Purdue University

Located in West Lafayette, Indiana, in the United States, Purdue University is one of the country's most prestigious public universities. Founded as a land-grant college in 1869 under The Morrill Act, a program championed by President Abraham Lincoln, the university delivers a world-class education to more than 37,000 undergraduate and 12,500 graduate students across several campuses.

Industry

Education

VMware footprint

• VMware® Greenplum®

Purdue University Uses Big Data to Defeat Pandemic Shutdowns

Based in West Lafayette, Indiana, Purdue University is one of the largest public universities in the U.S., delivering undergraduate and graduate instruction to more than 49,000 students. Following the onset of the COVID-19 pandemic, Purdue University used VMware Greenplum, a big data analytics and warehouse solution, to create a powerful contract tracing system that helped the university bring students back to campus for a safe, in-person learning experience.

## Delivering excellence and equity in education

Since its founding in 1869 under the public land-grant college system signed into law by President Abraham Lincoln, Purdue University has become one of the most prestigious public universities in the United States. From its humble beginnings—the inaugural class of 39 students met in 1874—the university today welcomes more than 37,000 undergraduate and 12,000 graduate students to its network of campuses.

Highly respected for colleges of agriculture, engineering, business, pharmacy and veterinary medicine, Purdue University offers a well-rounded liberal arts education to thousands of students from diverse socio-economic backgrounds. Deeply committed to making higher education affordable, the university has, for the past decade, frozen tuition at 2012-2013 rates, saving students more than USD $1 billion compared to the national average for college costs.

"We work every day so students can have a more affordable education, so we can be good value for them," says Ian Pytlarz, lead data scientist, Institutional Data Analytics + Assessment, Purdue University.

> "We were already using VMware Greenplum as a gray box for all sorts of data at the university. We used it to make a contact tracing system that considers all the wireless access point logs, to see where and when students are co-located for long periods of time."
>
> Ian Pytlarz, Lead Data Scientist, Institutional Data Analytics + Assessment, Purdue University

vmware®

WATCH HOW PURDUE UNIVERSITY USES BIG DATA TO DEFEAT SHUTDOWNS
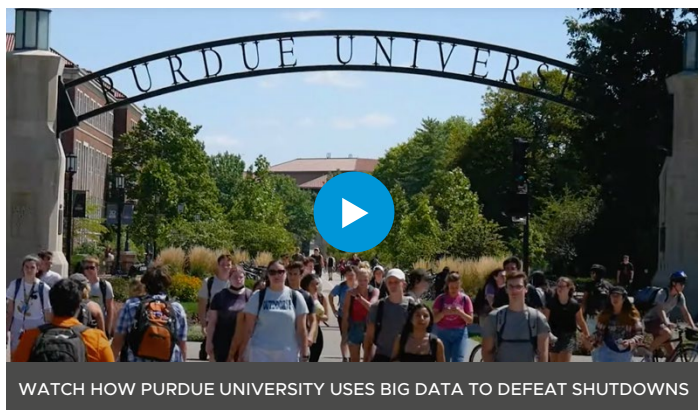
Confronting difficult choices with data science

The university's Institutional Data Analytics + Assessment (IDA+A) department, part of the office of the provost, performs research, statistical and predictive analysis to support evidence-based decisions on everything from enrollment to retention, academics to campus operations.

When the COVID-19 pandemic turned higher education upside-down in March 2020, universities and colleges across the country closed their doors, sent students home, and shifted to virtual learning. "Like everyone else, we shut down in March 2020 for the rest of that spring semester. When we shut down, I knew how serious it was," says Pytlarz.

As the months dragged on, university administrators understood that to support high standards for educational excellence, Purdue needed to bring students back to campus for safe in-person learning. As lead data scientist at IDA+A, Pytlarz knew data was the key to creating a safe environment. As soon as the university closed its doors, Pytlarz and his data science team got to work. "I knew we needed to do something big, or this was going to be a mess," says Pytlarz. "And I would say the vast majority of the university reacted the same way. Everybody was very proactive, part of a vast effort to keep students in-person and safe."

Mining big data to ensure student safety

The uncertain trajectory of the pandemic made it difficult to know how—or how long—COVID-19 would affect the university. When students arrived on campus in January 2020, no one had heard of the virus.

Johns Hopkins University & Medicine reports that by the end of May, COVID-19 cases had skyrocketed across the U.S. to more than 1.8 million cases and more than 100,000 deaths. The pandemic was the definition of a worst-case scenario.

