bit 9 + Carbon Black, data-driven cybersecurity company, provides a comprehensive set of threat-protection solutions. The company has raised a total of $120M in funding since it started in 2002. A quarter of its funding since 2013 comes from eight venture capital firms including the Boston-based groups .406 Ventures, Atlas Venture, Highland Capital Partners, and The Blackstone Group.[21] In September 2015, the company launched another round of funding and one investor estimates the company may raise an additional $50M from this investment round, a confident estimate due in part to the fact that cybersecurity firms are doing well despite unease in the venture capital community. The company also plans to open a new office in Boston to house its technology development team, which would be its second office in the Commonwealth.[22]

> One hundred and thirty four Massachusetts big data companies received investment funding since 2013, a quarter of which are pure play big data and machine learning companies.

RapidMiner, founded in 2006, has raised $20M in funding for its modern analytics platform for big data.[23] RapidMiner is unique in that it was founded as an open source data mining project out of the University of Dortmund, Germany, and is now a for-profit company with offices in Germany, London, and Cambridge (this office opened in 2012). The company hopes to

21 http://pitchbook.com/
22 http://fortune.com/2015/09/23/cyber-security-investing/
23 http://techcrunch.com/2015/02/20/rapidminer-picks-up-15m-more-to-scale-its-business-source-data-analytics-service/

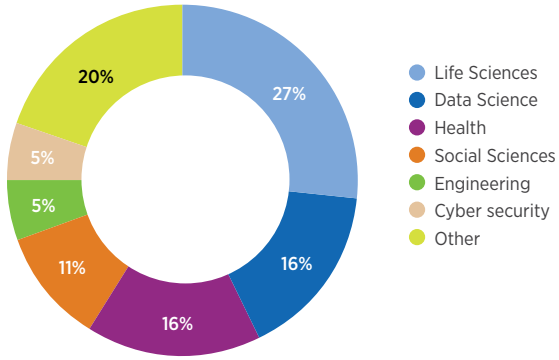
empower the next generation of data scientists through its groundbreaking work in predictive analytics for a wide variety of verticals. RapidMiner is a leader in the big data community, spreading the message of big data's growing role in the expanding technology sector.[24]

InsightSquared, founded in 2010, brings business intelligence tools and software, including data analytics, financial reporting, and sales forecasts, to small and mid-size companies.[25] The company's software allows its users to find and visualize links between data sets, and create customizable data dashboards facilitating data-driven business decisions. InsightSquared has received almost $30M in funding, with its most recent fundraising round toward the end of 2014.[26] The company used the money it raised in its latest funding round to spread the word about its services to smaller companies, create more advanced analytics tools, and develop its mobile application.

RESEARCH CENTERS

Many of the world's leading research universities are located in Massachusetts; Boston, in particular, is known for its high concentration of colleges and universities. Additionally, the Commonwealth boasts nationally recognized hospitals and medical centers. These universities and hospitals house a variety of research centers which conduct data-driven studies and/or develop new algorithms and computational methods to analyze and interpret big data. There are 44 research centers at eight universities and five hospitals, in addition to 12 standalone research institutions across the Commonwealth.[27] Sixteen percent of the centers in Massachusetts conduct research in the field of data science; other common topics include life sciences, health, and social sciences (Figure 9). The data science research centers develop algorithms for big data as well as curricula for data science degree programs, whereas the other centers apply computational methods to a variety of fields (Figure 10).

FIGURE 9: MA RESEARCH CENTERS BY SECTOR



Source: MassTech 2015 Mass Big Data Industry Survey

Sixteen percent of the research in Massachusetts is conducted in the field of data science; other common topics include life sciences, health, and social sciences

24 http://opensourcebi.hu/letoltes/osbi2012/open_source_bi_2012_rapidi.pdf

25 http://www.insightsquared.com/

26 http://www.betaboston.com/news/2014/11/17/insightsquared-gets-13-5-million-to-continue-growing-its-business-intelligence-platform/

27 See Appendix 5, Research Centers

FIGURE 10: NOTABLE MA RESEARCH CENTERS

Title	Associated Institution	Description
Geospatial Data Center	MIT	Analyzes geospatial data in the context of national security with a number of programs including cyber-physical security, the big data lab, and a data center infrastructure simulator.[28]
Center for Computational Science	Boston University	Coordinates and promotes computationally based research, fosters computational science education, and provides for the expansion of computational resources and support.[29]
Center for Cancer Computational Biology	Dana Farber Cancer Institute	Finds new ways to understand cancer by improving analysis and interpretation of genomics and computational biology. Also supports analysis and interpretation of genomic and other large-scale data to further basic, clinical, and translational research.[30]
Institute for Applied Computational Science	Harvard	Developed masters degree curriculum in applied computational science and supports the development of new courses in applied mathematics and computational science across the university.[31]
Woods Hole Oceanographic Institution	N/A	Explores and works toward a better understanding of the ocean; many of its research projects are heavily data-driven and require the use of big data techniques.[32]

28 http://geospatial.mit.edu/
29 http://bu-ccs.com/
30 http://cccb.dfci.harvard.edu/home
31 http://iacs.seas.harvard.edu/
32 http://www.whoi.edu/

FEDERAL INVESTMENT

Since 2013, 194 big data projects in Massachusetts have received over $115 million in prime grants from the US government.[33] Seventeen of these projects each received over $1 million, one of which received over $17 million. The number of big data projects has been increasing each year since 2013: Forty-three projects were federally funded in 2013, 55 in 2014, and as of July, 29 projects had received funding in 2015. The department of Health and Human Services and the National Science Foundation provide almost 90% of the funding for these big data projects (Figure 11), given the fact that many of the projects (and the most highly funded projects) are in health, medicine, and the life sciences (Figure 12). Massachusetts Institute of Technology has conducted the most data-driven research projects since 2013 and the Broad Institute has received the most funding (Figures 13 and 14). These projects tackle complex problems that will lead to discoveries in the fields of medicine, education, and transportation (Figure 15).

FIGURE 13: TOP RECIPIENTS OF FEDERAL GRANTS BY NUMBER OF PROJECTS

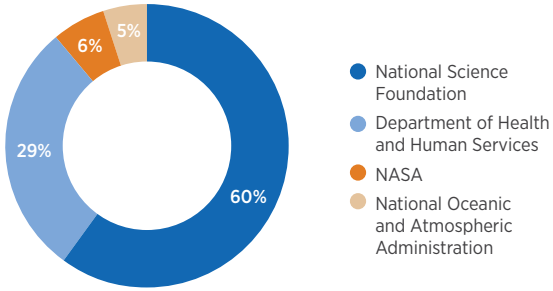
Massachusetts Institute of Technology	23
Harvard University	22
Woods Hole Oceanographic Institution	18
Boston University	13
University of Massachusetts	10

(50% of projects were at these five institutions)

Source: Innovation Institute at MassTech

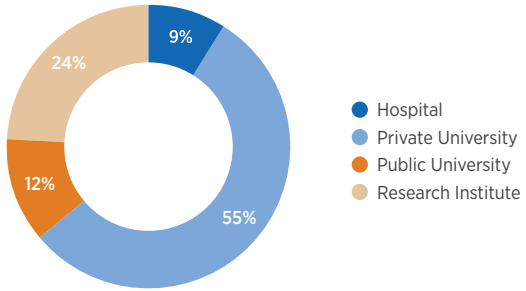
33 https://www.usaspending.gov/Pages/Default.aspx; http://www.nsf.gov/statistics/data.cfm; http://www.hhs.gov/

FIGURE 11: SOURCES OF FEDERAL INVESTMENT



Source: usaspending.gov

FIGURE 12: RECIPIENTS OF FEDERAL INVESTMENT



Source: usaspending.gov

FIGURE 14: TOP RECIPIENTS OF FEDERAL GRANTS BY DOLLAR AMOUNTS

Broad Institute, Inc.	$22.5M
Harvard School of Public Health	$14.7M
Harvard University	$10M
Massachusetts Institute of Technology	$9.4M
Brigham and Women's Hospital	$7.7M

(Over 50% of grants went to these five recipients)

Source: Innovation Institute at MassTech

FIGURE 15: NOTABLE FEDERALLY FUNDED DATA SCIENCE AND DATA-DRIVEN RESEARCH PROJECTS

Protect Title	Institution	Description
The Genotype-Tissue Expression (GTEx) project	The Broad Institute	Creation of a database and tissue bank so scientists can study the relationship between genetic variation and gene expression to understand more about the genetic basis of common diseases.[34]
AIDS Clinical Trial Group	Harvard University School of Public Health	Development of the Statistical and Data Analysis Center (SDAC) which provides biostatistical and data management support to the AIDS Clinical Trial Group and conducts data-driven research on the design and analysis of clinical trials.[35]
Mobility Electronic Market for Optimized Travel (MeMOT)	Massachusetts Institute of Technology & UMass Amherst	Using real and simulated individual-level travel data, the researchers will design a system for Massachusetts residents to conserve energy by rewarding them with travel-related benefits based on how much energy their travel decisions conserve.[36]
Long Term Multidisciplinary Study of Cancer in Women: The Nurses' Health Study	Brigham and Women's Hospital	Continue data collection for this longitudinal study that started in 1976, which follows over 100,000 women to advance cancer survival research. Improve data management and statistical support for the study and expand analysis on the data.[37]
Large-Scale Research on Engineering Design based on Big Learner Data	Concord Consortium	Analyze large educational data sets with advanced analytics and data mining to determine how to improve engineering design and science education.[38]
High-Dimensional Statistical Machine Learning for Spatio-Temporal Climate Data	Northeastern University	Advance machine learning techniques to analyze complex spatio-temporal data, common in the field of climate science. The algorithms, models, and analysis from this research will greatly improve assessments of impacts, adaptation and vulnerability in several fields including urban planning, energy, coastal systems, and health.[39]
Visualizing Healthcare System Dynamics in Biomedical Big Data	Harvard University Medical School	Create visualization of clinical data, in the form of electronic health records and administrative claims, to further biomedical research and health care policy.[40]

