

Cloud-Based Open-Platform Data Solutions: The Best Way to Meet Today's Growing Health Data Demands

Jared Crapo
Senior Vice President Sales,
Health Catalyst

Linda Simovic
Principal Program Manager,
Azure for Health and Life Sciences

As healthcare transitions from fee-for-service to value-based reimbursement models and the need for data-driven improvement becomes more urgent, health systems are finding that their current approaches to data management and [analytics](#) are failing to keep up. According to a Gartner [report](#), “Because traditional [data warehouse](#) practices will be outdated by the end of 2018, data warehouse solution architects must evolve toward a broad data management solution for analytics.” Traditional on-premises data warehouses lack the scalability, elasticity, and analytical agility to support today's data and analytic requirements. They can't satisfy healthcare's increasing demands for data storage and analytics and near-real time insights.

Organizations must now look beyond traditional on-premise data warehouses toward cloud-based, open-platform data solutions and analytics as a service (AaaS), the provisioning of analytics software and operations through web-delivered technologies.

Cloud-based, open-platform data solutions, such as the Health Catalyst [Data Operating System](#) (DOS™), aim to answer healthcare's growing data needs by combining the features of data warehousing, clinical data repositories, and HIEs in a single, common-sense technology platform. Such open-platform solutions layered on top of existing

source systems can support real-time processing and movement of data from point to point, as well as batch-oriented loading and computational analytic processing of that data. This approach allows real-time integration of data from multiple sources and enables the delivery of the right knowledge at the right time in the workflow. The combination of real-time, granular data, domain-specific, and reusable analytics and open-source APIs in an open-platform solution also facilitates the rapid development of new applications. These capabilities give organizations the insights needed to improve care and patient experiences while simultaneously reducing their costs.

With Growing Data Demands, Cloud Hosting for Healthcare Will Become Critical

The future of health IT resides in the cloud. Health systems will need far more data than they have today—much of it originating outside their EHRs—to navigate value-based care and new forms of reimbursement. They will also need increased storage capacity as well as sophisticated analytical and machine learning capabilities to derive insight from this data.

Emerging sources of data and greater data demand make cloud-based solutions a critical supplement to on-premises data warehouses, as they serve key needs:

More Patient Data

Health systems have a relatively small slice of information about their patients—mostly stemming from office and ED visits and hospital admissions. Innovative smartphone applications, home monitoring equipment, genomic sequencing, and social determinants of health data will add to the overall picture of patient health. This data will have to be aggregated, normalized, and analyzed, along with ever-growing clinical and claims data sets, to provide optimal care.

[Genomic medicine](#), which enables highly individualized healthcare, is also poised to break into mainstream healthcare and further expand health data. Philadelphia's [Geisinger Health System](#), for example, announced in 2018 plans to provide [DNA sequencing](#) as part of routine medical care to 1,000 patients as a pilot program, then to all 3 million of its patients. [Geisinger](#) will look for mutations in each person's [exome](#) (the known functional parts of the genome), which is about 1 percent of the total genome.

Progress in genomic medicine will add more patient data to already burgeoning health data stores, and on-premises data warehouses can't accommodate this additional information. To keep up with the data windfall, health systems will have to supplement on-premises data warehouse capacity with cloud-based solutions. Cloud-based platforms

also offer sophisticated analytical and machine learning capabilities, which will help health systems derive actionable insights from the expanding data, a burden that would otherwise fall on already overextended clinicians who lack the required analytics training.

Patient-Driven Demand for Data

As health systems obtain more types of data, consumers want information about their health status and care. Healthcare delivery is not transparent—patients don't know the details of their doctors' orders in the hospital, how much care will cost them, or what role comorbidities, such as diabetes, might play in their recovery time. Patients tend to have a hard time obtaining their own health records in a form they can understand and can combine with records from other providers. As consumers demand more information about their health, cloud-based data platforms and AaaS will enable health systems to better deliver that data easily and cost-effectively.

Transitioning to the Cloud: Healthcare Is Making Headway

Though healthcare has been slower to adopt the cloud than other industries (e.g., [financial services](#), [hospitality](#), and [energy](#) industries), health systems are starting to adopt cloud-based solutions at a quicker pace:

- According to a [2018 report](#), health system skepticism about the cloud is dissipating as more cloud vendors offer HIPAA-compliant services with strong security.
- By 2021, Gartner estimates, public cloud vendors will process more than [35 percent](#) of healthcare providers' IT workloads. [Cloud computing](#) is becoming the standard for health IT infrastructure with the growth of offerings driving adoption:
 - Infrastructure as a service (IaaS)—a form of cloud computing that provides virtualized computing resources over the internet.
 - Platform as a service (PaaS)—third-party providers deliver hardware and software to users over the internet.
- According to a recent survey, leading healthcare organizations expect to have [85 percent](#) of their applications in the cloud within three years.

