Dr. David F. McQueeney
Vice President, Technical Strategy and Worldwide Operations
IBM Research

House Science Research Subcommittee and
House Science Technology Subcommittee
Next Generation Computing and Big Data Analytics Hearing
April 24, 2013

Good morning, Chairman Bucshon, Chairman Massie, Ranking Member Lipinski, Ranking Member Wilson, and members of the subcommittees. Thank you for the opportunity to speak with you about big data, analytics, and the new opportunities they present.

My name is David McQueeney and I am the Vice President for Technical Strategy and Worldwide Operations at IBM Research. I am responsible for setting the direction of IBM's overall research strategy and I lead the creation of IBM's Global Technology Outlook. This is an annual effort which guides IBM's R&D directions and its acquisition strategies. I have held a variety of senior research and business unit leadership roles throughout my twenty-five years at IBM.

My testimony today draws on this and the wider IBM experience with Big Data and Analytics and associated technology and skills requirements. I will focus on next generation computing, the nature of big data and analytics, opportunities and issues these technologies present, and how best to promote their growth.

What is Big Data and Analytics?

We are entering a new era of computing which is causing profound change in the IT industry. We are moving from computing based on processors that are programmed to follow a pre-designed sequence of instructions, to the cognitive computing era based on massive amounts of data and systems evolving into systems that can "learn". Cognitive systems can modify and optimize projections or weigh the value of information based on experience and results. This new approach to computing requires new strategies and skills to maintain U.S. leadership.

Cognitive systems will digest and exploit the massive data volumes being generated today. The data is coming from the technologies which mark our age: mobile phones, cloud computing, social networks and what we call the "internet of things", including everything from your car to your refrigerator to the thousands of texts your son or daughter sends each month. Imagine the data volumes generated by just two popular sources, the two billion videos watched daily on YouTube and the 293 billion emails sent every single day. As much data is generated in two days in 2013 as in all of human history prior to 2003.

Why is Big Data Analytics important?

Advanced analytics can be thought of as tools for using data to make decisions based on facts rather than intuition. These tools are applied to unstructured data from the Web, communications networks, governments, businesses, and tens of millions of sensors, to help us understand how the world works better than ever before. The challenge for business and government alike is to transform latent data into meaningful, actionable information. Analytic tools enable predictive insights and inform decision making to prevent bad outcomes and define helpful courses of action.

For example, the Memphis Police Department is using data analytics to create maps of crime hot spots and find patterns that police could not see themselves. They have correlated crimes committed at night with presence of pay phones outside convenience stores. As a result, the police suggested moving the phones inside the stores. The outcome – crime has fallen. Memphis has increased targeted, effective police presence and eliminated random patrolling. The city has seen a 30 percent reduction in crime with no increase in overall manpower using new knowledge and skills based on data analytics. There are numerous examples around the world from all sectors of society.

In sum, advanced analytics are tools for using data to improve organizational effectiveness and create value. Combining analytics with massive data flows enables organizations to make decisions based on facts rather than intuition. This factual, analytic decision making can revolutionize industries and help society successfully address challenges such as energy conservation, health care, and transportation, as well as rooting out waste, fraud and abuse.

Trends in technology development for Next Generation Computing and the Data Explosion

The most powerful computing systems today are essential in both government and industry. They are used in applications ranging managing the nation's nuclear stockpile to automobile, aircraft and semiconductor design to the development of new tires and race car aerodynamics to oil exploration and oil recovery. However, they are reaching a performance level that will stagnate without significant innovation. We must move to the next generation of large-scale computing, exascale computing – 1000 times faster than today's petascale machines.

An exaflop is a quintillion floating point operations per second. Exascale computers will have performance speeds at or above 10 to the 18^{th} power of floating point operations. That's fast. We are at a major inflection point in developing next generation, high end, exascale systems.

The United States must invest now to maintain its economic leadership and competitiveness. Government funding and domain skills, especially in our national labs, are required. Investment should be targeted at developing innovative and composable systems for modeling, simulation and analytics on big data. The new imperative is to design for data, not for processor performance. New systems and software concepts must be developed to continue U.S. strategic leadership and to support and enhance the economic competitiveness of U.S. Industry.

To put this in context, before 2005, the U.S. led decidedly in the global supercomputing race, with Europe and Japan following. Today, however, and in the future, additional regions are making important sovereign strategic investments to compete aggressively.

