Great. Thank you, Tyler. Thanks everyone. We appreciate you joining us and we'll do our best to make this a good use of your time. This is a fun topic and it's an interesting topic. I was chatting with a friend about it yesterday and I said, you know, big data is a little bit like the Civil War. Just when you think you know something about it, you meet somebody who really knows something about it and you feel like you need to do a little more studying. And that’s the pace of change, and the activities surrounding big data are almost impossible to keep up with as a
single individual. I am doing my best to keep up with it as an executive vice president for product development but the truth is, you know, my depth of knowledge about big data is kind of at the strategic level and that's what you'll see today in the slides, my thoughts that are affecting our decisions about our big data strategy, not so much about nitty-gritty details of the technology, although we will get into some level of detail. But it's more about how do you decide, when and how and where to deploy big data.

So, let's get started here.

---

**Agenda [01:19]**

We've got basically three sections to the webinar today. The destination, kind of where are we headed, where should we be going with healthcare IT in general and how does Big Data fit in to that, what are some of the guiding concepts and influences affecting my vision around big data. We'll talk a little bit about where we are now both in the industry, as well as in Health Catalyst®, and then also what should we do going forward. And by the way, I want to mention right now, I spend a little more time talking about Health Catalyst® in this webinar more than I usually do in these webinars. We try to avoid any kind of notion of a sales message. They are making this purely educational. The reason I will go into a little more detail about Health Catalyst® in this
topic is because I am going through exactly the same thing that a lot of organizations in healthcare are right now. A lot of CIOs are going through the same kind of thing I am, and that is trying to decide how, where, and when to deploy Big Data and what point do you start embracing it. So, I'll go into my thinking in kind of the process that I'm using within Health Catalyst® as it relates to our product line.

Where are we headed? [02:28]

So, where are we headed?
What's the real question we should ask?

- Is Big Data a big deal or not?
- Or maybe...
- Is the Cloud a big deal or not?
- The Cloud is making Big Data accessible, affordable, and transparent for everyone

While I was putting these slides together, it struck me that, you know, is the real question about Big Data and whether it's a big deal or not. Or is it about the Cloud? And the reason I started kind of matching these two thoughts together is because they are related, and I believe that the cloud is making big data accessible, affordable, and transparent for everyone, and I think that most organizations will probably find the greatest value from big data by taking advantage of what's emerging in the Cloud from a Big Data perspective. I will talk a little bit more about the timing of that but I believe that the two kind of have to come together as a strategy. And you'll see how I'm approaching that at Health Catalyst® to how we're talking about the transition of Big Data and the same context we're talking about our transition to the Cloud. And when I talk about the Cloud right now, friends, I'm talking mostly about the work that's being accomplished with Microsoft in the Azure platform, Amazon Web Services, and in the Google Cloud.

And by the way, we have on the line today that can join us during the Q&A session, Shawn Weisfeld from Microsoft. He is a technical evangelist and systems architect at Microsoft. I brought him on to answer questions that you might have specific to Microsoft towards this slide deck. And you'll see later on why I invited Microsoft specifically. We're in the middle of our evaluation of all three of those vended options and Shawn was with us in a meeting
yesterday towards that evaluation of Microsoft. So I said, "Why don’t you join me on this webinar and if folks have particular questions about Microsoft, this will be an opportunity for you to explain and share the Azure vision."

So, lots of overlap between Big Data strategy and the Cloud for us. I just want to mention, you know, that as most of you know, commoditization in the technology stack is going up. It's amazing to me how little time now as a CIO I would have to spend on all of those layers in the stack below data content. All of that is becoming a complete utility and commoditized stack. And so, it's wonderful. You don’t have to worry about data centers anymore. And so, what I would argue is that at some point we would probably need to stop worrying about what a technology is and whether it's big data or not and more and more about the services that we are getting back from the Cloud. And that's the direction that my mind is headed. I want to worry about the data content and applications layer and worry less about what's underneath those layers of the stack. Turn that over to Microsoft, Amazon, or Google.
Deep Shift [05:30]