Even with many health systems moving towards the cloud, there are still concerns about adopting cloud-based solutions. A 2014 [HIMSS Analytics](#) survey found that the major reasons [for cloud avoidance](#) were security concerns (62 percent), availability and uptime concerns (39 percent), and a continuing focus on in-house IT operations (42 percent). However, much has changed since the 2014 survey, with advances addressing these concerns:

- With regard to cyber security, [a 2018 KPMG report](#) observed, “Moving to the cloud is an opportunity to improve your security profile, as most cloud vendors have more robust cyber-security capabilities than hospitals could build themselves.”
- Availability and uptime concerns have largely been assuaged as cloud infrastructure has matured and become more resilient.
- Many health systems continue to focus on in-house IT operations, having made substantial investments to develop and maintain their own data warehouses. In spite of these meaningful efforts, many institutions are realizing that their demand for analytics is outpacing their current capabilities. CIOs must choose between investing more in current platforms or recommending a shift to a cloud-based analytics strategy. The good news is that cloud-based data platforms can complement currently deployed analytical tools, providing additional capacity while ensuring a smooth transition to next-generation functionality.

Five Common Challenges with On-Premises Data Warehouses

Whether health systems use the homegrown data warehouses they’ve built over many years to fit their unique needs or use purchased data warehouses from major vendors, traditional data solutions have five key limitations in an increasingly data-driven era:

1. Predicting Future Demand Is Difficult

Most large organizations tend to adopt data-driven decision making unevenly. Capital purchases of data warehouse infrastructure (hardware and software) from traditional analytics vendors (e.g., [SAP](#), [Oracle](#), or [Teradata](#)) require accurate predictions of future demand. Organizations that underestimate the demand may seem unprepared, whereas those that overestimate demand may appear wasteful.

2. Infrastructure Scaling Is Lumpy and Inelastic

When organizations buy their own computes, memory, and storage, they typically do it in fairly large chunks to align with budget cycles. Once they acquire the equipment, it’s theirs until they retire it. This structure makes it difficult to support several common scenarios:

- Temporary increases in infrastructure required for software or system upgrades.
- Short-term analysis of large datasets for mergers and acquisition modeling, payer contract negotiations, or research.
- In-house development of machine learning models.

3. Security Risk Mitigation Is a Major Investment

Analytics platforms are attractive security targets because they contain a lot of high-value, sensitive data. The more analytics a health system does on its own, the farther it has to stretch its limited cyber security budget. Major software firms may invest more than [\\$1 billion](#) annually in cyber security research and development; by using cloud-based analytics platforms, health systems leverage this investment, versus trying to duplicate it themselves.

4. Data Architectures Limit Flexibility and Are Resource Intensive

Healthcare datasets are diverse in breadth and depth. Cultural adoption of data-driven decision making is creating growth in analytics requests, as are annually updated regulatory and reporting requirements. This progression requires a data architecture that can support two key requirements:

- Reusing logic and calculations.
- Adapting to changing use cases.

The enterprise data model architectural pattern, which has been successful in many other industries, isn't ideal for healthcare datasets. As analytics demands increase in quantity and variety, analysts end up spending more time trying to extend or adapt the enterprise data model to meet new analytical use cases than they spend on actual analysis. Data lake architectures bypass the rigid enterprise data model and push all logic into the query. This approach eliminates the resource overhead of maintaining an enterprise model, but limits opportunities for reusability.

5. Analytics Expertise Is Misallocated

Because they often work with outdated architectures and lack modern tooling, data analysts tend to spend more time working on infrastructure or data acquisition problems than they do on higher value analysis, modeling, and machine learning work. By fixing the structural issues with the analytics platform, analysts will be able to work at the upper end of their skillsets and deliver critical insights.

The Benefits of Analytics as a Service in a Cloud-Based Solution

While organizations can apply a cloud-based data solution to an on-premises IT system to meet the challenges of healthcare's increasing data demands, solutions are most successful with advanced analytics in the cloud (AaaS). AaaS brings six key benefits to healthcare data work:

1. Aligns Infrastructure Costs with Value Delivered

With scalable cloud-based solutions, health systems don't need to worry about making capital investments that are too small or too big for their future needs or too expensive. Cloud-based platforms allow health systems to pay as they go, matching their expense with their current needs. This utility style economic model (i.e., paying the water company for water used instead of investing in a water treatment plant for the back yard) also makes it much easier to compare the costs for a specific analytics initiative with the value delivered by that initiative.

2. Adds Elasticity

Many IT processes are designed around the assumption that infrastructure is relatively static. Cloud-based infrastructure, however, is elastic, allowing IT leaders to implement more efficient and productive processes.

