MEANS BIG POTENTIAL FOR LSU

Everything we do generates data these days. When you use a search engine like Google or Bing, you leave behind data. Your purchasing habits are recorded as data. Clinical studies run by universities and biomedical research institutions create masses of valuable data. Weather forecasts leave behind waves of data. Even your annual checkup at the doctor’s office creates data.
What if all of this information from many multiple sources could be synthesized and analyzed for overarching themes, trends, or anomalies? What if, instead of looking at a sample group of 100 or even 1,000, you could gather data for tens of thousands—even millions? We’ve already developed vaccines, launched satellites, and created supercomputers. What could we do next if researchers had a better way to look at the massive amounts of data just lying around?

That’s where “big data” comes in.

“Big data is an all-encompassing term that refers to the handling of vast amounts of data,” explained LSU Associate Vice Chancellor of Research & Economic Development K. Gus Kousoulas. “It allows for storage and mining of data that enables more informed strategic decision-making at the highest levels.”

That’s why the movement is so popular. There are big data consortiums, big data conferences, and a rash of big data grants. The term is peppered throughout speeches and applied liberally to proposals, presentations, and meetings. But, despite its popularity, there are many of us who feel a little left in the dark when it comes to big data applications, both current and projected.

**LSU’s Big Data Foundation**

Big things are happening in big data at LSU. There’s never been more excitement surrounding computer science and its applications in the capital city region and beyond. Last year saw the announcement of an innovative public/private partnership between IBM, LSU, and Louisiana Economic Development, which spurred enthusiasm in response to the creation of 1,342 new, permanent jobs in the region as well the opportunity IBM’s support center symbolizes to Louisiana citizens. LSU designed a streamlined computer science degree program to prepare students for the new jobs the center will create, offering a fast-track to higher education—and a job—for Louisianans.

“LSU’s Computer Science faculty have embarked on an unprecedented initiative to create a program that will serve as a major force in positioning Louisiana as one of the most tech-savvy states in the nation,” said LSU College of Engineering Dean Richard Koubek. “We’ve not only produced a streamlined, three-year degree program, but introduced concentrations in cloud computing and data analytics, making LSU one of only a handful of universities offering these studies at the undergraduate level.”

While all these advances are great accomplishments, the work isn’t done, and things aren’t slowing down at LSU. In February 2014, more than 70 representatives from IBM and LSU gathered in the new Digital Media Center—another example of a successful public/private investment focused on digital technology and supercomputing—for a summit on focal research areas and major developments that could spur future big data collaborations.

“Supercomputing and high-performance computing are key research areas at LSU,” said Vice Chancellor of Research & Economic Development Kalliat T. Valsaraj. “Working with industry giants like IBM is the perfect complement to our faculty’s basic and applied research. We’re looking forward to expanding this collaboration to the benefit of our researchers, our students, and our state.”

High-performance computing and big data capabilities also serve as the backbone of all of LSU’s other research strengths. It is necessary to study coastal processes and develop better, more integrated methods of hurricane protection. It is a driving force behind the search to find safer, more efficient ways to extract energy from the Earth and the development of new alternatives to fossil fuels. And it is vital to identifying better methods to treat disease and illnesses and to determine patterns of healthcare outcomes in our citizenry.

“The capacity is here, and this movement is only going to gain momentum in the days and years to come,” said LSU President and Chancellor F. King Alexander. “Just the very fact that so many of us from LSU and IBM came together to identify new ways to work together is proof that we are all committed to collaborations that will have real impact and provide real solutions for Louisiana.”

**LSU’s Big Data Needs**

One of the challenges Louisiana—and many other states face is that every important industry generates massive amounts of data that require extensive amounts of storage. Supercomputing and big data capacity is needed not only to support the growing needs of industry, but also to explore data-sharing in fields that could benefit from cross-referencing information. For example, LSU Health Sciences Centers in Shreveport and New Orleans generate large amounts of bioinformatics data, which is stored at their individual locations. Meanwhile, the LSU School of Veterinary Medicine and the Pennington Biomedical Research Center likewise produce enormous amounts of biomedical data. With the proper connections and support, researchers across LSU’s campuses could combine their data and expertise using the Louisiana Optical Network Initiative (LONI), which is the fiber optics network connecting Louisiana research universities to one another, as well as to Internet2 and the National LambdaRail, to look at healthcare outcomes, for instance, in an entirely new way.
One such example is genome sequencing, which involves determining the exact sequence of an organism’s DNA and has many applications in biological and medical research. However, it requires expansive amounts of data processing.