So, kind of back to the things that are affecting the way we look at product development. For better and worse, not better or worse, but better and worse, it's pretty clear that software has overtaken the impact of heroic leadership as the greatest agent of change in human behavior. You're seeing the software doing great things for humanity, doing horrible things for humanity right now, and it is the reach the software has now not only on a population basis but also the effect it has on behavior is overwhelmingly larger now than what heroic leadership used to be as the greatest agent in change in the society. So we have to think about this as we build our software. I am actively suggesting to all of you and we are definitely doing this in Health Catalyst®, going through each of these software products and ecosystems that you see the logos on this slide, and asking, what's happening with that software? What's going on with that software that is affecting human behavior? And what can we do to borrow those ideas and instill those in the software that we provide to healthcare? Where can we borrow ideas that are changing human behavior (06:54) from all these different sources, bring that into products we develop for healthcare. And we're doing that at Health Catalyst®. We're actively very precisely looking at each of these applications, asking ourselves what can we learn. If you haven't had a chance to read this document, the Deep Shift, 21 Ways Software Will Transform Global Society, this is a great report. It's fascinating. It's not that long. Produced by the World Economic Forum and it's just very very interesting. We stand at an amazing time in history right
now as IT professionals and software professionals. We're finally reaching the point where we can do amazing things in an amazingly short period of time for very little might.

The Facebook and Amazon EMR
From a blog I wrote in 2010 [07:41]

About six or seven years ago, I wrote a blog along this theme of borrowing ideas from these software apps and these ecosystems that were affecting change in society. So I went through my Facebook page in 2010 and I marked it up and I said, what if Facebook, what if this thing was in EMR, what would it look like? It would deliver multimedia content, it would encourage social interaction, you could add members of your care team to your formal care team at the organization, it would have some indication of trends and things. So, I'm just sharing this with you. Go out and find it out on the web, if you're interested in reading the details. But it really is not that hard to draw parallels between these software ecosystems and what we should be doing better in healthcare to use software to positively affect both patients and providers.
Amazon is basically a sophisticated order entry system. So I marked it up as if it were embedded in an EMR and what would it look like and what could we learn from Amazon as an order entry system.
The Facebook and Amazon EMR [08:55]

So you'll see some of these here.
I want to point out that 90 percent of the screen space in those applications is driven dynamically, by context, through analytics and algorithms in the background that are nudging your decisions through suggestive analytics™ & collective intelligence. And behind the scenes of both Facebook and Amazon is the Cloud and Big Data technology, that's enabling the fusion of data and workflow and data collection all in the same user interface. And this term "suggestive analytics™" is one that I'm using. It's not predictive analytics, it's not prescriptive analytics, it's very simple algorithms actually that make suggestions to the end user in a way that affects a positive change. Of course in the case of Amazon, they are trying to cross-sell you and upsell you, and with Facebook they're trying to push products through ads and things like that. But we can borrow from these ideas and do the same thing in Healthcare.

And we did this at Intermountain Healthcare while I was there. So the credit of Stan Pestotnik and Scott Evans and David Classen and the things I learned from them in the medical informatics program there at Intermountain. We did this sort of thing there. We know it can be done. It can be done. It requires the technology to do it. At Intermountain, we had the complete ecosystem of electronic health record and analytics all within our control. Not as easy to do in an environment where you don’t control all of that technology. But we can do it,
and things like SMART and SMART on FHIR in the EMR world are going to enable more and more of this kind of thing going forward, and that's a big part of our strategy as well.

So if you kind of look at the state of analytics and kind of look at where we are at Health Catalyst® right now at this moment, it's data in on the left, workflow, data gets loaded into Epic, Cerner, other workflow information systems. We pull data out of that, we put it in a data warehouse, we analyze it, we produce very cool reports and prescriptive algorithms, predictive algorithms and things like that. But then unfortunately there is this big Cloud that our providers have to walk through to try to remember what they learned here from the output of that analytics system and how that affects their workflow on the left. And that's the Cloud of opportunity that we have to close. That's what we call closed loop analytics at Health Catalyst®. We need to close that loop so there's not that Cloud of mystery between the analytics and the knowledge that comes from massive quantities of data and the delivery of that to the point of decision-making.
So where we are headed, and again, this is Health Catalyst® but where we need to go as an industry is more where other industries like Facebook, Google, Amazon, everyone else has been going for the last several years, and that is you lay over the top of these workflow and analytic systems, a skin that includes web and mobile applications, so that the end user does not really know that they are interacting with both. They are interacting with workflow and analytics to use together in the same user interface. We will always produce the traditional reports that we do right now for offline consumption, but more and more we have to start producing reports in this paradigm as well across both workflow and analytics. And so, that's what we're building out right now at Health Catalyst®. And once again, that's where we need to go, I believe, as an industry.
To be or not to be...

At Health Catalyst, I’m asking the same thing you are: To what degree do we go ‘Big Data’ and when?