Source: Innovation Institute at MassTech

34 http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4010069/
35 https://actgnetwork.org/Leadership/Michael_Hughes
36 http://news.mit.edu/2015/mit-awarded-department-energy-grant-travel-information-incentives-system-0806
37 http://grantome.com/grant/NIH/UM1-CA186107-01
38 http://www.nsf.gov/awardsearch/showAward?AWD_ID=1348530
39 http://nsf.gov/awardsearch/showAward?AWD_ID=1447566
40 https://projectreporter.nih.gov/project_info_description.cfm?aid=8875287&icde=0

MEETUPS

Meetup groups are critical to building the extended big data community throughout the Commonwealth. These groups are gaining traction as data scientists and big data experts wish to hone their skills, learn from each other, and solve multifaceted social problems. Several of these groups focus on big data programming languages such as Python, Hadoop, and R. Others focus on specific topics related to big data such as data visualization, artificial intelligence, data mining, machine learning, and business intelligence. In addition to coding and big data related topics, there are also groups that focus on specific verticals such as education, social issues, and pharmacology.

There are currently 52 active meetup groups in Massachusetts, 36 of which have been founded since 2013, and 8 of which have been founded in 2015 alone.[41] Since 2013 there has been a 50% increase in the number of members of the groups, on average, and a 17% increase in the number of Meetup events, on average.[42] Events include presentations from data scientists at companies in the Mass Big Data ecosystem, discussions with leaders in big data, networking get-togethers, interactive skills sessions, and hackathons. Most meetup groups meet once every other month, with some meeting more or less frequently. These groups spur innovation in Massachusetts in response to new big data programs and developments. Furthermore, meetup groups stoke data scientists' passion and dedication, allowing communities to thrive outside the office or research center.

FIGURE 16: NOTABLE MEETUP GROUPS

Name	Founded	Description	Members	Events
Greater Boston useR Group (R Programming Language)	2011	Shares ideas, discusses R related topics, and provides direction for new and experienced users of this language for computing, data analysis, and visualization.[43]	1,517	21
The Data Scientist	2012	Concentrates on understanding the tools and skill-sets needed to become an effective Data Scientist. Explores all topics related to the data lifecycle including acquiring new data sets, parsing new data sets, filtering and organizing data, mining data patterns, advanced algorithms, and visually representing data.[44]	2,314	31
Data Science for Social Good	2015	Brings the community together to apply data science to improve society in areas such as crime, healthcare, transportation, clean energy, and income disparity.[45]	386	5
Data Science Professional Development Boston	2015	Members of this Meetup group will get and give personalized career advice, learn content and receive content recommendations from peers, and intelligently grow their professional network.[46]	768	13

Source: Innovation Institute at MassTech

41 http://www.meetup.com/Boston-useR/
42 http://www.meetup.com/The-Data-Scientist/
43 http://www.meetup.com/Data-Science-for-Social-Good/
44 http://www.meetup.com/DS-ProD-Boston/
45 http://www.meetup.com/
46 http://www.meetup.com/ds-prod-boston/

PATENTS

Patents, as a straight forward metric of innovation, represent new ideas in data science coming from the Commonwealth's researchers, companies, and universities. 23 data-related patent categories, including data processing, electrical computers, and information security and storage, have been identified. 3,191 patents were awarded in these 23 categories in Massachusetts in 2013 and 2014 combined.[47] The 1,724 patents awarded in 2014 represent a 46% increase from the amount awarded in 2010. The top patent-producing categories in the Commonwealth are finance and business, database management, and data transfer.

The location quotient data in Figure 17 gives a sense of where Massachusetts stands in patent production, compared with the average US state. A location quotient equal to 2.406, for example, means that Massachusetts produces almost two and a half times more patents than the average US state in that particular category. The "percent" in U.S. column in Figure 17 lists the percent of total patents in each category that come from Massachusetts.

The 1,724 patents awarded in 2014 represent a 46% increase from the amount awarded in 2010. The top patent-producing categories in the Commonwealth are finance and business, database management, and data transfer.

Figure 17 on the facing page shows that Massachusetts produces at least 20% more patents than the average U.S. state in select big data categories.[48]

47 http://www.uspto.gov/web/offices/ac/ido/oeip/taf/stctec/mastcl_gd.htm

48 http://www.uspto.gov/web/offices/ac/ido/oeip/taf/cbcby.htm

FIGURE 17: PATENTS IN 2013 AND 2014

Class	Title	Number of Patents in MA	Percent in U.S.	Location Quotient	Percent Growth 2010 to 2014
700	Data Processing (DP): Generic Control Systems or Specific Applications	148	7.19%	1.461	170%
701	DP: Vehicles, Navigation, and Relative Location	72	1.84%	0.373	118%
702	DP: Measuring, Calibrating, or Testing	99	5.12%	1.039	-6%
703	DP: Structural Design, Modeling, Simulation, and Emulation	119	11.85%	2.406	60%
704	DP: Speech Signal Processing, Linguistics, Language Translation, and Audio Compression/Decompression	102	6.34%	1.287	61%
705	DP: Financial, Business Practice, Management, or Cost/Price Determination	328	3.65%	0.741	21%
706	DP: Artificial Intelligence	77	5.79%	1.175	65%
707	DP: Database and File Management or Data Structures	396	5.35%	1.085	30%
708	Arithmetic Processing and Calculating	30	6.67%	1.353	60%
709	Multicomputer Data Transferring	427	5.22%	1.059	38%
710	Input/Output	84	6.16%	1.251	51%
711	Memory	206	6.70%	1.360	20%
712	Processing Architectures and Instruction Processing, e.g., Processors	18	3.73%	0.757	-17%
713	Support	161	4.58%	0.929	45%
714	Error Detection/Correction and Fault Detection/Recovery	95	3.21%	0.652	-33%
715	DP: Presentation Processing of Document, Operator Interface Processing, and Screen Saver Display Processing	157	4.70%	0.953	45%
716	Computer-Aided Design and Analysis of Circuits and Semiconductor Masks	39	2.57%	0.521	31%
717	DP: Software Development, Installation, and Management	200	7.58%	1.539	77%
718	Virtual Machine Task or Process Management or Task Management/Control	61	4.60%	0.934	83%
719	Interprogram Communication or Interprocess Communication	45	6.47%	1.314	44%
720	Dynamic Optical Information Storage or Retrieval	2	12.50%	2.538	N/A
725	Interactive Video Distribution Systems	94	3.81%	0.774	556%
726	Information Security	231	5.13%	1.041	138%
Total		3,191	4.93%		

Source: U.S. Patent and Trademark Office, 2010-2014 Utility Patent Grants as Distributed by Calendar Year

EDUCATION AND TALENT

Big Data programs in higher education

Demand for professionals with skills to work in the big data sector has significantly increased with the expansion of the big data ecosystem and the availability of data across the Commonwealth. Data science degree programs are a crucial way to ensure that there is a sufficient supply of these professionals to match this growing demand. Massachusetts is currently home to 10 data science programs (two Bachelors, three Masters, one combined BS/MS, three certificates, and one doctorate) at five universities across the state, all of which began in the past three years (Figure 18). In addition, three universities offer data science concentrations as a part of their computer science and MBA programs.

Bachelors programs provide an overview of the discipline and allow students to develop technical skills through a broad range of courses including artificial intelligence, data mining, machine learning, data warehousing, data structures, algorithms, and several advanced statistics courses. Masters programs allow students to dive deeper into specific fields within data science and to gain a better understanding of the theory underlying computer science and advanced analytics. These programs also encourage or require students to partner with Massachusetts-based companies or government departments to tackle real world problems. Certificate programs give a taste of data science with four or five introductory and elective courses. See a list of all ten data science programs in Figure 18.

Forty percent of the programs have been founded in the past three years; academic institutions in Massachusetts are responding to the growing demand for skilled professionals to enter big data careers in a variety of verticals.

