Resultant

MANAGEMENT
PERFORMANCEHUB

Enhanced Research Environment Unlocks Secure Collaboration and Machine Learning for Indiana Management Performance Hub

> ❝
> We came to the realization that we needed to start moving data to the cloud to enable our researchers to have all the data and tools they need in one place."

OWEN BOBERG, PhD
MPH DATA SCIENTIST

BACKGROUND

For most state governments, measuring the long-term impact of programs intended to benefit citizens is challenging. Siloed, heavily protected data makes it nearly impossible to know whether participants in the Women, Infants and Children (WIC) program, for example, achieved intended outcomes like longer gestation periods and lower infant mortality. Often, private sector organizations—including grocery vendors and distributors—assist in delivering services for these programs but lack the context to fully understand program outcomes.

The Indiana Management Performance Hub (MPH) was created in 2014 to address this need. MPH's mission is to facilitate data-driven decision making and data-informed policy making, helping agencies manage performance; saving Indiana residents' lives, time, and money; and facilitating governmental and non-governmental data partnerships.

THE SITUATION

Granting third-party researchers access to data is not without drawbacks. When sharing files that recipients download locally, sending a researcher a particular dataset granted that researcher permanent and irrevocable access. Additionally, researchers had no way of knowing when a particular dataset was out of date. In order to maintain the security of sensitive datasets, MPH would require researchers to undergo a thorough vetting process that typically took several weeks to complete.

"We're seeing that, across all realms of research, datasets are either getting too big for people to do research on their own devices, or in our case, the datasets are too big and they're too sensitive to be downloaded locally," said Owen Boberg, MPH data scientist. "We came to the realization that we needed to start moving data to the cloud to enable our researchers to have all the data and tools they need in one place."

In April 2019, MPH cast a vision for a "virtual data warehouse" containing "rooms" where researchers and analysts could meet to collaboratively share and analyze data and called upon Resultant to bring that vision to fruition.

THE RESULTANT APPROACH: LEADING WITH EMPATHY

Resultant sought to deeply understand what would most benefit the researchers who used the platform as well as MPH's security concerns and desired use cases. Our team developed user stories by interviewing external researchers and asking questions like, "What are you hoping to gain from a research environment?" and "What tools do you need within the platform?"

Resultant then compiled the key requirements for the environment, including:

1. Prevention of unauthorized access to data by segregating an individual's access to specific projects within the environment

2. The ability to monitor user behavior and use of the environment

3. Enhanced transparency surrounding the analysis and use of sensitive data, including administrative oversight and approval of any exports from the environment

4. Joint analysis of data from multiple sources by multiple analysts in a secure manner

5. Not only facilitation but acceleration of advanced analysis through best-in-class data science and analytical tools

6. Secure sharing of analysis and code

7. The ability to scale to a nearly limitless number of users and volume of data by dynamically flexing to meet computational and user needs in near real time

Once all parties were aligned on requirements for the environment, the project team quickly discovered that no existing solutions met MPH's requirements. Fulfilling the needs of MPH and best serving the state required custom development.

> Evidence-based policies informed by data and research is an essential part of ensuring the health and well-being of Hoosiers. The Enhanced Research Environment revolutionizes sensitive data-sharing and analysis for the State of Indiana and its academic partners. It has unlocked secure collaboration and facilitated complex data analysis between the state and experts in the research community."