To be or not to be...[12:40]

So as I mentioned earlier, I'm in the exact same boat that you folks are. By the way, my wife told me that my cartoon looked like this gentleman was holding a Big Data potato. She asked me what a Big Data potato was. So, part of the hard work. But I'm going through the exact same thing. I'm wrestling with all these exact same questions that you are, when do we go into big data and what role will relational database management systems have in the future.
Now we're kind of in a classic life cycle path for technology, where there's been a huge amount of R&D without much business gain, but now Hadoop is on the upswing. Its Vital Life is ahead of it. Relational databases that come with the analytics are on the down swing. But it's too early to go all in on Hadoop and the Hadoop ecosystem and it's too early to ditch entirely relational databases. They will be around for a while. But their utility in analytics is declining as we see the power, the amazing power and the amazing ecosystem that's emerging around Hadoop and the whole NoSQL world.

Now, the relational database vendors, like Microsoft and IBM and Oracle, they're doing some interesting things to kind of extend this tail a little bit with RDBMSs, adding support for open systems technology that the Hadoop world is benefiting from, like supporting R and Python as an embedded part of the product in the relational world. But the reality is relational database systems were never really designed for the mass analytics that we're trying to torture them through right now. They were designed for data collection and very precise, commits those transactions across all of those tables. And if you go back to the early days of CAD and the folks who originated relational thinking, it was about modeling workflow, collecting data on a single record basis, pulling single records back up. Relational databases are not that great for analysis.
And so, they are starting to come to the end of their life span especially as the volume and velocity and variety of data overtakes their capabilities.

What is Big Data, anyway? [15:10]

So speaking of that, what is Big Data, anyway? Well, I have a Big Data equation here, and that is if the volume, velocity, variety, limits of the database management system, relational database management system, and the costs of that relational database management system reach some threshold, they may need to look at Big Data as an option. And Big Data is equal to the Hadoop ecosystem plus the other things surrounding it in the entire NoSQL world, like the Mongo and document-based, graph-based data systems. So again, it's volume of data, velocity of data, the variety, the limitations of relational systems and the license like these. And I'm going to editorialize a little bit. Shawn from Microsoft is on the call. About 10 years ago, I wrote an article for Information Week in which I suggested that Microsoft needed to start giving away its operating system and its desktop software because they were going to be commoditized. They might as well start giving it away for free now and beat the market to that. Well, literally Microsoft (16:24) in that and I got all sorts of laughs and criticisms for that, but look at what's happening now. And I would suggest they need to do the same thing and think about it in the same way with relational databases when it comes to analytics.
The costs of SQL server are becoming prohibitive and when you compare the capability and what you can do with the Hadoop ecosystem for free, relational database management systems in the analytics world have a very short life span. So it would be in the interest of the vendors to start giving up and basically getting that product away and take advantage of the new products and make your margins on something else in the Big Data world.

**What are the numbers? [17:04]**

So, if you are going to go back and ask yourself, well alright, I've got those variables, but what kind of numbers are going to help me and form a better decision? At the volume level, I would suggest that you have to be talking about hundreds of Terabytes into the Petabytes range of data to start seriously considering Big Data. Now, the reason I say that is that current relational database management systems can actually scale pretty well. You can get into the hundreds of Terabytes, the low hundreds of Terabytes with SQL Server, for example, that's our platform, same with Oracle, same with IBM. So relational databases are a lot more scalable than they used to be and they can handle a lot of data. So you need to start and the only reason you wouldn't loop faster is because we have some many skills tied up with relational database management systems now. So if you get up into these really large data sets, it might be a little
early to think about Big Data. If your velocity of data is coming in at hundreds of Gigabytes up to a Terabyte per day, you're definitely in the realm of a Big Data decision and transition. If your variety of data includes hundred thousand, maybe even millions of source systems, structured, text, tagged data, images, and video, you're definitely in the realm of big data. So I want everyone to kind of pause and think about those three variables in the context of your healthcare systems and think about where you rate in terms of volume of data, velocity, and variety, and then we'll see if it makes sense for you to go into the Big Data world or not.

What are the numbers? [18:47]

In terms of the limits of the relational database management system and the overhead that those engines put on data analysis, if your queries and your load times are taking hours, then you probably ought to start thinking about Big Data. Of course, there's a lot of things you can do to scale that, but it costs a lot to scale up to better performance in the relational world. And when your costs for technology, pure technology, start to exceed $2 Million, you're definitely in the realm of big data, if not, sooner.