FIGURE 18: DATA SCIENCE DEGREE PROGRAMS

Institution	Degree Type	Year Founded
Becker College	BS	2014
Harvard Extension School	Certificate	2014
Harvard University	MS	2013
Northeastern University	Certificate	2015
UMass Dartmouth	BS	2015
UMass Dartmouth	MS	2015
Worcester Polytechnic Institute (WPI)	Certificate	2014
WPI	BS/MS	2014
WPI	MS	2014
WPI	PhD	2015

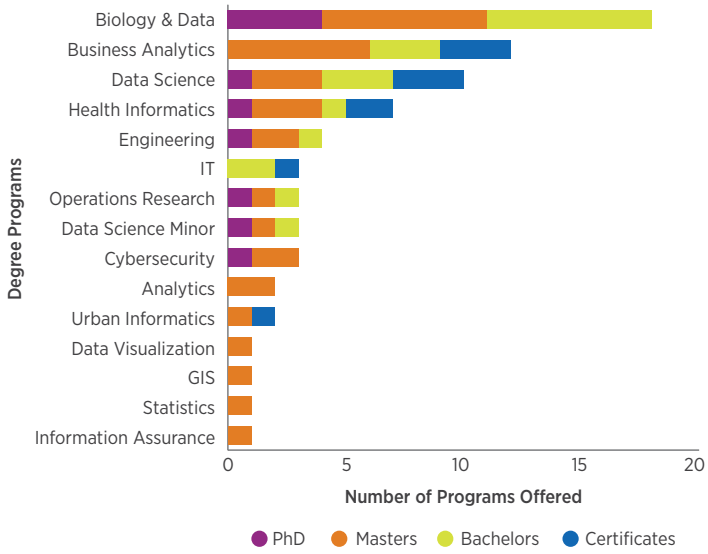
Source: See Appendix 4, Data Science & Related Degree Programs

In addition to data science programs, the Commonwealth is also home to a broad array of programs that train students for careers in big data, often for a specific vertical that big data companies address.[49] These programs include informatics, business analytics, and biostatistics, among others (Figure 19). Programs that mix biology and data make up a quarter of all programs, reflecting the large amount of research in health and life sciences throughout the Commonwealth. Programs that began before the term big data came into being have adapted their curricula to prepare students for the vast amounts of data now readily available to analyze. Many bioinformatics and health informatics programs, for example, began in the early 2000s and now include introductory data science courses in their curricula. Bentley University, in 1999, was the first university in Massachusetts to offer a degree

49 See Appendix 4, Data Science & Related Programs

in business analytics. Since then, the discipline has expanded with a steadily increasing number of programs—seven additional programs in the past two years—and graduates. Massachusetts is home to a selection of unique programs offered in few places in the country (Figure 20). About 40% of the programs have been founded in the past three years, proof that academic institutions in Massachusetts are responding to the growing demand for skilled professionals to enter big data careers in a variety of verticals.

FIGURE 19: BIG DATA DEGREE PROGRAMS BY CATEGORY & LEVEL AT MASSACHUSETTS UNIVERSITIES



Source: Innovation Institute at MassTech

FIGURE 20: NOTABLE DATA SCIENCE RELATED PROGRAMS

Title	University	Year Founded	Description
Quantitative Genetics	Harvard	2013	Combination of computation skills and biology to prepare students to analyze public health big data. Students complete a collaborative research thesis, in the area of genomics, at a Massachusetts research institute.[50]
Information Design & Visualization	Northeastern	2014	Prepares students to discover and communicate information across a range of socially relevant issues. One of the only MFA programs rooted in data.[51]
Marketing Analytics	Bentley	2004	Developed in close partnership with marketing companies in Massachusetts, the program prepares students to make marketing decisions based on quantitative analysis using big data.[52]
Learning Analytics	Brandeis	2015	Focus on business intelligence, data analysis and design management using educational datasets.[53]

Source: See Appendix 4, Data Science & Related Degree Programs

50 http://www.hsph.harvard.edu/sm-computational-biology/

51 http://www.northeastern.edu/camd/artdesign/academic-programs/mfa-in-information-design-and-visualization/

52 http://www.bentley.edu/graduate/ms-programs/marketing-analytics

53 http://www.brandeis.edu/gps/future-students/learn-about-our-programs/learning-analytics.html

Big Data Workforce: STEM Graduates

Undergraduates completing degrees in computer science, physical sciences, engineering, and mathematics are much more likely to enter advanced data science (and related) higher education programs. These graduates are also much more likely to apply for and accept jobs at big data companies, as they can quickly develop data science skills given their quantitative backgrounds. In fact, our survey of executives at fifty big data companies, revealed that mathematics/statistics and computer science are two of the required degrees for the most difficult positions they need to fill (Figure 22).

According to the latest available data, there were over 6,000 graduates in 2013 from selected STEM programs at 175 academic institutions throughout the Commonwealth, a 10% increase from 2012 (Figure 21).[54]

The STEM degrees with the most growth since 2013 are medical informatics and management sciences, which grew 47% and 27% respectively.

The STEM degrees with the most growth since 2013 are medical informatics and management sciences, which grew 47% and 27% respectively.[55] The STEM degree with the highest number of graduates was computer and information sciences with 2,653 graduates, representing almost half of all STEM graduates. There were 1,405 graduates from mathematics and statistics programs, which account for almost a quarter of all STEM graduates.

FIGURE 22: DEGREE REQUIREMENTS FOR MASS BIG DATA JOB OPENINGS

Area of Study	Required Degree	Percent of Total Responses
Computer Science	Master	12%
Mathematics/Statistics	Bachelor	12%
Data Science	Master	10%
Computer Science	Master	9%
Computer Engineering	Bachelor	9%
Data Science	Bachelor	8%
Mathematics / Statistics	Bachelor	8%
Data Science	PhD	7%
Mathematics / Statistics	PhD	5%
Electrical Engineering	Bachelor	5%
Computer Engineering	Master	4%
Computer Science	PhD	4%
Other	Bachelor	3%
Computer Engineering	PhD	2%
Electrical Engineering	Master	2%
Electrical Engineering	PhD	1%

Source: MassTech 2015 Massachusetts Big Data Industry Survey

54 http://nces.ed.gov/ipeds/datacenter/Default.aspx
55 http://nces.ed.gov/ipeds/datacenter/Default.aspx

FIGURE 21: SELECT STEM DEGREES GRANTED IN MASSACHUSETTS, 2013

Classification of Instructional Programs	Title	Number of Schools	Number of Graduates				
			BA/BS	MA/MS	PhD	Certificate (a)	Total
11	Computer and Information Sciences	47	1536	1011	101	5	2653
14.09	Computer Engineering	7	108	173	17	0	298
27	Mathematics and Statistics	44	1130	186	73	16	1405
40.08	Physics	30	384	81	128	0	593
26.0203 26.0206	Biophysics and Molecular Biophysics	3	5	1	12	0	18
40.0202 40.0403 40.0603	Astrophysics, Atmospheric Physics and Dynamics, and Geophysics and Seismology	6	8	18	5	0	31
26.11	Biomathematics, Bioinformatics, and Computational Biology	7	6	72	35	0	113
51.2706	Medical Informatics	4	1	45	0	1	47
45.0603	Econometrics and Quantitative Economics	1	43	0	0	0	43
14.37	Operations Research	2	0	22	11	0	33
52.12	Management Information Systems and Services	12	100	92	0	0	192
52.13	Management Sciences and Quantitative Methods	8	256	421	0	24	701
30.06	System Science and Theory	2	0	15	0	0	15
30.08	Mathematics and Computer Science	2	28	0	0	0	28
	Total	**175**	**3605**	**2137**	**382**	**46**	**6170**

Source: National Center for Education Statistics

THE 2015 MASS BIG DATA INDUSTRY SURVEY

Data analysis software has the greatest potential to grow the big data ecosystem, which follows from the growing number of business analytics companies tapping into previously unexplored datasets.

In July, 2015, MassTech surveyed close to 60 CEOs and leaders at Massachusetts data-driven and pure-play big data companies. Results from the 2015 Mass Big Data Industry Survey provide opinions on and a window into the future of the regional big data ecosystem. Companies generally feel that graduates from local colleges and universities have the skills they seek. Fewer companies agree, however, that the number of graduates from colleges and universities in Massachusetts is sufficient. This last opinion is logical given that many of the data science programs started within the past few years and are still developing their programs, alumni networks, and student bodies. Companies tend to think positively of the Mass Big Data ecosystem at large. Many believe that there is a strong demand for their products, a healthy rivalry among big data companies, and that government policies support their businesses (Figure 23).

Results from the survey give a clear picture of hiring at big data firms throughout Massachusetts. There is currently an average of five big data related job openings at each company across the Mass Big Data ecosystem. The hardest big data related positions to fill are data scientists and software engineers (Figure 24). Companies across the ecosystem plan to add an average of eight big data related positions in 2016, 15 in 2017, and 27 in 2018. This projected job growth speaks to the role of the big data industry in growing the regional economy. Companies believe that there are several products/services, verticals, and policies that offer the biggest opportunities for this growth (Figure 25). According to the surveyed companies, data analysis software has the greatest potential to grow the big data ecosystem, which follows from the growing number of business analytics companies tapping into previously unexplored datasets. Companies cite the need for policies that develop data science curricula and support workers in the big data industry, which reflects the importance for the supply of big data talent to keep up with the growing demand.

FIGURE 23: COMPETITIVE ADVANTAGE IN MASSACHUSETTS, INDUSTRY PERSPECTIVE

Statement	Level of Agreement Among Companies	Majority
There is strong demand for our types of products/services from customers in Massachusetts	High	Agrees
There is a healthy rivalry among competitors in Massachusetts	High	Agrees
Government policies in Massachusetts are supportive of our type of business	High	Agrees
There are strong collaborations between companies and universities in Massachusetts	Medium	Agrees
There are strong collaborations among companies in Massachusetts	Medium	Agrees
There is an extensive supplier network in Massachusetts	Medium	**Disagrees**
Venture capitalists in Massachusetts are interested in making investments in this sector	Low	Agrees
There are strong collaborations between companies and healthcare institutions in Massachusetts	Low	Agrees
The cost of doing business in Massachusetts is relatively low	Low	Agrees

Source: MassTech 2015 Massachusetts Big Data Industry Survey

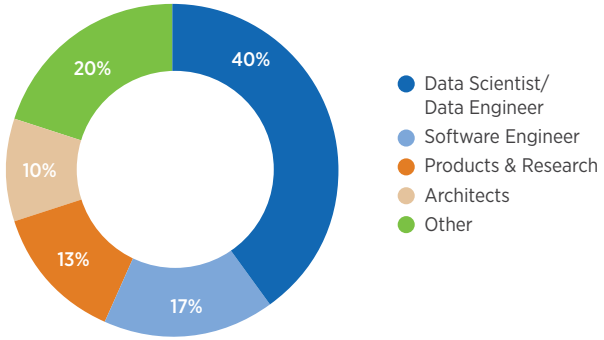
Several Boston-based companies have recently appeared in the news, as they are rapidly expanding and thus focused on hiring in the data science field.

