The Homegrown Versus Commercial Digital Health Platform: Scalability and Other Reasons to Go with a Commercial Solution

By Dale Sanders   President of Technology   Health Catalyst

There's a growing sense of confidence among CIOs and CTOs that they can now build their own enterprise analytics and application development platforms. Some of that confidence is well-founded, but much of it is not…that's the opinion of Dale Sanders, President of Technology at Health Catalyst. For most of his 30-plus-year career, Dale designed, built, and operated those enterprise platforms because no commercial alternatives were available.

The public cloud…Azure, Google, AWS…makes it incredibly easy for CIOs and CTOs to spin-up an incredibly rich IT infrastructure. In the past, it took months, if not years, to design and implement the infrastructure that can now be activated in one to two days. It's amazing and is fueling a long-awaited renaissance in data management and software development. While the public cloud makes the infrastructure easier and more capable than ever, the vast majority of the long-term costs and challenges with scalability reside in the data curation and API layers that reside on top of the infrastructure. The allure of the public cloud is "clouding" the long-term awareness of CIOs and CTOs.

We sat down with Dale for an interview about this trend.

What is the current state of the digital healthcare market?

After a market slowdown in 2017, due mostly to uncertainty with the new presidency and changes in CMS quality measures, sales picked up in 2018, at least for us. As the hesitation has faded, the market is returning to enterprise analytics to support financial risk management models coming from private payers and CMS. I'm seeing an uptick in emphasis on more precise cost

understanding and management, and, of course, adherence to compulsory quality measures. Some of the more leading-edge organizations are also starting to look at what's next, after their EHR implementation, and looking towards the true digitization of health and the patient.

Who are the main players in healthcare digital platform development?

Health Catalyst's competition in the digital health marketplace is primarily EHR vendors; organizations are naturally inclined to leverage big investments they've already made. Homegrown systems are now becoming competitors, too. The ease of homegrown infrastructure, thanks to the public cloud, makes this approach more attractive than ever. Early in my career, the technical infrastructure required for an app and analytics development platform cost 80 to 85 percent of the budget. This left limited money to apply the platform's value to analytics and app development, the actual client-facing products to support decision making.

In that context, what are you hoping to achieve with Health Catalyst's technology?

At a high level, we have two simple goals: Number one, we want to simplify and reduce the costs of IT infrastructure at the data and application layers of infrastructure. Number two, we want to provide a platform that enables data-first application development, whether the application is purely analytic or workflow or some combination of the two, hence the term data-first application development; and we want to make it easy to build those applications so that our clients can differentiate themselves with software and data.

I've been a CIO for something like 25 years out of my 35-year career. The other 10 years, I was a vendor, building data-intense software applications, first in the space and defense sector, then in healthcare. I was the sort of CIO who always had a software development team, building tools and products to help differentiate the organizations that I served. I've always had to build these data-rich development platforms from scratch. With Health Catalyst, I want to provide an affordable, scalable platform to the industry so that every CIO can have the same options that I had in previous positions as Intermountain Healthcare and Northwestern Medicine, two places where we could afford to build those development environments on our own. We have to make that capability more affordable and accessible for the entire industry.

All businesses now run at the speed of software, either slow or fast, depending on the businesses' ability to afford and build differentiating software. Intermountain differentiated itself for decades because it had its own application development platform called HELP, but it became unsupportable and unaffordable over time. And we had one of the first enterprise data warehouses in healthcare that we had to build on our own, too. Now Intermountain has a market-leading EHR, but so do many other organizations; it will be interesting to see how Intermountain maintains their technology differentiation going forward. With Health Catalyst, we're trying to provide something in-between the HELP system—which enabled local differentiation but couldn't scale commercially—and commercial enterprise applications, like EHRs and ERPs, that are built for commercial scalability but make local differentiation more difficult. Something that's easier to program and support than HELP and more flexible and extensible than a commercial enterprise system.