“The human genome contains three billion molecular units. Assembling and analyzing this amount of data, or even assembling and analyzing many much shorter genome sequences like those of the West Nile or AIDS viruses or pathogenic bacteria that cause acute human disease, will require massive computational power and data storage capabilities, such as those produced through big data networks,” Kousoulas said.

Researchers at LSU are already investigating a cyberinfrastructure to satisfy these big data demands through a $1 million grant from the National Science Foundation. Seung-Jong Park, associate professor of computer science with a joint appointment in the Center for Computation & Technology (CCT), along with co-investigators Joel Tohline, Sean Robbins, Lonnie Leger, Gus Kousoulas, and other senior LSU faculty are involved in a campus-wide project aimed at bringing big data computational capabilities to various university research groups.

“Processing terabytes of data, such as a metagenome data set, has been a headache for researchers using their own equipment. But that kind of problem can be solved with our technology based on Hadoop and cyberinfrastructure that includes a 10Gbps high-speed network and solid-state drive storages,” Park said. “We can help researchers transfer their data to, and analyze that data on, bigger machines, not just one or two computers.”

While Park’s group focuses on the logistics and technical sides of such connectivity, work in specific fields—such as biomedical, environmental and coastal sciences, materials and manufacturing, physics and astronomy—is already making use of existing capacity.

Coastal Issues

A state like Louisiana faces many unique geographical challenges, such as threats from hurricanes and coastal erosion. As a hurricane forms and begins moving toward land, emergency officials in coastal states tap into groups like those at LSU for the most up-to-date modeling, tracking, and monitoring systems available so that they can have the best data to inform their decisions. LSU has significant expertise in these areas—expertise necessary to allow government emergency preparedness officials to make the right call as early as possible for public safety.

One example of the impact big data can have on public safety is the Coastal Emergency Risks Assessment (CERA) interactive website, a collaboration between the supercomputing power of LSU’s CCT, LSU’s School of the Coast & Environment, and the Louisiana Sea Grant.

“CERA makes use of the ADCIRC Coastal Circulation and Storm Surge Model, which takes an incredible amount of supercomputing power to generate real-time hurricane models,” said Robert Twilley, director of Louisiana Sea Grant and CERA project manager. “During the last hurricane to make landfall in Louisiana, Hurricane Isaac in 2012, our projections were the only ones that identified its specific path accurately.”

The ADCIRC model used by the CERA project is impressive because of its speed and detail. When Hurricane Katrina made landfall in 2005, computer models used about 300,000 nodes—locations where the computer makes calculations to determine how water levels will change during a storm—and took four hours to run.

More nodes result in more detail. For example, a typical grid used for modeling employs a million nodes. During Hurricane Isaac, 1.1 million nodes were used on three different storm tracks, and the simulations were completed in two hours using a suite of high-performance computers.

“State emergency officials start the decision-making process five days before landfall and need constantly updated models throughout a weather event,” said Twilley. “Our software was built to configure big data and turn it into a useable framework for quick evaluation by these officials, who literally are responsible for people’s lives during a hurricane.”

Since Katrina, Louisiana has had four hurricanes make landfall, and the system has seen some failures and some successes.

“That’s what we built this system for,” said Carola Kaiser, an IT consultant at LSU who built and now operates CERA. “We knew it would be a trial-and-error approach, but in the end, there’s no way to get accurate models to provide data driving the most informed decision possible without supercomputers and big data.”