Kayak, a leader in the travel industry, is changing the way online customers arrange travel plans. With the development of data-driven products and features, the company hopes its product can serve as a virtual, travel-tailored personal assistant. Kayak launched its Cambridge office in 2015, about ten years after it was founded in Stamford, CT. The office has almost 100 employees and is looking to expand quickly, with 55 open engineering, data science, and user interface jobs in both of its MA offices — it also has an office in Concord — as of April 2015.[56]

Kyruus, an digital health company founded in 2010, delivers a software system that helps healthcare providers match patients with the best suited doctors. With its sights on serving more clients and expanding its service line, Kyruus is increasing its number of employees, most notably in its Boston office. 20,000 healthcare providers across the country currently use Kyruus' software.[57]

IBM opened a brand new office in Cambridge this September to house its new IBM Watson Health program.[58] It aims to bring together personal health data into one data sharing hub so professionals from the medical and academic fields can conduct analysis of health data more efficiently. The program will involve hiring hundreds of employees in Massachusetts, many of which will be in quantitative disciplines, including data science. The program also includes several partnerships with leading health and technology organizations, such as Boston Children's Hospital, and companies such as Apple.[59]

FIGURE 24: HARDEST BIG DATA RELATED POSITIONS TO FILL



Source: MassTech 2015 Massachusetts Big Data Industry Survey

FIGURE 25: MOST PROMISING AREAS FOR GROWTH

Segments	Verticals	Actions for Growth
Data analysis software	E-commerce	Develop new data science curricula in colleges
Applications geared to specific verticals	Healthcare	Provide financial incentives
Data management	Financial services	Support efforts to increase supply of workers with needed skills
Data integration	Energy	Make government data readily accessible to companies
Business intelligence software	Education Cybersecurity	

Source: MassTech 2015 Massachusetts Big Data Industry Survey

56 http://www.xconomy.com/boston/2015/04/02/kayak-joins-growing-field-of-tech-giants-hiring-in-cambridge/

57 http://www.xconomy.com/boston/2015/09/17/kyruus-gets-25m-more-for-moneyball-approach-to-doctors-and-patients/; http://www.bizjournals.com/boston/blog/health-care/2015/09/health-software-company-to-hire-dozens-of.html

58 MassTech 2015 Mass Big Data Industry Survey

59 http://techcrunch.com/2015/09/10/ibm-watson-health-unit-begins-to-take-shape/

CONCLUSION

Mass Big Data Tech Trek,
Cambridge Innovation Center
Cambridge, MA

Overall, the Mass Big Data ecosystem has grown significantly since 2013. A quarter of the companies launched since 2013 are pure play big data companies and they make up the largest share of new companies, which is crucial to the Mass Big Data ecosystem. These companies facilitate big data processing, ensuring that other companies and organizations can make the most of their data. It is the development of big data technologies that allow for the analysis of big data sets across a variety of verticals. Data-driven companies across the Commonwealth focus on determining the best questions through which to explore datasets. Both pure play and data-driven companies are coming up with new approaches to working with big data: the former focuses on algorithms and the latter focuses on innovative ways to use these algorithms.

New big data technologies and research centers are driving innovation throughout the Commonwealth. The number of patents awarded to data-related technologies has continued to increase in Massachusetts, and the state continues to exceed the country average in most data-related categories. It is no surprise that Massachusetts is a leader in data-related patents, as employees of big data companies are often the ones who develop new technologies. The number of federally funded research projects using data-driven techniques continues to grow as well. These projects speak to the Commonwealth's strength in the life sciences, energy, and health, as many of the grants are awarded to projects that apply big data to these verticals in which Massachusetts leads the country. Big data can further our understanding of such complex fields as cancer research and energy, proving its importance to both society and the economy.

This growth of the big data ecosystem, through industry and innovation, has created a demand for employees who are trained in data science and data-driven analysis. Universities and colleges in Massachusetts have rapidly responded to this demand for data science, not only through the creation of degree programs, but also by building data science research centers. These research centers work to build connections between academia and industry, a crucial part of growing the big data sector of the economy, in addition to training students for the increasingly data-driven workforce. Massachusetts is now seen as a hub for big data, attracting top talent from schools outside of the Commonwealth, in addition to students who graduate from regional institutions.

It is clear that big data's role in the Massachusetts economy will only continue to grow, as its impact continues to reach more and more industries and research areas. The three ways in which we observe how big data impacts the Massachusetts economy—industry, innovation, and talent—work together to solidify Massachusetts as a global leader in big data.

APPENDIX

Massachusetts Green
High-Performance Computing Center,
Holyoke, MA

1. MASS BIG DATA COMPANIES BY INDUSTRY

ADTECH & MARKETING

COMPANY	LOCATION
89 Degrees	Burlington
Acquia	Boston
Adelphic	Waltham
AMP Agency	Boston
Awareness Hub	Boston
Black Ink ROI	Boston
BlueConic	Boston
Boost3	Boston
CaptureCode	Burlington
Celtra Inc.	Boston
Charles River Interactive, Inc.	Lexington
ChoiceStream	Boston
ClickFuel	Wakefield
Clypd	Somerville
Comlinkdata	Boston
Concentric Marketing Analytics	Cambridge
Constant Contact	Waltham
Cortex	Boston
Curata	Cambridge
Dataxu	Boston
Digilant	Boston
Digital Bungalow	Salem
Digitas	Boston
Epsilon	Wakefield
Fiksu	Boston
ForwardJump	Boston
Genuine Interactive	Boston
Grapevine	Cambridge
GrowEpic	Cambridge
HapYak Interactive Video	Boston
Havas Media	Boston
Hill Holliday	Boston
Howler	Boston
Hubspot	Cambridge
Infinite Analytics	Cambridge
InStream Media, Inc.	Wellesley
Intlock	Brookline
iProspect	Boston
ItsMyURLs	Boston
Jebbit	Boston
Klaviyo	Boston
Kuldat	Boston
Localytics	Boston
Market Vantage LLC	Groton
Millenial Media	Boston
Mobee	Boston
Mullen Advertising	Boston
Nanigans	Boston
NetProspex	Waltham
New Perspective Web Solutions	Worcester
Nexage	Boston

ntoggle	Boston
OmniLync	Cambridge
OwnerIQ	Boston
Pegasystems	Cambridge
Pepper Gang	Boston
Pixability	Cambridge
Prism Marketing Group (Prism Data Services, Prism Web Services)	Boston
Privy	Boston
Prognosys Direct	Cambridge
Pursway	Waltham
Search Benefit SEO Consultancy	Cambridge
Semcasting, Inc.	North Andover
Session M	Boston
Shareaholic, Inc.	Boston
Skyword	Boston
Small Army	Boston
Social Fulcrum	Boston
Tapjoy	Cambridge
True Fit	Woburn
UpSnap	Williamstown
Velir	Somerville
Visible Measures	Boston
VisualIQ	Needham
Wistia	Somerville
Yieldbot	Maynard
Zaius	Cambridge

BIG DATA

COMPANY	LOCATION
Actifio	Waltham
Affectiva	Waltham
Analytika	Boston
Attunity	Burlington
Basho Technologies	Cambridge
Bedrock Data	Boston
BigR	Boston
Cambridge Semantics	Boston
Cazena	Waltham
CBIG Consulting	Boston
Cognitive Electronics	Boston
Cometrics	Boston
Courion Corporation	Westborough
Curate Partners	Boston
Darwin Ecosystem LLC	Arlington
DataKitchen	Cambridge
DataRobot	Boston
Deep Information Sciences, Inc.	Boston
Digital Contact	Boston
Driven Data	Boston
DropFire Inc.	Boston
Exaptive	Cambridge
Experfy	Boston
Fathom	Boston

GFK	Boston
Hermes IQ	Braintree
HP Vertica	Cambridge
Indico	Boston
Integra Technology Consulting	Waltham
Intersystems	Cambridge
Location Inc.	Worcester
LogMeIn	Boston
Machine Analytics	Belmont
NewVantage Partners	Boston
NorthBay Solutions	Cambridge
OpenLink Software	Burlington
Opera Solutions	Newton
Paradigm4	Waltham
Pixel Forensics	Burlington
Podium Data	Lowell
Quantiphi, Inc.	Marlborough
QuantUniversity	Boston
Rocana	Boston
Soft10, Inc.	Cambridge
Tamr	Cambridge
Teradata	Boston
Third Channel	Cambridge
VoltDB	Bedford
Xtream IT Solutions	Marlborough
Zettics	Concord

BIOTECHNOLOGY & LIFE SCIENCES

COMPANY	LOCATION
Anoixis Corporation	Natick
Ariana Pharma	Cambridge
Arxspan	Southborough
Beckman Genomics	Danvers
Boston Health Economics	Waltham
Boston Scientific Corporation	Marlborough
Broad institute	Cambridge
Cadre Research Labs	Acton
Cell Signaling Technology	Danvers
DRG Analytics	Burlington
Expert Biosystems	Burlington
GenomeQuest, Inc.	Westborough
GenoSpace	Cambridge
InfoBionic	Lowell
Innerscope Research	Boston
inviCRO, LLC	Boston
KNODE, Inc.	Boston
Knome, Inc.	Cambridge
Milennium Pharmacueticals	Cambridge
Mirabo Systems International	Boston
Moderna Therapeutics	Cambridge
Novartis Institutes for BioMedical Research (NIBR)	Cambridge
PerkinElmer Inc.	Waltham
Physicians Interactive	Marlborough