In a recent case study I read on a cost per Gigabyte basis for relational database management data warehouse, they were looking at $19 per Gigabyte, total cost of ownership, in comparison
to 23 cents per Gigabyte for Big Data solution, just at the technology level. So when you get to that kind of order of magnitude, difference in costs, it becomes pretty compelling, right? You have to start thinking about these new platforms.

What are the numbers in healthcare? [19:48]

So let's take a look at Health Catalyst® numbers just to give you some idea of where we are. We have about 40 current clients of all sizes right now from giant healthcare systems down to one or two hospital systems. What's interesting in there is so there is quite a bit of sigma there. There's a big standard deviation because of that distribution of the size. But on average, about every one of our clients has somewhere in the 10 Terabytes of storage. They accumulate about a few tens of Gigabytes of data, the velocity. And on general, there's about 20 source systems per client.

Now, again, there's quite a big variation. We have clients with over 100 Terabytes and we have clients that are in the low hundreds of Gigabytes of data per day and then some clients up to a hundred source systems. But overall, this is kind of the average across our client base and probably average across the industry. So, you can see that from a volume, velocity, and variety perspective, we don't stack up in healthcare to the numbers that I talked about earlier, that
the origins of Big Data came from Silicon Valley because they were dealing with a completely
different volume, velocity, and variety of data than we are right now in healthcare.

So, I argue that we're not "Big Data" in healthcare yet and I use this graphic in quite a few lectures these days to emphasize that. I know that hundred Megabytes per year for patient storage because I used to have to calculate and manage that as a CIO. We don’t collect much on a per patient basis now. Clearly, there are patients with really Big Data footprints and clearly there are patients like me who have like zero data footprint in their healthcare systems, as I avoid the doctor and the hospitals and clinics at all cost. I try to take complete control of my healthcare. A Boeing 787 is collecting 500 Gigabytes of flight performance and flight health data in a six-hour flight.

So we're just not "Big Data" in healthcare yet. We're "Big Data" from an imaging perspective. When we pull images, CTs, and MRIs and things like that, there's a lot of data there. But again, it's a very small slice of the overall data ecosystem of a patient. It's only when that patient comes in for an encounter. So, from a volume, velocity, and variety perspectives, the reality is that Big Data story is not that compelling yet in healthcare.
Just Beginning: Digitization of Health

“EMR data represents ~8% of the data we need for population health and precision medicine.” -- Alberta Secondary Use Data Project

But, that will change. This is a description of all the data that's kind of associated with a human being from a healthcare perspective. And this is the sort of data ecosystem that we need to understand if we are going to achieve population health. Right now, we're pretty well stuck around the left side of this diagram. I'm lucky enough to be sharing a project in Canada in Alberta right now, called the Secondary Use Data Project and it boils down to their precision medicine initiative and trying to collect all of the data that you see here across Alberta for all patients and put that in one repository for clinical outcomes research. And we have concluded that EMR data represents about 8 percent of the total data we need for population health in that context. So we're just beginning and we're still relatively small data, but we need to start planning our architectures and our data collection strategies to round out these other gray areas to the right. The strategic acquisition of data is the key battleground of the future for success and failure in healthcare. It's not the acquisition of facilities, clinics, or physician groups. It's the acquisition of data and what you do with it. That's the battleground of success and failure for healthcare.
Dear Data...