If I were still a practicing CIO or CAO, and Health Catalyst wasn't in the market, I'd build my own platform, again, to complement my core EHR and ERP vendors' solution. The good news is, Health Catalyst does exist now. We've created a very affordable and scalable platform in DOS that I couldn't possibly afford to build and support on my own as a CIO. And it has at least 20 years of agile evolution in its future. I have a pretty good track record for being right about these sorts of things. And the really good news is, we've just started. DOS is barely 18 months old. It's on a path to become the platform for healthcare's digital future. I know that sounds salesy, but I have to feel good about selling it to my fellow C-level friends and colleagues in healthcare. It's like selling to family. I sincerely believe in what we're selling. I believe deeply in the value of it, even with its 18-month adolescent awkwardness. DOS and our apps will get better and better. We are onto something pretty big for the industry.

How does all of this simplify and reduce the costs of an IT architecture?

Good question. The short answer is through consolidation of data and data services. At the high level, and I'm going to oversimplify things a bit, there are essentially three big categories in a healthcare IT architecture: Number one, you have workflow systems, such as EHRs, ERPs, and office workflow; number two, you have analytic systems like a traditional data warehouse; and number three, you have data exchange engines, like an HIE, that push data around the organization and to its partners.

First and foremost, EDWs, enterprise data warehouses, are expensive to operate and maintain over time, so we've created a very affordable, scalable, next generation EDW in the Health Catalyst® Data Operating System (DOS™). We saved a bunch of money by building our own tools for data movement and management. These tools are tailored specifically for healthcare data, as opposed to cross-industry products, such as Informatica, Talend, and Collibra, that don't offer any functionality that is unique to healthcare data. Also, I always found them to be too expensive, as a budget-constrained CIO.

Again, my DNA is essentially as a CIO who operates at the data and application layers of the stack, so our data movement and management tools are built for healthcare and built to be orders of magnitude less expensive than cross-industry products. Our tools reflect years of operational experience in the trenches of healthcare data—analyzing data, moving it around, curating it, securing it, you name it. We bundle those in DOS; if you had to buy them separately, they'd cost hundreds of thousands of dollars per year and offer nothing uniquely valuable to healthcare data. The same applies to open source, which is not free, by any means. What you don't pay in licensing, you definitely pay for in very expensive labor to make those open source tools work in your environment. There's a reason Red Hat just sold to IBM for $34B.

Speaking of moving data around, we acquired Medicity this year, now known as Health Catalyst Interoperability (HCI). Over the next two years, we will build a unified platform so that our clients will have three capabilities in one architecture. First, they will have a next-generation platform for enterprise analytics and digital health—a platform that can handle all the new and non-traditional types of data that's required for personalized care and precision medicine. Second, this unified platform under DOS will allow single-record data sharing like an HIE, but through APIs, not

messages and CDAs. And third, the unified DOS platform will allow data-first application development to support very specific and personal workflows and decision support.

By peeling the data out of the traditional workflow systems, like the enterprise application vendors, clients can repurpose that data to build their own applications. Look at your mobile phone. The world of software is now characterized by lots of small, special-purpose, agile applications that are fundamentally about data. Applications will come and go, but data is here forever. With DOS, we are providing an analytics platform, an HIE-data sharing platform, and a data-first application development platform all in one modern architecture. Thanks to Silicon Valley and the public cloud, we now have the technology and the engineering patterns to provide this unified platform. It's entirely within grasp.

So homegrown platforms seem easy to build right now, but you're saying they won't scale over time?

There are direct parallels to the homegrown EHRs of the past. Organizations such as Intermountain, Veterans Affairs, Partners, Mayo, and Vanderbilt didn't have a commercial vended choice in digital healthcare solutions many years ago. They had a vision for computer-aided healthcare, but no commercial alternatives were available, so they built their own. Over time, when those organizations realized that their homegrown systems were not financially scalable or technically supportable, they had to turn to commercially scalable EHRs. I see the same thing happening with homegrown data warehouses at some of these same organizations. They had to build their own, there were no commercial options, but those homegrown systems are running into the glass ceiling of affordability.