Pion Inc.	Billerica
Quanterix	Lexington
Scilligence	Burlington
Seven Bridges Genomics	Cambridge
Shyft Analytics	Waltham
Spasmotica	Boston
Symmetric Computing	Boston
Thermo Fisher Scientific	Waltham
Waters Corporation	Milford
WuXi NextCODE	Cambridge

BUSINESS ANALYTICS

COMPANY	LOCATION
908 Devices	Boston
Acmeware	Dedham
Akamai Technologies, Inc.	Cambridge
Analysis Factory	Boston
Analytics Operations Engineering	Boston
Analyx	Wellesley
Applause	Framingham
Apptopia	Boston
Archetype Consulting	Boston
Aternity	Westborough
Athena IT Solutions	Maynard
Attivio	Newton
Avanade	Boston
Avention Inc	Concord
Azima DLI	Woburn
BlueGranite	Boston
BlueMetal Architects	Watertown
Booz Allen Hamilton	Boston
Boston Decision, LLC	Cambridge
Brightleaf	Brookline
Burning Glass	Boston
Canback Dangel	Somerville
Cangrading	Cambridge
Celect	Boston
Centius	Cambridge
Cisco	Boston
Cleartelligence, Inc.	Norton
Cloud Square	Norwood
Cognizeus	Boston
Coherent Path	Arlington Heights
Compete	Boston
Composable Analytics	Cambridge
Comptryx	Hopkinton
Corporate Technologies	Burlington
Crimson Hexagon	Boston
Curiositybits Collective	Boston
Data Intensity	Bedford
Databox	Boston
Dataupia	Cambridge
Datawatch Corp	Chelmsford
Deloitte (Oco)	Boston
Diamond Technologies	Maynard
Digimind	Cambridge
Dimensional Insight	Burlington
Ditto Labs	Cambridge
Dotmatics	Woburn

Dunnhumby	Boston
EMC Corporation	Hopkinton
Empirix	Billerica
Evergage	Somerville
Evertrue	Boston
galaxyadvisors	Cambridge
GoodData	Duxbury
Google	Cambridge
Ground Signal	Boston
Group Visual IO	Cambridge
Halexi, LLC	Arlington
HARTE-HANKS, INC.(Trillium.Software)	Billerica
Hitachi Data Systems	Waltham
Hub'Scan Inc.	Boston
Humanyze	Boston
Hypatia Research Group	Lexington
Icosystem	Cambridge
iKnowtion	Burlington
InforcePro	Cambridge
Information Collaboration LLC	Reading
Infosys	Quincy
Insight Squared	Cambridge
Internet Time Machine Project	Woburn
Ironside Group	Lexington
iSoftStone	Waltham
Jana Mobile (FKA Txteagle)	Boston
Janys Analytics	Cambridge
Jenzabar	Boston
Knowledgent	Boston
Krux Digital	Waltham
Lattice Engines	Boston
Lavastorm Analytics	Boston
LeadKarma LLC	Cambridge
Leaf	Cambridge
Liberty Mutual	Boston
Libring	Cambridge
LinkeDrive, Inc.	Boston
Liquid Hub	Waltham
Logentries	Boston
Lucidel	Cambridge
Luminoso Technologies, Inc.	Cambridge
Lux Research Inc.	Boston
Machine Metrics	Northampton
Magnitude Software	Burlington
MarketMuse	Cambridge
Mathworks, Inc.	Natick
Maven7	Boston
MERCURY COMPUTER SYSTEMS INC	Chelmsford
Microsoft New England	Cambridge
Minubo	Boston
Mobiquity	Wellesley
Moviri	Boston
Nara Logics	Cambridge
Nautilytics, LLC	Boston
NetApp	Waltham
New England Sql Server Experts Group, Inc.	Ipswich
New England Systems	Ludlow
nighthawk apps, LLC	Boston

Nimbata	Boston
NodeSpan	Cambridge
Nutonian	Somerville
ORB Analytics, Inc.	Concord
OzVision	Woburn
Panjiva	Cambridge
Peloton Group	Boston
Personali	Waltham
Pivotal	Boston
Prelert	Framingham
Price Intelligently	Boston
Profitect, Inc.	Waltham
Progress Software	Bedford
PTC	Needham
Qlik	Newton
Qstream	Burlington
qunb	Boston
Qvidian	Chelmsford
RAGE Frameworks	Dedham
RapidMiner	Cambridge
Rifiniti	Boston
Riparian Data	Watertown
Rocket Software	Waltham
ScaleBase	Newton
Semantria	Amherst
Senexx	Cambridge
Sky Analytics	Framingham
Slalom	Boston
Software AG	Cambridge
Strategy Analytics	Newton
synapbox	Cambridge
Talent Analytics	Cambridge
Targit	Boston
Techlogix	Woburn
Teikametrics	Boston
TesoraCorp	Cambridge
TIBCO Spotfire	Somerville
ToolsGroup	Boston
TR3 Solutions, Inc.	Stoneham
Ve Interactive	Boston
Via Science	Cambridge
Visionware	Newton
Vistaprint	Lexington
VMTurbo	Boston
Wall Street Horizon	Woburn
Weft	Cambridge
Westbury	Cambridge
Westernacher	Wellesley
Yesware	Boston
Yottaa	Waltham
Zensight	Boston
ZoomInfo	Waltham

CLEAN TECHNOLOGY & ENERGY

COMPANY	LOCATION
1Efficiency	Newton
Digital Lumens	Boston
EnerNOC	Boston
Essess, Inc	Boston

FirstFuel Software	Lexington
GreatPoint Energy	Cambridge
Interval Data Systems, Inc.	Waltham
ISO New England	Holyoke
Loci Controls	Somerville
National Grid	Waltham
Newton Energy Group	Cambridge
Peregrine Energy Group	Boston
Powerhouse Dynamics	Newton
Promethean Power Systems	Boston
Schneider Electric	Andover
Vaisala	Woburn
Ventana Systems, Inc.	Harvard
WegoWise, Inc.	Boston
WindPole Ventures	Arlington

CYBERSECURITY

COMPANY	LOCATION
Apperian	Boston
Aptima, Inc.	Woburn
Arbor Networks	Burlington
Aware	Bedford
BAE Systems	Boston
BioCatch	Boston
Bit9 + Carbon Black	Waltham
BitSight	Cambridge
Boston Fusion	Burlington
Bradford Networks	Boston
Cigital, Inc	Boston
CloudLock	Waltham
Co57 Systems, Inc	Cambridge
Conjur	Waltham
Content Raven	Marlborough
Core Security	Boston
Corero	Hudson
CounterTack	Waltham
Cryptzone	Waltham
CyberArk	Newton
Cybereason	Cambridge
Cybric	Boston
DCS Corp	Harvard
Digital Guardian	Waltham
EiQ Networks	Boston
Forum Systems	Newton
FST21	Southborough
General Dynamics Fidelis Cybersecurity Solutions	Waltham
GreatHorn	Belmont
Kaspersky Lab	Woburn
Lockheed Martin Industrial Defender Inc.	Foxborough
MaxMind	Waltham
Milcord	Waltham
Mimecast	Watertown
NetScout	Westford
Promisec	Needham Heights
Rapid7	Boston
Raytheon Co.	Waltham
Recorded Future	Somerville
Red Bend Software	Waltham
Resilient Systems	Cambridge

ReversingLabs	Cambridge
Sophos	Burlington
sqrrl	Cambridge
Symantec	Waltham
Systems & Technology Research	Woburn
Threat Stack, Inc	Boston
TowerWall	Framingham
TrustLayers	Cambridge
Viewfinity	Waltham
Wave Systems	Lee

DIGITAL HEALTH

COMPANY	LOCATION
5AM Solutions	Boston
Arcadia Healthcare Solutions	Burlington
athenahealth	Maynard
Berg	Framingham
Boston Advanced Analytics	Boston
Capsule Technologies	Andover
Cliniworks	Cambridge
Connance	Waltham
Converge Health	Newton
CoPatient	Boston
Curoverse, Inc.	Boston
Deerwalk, Inc.	Lexington
Definitive Healthcare	Framingham
Diameter Health	Newton
Eliza Corporation	Danvers
Epidemico	Boston
ERT	Boston
Foundation Medicine	Cambridge
GameMetrix Solutions	Salem
Gennius, Inc.	Waltham
Ginger.io	Cambridge
GNS Healthcare	Cambridge
Health Dialog	Boston
HealthEdge	Burlington
HospitalIQ	Newton
Humedica, Inc.	Boston
Iatric Systems	Boxford
IBM Watson Health	Cambridge
ilumivu	Cambridge
InsideTracker	Cambridge
JEN Associates, Inc.	Cambridge
Kyruus	Boston
Lincoln Peak Partners	Newton
LuminaCare Solutions	Cambridge
Maxwell Health	Boston
McKesson Health	Northborough
MedAptus	Boston
MediCollector	Cambridge
MedPanel	Cambridge
meQuilibrium	Boston
Mobile Integrity Consulting	Boston
Neumitra	Boston
Objective Health (division of McKinsey)	Waltham
Optum	Cambridge
Ovuline	Boston