With university leadership determined to avoid a long-term shutdown, Pytlarz felt the only way forward was to devise a bold, creative solution. "When we shut down, I began work developing what would become one of the most sophisticated digital contact tracing tools in the world," says Pytlarz.

The volume of data Pytlarz and his team needed to analyze was gigantic. "We have access to virtually everything," says Pytlarz. "We can view grades, who's in class and when, class schedules. We have every card swipe in every door, every dining swipe, every gym swipe, fraternity and sorority housing records and wireless transactions on our Wi-Fi network."

There was no shortage of data—universities are awash in information—but the team needed a powerful analytics tool to analyze the data to map student behavior and understand patterns of contact. Pytlarz theorized that the university could mitigate the threat by understanding who spent time with whom, and where, and accurately predict where virus transmission might occur.

"We were already using VMware Greenplum as a gray box for all sorts of data at the university," says Pytlarz. "We used it to make a contact tracing system that considers all the wireless access point logs, to see where and when students are co-located for long periods of time. This meant that our executive teams and our medical teams had access to all the data they could ever want."

> "We work every day so students can have a more affordable education, so we can be good value for them."
>
> Ian Pytlarz, Lead Data Scientist, Institutional Data Analytics + Assessment, Purdue University

Data-driven, safety-forward learning

The university launched its proximity contact tracing system in August, just as students returned for the Fall semester.

Because the system's success relied on the collection and analysis of personal data, full transparency around confidentiality was critical. "There were enormous privacy implications, so we were very open and honest throughout," says Pytlarz.

Advised by privacy experts on the computer science faculty, Pytlarz and his team used only enough information for the medical staff to do their jobs while protecting individual privacy. "We explained what we were doing with the data and why it would benefit students to opt in," says Pytlarz. "And because we have very strong data governance, we were able to make well-informed ethical decisions about data use.

"When we got hourly COVID-19 test results from our onsite testing facility, we fed them into our system, which extracted every significant contact and sent that information to the medical team, all within five minutes of receiving a positive test," says Pytlarz. "The medical team always had the most up-to-date information on who was at risk, who needed to be tested and who needed to be pulled out of their dorm and put into isolation housing."

Using VMware Greenplum, the team created reporting tools to summarize, in real or near-real time, COVID-19 infection numbers for each residential area within the university. The team shared this data daily with the Purdue University leadership team to ensure that all administrative decisions were data-driven and grounded in sound science.

Based on open source technologies PostgreSQL and Greenplum Database, VMware Greenplum is a massively parallel data platform (MPP) which enables organizations to ingest big data sets and perform deep analytics—using machine learning and AI—at extreme scale and speed.

This unprecedented data gathering project has recorded a staggering 4.9 billion potential contacts since the beginning of 2020.

"We use Greenplum for everything," says Pytlarz. "It's one big box, and we just throw all the stuff into the box. And if you have a wild question that's based on data that's already in there, we can start working on it tomorrow. There's no rigmarole. There's nothing holding us back."

According to the National Conference of State Legislatures, more than 1,300 colleges and universities across the U.S. remained closed for the rest of the 2020 academic year. Some reopened in the Fall with hybrid modes of in-person and virtual instruction, interrupted periodically by shutdowns ignited by spikes in infections. But Purdue University used successful contact tracing to keep its campus fully open. "After August 2020, we never shut down. We never had a major outbreak, and we were able to stay in-person, which is a value proposition that our students really loved," says Pytlarz.

Charting a bold future with in-house knowhow

By building on in-house data science knowhow, Purdue University protected in-person learning, continuing its commitment to equity in higher education. And the promise has paid off. "We've seen unprecedented interest from students interested in attending Purdue," says Pytlarz. "This is in part because we took the challenge seriously to be open, in-person and safe. We put our money where our mouth is and invested heavily to make that happen."

What does the future hold? "We are working actively with VMware to develop new machine learning methods that will work natively in Greenplum," says Pytlarz. "And that will enable us to work more quickly, because if VMware writes the custom code for us, we can just focus on the data science."

As the pandemic winds down and enrollment steadily rises, Purdue University data scientists continue to envision new ways to improve the student learning experience. To say the project has been popular is an understatement. "COVID-19 really broke down a lot of barriers, and everybody learned to do something new and interesting," says Pytlarz. "Now, there are fewer barriers to innovation across the university, and we're seeing a lot of success because everyone is on that same page. So, let's get creative, let's do something weird and interesting, and let's change the world."