You could argue that, with commercial vendor solutions, these organizations gave up the custom benefits of their locally developed systems; those advantages, however, couldn't overcome the financial scalability problem. Still, the ROI case wasn't there anymore, and the cost of conversion—culturally and technically—is enormous. I see the same thing happening in the analytics space. The same early adopters of EHRs were also pioneers in analytics, so they built their own EDWs. Those home-grown data warehouses are now facing the same problems of economic scalability, and the cultural and technical costs will also be significant.

Today's public cloud and open-source infrastructure cuts analytics technology fees dramatically, making homegrown systems very appealing, initially. But, it's the long-term total cost of ownership where economic scalability becomes a problem. An easy, accessible infrastructure doesn't necessarily translate into a scalable platform for the next 20 years. And you're facing at least a 20-year commitment to an analytics platform when you make that decision.

Intermountain is still using the core platform that my team and I developed in 1998. The public cloud makes the technical infrastructure easy to spin up, but economic scalability becomes unaffordable at the data management, curation, analytics logic, and application development layers. As homegrown systems move up the stack of data content and data value, they may be able to ingest raw data, but they add nothing of value to more complex and expensive tasks, such as adding value through APIs, data logic, and binding analytics on top of it.

Why is a commercial vendor better prepared for scalability challenges?

A metaphor that applies here is the build-your-own-PC fad of the 1990s. Consumers could build their own PCs for less money than they could buy them commercially, but then commercial vendors got scalable and offered competitive products to homegrown PCs with no labor. Today very few people build their own PCs.

I think we'll see a similar progression with the data cloud and interest in homegrown data platforms. While there are a lot of attractive reasons to build your own, for example, a sense of local control and customization, once you get past the infrastructure and onto the enhancement and frontier layers on top of the data content layers, homegrown systems become very hard to financially scale.

We have 300 engineers and other staff at Health Catalyst tasked with the specific mission of offering a commercially scalable but locally flexible analytics and app development platform. We use the public cloud, Azure, which homegrown systems want to use, so the benefits of the public cloud become a wash. Those 300-plus engineers and healthcare data domain experts are applied at the data management, curation, terminology, API, and application levels. You simply can't scale that capability at the local level. As a former CIO, I know that firsthand.

How effectively can niche vendors address AI and machine learning for healthcare?

Niche vendors are all over the market and addressing artificial intelligence (AI) and machine learning. IBM Watson is the biggest and most visible entrant. Now that AI models and machine learning code are commoditized, thanks to the open source community around AI, organizations no longer need teams of PhD data scientists.

What I call the citizen data scientist is emerging—an individual with relatively low data skills and rudimentary statistics skills who can access machine learning and predictive models, run data through them, and get meaningful results. We've taken care of all the data preprocessing for these citizen data scientists. We've commoditized the healthcare data management layer. Combine that with the commoditization of AI models, and almost anyone can now become a data scientist. Also, niche AI vendors underappreciate how difficult and costly it is to ingest, understand, curate, and utilize healthcare data for AI purposes. There's naivety on their part and naivety on the part of healthcare executives, who are attracted to the small shiny AI object on top but don't understand the need for data infrastructure and development and test processes underneath.

Google published a great paper a few years ago, The Hidden Technical Debt of Machine Learning. Every healthcare executive who falls in love with a shiny AI object or listens to their new CTO from outside healthcare, should read that paper before losing millions of dollars and years of setback.

This past summer, Google published another important paper, Opportunities in Machine Learning for Healthcare. It puts a realistic spotlight on both the exciting potential, as well as the very real pitfalls that are subtle and easily missed by people who don't understand healthcare data and the science of healthcare. For example, the paper describes a highly-regarded

healthcare system that used an AI model to risk stratify patients who were admitted to a hospital with pneumonia. The AI model concluded that patients who were asthma co-morbid had a much lower mortality rate, identifying asthma as a protective factor to pneumonia. Those asthma patients were only lower risk because they received much greater attention and aggressive risk management. Don't go out and contract asthma to lower your pneumonia mortality risk. Luckily, the folks who were involved in this experiment understood healthcare and the AI model's limitations. There are lots of niche vendor who are flooding into the AI space in healthcare who don't have the knowledge of healthcare as a science and what amounts to experimental design.