PatientsLikeMe	Cambridge
Persivia	Lowell
Philips	Andover
PointRight, Inc.	Lexington
Predilytics	Burlington
QPID, Inc.	Boston
Quanttus	Cambridge
Radial Analytics	Cambridge
RGI Informatics, LLC	Boston
Smart Scheduling	Cambridge
Truven Health Analytics	Cambridge
Verisk Health	Waltham
Wellframe	Cambridge
Zakipoint	Cambridge

FINTECH

COMPANY	LOCATION
Advanti	Boston
Alivia Technology	Woburn
Alt Options	Boston
Altisource Labs	Boston
Bison	Boston
Blueleaf	Boston
Cerulli Associates	Boston
Cignifi	Cambridge
Domeyard LP	Boston
EFL	Cambridge
Elsen Inc.	Boston
EverVest	Boston
Fidelity Investments	Boston
Hanover Insurance	Worcester
Kensho	Cambridge
Kuberre Systems	Wilmington
LevelTrigger	Cambridge
MassMutual	Springfield
Mavenomics	Cambridge
Micronotes, Inc	Cambridge
Milliman Analytics Exchange	Cambridge
NorthPoint Solutions LLC	Boston
Pro Teck Valuation Services	Waltham
Quantopian	Boston
Santander	Boston
Saylent Technologies, Inc	Franklin
Spinnaker Analytics	Boston
State Street	North Quincy
Thomson Reuters	Boston
Trefis	Boston

OTHER

COMPANY	LOCATION
Ab Initio	Lexington
Advanced Visual Systems	Waltham
Amazon	Cambridge
Azavea	Cambridge
B2BGateway.net	Brockton
BackOffice Associates	South Harwich
Basis Technology	Cambridge
BeON Home	Cambridge
BlockAvenue	Boston
Bridj	Boston

Briefcase Analytics, Inc	Cambridge
Censio	Allston
Cogito	Boston
Cognika Corporation	Brookline
Cogo Labs	Cambridge
DataCamp	Cambridge
DiscoverText	Amherst
DraftKings	Boston
Drizly	Boston
EchoNest	Somerville
Ecovent	Cambridge
EWB Analytics LLC	Natick
Facebook	Cambridge
Fleetmatics	Waltham
Forrester Research	Cambridge
HayStackID	Boston
Hopper	Cambridge
IBM Cloudant	Boston
II-VI Photonics	Woburn
Inex Advisors	New Bedford
intuVision, Inc.	Woburn
iRobot	Bedford
Kayak	Cambridge
LeafLabs	Cambridge
LevelUp	Boston
Lexalytics, Inc	Amherst
Linguamatics	Westborough
metacog, Inc	Worcester
NDensity	Cambridge
Neurala	Boston
NuoDB	Cambridge
Objective Logistics	Boston
OnCorps	Cambridge
Oracle	Cambridge
Panorama Education	Boston
QuadWrangle	Boston
QuikForce	Boston
Seven10 Storage Software	Lawrence
Skyhook Wireless	Boston
Sonamine	Woburn
STAR Analytical Services	Bedford
Taneja Group	Hopkinton
Tarmin	Boston
Trip Advisor	Needham
Uber	Boston
uberVU	Cambridge
Understory, Inc.	Somerville
Verizon	Littleton
Vlingo	Cambridge
Wayfair	Boston
Weather Source	Amesbury

2. MASS BIG DATA COMPANIES BY INDUSTRY, FOUNDED SINCE 2013

ADTECH & MARKETING

- Awareness Hub
- Cortex
- Howler
- ntoggle

BIG DATA

- BigR
- Cazena
- Curate Partners
- DataKitchen
- Driven Data
- Experfy
- Hermes IQ
- Indico
- Podium Data
- Quantiphi, Inc
- QuantUniversity
- Rocana
- Tamr

BIOTECHNOLOGY & LIFE SCIENCES

- Cadre Research Labs
- DRG Analytics

BUSINESS ANALYTICS

- Cognizeus
- Curiositybits Collective
- Ground Signal
- Hub'Scan Inc
- Machine Metrics
- MarketMuse
- Minubo
- NodeSpan
- Pivotal
- Weft
- Zensight

CYBERSECURITY

- CounterTack
- Cyber 360 Inc
- GreatHorn
- TrustLayers

DIGITAL HEALTH

- Diameter Health
- HospitalIQ
- IBM Watson Health
- LuminaCare Solutions
- MediCollector
- Persivia
- QPID, Inc
- Radial Analytics

FINTECH

- Alt Options
- Elsen Inc
- EverVest
- Kensho
- LevelTrigger

OTHER

- Bridj
- Briefcase Analytics, Inc
- DataCamp
- metacog, Inc
- NDensity
- QuikForce

3. MEETUP GROUPS

NAME	FOUNDED	DESCRIPTION
ACM Boston	2014	ACM Boston is a non-profit professional group that meets regularly to discuss diverse topics in computer science such as predictive analytics, applied machine learning, statistical modeling, open data, and data visualization, user experience, user research, and artificial neural networks.
Action Design Boston	2014	A group at the intersection of technology (UI/UX, data science), psychology (behavioral economics, positive psychology), and advocacy, and we focus on exposing our members to cutting edge tools, research, and practitioners in each of these areas so they can apply the latest learning in action design in their respective fields.
Analytics Club MetroWest Boston	2014	Analytics Club is an 'All Things Analytics' community hub for bringing together top talent, tools, technologies and techniques around data professionals and businesses.
Big Data Analytics, Discovery & Visualization	2013	This group is relevant for data scientists, business executives seeking to learn what big-data discovery could do to their businesses & professionals who are interested to learn what data discovery is all about.
Big Data Behavioral Analytics Boston	2015	A group that meets to discuss A/B testing, cohort analysis, path analysis, funnels, CAC and KPI to improve churn, conversion, customer Life Time Value, and any other topic, related to user acquisition, conversion and monetization.
Big Data Boston	2012	Big Data Boston is for people with a passion for analytics & insights that are derived from the extreme information generated today.
Big Data Developers — Boston	2013	This is an IBM sponsored Big Data meetup group, geared towards developers, data scientists and all Big Data enthusiasts, our meetups provide an opportunity to work hands on with the solutions and tools in our Big Data portfolio.
Big Data Developers Boston	2013	Hands-on software development experience on all topics related to big data.
Boston #ODSC	2013	The focus of this meetup group is to present informative lectures, hands-on tutorials, and networking events to help grow the use of open source languages and tools within the data science and data-centric community.

Boston Algorithmic Trading	2012	Boston Algorithmic Trading is for anyone interested in creating and using algorithms in the financial markets.
Boston Apache Spark User Group	2014	This is a group for people in the Greater Boston area who are interested in Apache Spark — what it is, what it does, and what other people are doing with it.
Boston Big Data & Hadoop Learning Group	2015	This group is dedicated towards aspiring Big Data & Hadoop professionals as well as anyone who are keen to understand data analysis and its real-world applications.
Boston Big Data 2.0 Group	2013	This is a group for people and technologists who want to explore how to make sense of the information overload that comes with Big Data 2.0 (the Big Data of open social expressions, news, blogs and microblogs).
Boston Business Intelligence	2013	This group is for those interested in learning the latest technology and tools in Microsoft Business Intelligence Platform: Analysis Services (SSAS), Repoting Services (SSRS), Integration Services (SSIS), MDS and languages (MDX, DAX, DMX) as well as Client side Data Analysis
Boston Cloud Services — All things Cloud, SaaS, PaaS, XaaS	2009	A group dedicated to sharing, evangelizing and promoting the next big wave in technology, Cloud based services: software (SaaS), Platform (PaaS) as-a-service etc.
Boston Data Mining	2013	The Boston Data Mining meetup focuses on making data mining accessible to everyone by focusing on the practical side of data mining and analytics.
Boston Ember.js Group	2012	This group focuses on the data science applications of Ember.js, an open source Javascript application.
Boston Entrepreneurs and Advanced Degrees	2013	The purpose of this group is to help professionals develop an entrepreneurial mindset, while giving entrepreneurs access the highly specialized knowledge that these professionals possess.
Boston Hacking Predictive Analytics App	2009	This is your chance to prototype your ideas before going to VCs or for seed funding.
Boston Hadoop User Group	2009	This group aims to build data models that attendees can use themselves; make data mining and data analytics accessible to everyone; and increase awareness of open source data mining tools.
Boston Machine Learning	2014	A group with a goal to help connect people with machine learning practitioners from academia and industry.
Boston Mesos User Group	2015	This group is focused on promoting Apache Mesos as a platform for creating distributed systems.

Boston Mongo DB	2011	This group is a place for developers to learn more about the non-relational, open source, document-oriented database MongoDB
Boston Predictive Analytics	2010	The goal of this meetup is present informative lectures, hands-on tutorials, networking events, etc, towards helping the local community further its understanding and proficiency regarding Predictive Analytics.
Boston Quantitative Systems Pharmacology Meetup	2015	Through iterative process of — systems model-based analysis of the data, evolving understanding of the underlying system generating the data, further data generation, and so on — we can get a much cleaner and quantitative understanding of pharmacology.
Boston Smart Data Meetup Group	2013	Smart Data refers to the high-value data used to inform business decisions. Our interest is in comprehensive enterprise solutions, from ingestion of data through to business insight.
Boston Storm Users	2012	Boston Storm Users is a group for developers using or hoping to learn about Twitter's Storm real-time data processing framework.
Boston Streams	2015	Learn how to uncover the mysteries of continuous data flows and build robust stream computing environments.
Boston Vowpal Wabbit Meetup	2015	This group is targeted at industrial uses of vowpal-wabbit machine learning library. The meetup is designed to share ideas, tips and showcasing projects based on Vowpal.
Cambridge Dataflow Analytics Meetup	2014	We're a group of analysts, developers, IT pros, and hackers that are interested in using dataflow-based technologies to solve big data problems and automate business processes.
Cambridge Area Julia Users Network	2013	A meetup group dedicated to users and developers of the Julia language for high performance technical computing.
Code for Boston	2013	Code for Boston is a Code for America Brigade — a volunteer civic innovation organization created by Boston-area developers, designers, urban planners, data geeks, and researchers with an interest in solving civic and social problems through the use of creative technology.
Cognitive Computing	2014	Cognitive Computing (CC) is the next wave. Large companies like IBM, Microsoft, Intel, SAS, HP, Oracle, etc. are investing millions of dollars automatically linking Big Data analysis and Affective System deployment, to better find and serve new prospects and existing customers.

