

## IBM Research to Accelerate Big Data Discovery

**New lab unifies data, expertise and novel analytics to speed discovery in industries including retail, medicine and finance**

PR Newswire  
SAN JOSE, Calif.

SAN JOSE, Calif., Oct. 10, 2013 /PRNewswire/ -- Scientists from IBM (NYSE: [IBM](#)) today announced the [Accelerated Discovery Lab](#), a new collaborative environment specifically targeted at helping clients find unknown relationships from disparate data sets.

(Logo: <http://photos.prnewswire.com/prnh/20090416/IBMLOGO>)

The workspace includes access to diverse data sources, unique research capabilities for analytics such as domain models, text analytics and natural language processing capabilities derived from Watson, a powerful hardware and software infrastructure, and broad domain expertise including biology, medicine, finance, weather modeling, mathematics, computer science and information technology. This combination reduces time to insight resulting in business impact -- cost savings, revenue generation and scientific impact -- ahead of the traditional pace of discovery.

The notion of Moore's Law for Big Data has less to do with how fast data is growing, and more with how many connections one can make with that data, and how fast those connections are growing. While companies could utilize data scientists to analyze their own information, they may miss insights that can only be found by bringing their understanding together with other experts, data sources, and tools to create different context and discover new value in their Big Data.

"If we think about Big Data today, we mostly use it to find answers and correlations to ideas that are already known. Increasingly what we need to do is figure out ways to find things that aren't known within that data," said Jeff Welser, Director, Strategy and Program Development, IBM Research Accelerated Discovery Lab.

"Whether it's through exploring thousands of public government databases, searching every patent filing in the world, including text and chemical symbols, to develop new drugs or mixing social media and psychology data to determine intrinsic traits, there's a big innovation opportunity if companies are able to accelerate discovery by merging their own assets with contextual data."

With much of today's discovery relying on rooting through massive amounts of data, gathered from a broad variety of channels, it is painful for many businesses and scientists to manage the diversity and the sheer physical volumes of data for multiple projects and to locate and share necessary resources and skills outside their organizations.

Leveraging the best research and product technologies for analytics on a scalable platform, the Accelerated Discovery Lab empowers subject matter experts to quickly identify and work with assets such as datasets, analytics, and other tools of interest relevant to their project.

At the same time, it encourages collaboration across projects and domains to spark serendipitous discovery by

applying non-proprietary assets to subsequent projects. This collaboration can occur whether the experts are co-located in the same physical location or are geographically distributed but working within the same system infrastructure.

"The history of computing shows that systems commoditize over time," said Laura Haas, IBM Fellow and Director, Technology and Operations, IBM Research Accelerated Discovery Lab. "Moving forward, people and systems together will do more than either could do on their own. Our environment will provide critical elements of discovery that allow domain experts to focus on what they do best, and will couple them with an intelligent software partner that learns continuously, increasing in value over time."

## **Industry Applications**

**Drug Development:** The process of drug discovery today spans an average of 12 to 15 years, with billions of dollars invested per drug, and a 90+% fallout rate. Working primarily with pharmaceutical companies, IBM Research is using machine-based discovery technology to mine millions of published papers, patents and material properties databases. Then using advanced analytics, modeling and simulation to aid human discovery, IBM is able to uncover unexpected whitespace and innovation opportunities, and predict where to make the most profitable research bets. The inability to discover the next "new thing" quickly is a huge shortcoming faced by companies today across multiple industries including retail, medicine and consumer goods. A diverse set of skills and tools were needed to integrate and analyze these many sources of data, from deep domain knowledge of chemistry, biology and medicine, to data modeling and knowledge representation, to systems optimization. The data sets, skills and infrastructure provided by the Accelerated Discovery Lab not only enabled this work, but also are allowing the re-use of the tools in domains from materials discovery to cancer research.

**Social Analytics:** Marketers gather terabytes of data on potential customers, spend billions of dollars on software to analyze spending habits and segment the data to calibrate their campaigns to appeal to specific groups. Yet they still often get it wrong because they study "demographics" (age, sex, marital status, dwelling place, income) and existing buying habits instead of personality, fundamental values and needs. Recognizing this, scientists at IBM Research are helping businesses understand their customers in entirely new ways using terabytes of public social media data. They are able to understand and segment personalities and buying patterns from vast amounts of noisy social media data and do so automatically, reliably and after as few as 50 tweets. This is data that marketers never had before, permitting much more refined marketing than traditional approaches based on demographics and purchase history alone. The Accelerated Discovery Lab brought together the expertise in text analytics, human-computer interaction, psychology and large-scale data processing to enable these new insights. Because clients from multiple industries including retail, government, media and banking are exploring different applications of social analytics in this common environment, the opportunities for unexpected discoveries abound as new analytics are applied to diverse challenges.

**Predictive Maintenance:** Natural resources industries, such as oil and gas, mining and agriculture, depend on the effectiveness and productivity of expensive equipment. Most maintenance processes result in costly in-field failures, which can cost a company \$1.5M for one day of downtime on a single piece of equipment. In order to have a real bottom-line impact, analytics and modeling need to be integrated with current processes. IBM developed an intelligent condition monitoring technology using the most comprehensive data set ever assembled in this domain. This system proactively presents decision support information to drive actions that reduce downtime, increase fleet productivity, and minimize maintenance costs – in fact, one estimate suggests

that a \$30B company can save \$3B a year by implementing predictive maintenance technology. The Accelerated Discovery Lab brought together experts in the domain, the systems and mathematical modeling and provided systems infrastructure and expertise that freed the domain researchers and mathematicians to focus on the client problem and sped up the execution of the resulting models by a factor of 8.

### **About IBM Big Data and Analytics**

Each day we create 2.5 quintillion bytes of data generated by a variety of sources -- from climate information, to posts on social media sites, and from purchase transaction records to healthcare medical images. At IBM we believe that Big Data and analytics are a catalyst to help clients become more competitive and drive growth. IBM is helping clients harness this Big Data to uncover valuable insights, and transform their business. IBM has established the world's deepest and broadest portfolio of Big Data technologies and solutions, spanning services, software, research and hardware. For more information about IBM and Big Data and analytics, visit <http://www.ibmbigdatahub.com> Follow IBM and Big Data on Twitter @IBMbigdata and #ibmbigdata.

Contact:

Ari Entin

408-927-2272

[aentin@us.ibm.com](mailto:aentin@us.ibm.com)

SOURCE IBM

Web Site: <http://www.ibm.com>

---

<https://newsroom.ibm.com/2013-10-09-IBM-Research-to-Accelerate-Big-Data-Discovery>