Volume, Velocity, and Variety aren’t the only reasons to move

So, from another interesting kind of perspective, volume, velocity and variety aren’t the only reasons to moving the Big Data. The reality is you can do things in Big Data for analytics that are harder to do, if not impossible to do in a relational database management system. So typically in relational database management system, we know the questions. Dear Data, we know the question, give us the answer. It's the world of SQL programming. In the Big Data world, we're asking the data tell us what you know. We don't know all the questions. It's the land of procedural programming and machine learning. And you can sort of torque through relational database management systems to support that kind of open-ended knowledge discovery but it's not very easy. The whole big data world is evolving to support the machine learning environment, unsupervised, you don’t know what the questions are, tell us what's in the data. And so, that's another reason and factor to think about besides the volume, velocity and variety, what's the nature of the sophistication of your analytics.
I love this diagram from Gartner. It talks about the four categories of analytics and the degree of human input associated with each of those four categories. So, in descriptive analytics, you're looking backwards what happened. The pathway to decision and action requires significant human input. That's the Cloud I was talking about earlier, where you have to make all sorts of inferences and you have to remember things back to the point of care that we'll produce in a conference room somewhere. Diagnostic, when you're starting to ask why did it happen. Still has a huge human input right now. Predictive analytics starts to narrow the human impact. Right now we're starting to offload some of the data analysis to the computer from the human, making it closer from the algorithm to the decision of the action. And then finally, with prescriptive analytics and what we call decision automation and true closed loop analytics, the sort of thing that we did at Intermountain, you're actually giving the algorithm the ability to make a decision and take action, and that's where we have to go. As a healthcare industry, that's the direction we have to go. Again, we will always have these four categories of analytics. They all have their place, but we have to do more and narrow that gap of human input and offload what we're now putting on physicians, nurses, and other folks and offload that thinking to a computer and dense it from that.
This is a diagram that dear friends and colleagues of mine in Canada were writing this paper. In fact, I saw this morning that we submitted the final to publication today. And it rates decision making in healthcare in these three simple loops. At the population level, you're dealing with millions of patients affecting change over years and decades. At the protocol level, you're dealing with cohorts and subsets of patients numbered in the thousands, maybe tens of thousands of patients, hundreds of thousands maybe. You're trying to optimize protocols for general applicability to all patients in those cohorts. And then at Loop A, you're precisely delivering information and you're treating patients very personally with quite a bit of variability actually. There's this belief who had to squeeze all sorts of variability out of healthcare, but the reality is protocols that are developed for cohorts only apply 80 to 90 percent of the time to patients. And so, you have to tailor and have great variability and the data that support that variability at the patient level. All of these loops should be using the same data but aggregated and analyzed differently. You have to have different interventions for population health strategies that you have for intervention at the protocol and at the patient level. But everything, you should be using the same data ecosystem. And the problem right now we're using different data ecosystems to inform all of our strategies and interventions in each of these loops.
We are progressing as an industry in the population loop. We are pretty good and I would say that Health Catalyst® has excelled at Loop B. That's kind of our sweet spot, helping organizations understand what they should be doing that they are not doing and filling in those care gaps. We're building out products and services and the data ecosystem for population health management. But the land of opportunity is at Loop A and nothing really is happening significantly at Loop A to help inform patients and physicians with better decision support at Loop A. Vendors like Health Catalyst® operate at Loop C and Loop B with their analytics. EMR vendors are getting into this space. EMR vendors on Loop A, the workflow, but they have been very good with decision support. And so, the tension in the future will be what can we do at Loop A that is choreographed well with Loop B and Loop C? How do we fold population health management strategy at Loop C all the way down to personalized care for an individual patient.

Physicians are 15x more likely to change their ordering and treatment protocols if presented with substantiating data at the point of care versus presented with the same data in a clinical process improvement meeting.

Physicians are 15 times more likely to change their ordering and treatment protocols if present with substantiating data at the point of care versus presented with that same data in a clinical process improvement meeting. So that Loop A is the Holy Grail. That's what we all have to go for. We found this to be the case at Intermountain where we had the opportunity to watch this
in action. So, we've got to come together and we've got to apply the data behind the scenes, the large data that's – let me back up...

The Three Loops of Clinical Decision Support [30:24]

Millions and millions and millions of records, billions of records of data at the population level, analyzing that, and delivering that knowledge precisely to that patient. That's where we're headed. That's where we have to go, I would say.
We are evolving from an offline, data aggregator and analysis company, to a **real-time data production and decision support company**, integrating the knowledge derived from analytics into the workflow of our clients and their patients, wherever that decision workflow occurs.

So Health Catalyst®, and I would argue that – every vendor in this space, I would argue with every organization out there, you have to evolve from being this offline, data aggregator or data warehousing analysis company to a real-time data production and decision support company. You have to integrate the knowledge derived from analytics back to the workflow of clients and patients, wherever that workflow occurs – in the clinic, the hospital, the home, or the workplace.
There's not a cultural change problem among physicians in healthcare. There's a software and data problem. We send physicians out to drive without a speedometer—while CMS, insurance companies, and administrators have a radar gun—then we penalize them when they drive too fast or too slow.