The U.S. is still ahead, but others are catching up fast. Japan is continuing its long term effort in high performance computing (HPC) as well, with a focus on building "Big Iron" systems. India is making a $1B+ investment to pursue exascale computing with an emphasis on applications and user capability. And in Europe, which has been a leader in software and applications for industrial use, we now see strong HPC support for small and medium enterprises. There is an aggressive effort to build a European capability in HPC with support for companies such as Bull and in complementary infrastructure efforts with a number of "Big Science" projects ($1B+) including Blue Brain, Graphene, Robotic, and Health Care.

In sum, the global race is on. The necessity for leadership in high end computing for a highly competitive economy has been recognized around the world and regions are stepping up to the plate.

Work Force Development

Skills are a major inhibitor to the growth of next generation computing and Big Data. Big Data requires new skills, knowledge and new types of decision-making. Today's employers are seeking job candidates who can analyze and build strategy around Big Data, or the 2.5 quintillion bytes of information generated daily. Indeed, the number one barrier to the adoption of business analytics technologies is insufficient skills and experience, according to IBM's 2012 Tech Trends Report.

Much has been written about the urgency of this skills gap. McKinsey Global Institute reports that over the next seven years, the need for skilled business intelligence workers in the U.S. alone will dramatically exceed the available workforce -- by as much as 60 percent. There is not a job crisis in the U.S.; there is a skills crisis. Nearly two million information technology jobs will be created by 2015 in the U.S. to support Big Data, according to research firm Gartner, Inc. Analytics skills will be a key differentiator for candidates seeking to fill those jobs.

Further, in our global society, applying analytics to Big Data will be a key factor in determining which countries pull ahead economically and which ones fall behind. Industry is responding to this situation with a variety of programs to build new skills. I will describe several which IBM has created. We are addressing skills requirements from grade nine through graduate school and into the working population around the globe.

In fact, just today IBM is announcing a new partnership to prepare business students for the expanding scope of careers requiring Big Data analytics skills. IBM and Rensselaer Polytechnic Institute are combining forces to offer a new Lally School of Management and Technology graduate degree program in fall 2013: the Master of Science in Business Analytics.

This new Master of Science in Business Analytics is a one year, 30-credit graduate degree program that will provide students and career professionals with the hands-on experience and knowledge required to succeed in analytics jobs. It will feature a three-part curriculum comprised of:

- A business core to ensure students understand where Big Data fits into a business' strategy and operations, as well as how analytics-driven decisions can impact a business' growth, competitive standing and bottom line.
- An analytics core that includes hands-on training in predictive modeling to help businesses identify profitable data patterns, focusing also on data management, statistical analysis and leading-edge techniques in harnessing Big Data.
- An experiential core with project-based courses that allow students to apply their newly gained skills to real-world problems faced by businesses spanning a range of industries.

This new collaboration builds on experience gained from our numerous other engagements. Some highlights of those are below:

Early STEM Development - IBM collaborates with the New York City Department of Education, the City University of New York, and NYC College of Technology to provide a single public school for grades 9-14 and a new public school program to be replicated in other schools. The school, called P-TECH (Pathways in Technology Early College High School) is preparing students to fill entry-level jobs in technology fields or provide them with the foundation for ongoing learning in a four-year college in the STEM disciplines.

Higher Ed - IBM has 200 academic partnerships focused on Big Data analytics with schools such as Yale, Northwestern and Michigan State universities (in addition to our 30,000 academic partnerships overall). The mission: develop Big Data analytics curriculums primarily in their business schools. IBM works closely with professors in support of curriculum materials and case studies and provides guest speakers and faculty awards to accelerate degree program development. A key example of this curriculum: Working with Fordham University in New York City, IBM partnered with faculty to create a Center for Digital Transformation and two degree programs in analytics: the Masters of Science in Business Analytics and Marketing Intelligence.

Advanced Analytics Center - IBM recently unveiled a first of its kind Advanced Analytics Solutions Center in Columbus, Ohio, which aims to create 500 new analytics jobs. The center will serve as an innovation hub to advance analytics skills, drawing on the expertise of local businesses, educational institutions and industry partners. IBM is partnering with Ohio State University to develop new analytics curricula at the undergraduate, graduate and executive education levels. The Columbus center is the most recent addition to our existing centers, which include both Dallas and Austin.