How can commercial vendors better manage healthcare's variety of data sources?

Another paper from Google addressed a debate over which was more important in analytics work: sophisticated AI models or data volumes and feature sets. The school of thought a couple years ago was that sophisticated models could compensate for lack of data or fewer features. When the Google investigators tested different models and different data sets and features, however, they concluded that the most important part of machine learning and AI was to have a lot of data and features.

Data volume trumped AI model complexity; in other words, the most sophisticated AI models and programmers could not overcome a lack of data and feature sets. At Health Catalyst, we've created the ultimate data platform for AI experts who want to make a difference in healthcare. We've done all the hard work and heavy lifting to provide the data infrastructure so that the AI experts can apply their skills at the top of their capability, not in data plumbing. If you try to build your own data platform for AI, you're going to spend at least 70 percent of your money on the plumbing, not on high value AI, I guarantee it. Again, this is a budget-tight CIO talking now, who also has a deep background in healthcare data, not my vendor voice talking.

How do data challenges specific to healthcare impact health analytics work?

Let me mention one very subtle but enormously important issue, a detail that people don't appreciate unless they've been in the trenches of healthcare data. The topic is change data capture. Silicon Valley doesn't have the same change data capture challenges as we do in healthcare; that is, a patient record that is constantly updated. Silicon Valley has been motivated to address data streams with big data platforms, Hadoop and Apache ecosystem, but these are web events that, unlike healthcare, never change. Those transactions are never updated. The architecture of Hadoop and Apache products doesn't accommodate change data capture very well, making this a domain-specific problem for healthcare to address. In addition, many healthcare data source systems didn't track changes in their logs in a way that makes it easy to analyze the data to understand the change data capture process.

As crazy as it sounds, change data capture is a very big deal in healthcare data, and if you don't have the skills and knowledge, you will make a lot of critical mistakes and spin your wheels for years and years trying to figure it out. No other company understands this level of detail like our team at Health Catalyst. It is a quiet but very valuable knowledge base.

We've run out of time. We've covered a lot of topics here. Any last thoughts or advice for the readers?

It's easy to understand the mirage appeal of a homegrown platform like we've just discussed. Big EHR and ERP vendors offer the same sort of appeal. The public cloud has certainly made homegrown data platforms easier than they were earlier in my career, but the long-term costs of these homegrown platforms is not rooted in the platform-as-a-service, PaaS, infrastructure. The costs emerge in the data-as-a-service, DaaS, layer. That's the hard part and always has been, especially in healthcare. As much as I admire the EHR vendors for what they've done, analytics and data-first applications are not their forte and never has been.

About the Author

Dale Sanders has been one of the most influential leaders in healthcare analytics and data warehousing since his earliest days in the industry, starting at Intermountain Healthcare from 1997-2005, where he was the chief architect for the enterprise data warehouse (EDW) and regional director of medical informatics at LDS Hospital. In 2001, he founded the Healthcare Data Warehousing Association. From 2005-2009, he was the CIO for Northwestern University's physicians' group and the chief architect of the Northwestern Medical EDW. From 2009-2012, he served as the CIO for the national health system of the Cayman Islands where he helped lead the implementation of new care delivery processes that are now associated with accountable care in the US. Prior to his healthcare experience, Dale had a diverse 14-year career that included duties as a CIO on Looking Glass airborne command posts in the US Air Force; IT support for the Reagan/Gorbachev summits; nuclear threat assessment for the National Security Agency and START Treaty; chief architect for the Intel Corp's Integrated Logistics Data Warehouse; and co-founder of Information Technology International. As a systems engineer at TRW, Dale and his team developed the largest Oracle data warehouse in the world at that time (1995), using an innovative design principle now known as a late binding architecture. He holds a BS degree in chemistry and minor in biology from Ft. Lewis College, Durango Colorado, and is a graduate of the US Air Force Information Systems Engineering program.