Data Science for Social Good	2015	The goal of this Meetup is to bring the community together to apply data science to improve society. Some areas that will be addressed include crime, healthcare, transportation, clean energy; income disparity, among others.
Data Science Group	2012	This group will concentrate on understanding the tools and skill-sets needed to become an effective Data Scientist.
Data Science Professional Development Boston	2015	Members of this Meetup group will get and give personalized career advice, learn content and receive content recommendations from peers, and intelligently grow their professional network.
Data Visualization in Metrowest Boston	2013	This group is for individuals who are statisticians, designers, programmers, journalists or even car salesmen with an interest in data visualization.
Elasticsearch Boston	2012	Elasticsearch is picking up steam in Bean Town. Let's share our experiences with it and strengthen the local community around it by introducing newcomers to the great features it has to offer.
Free Big Data Hands-On Workshops	2014	This is for people who want to learn and develop in the Big Data domain with hands-on workshops.
Girls Develop It Boston	2013	Our courses focus on coding, leveraging existing technology, and having something to show for it (aka building sweet websites), while supporting women in data science.
Graph Database Boston	2012	Developers interested in learning about and working with graph databases for social, spatial, hierarchical or other highly connected data sets.
Greater Boston useR Group (R Programming Language)	2011	R is a free and open programming language for statistical computing, data analysis, and graphical visualization. The Greater Boston useR Group seeks to bring this community together to share ideas, discuss R related topics, and provide direction for new and experienced users.
Learning Analytics Boston	2012	This group will provide a place to meet and discuss potential use of existing technologies and practice in the service of improving education.
New England Artificial Intelligence	2011	Our group is for those interested in AI, machine learning, forecasting, recommendation systems, and building smarter applications.
New England SQL Server	2013	This group is dedicated to learning about and practicing with the SQL server.

Oracle NoSQL & Big Data — Boston	2013	Meetup group focused on learning about Oracle Big Data/NoSQL and related technologies, including use cases, new features and exciting networking with other professionals.
PyData Boston	2013	Intersection of Python and Data Science.
PyLadies Boston	2013	PyLadies Boston is for women who are either Python programmers or aspiring Python programmers in the Boston area.
The Boston Python User Group	2007	Boston Python is the world's largest local Python user group. Meet other local Python developers, learners, employers, and enthusiasts of all kinds.
The North American VoltDB Meetup Group	2012	This is a group for application developers who aspire to make the impossible possible.
Western Mass Data Science, Stats, and R	2013	This is an R/RStudio, Statistical and Data Science user group for the Pioneer Valley and Five College Area.
Women Who Code	2014	Women Who Code is a global nonprofit organization dedicated to inspiring women to excel in technology careers by creating a global, connected community of women in technology.

4. DATA SCIENCE & RELATED PROGRAMS

ACADEMIC INSTITUTION	DEGREE TYPE	YEAR FOUNDED
BIOLOGY & DATA		
Simmons College	Bachelors	2011
UMass Lowell	Bachelors	
UMass Lowell	Bachelors	
Worcester State University	Bachelors	1993
Worcester State University	Bachelors	2005
WPI	Bachelors	1980
WPI	Bachelors	2014
Boston University	Masters	1999
Boston University	Masters	1987
Brandeis University	Masters	2002
Harvard University	Masters	2013
Northeastern University	Masters	2000
WPI	Masters	1982
WPI	Masters	2014
Boston University	PhD	1999
Boston University	PhD	1987
MIT	PhD	2004
WPI	PhD	1985
BUSINESS ANALYTICS		
Boston College	Bachelors	2015
Suffolk University	Bachelors	2014
Western New England University	Bachelors	2015
Bentley University	Certificates	1999
Bentley University	Certificates	1999
Wentworth Institute of Technology	Certificates	2014
Babson College	Masters	2014
Bentley University	Masters	2014
Bentley University	Masters	2004
Bentley University	Masters	2005
Boston College	Masters	
Northeastern University	Masters	2015
CYBERSECURITY		
Boston University	Masters	2012
Brandeis University	Masters	2008
Northeastern University	PhD	2011
DATA SCIENCE		
Becker College	Bachelors	2014
UMass Dartmouth	Bachelors	2015
WPI	Bachelors	
Harvard Extension School	Certificates	2014
Northeastern University	Certificates	2015
WPI	Certificates	
Harvard University	Masters	2013
UMass Dartmouth	Masters	2015
WPI	Masters	2014
WPI	PhD	2015

DATA SCIENCE MINOR

UMass Amherst	Bachelors	2014
Boston University	Masters	2016
UMass Boston	PhD	2016
Northeastern University	Masters	2014

ENGINEERING

Boston University	Bachelors	2009
Boston University	Masters	2009
Boston University	Masters	2009
Boston University	PhD	2009

HEALTH INFORMATICS

Simmons College	Bachelors	2010
Northeastern University	Certificates	2013
UMass Lowell	Certificates	
Brandeis University	Masters	2010
Northeastern University	Masters	2006
UMass Lowell	Masters	
Northeastern University	PhD	2012

OTHER

ACADEMIC INSTITUTION	PROGRAM	DEGREE TYPE	YEAR FOUNDED
Bristol CC	Computer Information Systems	Bachelors	
Simmons College	Information Technology	Bachelors	2000
Suffolk University	Information Systems and Operations Management	Bachelors	2010
Bunker Hill CC	Data Management	Certificates	2014
Northeastern University	Urban Informatics	Certificates	2015
Brandeis University	Strategic Analytics	Masters	2013
Brandeis University	Learning Analytics	Masters	2015
Clark University	GIS for Development and Environment	Masters	2000
Northeastern University	Information Assurance	Masters	2005
MIT	Operations Research	Masters	
Boston University	Statistical Practice	Masters	2015
Northeastern University	Urban Informatics	Masters	2015
MIT	Operations Research	PhD	

5. RESEARCH CENTERS

INSTITUTION	CENTER	SECTOR	DESCRIPTION
Atmospheric and Environmental Research	N/A	Geography	Researches weather and climate related risk and delivers solutions in the form of databases, research reports, and software.
Boston Children's Hospital	N/A	Health	World's largest pediatric, medical research center with a wide variety of centers and research areas, many of which are heavily data-driven.
Boston University	Center for Computational Science	Data Science	Focuses on efforts to coordinate and promote computationally based research, foster computational science education, and provide for the expansion of computational resources and support.
Boston University	Center for Reliable Information Systems and Cyber Security	Cybersecurity	Established to promote and coordinate research on reliable and secure computation and on information assurance education by providing increased opportunities for collaboration among researchers from cognate fields.
Boston University	Rafik B. Hairi Institute for Computing and Computational Science & Engineering	Data Science	Aims to initiate, catalyze, and propel collaborative, interdisciplinary research and training initiatives for the betterment of society by promoting discovery and innovations through the use of computational and data-driven approaches, and by supporting advances in the science of computing inspired by challenges in arts and sciences, engineering, and management disciplines
Boston University	Center for Computational Neuroscience and Neural Technology (CompNet)	Life Sciences	Brings together researches from a variety of disciplines to apply advanced computational methods and technological solutions to the field of neuroscience.
Boston University	BioMolecular Engineering Research Center	Engineering	Develops new, analytic methods to address problems in the bimolecular field using computer-based technology. The center hopes to contribute to the understanding of biology, engineering, and medicine through the computational methods it develops.
Boston University	Clinical and Translational Science Institute	Life Sciences	Aims to improve health, healthcare, and delivery of this care through community engagement, informatics, biostatistics, clinical research, and research education. The center analyzes health data using high through-put technology, while also providing connections to researches and emphasizing community needs.

Boston University	Center for Information and Systems Engineering	Engineering	Use data-driven techniques to create software and hardware that advances human intelligence in a wide variety of verticals including health care, communications, and national security.
Brigham and Women's Hospital	N/A	Health	Conducts top level research in several areas including genomics and bioinformatics.
Broad Institute	N/A	Life Sciences	A large-scale scientific collaboration in genomics and chemical biology that grew out of major initiatives at Harvard and MIT. Collectively, these projects aim to assemble a complete picture of the molecular components of life, define the biological circuits that underlie cellular responses, uncover the molecular basis of major inherited diseases, unearth all the mutations that underlie different cancer types, discover the molecular basis of major infectious diseases, and transform the process of therapeutic discovery and development.
Concord Consortium	N/A	Education	Brings technology and data-driven analysis to the field of education.
Dana Farber Cancer Institute	Center for Cancer Computational Biology	Health	Focuses on genomic and computational biology approaches that open new ways of understanding cancer by improving analysis and interpretation of genomic data through integration with information derived from other sources, including publicly available data. Also supports analysis and interpretation of genomic and other large-scale data to further basic, clinical, and translational research.
Draper Labs		National Security	Design and develops technological solutions to solve problems in the fields of security, energy, space exploration, and healthcare.
ECOG-ACRIN Cancer Research Group	Biostatistics and Data Management Center	Life Sciences	Improves quality of care and survival rates for cancer patients through data-driven methodologies and research.
Education Resource Strategies		Education	Conducts data-driven studies on how to best allocate resources in school systems across the country.
Harvard University	Center for Research on Computation and Society	Social Sciences	Focuses on development of new ideas and technologies designed to address fundamental computational problems arising from societal issues, such as privacy, security, and crowdsourcing. The Harvard Center for Research in Computation and Society's integrative approach combines research on computer science and technology informed by societal events to reach their research goals.