Watson Joins Rensselaer Research Team – IBM is providing a modified version an IBM Watson system to Rensselaer Polytechnic Institute, making it the first university to receive such a system. The arrival of the Watson system will enable new leading-edge research at Rensselaer, and afford faculty and students an opportunity to find new uses for Watson and deepen the systems' cognitive capabilities.

Watson Case Competitions – IBM works with professors and local businesses to create project-focused case studies for students to gain hands-on experience in Big Data, analytics and cognitive computing. For example, the University of Rochester's Simon School of Business launched the first-ever Watson academic case competition where MBA students identified critical areas in which Watson technology could be beneficial such as crisis management, mining and transportation. This was followed by competitions at Cornell University, the University of Connecticut and the University of Southern California, where more than 100 business and engineering students combined their skills to recommend new uses for Watson, including an innovation that helps doctors identify people who may be suffering from Post-Traumatic Stress Disorder.

Interning with Watson - This past summer, IBM brought students into its labs to learn about and develop applications for Watson's ground-breaking analytics technology. IBM Watson interns worked directly with clients and IBMers on real-world projects.

Professional Training - IBM actively trains current workers on Big Data skills. This past year, IBM held 1,200 Big Data boot camps at client, partner and university sites and trained over 2,400 IT professionals and students on the latest data management techniques; this year IBM launched BigDataUniversity.com to help students learn Hadoop, stream computing, and Big Data analytics skills. Over 13,000 students have enrolled over the past six months.

These collaborations between IBM and top U.S. universities are building a workforce of professionals and are creating jobs now. Consider North Carolina State University, whose Master of Science in Analytics program has generated the ultimate outcome – 90 percent of its graduates receiving job offers from data-hungry employers.

Privacy matters

Privacy must be considered in the design of Big Data systems. Importantly, realizing the promise of Big Data does not require the sacrifice of personal privacy. In many cases in Big Data projects, the data being aggregated is non-regulated, de-identified information with no need to re-identify to derive value. When personal information is used, organizational processes and technology can protect privacy.

Organizational practices include a systematic way of thinking and acting proactively and responsibly about the use of data. When organizations using personal information take privacy into account from the start and design in privacy protection practices, they act as better stewards of information and help individuals make more informed choices. For example, IBM practices "privacy by design" and has included privacy considerations as we have developed our new sense-making analytics technology. We welcome the growing ranks of organizations working to adopt this approach.

Furthermore, IBM and other companies are working with the Future of Privacy Forum on a consumer trust seal, authenticated by a third party, to give consumers confidence in smart grids. IBM believes that industry codes of conducts should promote transparency with customers. We have been urging better communication for a long time – as far back as 1999 we decided to withhold advertising dollars from North American websites that did not post their privacy policies.

Although these considerations are critical as technology is changing rapidly, a major trend in consumer expectations is also evolving. Indeed, Americans' views of online privacy are shifting as they use multiple social networking tools every day to describe their personal thoughts, actions and emotions very publicly, such as through online profiles, tweets, blog posts, and photos.

Our experience is that when people understand how information is used, have the ability to set data usage policies, and enjoy the benefits of the analysis, they tend to see a helpful tool rather than a privacy violation. Information – including personal information – is becoming a helpful tool on a grand scale. When combined in great quantities and put through sophisticated analysis, general information can be used to address some of society's most pressing problems. For example, the Center for Medicare and Medicaid Services uses big data in a new program to prevent health care fraud. A sophisticated and evolving set of algorithms is used to highlight problem areas and generate fraud alerts that allow the agency to direct attention and resources, and take appropriate action. Increasingly, CMS will be in a position to interrupt claims that should not be paid. The need to deter and detect is great, with health care improper payments totaling in the hundreds of millions of dollars.

What is the role of government in research?

Federal research investment in high performance computing is critical to Big Data. Research both creates new ideas and insights and trains students with critical skills for later employment. Industry needs fundamental, exploratory research as we push the boundaries of programmable systems with our high performance systems. We need research into numerous areas including: system design for optimized handling of the volume, velocity, and variety of data described earlier; software research to understand how to create dynamic and flexible software-defined environments; IT infrastructure research to build programmable, optimized and automated environments. New knowledge, practices, and infrastructures will enable the discoveries and innovation which are the foundation of U.S. competitiveness.