The process of training machine learning algorithms requires a large data set and many iterations of feature engineering (in which a data scientist applies domain knowledge to a data set to select appropriate inputs for a machine learning model). Once the model has been developed, those computing resources are no longer required. The elasticity of resources offered by cloud computing are ideally suited to the wide variations in intensity of machine learning workloads.

Many test or training environments are not used outside of business hours. With cloud-based services, organizations can turn off those environments when they're not using them and request resources and build environments on demand. Both can yield meaningful savings over time compared with static on-premises infrastructure.

By turning off the servers in a development or training environment, users save the cost of running the servers. For example, if an organization needs a new environment but will only need it for a few months, in the cloud, it can add a new environment in minutes and only pay for the environment while it's using it.

3. Makes Resources More Available

The worldwide footprint of cloud infrastructure available from Microsoft, Amazon, and Google makes it more economical to deploy high availability solutions. Instead of leasing space in a secondary data center and filling it with equipment that's rarely used, organizations can configure a minimal footprint in one or more places, then access the

rest of the infrastructure only when it's needed. This cloud-native approach to high availability offers a broader range of both technical and economic options.

4. Enables Machine Learning for More Accurate, Faster Insights

Health systems increasingly rely on [machine learning to identify opportunities](#) for improvement. More sophisticated forms of machine learning require a broader range of data (e.g., claims data and information from multiple EHRs). All of this data can be manipulated in a [data lake](#), an open reservoir for the vast amount of data inherent with healthcare that reduces the time and resources required to map data. Manipulating data in the data lake generally requires a cloud-based database framework, such as [Hadoop](#).

With cloud hosting for healthcare, organizations can use machine learning to supply valuable, timely analytics insights to their clinicians and financial managers. For example, health systems need ways to predict sepsis and intervene before it becomes life threatening. Managing sepsis involves many data points and a variety of clinical protocols; to reduce sepsis mortality, health systems must be able to analyze vast amounts of data with machine learning tools. Skilled data analysts can do this effectively in the cloud with access to all of the health system's data, as well as external sources.

5. Access to Specialized Resources

Specialized analytics skills are useful in certain situations, but health systems don't generally have enough sustained demand to justify full-time resources. For example, an analyst may require particularly complex geospatial visualization that is beyond his or her current skills to implement. An AaaS provider can offer that expertise, as well as a broad range of specialized expertise, for customers who don't need or can't justify these roles in house.

6. Supports Outsourcing Analytics

Over the last few decades healthcare organizations have moved toward outsourcing major business functions. In keeping with this trend, health systems can look to an AaaS provider to meet both their technological and staffing needs. These technology firms can often recruit and retain people with analytics and data science expertise, job skills that are in high demand. Health system CIOs should expect their service providers to contractually commit to a high level of performance, tying payments to savings realized through analytic initiatives.

The Cloud Is the Platform of the Future

With the growing demand for analytics and data storage capacity, [healthcare data](#) warehousing must become cloud-based, and health systems must host more of their analytics work in the cloud. The cloud is the only way health systems can afford the infrastructure and the IT talent to manage the requirements of value-based care, the upsurge of new kinds of data, and consumers' increasing demand for healthcare information. Because of the strict limit to the [machine learning](#) tools health systems can apply on premises, most organizations are best served by placing their data and applications in the cloud and using AaaS platforms.

With the move towards cloud-based services, the structure that data warehouses have traditionally used has to change. Rules-based transactional systems that take months to reprogram for even simple reports can't provide the real-time insights that clinicians need to improve the quality of care and reduce its cost. To provide ad-hoc reporting, quickly aggregate and normalize data from many disparate sources, and facilitate the development of new apps, organizations can layer an open-platform, cloud-based solution on top of existing data sources. When a health system operates this open-platform solution in a public cloud with AaaS, it's poised for success.

Authors



Jared Crapo joined Health Catalyst in February 2013 as a Vice President. Prior to coming to Health Catalyst, he worked for Medicity as the Chief of Staff to the CEO. During his tenure at Medicity, he was also the Director of Product Management and the Director of Product Strategy. Jared co-founded Allviant, a spin-out of Medicity, that created consumer health management tools. In his early career, he developed physician accounting systems and health claims payment systems.



Linda is a product manager with a proven track record for delivering disruptive and innovative solutions in technology. At Microsoft, Linda has delivered an entirely new flexible unified interface platform for Dynamics 365 that adapts to web, tablet and mobile. This innovative technology allows for one configuration to now run on all sizes of screens, including offline mobile. After completing the first delivery of the interface, she joined Microsoft Azure. With Microsoft Azure, Linda is now combining her career of delivering exciting technologies to the Health and Life Sciences market with Cloud innovation. She is responsible across all 100+ services in Azure to ensure that Health and Life Sciences can innovate and build the next big thing.