Harvard University	Center for Systems Biology	Life Sciences	Overall goal is to find general principles that help explain the structure, behavior, and evolution of cells and organisms.
Harvard University	Institute for Quantitative Social Science	Social Sciences	Focuses on quantitative research in the social sciences across many disciplines.
Harvard University	Institute for Applied Computational Science	Data Science	Runs Harvard's Master Degree in applied computational science and supports the development of new courses in applied mathematics and computational science.
Harvard University	Berkman Center for Internet and Society	Social Sciences	Studies norms, dynamics, and standards of cyberspace while simultaneously figuring out the challenges and opportunities with which it presents society.
Harvard University	Center for Health Decision Science	Social Sciences	Applies the interdisciplinary field of decision science—how people make decisions and how they can be improved—to health related topics including health care and the environment. Synthesizes and analyzes data to figure out the economic costs and health consequences of various policies and practices.
Harvard University	Center for Biostatistics in AIDS Research	Health	Studies HIV by applying statistical analysis to research on clinical trials and study design, and promotes innovative approaches to studying HIV.
Massachusetts General Hospital	Center for Computational and Integrative Biology	Life Sciences	The study of biology, underpinned with data and statistics, in a variety of research areas from the origins of life to physiology.
Massachusetts General Hospital	Analytic and Translational Genetics Unit	Life Sciences	Determines how genome sequencing can be used to improve clinical decision making and human disease.
Massachusetts General Hospital	Center for Computational Discovery	Health	Analyzes complex biological data using biostatistics, bioinformatics, and advanced computational techniques. Specifically, the center aims to uncover new biomarkers and drug targets to improve cancer treatment.
McLean Hospital	Psychiatric Biostatistics Laboratory	Life Sciences	Biostatisticans consult on research projects throughout the hospital by providing insight related to experimental design studies, statistical methodologies, and result analysis.
MGH, Harvard	Center for Human Genetics Research	Life Sciences	Focuses on the underlying mechanisms in human diseases by studying genetics.
MGHPCC	N/A	Data Science	Infrastructure for data-driven research requiring intense computation techniques.

MIT	Computer Science and Artificial Intelligence Lab	Artificial Intelligence	Focuses on artificial intelligence, systems, and theory. Goal is to apply knowledge on human intelligence, extending functional capabilities of machines, human/machine interactions to engineer solutions with global impact. The Big Data group focuses on identifying and developing technologies to solve the next generation data challenges.
MIT	Media Lab	Technology	Focuses on efforts that combine seemingly disparate research areas to uncover ways to radically improve the way people live, learn, express themselves, work and play.
MIT	Operations Research Center	Operations Research	Aims to apply advanced analytical methods to help make better decisions. The center's research activities cover both methodological research (i.e. mathematical programming and combinatorial optimization, cluster analysis, network design) and application domains (i.e. flexible manufacturing systems, air traffic control, epidemiology).
MIT	Institute for Data, Systems, and Society	Social Sciences	Applies analytical methods in information and decision systems to address social issues in a variety of verticals including health, energy, finance, and social networks.
MIT	The Center for Computational Engineering	Engineering	Develops innovate computational methods and applications of these methods in the field of engineering.
MIT	Center for Computational Research in Economics and Management Science	Social Sciences	Researches algorithms and software development related to today's most advanced modeling techniques in economics, finance, statistics, and management.
MIT	Lincoln Laboratory	Cybersecurity	Federally funded research center that approaches national security with technological advancements and solutions.
MIT	Laboratory for Information and Decision Systems	Data Science	Aims to advance research and education in the analytical and information sciences which includes systems and control, communication and networks, and statistical data processing.
MIT	Geospatial Data Center	Geography	Analyzes geospatial data in the context of national security with a number of programs including cyber-physical security, the big data lab, and a data center infrastructure simulator.

MIT	The Center for Brains, Minds and Machines	Life Sciences	Works to bring together experts from a variety of disciplines to advance and create the science and engineering of intelligence. The Center's goal is to figure out how intelligence relates to computation in a variety of different contexts.
MIT	MIT Hacking Medicine	Health	Solves health problems around the world and teaches health entrepreneurship using digital strategies.
Northeastern University	Center for Complex Network Research	Networks	Studies ow networks emerge, what they look like, and how they evolve; and how networks impact on understanding of complex systems.
Northeastern University	Center for Interdisciplinary Research on Complex Systems	Life Sciences	Aims to elucidate fundamental aspects of the structure and function of complex physical and biological systems across multiple levels of organization using a combination of quantitative state-of-the-art experimental and theoretical research tools.
Northeastern University	Institute for Information Assurance	Cybersecurity	Examines cybersecurity security from three perspectives: (1) Network security spanning multiple network communication layers, such as sensors and wireless devices; (2) Information integrity, including threats such as viruses and insider attacks; (3) Hardware and software system vulnerabilities in information infrastructures.
Partners Healthcare	Informatics for Integratring Biology and the Bedside	Life Sciences	Focuses on developing software and methodologies to enable clinical researchers to accelerate translation of genomic and “traditional” clinical findings into novel diagnostics, prognostics, and therapeutics.
Pioneer Institute	N/A	Public Policy	Evaluates and researches public policy in a data-driven way on a variety of topic areas.
Public Lab	N/A	Health	Develops and applies open source software and hardware to address community environmental health issues.
Silent Spring Institute	N/A	Life Sciences	Studies health risks involved in women’s health, and our environment.
Tufts University	Initiative for the Forecasting and Modeling of Infectious Diseases	Health	Conducts research in the fields of computational epidemiology, bioinformatics, and biostatistics to develop innovative computational and analytical tools to be used in the health care and medical fields.

UMass Amherst	Center for Intelligent Information Retrieval	Technology	Focuses on developing technology that provides effective and efficient access to large networks of heterogeneous, multimedia information.
UMass Amherst	Institute for Computational Biology, Biostatistics, and Bioinformatics	Health	Aims to apply computational, biomedical and translational research to the life sciences through high-level analytic methods. Activities focus on catalyzing intellectual exchange and connections among participating departments, pursuing extramural funding to support the development of educational opportunities, and engaging life science and IT companies to identify shared interests for future collaborations.
UMass Amherst	Institute for Applied Life Sciences	Life Sciences	The institute's three centers — Bioactive Delivery, Models to Medicine, and Personalized Health Monitoring — aim to develop new technologies to advance the health care field.
UMass Amherst	Center for Data Science	Data Science	Home to a variety of research labs that collect and analyze big data to reach important decisions.
UMass Lowell	Center for Advanced Computation and Telecommunications	Data Science	Compute-intensive modeling of physical and information systems. Members of the Center have undertaken research in the areas medical imaging, acoustics, fluid dynamics, heat transfer, control, probabilistic modeling, information processing, and communication networks.
UMass Lowell	Center for Computer Machine/ Human Intelligence Networking and Distributed systems	Data Science	Research, training and education to help advance research in the analytical, experimental and operational aspect of computer hardware and software, data engineering, and information technologies.
Woods Hole Oceanographic Institution	N/A	Life Sciences	Explores and works toward a better understanding of the ocean; many of its research projects are heavily data-driven and require the use of big data techniques.
WPI	Center for Research in Exploratory Data and Information Analysis	Data Science	Research in data exploration and knowledge discovery, and to the application of this research in scientific, industrial, and commercial domains. Verticals: bioinformatics; e-commerce; earth and space science; security; communication networks; healthcare. Specific areas: knowledge discovery in databases; data mining; information visualization; machine learning; pattern recognition; statistics; signal analysis.

ACKNOWLEDGMENTS

We would like to thank the Mass Big Data Advisory Committee, the respondents of the 2015 Mass Big Data Survey and the academic program directors whose engagement and support made this report possible.

Mass Big Data Advisory Committee 2015

Azer Bestavros
Professor, Computer Science Department
Director, The Hariri Institute for Computing, Boston University

Alex Cosmas
Chief Scientist and Principal, Booz Allen Hamilton

John Goodhue
Executive Director, Massachusetts Green High-Performance Computing Center

Tom Hopcroft
President & CEO, Massachusetts Technology Leadership Council

Cort Johnson
Venture Partner, Accomplice VC

Patrick Larkin
Director, the Innovation Institute at MassTech

Micah Remley
Senior Vice President of Product, EnerNOC

Michael Schrage
Research Fellow, MIT Center for Digital Business

Mona Vernon
Vice President, Data Innovation Lab, Thomson Reuters

Report Staff

Mary Rose Greenough
Director of Program Development, the Innovation Institute at MassTech

Daniel Haro
Program Specialist, the Innovation Institute at MassTech

Allison Wainer
Research Intern, the Innovation Institute at MassTech

Design

Opus Design

The Massachusetts Big Data Indicators 2015 was produced by The Innovation Institute at the Massachusetts Technology Collaborative. Published in the U.S.A.