In addition, IBM strongly supports the reauthorization of the Department of Energy High-End Computing Revitalization Act of 2004 to be offered by Representative Hultgren. This bill will improve high-end computing research and development at the DoE and strengthen government-industry partnerships for integrated research, development and engineering of exascale platforms. IBM has a long history of successful partnership with DoE. Through our joint work, computational simulation has been established as an essential, broadly available tool for scientific inquiry. World leadership for the U.S. in HPC has been grown and sustained, and HPC has become a true engine for innovation. The challenge ahead is to continue this growth, recognizing that Big Data is an intrinsic aspect of high performance computing. There is not an "either/or" choice between HPC and Big Data.

The explosion of data is having a significant impact on the focus of HPC. HPC workloads are evolving from single "heroic" calculations to complex simulations with different time and space scales involving multiple scientific disciplines and several types of computational algorithms. These complex simulations interact with each other using mountains of data. Research is needed into systems which can integrate computation, data and storage to successfully exploit Big Data to address many enormously important scientific, social and commercial workloads.

Past federal investments in HPC-related research, particularly at DOE's national laboratories in partnership with the industry and academia, have underpinned mission critical supercomputers in many Federal agencies, including the Department of Energy, for both the Office of Science and the National Nuclear Security Administration; the Department of Defense; the National Aeronautics and Space Administration; the National Oceanic and Atmospheric Administration; and U.S. intelligence agencies.

What is the role of government in workforce development?

The Professional Science Masters program supported by National Science Foundation is particularly relevant to the new era of computing. A Professional Science Masters (PSM) degree is a new graduate degree designed to allow students to pursue advanced training in science or mathematics, while simultaneously developing workplace skills valued by employers. According to the Council of Graduate Schools, enrollment in PSM degrees increased 22% between 2010 and 2012 and was dominated by four fields of study: computer/information sciences (21%), biotechnology (16%), environmental sciences and natural resources (14%), or mathematics and statistics (14%).

Additionally, Congress should reauthorize the Carl D. Perkins Act and the Federal Work Study Program to align to labor needs in Big Data. While this is unlikely to fall in the Science Committee jurisdiction, it is important for the growth of STEM (science, technology, engineering and math) education and jobs.

In the U.S., community college graduation rates hover at or about 25 percent. At the same time, there are 28 million middle skill jobs – those requiring postsecondary degrees – currently available in the U.S. which pay close to $40,000 per year on average. Over the next 10 years, 14 million jobs – a 50 percent increase – will be created for students with "middle skills."

With so many high school graduates in need of remediation rather than prepared with the skills and credentials needed to fill these jobs, we must refocus our efforts on strengthening the education system in order to make the U.S. more competitive.

In the U.S., career and technical education (CTE), once called vocational education, is the core program linking school to career. Federal funding under Perkins provides over $1 billion to schools. Unfortunately, Perkins is not currently structured to meaningfully address the skills gap that we face.

Similarly, the federal government invests nearly $1 billion in the Federal Work Study Program (FWSP) every year, providing nearly 800,000 undergraduates the opportunity to gain work experience while earning critical financial aid. To maximize its potential, FWSP should not be viewed solely in the lens of college affordability, but also as a way to prepare students to be successful in a career.

If 10 percent (or $200 million) of the $2 billion invested in these two programs annually were targeted in new ways, CTE programs could be reshaped to help significantly bridge the skills gap. For example, over a 10 year period, a new skills-based apprenticeship program could better prepare more than 1,000,000 young people for the jobs of the 21st Century and reinvigorate the American economy.

IBM supports reauthorizing Perkins funding to align to labor market needs in high-growth industry sectors; improve CTE programs with strong collaborations among secondary, postsecondary institutions and employers (like the P-TECH example cited earlier); and create accountability measures that provide common definitions and clear metrics for performance of CTE programs to improve academic outcomes while building technical and employability skills of participants.

Conclusion

There exists today an overabundance of data. Leveraging the capabilities presented by this new era of cognitive computing presents us with the opportunity to provide benefit in many areas. It will be a key factor in determining which countries pull ahead economically and which fall behind, which cities attract knowledge workers and business development and which businesses thrive. I believe in the value and power of information and technology. The United States should continue and increase its support of advanced computing and its investment in building its workforce to seize the value that Big Data, Analytics and next generation computing offer.

Thank you for the opportunity to appear before you today to provide this testimony. I welcome your questions.