



Altius Institute Accelerates Medical Breakthroughs with Igneous Data Protection as-a-Service

ALTIVS
INSTITUTE

Protecting and managing enormous datasets was an increasingly urgent problem for the Altius Institute for Biomedical Sciences, where their data is at the core of advancing discoveries that save lives. Legacy backup tools proved too expensive due to their large infrastructure and IT resource requirements, and just couldn't meet the scale requirements, leading Altius to choose Igneous for its scalability, simplicity, and long-term data management and distribution solutions.

Data Value and Growth Trigger Move from Replication to Data Protection with Igneous

Altius Institute for Biomedical Sciences, an independent, nonprofit research organization dedicated to pursuing discovery at the leading edge of modern biomedicine, has an ambitious mission: to create a new paradigm for catalyzing ground-breaking biological innovation, integrating molecular and computational science and engineering, and empowering fundamental technology development to radically accelerate the leap from basic to medical breakthroughs. While the physical plant for Altius is located in Seattle, its staff of 80 or so researchers and collaborators are located around the globe.

"What our lab does is take tissue and we turn it into data. Everything the lab does is about looking inside the nucleus of cells and trying to understand what's happening to molecules that are smaller than the wavelength of light," says Michael Cockrill, Chief Technology Officer at Altius.

The organization is driven by two sets of priorities: expand the body of knowledge and science in the field of genomics; and collaborate with leading pharmaceutical companies to better understand the influence of epigenetics on disease, particularly as it relates to new drug targeting and gene therapy. Altius' epigenetic approach to research enables them to more rapidly identify the right working sets of data, which then provide their pharmaceutical partner with the information they need to come to market with new drugs more quickly and effectively.

"The specific thing we focus on is DNA editing—changing the structure of the nucleus so cells act a different way. DNA is essentially a recipe for producing proteins," says Cockrill. "By changing the recipe, we can see if the proteins they produce are the ones that cause cancer, or if they stop causing cancer. Once the DNA has been modified, you measure the changes using genetic sequencing and super resolution microscopy."

KEY BENEFITS

- Reduces costs by eliminating complexity and unnecessary administrative overhead, allowing team to instead focus on moving genomic research forward
- Provides hands-off solution for both data protection needs today and data movement needs in the future
- Moves data programmatically via S3 or SMB interface from company data center to research partners to increase the efficacy of new drug targeting and gene therapies
- Scalable in capacity, performance, protocol, location and clouds, to support biomedical research goals well into the future

SOLUTION OVERVIEW

- Igneous DataProtect to manage close to a petabyte of unstructured data expected to double year over year.
- Integration with existing NetApp NAS filers via SMB and S3 protocols
- Preparation for tiering backup data to a public cloud provider
- Dedicated Igneous Customer Success team

Sequencing and analysis generates enormous amounts of data that require complex algorithms. Even a relatively small sequence of 10 - 15 million reads will generate between 15 - 20GB of data. Although Altius' high-performance computing (HPC) cluster has multiple lines and the top performing line can do about 1.3 teraflops, it can still take hours to process the data. "It takes a lot of compute, a lot of storage, and a lot of complex data," notes Cockrill. "The needle in the haystack cliché is always at work. You're looking through millions, sometimes billions, of data points to find four base pairs that are out of order, or might be out of order."

Legacy Data Protection Tools Too Cumbersome and Operationally Intensive

With numbers like that, it's no wonder that Altius faces a data management problem. Altius deals with a mix of structured data and two types of unstructured data: a small number of very big files (tens to hundreds of GB each) and a large number of small files associated with each of those those datasets. Altius needed to address some of these challenges.

While Cockrill tackled the challenge of modernizing Altius' IT infrastructure, he also confronted the looming problem of needing to evolve from replication to data protection using backup at scale. Ultimately, Altius' original decision to not back up their data came down to dollars. "It's not that we were ignorant of the value of the data; it's just that we didn't know which data was valuable." In place of backups, Altius was mirroring about 400TB of data across the same NetApp filer—but this data was growing rapidly. Given the trajectory of Altius' data growth, it was obvious to Cockrill that their replication strategy wouldn't work for much longer. "Putting all of your data in the same place is a fundamentally flawed strategy, in our case if the data gets lost, it is gone forever - you cannot recreate the data," comments Cockrill.

However, as data growth continues, the costs of implementing a data protection solution began to pale in comparison to the risk of data loss. With the cost-benefit analysis now in favor of eliminating that risk, Cockrill decided to look for a solution. Running without backups was a business risk for Altius, and risk gets evaluated in economic terms just like any other rational evaluation.