CONNOR NORWOOD, PhD, MHA
CHIEF DATA OFFICER OF INDIANA FAMILY AND SOCIAL SERVICES ADMINISTRATION
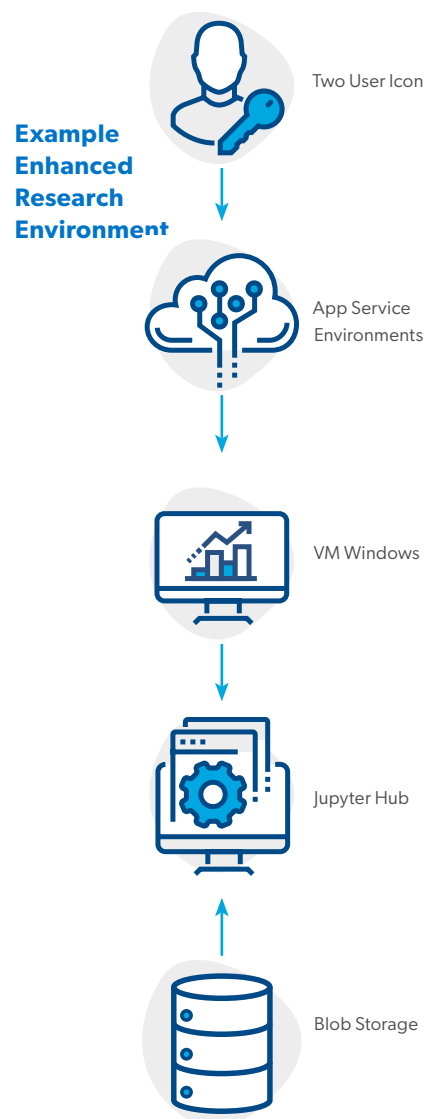
THE SOLUTION

Resultant embarked on building what would come to be known as MPH's Enhanced Research Environment (ERE) in a highly agile, iterative fashion. The team took an infrastructure-as-code approach to creating an elastic cloud environment largely built on Kubernetes.

Key features include:

- Windows-based desktop application containing Linux-based open-source tools

- Ability to be supported on multi-cloud environments so it can be stood up on AWS or Azure if needed

- More than 10 ready-to-use open-source data tools, including Jupyter Notebooks and R Studio

- Flexibility and capability for "bring-your-own-license" (BYOL), so data users can install familiar tools they would typically use for analysis, including SAS software, SSPs, IBM, Power BI, and Tableau, and Microsoft Office products

- Code repository so analysts can iterate on and keep track of their code as they develop

- Ability to support an unlimited number of users and projects

ERE was deployed on March 19, 2020.

"We initially rolled out ERE to just a handful of users," said Boberg. "Resultant helped us gather their feedback and then translate user stories into a meaningful backlog of features that would be useful to help researchers get the most out of ERE while making sure that it was fulfilling MPH's vision."

Example Enhanced Research Environment

Two User Icon

App Service Environments

VM Windows

Jupyter Hub

Blob Storage

THE OUTCOME

MPH deployed ERE only 14 days after Indiana confirmed its first case of COVID-19. ERE would prove crucial to the state's COVID-19 response, for the same reasons it would revolutionize sensitive data-sharing and analysis for the State of Indiana: it unlocked speedy and secure collaboration while facilitating complex modeling and machine learning.

Before ERE, the state had no centralized location where data could have been quickly consumed and analyzed, which would have made facilitating Indiana's COVID-19 response much more difficult. If data-sharing had been relegated to email, most of the computation would have been pushed to local environments, resulting in less secure, less accurate, and much slower analysis. The ability to quickly engage third-party partners like Regenstrief Institute, Indiana Health Information Exchange (IHIE), and Indiana University researchers was also critical to the speed of the response in the early days of the outbreak when every minute mattered. ERE enabled analysts and researchers to access critical datasets within hours rather than weeks.

Today, the COVID-19 project within the ERE has more than 90 concurrent users from organizations including Indiana State Department of Health, Indiana University, and Regenstrief Institute, far surpassing unique user key performance metrics.

> ERE has created a new ecosystem where people can just come and do meaningful work. They don't have to worry about getting data sent to them or building their tools, they just log in and hit the ground running to produce meaningful results for their communities and Hoosiers across the state."

OWEN BOBERG, PhD
MPH DATA SCIENTIST

COMPLEX INFECTIOUS DISEASE MODELING AND MACHINE LEARNING

To inform the COVID-19 response, analysts used natural language processing to engineer features from doctors' handwritten notes to determine how to decouple the pandemic signal from the endemic signal. To understand the endemic signals, the analysts characterized influenza-like illnesses from the chief complaint narrative and the syndromic surveillance system to model epidemic and endemic signals separately. ERE provides access to both the statistical programming languages and GPU processing power needed to facilitate machine learning.