The Future of Medicine

LSU has extensive biomedical expertise. One of its most unique assets is its connection with the Mary Bird Perkins Cancer Center (MBPCC). In 1980, this partnership resulted in the development of LSU’s medical physics program, the only program of its kind in Louisiana and one of only 40 in the nation accredited by the Commission on Accreditation of Medical Physics Educational Programs, Inc. The program leverages Mary Bird Perkins’ clinical team and facilities, treatment planning, and dosimetry laboratories with LSU’s expertise in imaging and medical physics within the College of Science.
We can help researchers transfer to, and analyze their data on, bigger machines, not just one or two computers.

— Seung-Jong Park
LSU professor of computer science
One of the (albeit many) challenges facing those tasked with developing treatment regimens for cancer patients is that, technically, cancer isn’t a single disease. It’s hundreds, joined together by one common reality—uncontrollable cell growth.

“Traditionally, cancer treatment options are based off of large-scale clinical trials,” said Wayne Newhauser, LSU’s Dr. Charles M. Smith Professor of Medical Physics and Chief of Physics at MBPCC. “So basically, you’re basing treatment on averages. Modern medicine is moving more toward personalized medicine, which requires big data in order to process and analyze your DNA and genetics.”

In other words, big data can provide individuals with the lowest-risk, highest-return type of treatment available based on a person’s specific genetic makeup.

“With the right kind of capabilities, we will be able to use imaging studies to build a genomic profile at even a basic physician’s appointment,” said Newhauser. “Some of the world’s leading institutions already have this capability. We’re only a few years away from it ourselves.”

Using information to diagnose and develop treatment options is known as bioinformatics, and it’s an up-and-coming field of study that stands to revolutionize the way we look at healthcare options.

“We’re not there yet in terms of applying all this information to the average cancer patient,” said Newhauser. “But we have the clues, and we know the pathways. This partnership between LSU and Mary Bird Perkins Cancer Center links Louisiana citizens with world-class treatment, university supercomputing and biomedical expertise, and clinical trials to bring us one step closer to the holy grail of treatment.”

In the Sky

One relatively unexpected way that big data is making an impact is in the field of astrophysics.

LSU is involved in the Laser Interferometer Gravitational-wave Observatory (LIGO) project, and in fact one of only two such observatories in North America is located in Livingston, Louisiana, only 30 miles away from campus. This group has been attempting to record the existence of gravitational waves for years.
“According to general relativity, gravitational waves are ripples in space-time that are produced when massive astronomical objects suffer violent processes, like black hole collisions,” said Gabriela Gonzalez, LSU professor of physics and astronomy and LIGO Scientific Collaboration spokesperson. “These ripples travel almost unperturbed through the universe, and when they pass through interferometers, they affect the fringes in them. Being able to view the universe through these ripples of space-time will open a complete new window to the universe.”

The entire LIGO system recently underwent upgrades to enhance sensitivity of measurement devices and won’t be brought back online until 2015. But it is expected that, when it is up and running again, the acute nature of its new mechanisms will record many astrophysical occurrences—far too small to detect using traditional tools—hiding in large amounts of data.

“The expectations are that Advanced LIGO will result in terabytes of data, with only intermittent bits relevant to the rare events our astronomers hope to better understand,” said Michael Cherry, LSU chair of the Department of Physics & Astronomy. “There’s significant difficulty in extracting real science from extremely large data sets, and it requires true big data capabilities.”

The New Worlds, New Horizons in Astronomy and Astrophysics Decadal Survey, which was commissioned by the National Research Council, identified the two highest-priority facilities in astrophysics as the Wide-field Infrared Survey Telescope (WFIRST) and the Large Synoptic Survey Telescope (LSST). These represent space and land missions, respectively, each of which will involve data sets on the petascale level.

“What this means is that astrophysicists can no longer be astronomers alone, but also have to be well-trained and exceptionally competent data scientists as well,” Cherry explained. “So, big data isn’t just a scientific tool … it’s a workforce development tool as well.”

While all these fields are still emerging and it’s too early to tell what the future will truly hold for big data, the implications are clear, and LSU has built a strong foundation in the area—one that will certainly yield impressive results in due time.