Data Protection and Data Movement as-a-Service Wherever Data Needs to Live

As Cockrill evaluated his options, he realized that he would have to look beyond the sticker price of each solution to assess the true cost to Altius as the organization scales.

While options such as backing up data in the cloud or other physical locations can initially be expensive, the costs will go down over time. What really concerned Cockrill, though, was the operational costs of these options: The time and resources you spend getting your data where it needs to be, backing it up, validating and checking it, is going to do nothing but go up over time. "That's going to cost me an additional \$1 million per year. And, what do I get for it?"

What Cockrill needed was a solution that could back up all of Altius' file and object data at scale and provide the option to tier to cloud, while also keeping his operational costs to a minimum. "Backup is not the killer app for Igneous in my vote. It's data distribution and the ability to programmatically manipulate data to move it to wherever it's needed. Luckily for me, in the not-too-distant future, I'm going to also need a data distribution strategy and Igneous provides just that."

"So, we ended up buying an Igneous subscription to protect 800+ TB's to start," notes Cockrill. This solution was based on the idea that soon, the bulk of his data is not going to physically reside in Altius' data center; rather, the bulk of his data will be out in the cloud.

“ Backup is not the killer app for Igneous in my vote. It's data distribution and the ability to programmatically manipulate data to move it to wherever it's needed.

Michael Cockrill
Chief Technology Officer at Altius

The primary business objective for Cockrill in choosing Igneous is to obtain a hands-off integration for his data management and data distribution problem. "That means my team gets to focus their time on epigenetics -- on moving the body of science forward. There's no value to anybody on my team understanding backup beyond 'I need that metadata over there'."

"I think of Igneous through my own lens, which is I have a data management and distribution problem, and they act like a router that routes network traffic—but instead, they do it with petabytes of data," notes Cockrill. "It's programmable, it's configurable, it can talk to a whole range of different endpoints, and that gives me all the flexibility I need to not get locked into any service provider, but also it allows me to be able to integrate with my down-the-line customers."

“It Was So Easy, I Napped Through It”

On the day of the Igneous installation in Altius' datacenter, Cockrill had planned to go over, see the gear, and talk with the Igneous technicians. Their plan was to start the install around 10:00am and be finished around 3:00pm. By the time Cockrill arrived at 12:30pm, nobody was in the datacenter around the Igneous system. “We were all asking ‘What happened?’” says Cockrill. The Igneous solution was already installed, up and running, and their technicians had left. “We couldn't believe how simple it was.” His experience with setting up and doing the initial backup with Igneous was similar: “It was so easy, I napped through it.”

“ I don't have to hire any more people, I don't have to take on any more risk, and I don't have to increase my management in any way.”

Michael Cockrill
Chief Technology Officer at Altius

In total, Altius needs to back up about 1.1PB of data; so far, they have backed up 700TB. Thanks to Igneous' data compression, which is running at 40% at Altius, their initial full backup of the total 1.1PB will only consume 812TB on Igneous, and will all be resident locally before Altius starts pushing anything out into AWS Glacier or Microsoft Azure. The whole process has been seamless for Cockrill and his team, and he already has bigger plans for his Igneous system.

API-enabled Data Movement and Dataset Collaboration

“While the use case of backup is important, it is not nearly as interesting as our next use case,” notes Cockrill. “I look forward to having our research partners be able to expose an S3 or SMB interface for me to push data from our Igneous system to them. I will have byte-level control of the data that leaves my data center and goes to theirs -- it will be guaranteed delivery, and it will just work.”

These data movement capabilities make Igneous exceptionally appealing to Cockrill, and he is confident that Altius has room to grow with Igneous. By the end of 2018, he estimates he'll have more than 2PB of data under management. “In the past, as the amount of data increased, it would require me to write a bigger check to address the issue. But now, because of Igneous, I don't have to hire any more people, I don't have to take on any more risk, and I don't have to increase my management in any way.”

Cockrill feels lucky to be in a situation where the physical IT infrastructure he inherited was overbuilt enough that he had time to make choices. “Because we had time to look around the market, we could just wait for a solution that made economic sense. Igneous was definitely the right solution at the right time.”

“I think the other noticeable thing about doing business with Igneous is it actually feels like a partnership. I didn't feel like they were trying to sell me something. It felt like they were trying to understand what my business problem was, and they told me how they could solve it.”

Contact Igneous

1-844-IGNEOUS / 206-504-3685 / info@igneous.io