And I would also argue that there is not a cultural change problem among physicians in healthcare. I get so tired hearing this, that these physicians are the reason that we're not progressing better because they just want to be left alone, don't tell me how to, you know, conduct myself or anything. Well, I would argue that there's a software and a data problem and it's equivalent to sending physicians out to drive without a speedometer. While CMS, insurance companies, and administrators have a radar gun—and then we penalize those physicians when they come back to the house and they have been driving too fast or slow and they've had to guess while they're out driving what their speed was. So, if you give physicians the right data at the right time in the right modality, they will change. They are data-driven people. So it's not a cultural issue that we have here. It's a software and a data problem and we have to step up to that as vendors and members of the community.
The Right Data to the Right Person at the Right Time in the Right Modality [32:04]

So our mantra is we're going to deliver the right data to the right person at the right time in the right modality. We purposely borrowed that from the five rights of medication administration. We have to do all four of those things.
And the modalities I'm talking about, for us, are listed on this slide. Traditionally we have deployed analytics through Qlik and Tableau. We have to start providing applications to support mobile and web and also bring your own tool. And so, now, and I would argue again, this is not just about Health Catalyst®. This is about your organization. This is about your companies. You have to start delivering content to your physicians and patients, members of care team, administrators, and these four modalities. We have to start mirroring the rest of the world. It's been doing us for a long time.
Where are we now? [33:02]

So where are we now?
If you go into the Hadoop community, you get a lot for free. It's incredible. It's just incredible. And this diagram does not even come close to capturing the amazing ecosystem that's surrounding Hadoop and the whole NoSQL world. By the way, a lot of people think that NoSQL means like no relational or no SQL programming. What it stands for is not only SQL. So it's not the exclusion of SQL. It's not only SQL. As I mentioned earlier, Oracle, Microsoft, and IBM might as well start giving their databases away for free because they cannot possibly keep up with the acceleration of the open source world when it comes to analytics. And by the way, you don't have to use Hadoop for just analytics. You can use the Hadoop ecosystem for data collection as well. So, relational database management systems are going to take a hit if they don't start making themselves a lot more affordable. They are going to be marginalized.
I've put this Hortonworks slide up because Hortonworks is kind of the only pure open source Big Data vendor out there. They are purely open source. What they offer comes straight from Apache. This is an overview of the architectural Hortonworks platform. And what I want to point out here is that Hadoop ecosystem of NoSQL evolved by a bunch of brilliant young folks as well as a couple of senior folks in Silicon Valley who had the opportunity to witness the evolution of data analytics and data management as a consequence of relational database management systems. And so, it's not just about the technology of storage and processing and algorithms. There's a whole ecosystem of data management that surrounds the Hadoop environment because those young folks, those senior folks got together and said, "What would we have done differently now that we know what's happened in the relational world? What should this environment look like?"

So, it's more of an operating system for data content management than it is a technology platform. You have governance and integration to help you understand the life cycle of data, data workflow that helps you collect data, understand the stream and where it came from so you can collect better data about it. Of course you have all these different methods for accessing the data that exist and reside in the Hadoop File System. They are in layers over the
top of that and it makes this kind of a plug and play environment for all these tools. There's a
great and amazingly capable security environment that surrounds it. Again, lessons learned
from the relational world. And then there’s all the operation – how do you administer the stuff,
how do you take care of it, how do you keep an eye on it, how do you schedule it.

So, the Hadoop world is a lot more than just raw technology. There's also all sorts of concepts
in here around data governance and systems management that make it incredibly valuable.

So, Gartner ran a survey a few months ago and it was kind of interesting and not a big surprise
because most organizations haven't crossed that volume velocity variety threshold yet and the
pain of the relational database management system from a cost and performance perspective
haven't hit people really hard. So, most folks don't have a firm plan for Hadoop investment
right now. A fair number of folks have, you know, in the 26 percent range, have some plans
and plan on doing it soon. And this is in my observation across the clients and the industry that
I worked into. The space and the industry I work in. There's not a compelling driving need for
this in healthcare yet, but it's getting there.
Some of the obstacles, and we feel the same thing in Health Catalyst®, it really comes down to the skills and capabilities required. Now, the good thing about that is the Hadoop ecosystem is making that skills transition a lot easier than it used to be through products like Hive that allow you to submit what amounts to SQL queries into the Hadoop ecosystem. So you've got those SQL programmers that can transition over. You don't have to program and Map-Reduce anymore. So the Hadoop folks and the Apache folks are overcoming the skills transition problem quite rapidly.
Our New Product Lines

- Health Catalyst Analytics Platform
- Operations & Performance Management
- Financial Decision Support
- Clinical Analytics & Decision Support
  - Research Informatics
  - Precision Medicine
- Population Health & Accountable Care
  - Care Management & Patient Relations
- Comparative Analytics (CAFÉ)
  - Collective Intelligence

I want to talk about our product lines because, again, kind of behind the scenes of all this is I believe this is the sort of thing that every organization has to embrace. You have to have an analytics strategy, I would call it a decision support strategy, in each of these categories. If I were still a practicing CIO, this would be how I would organize my analytics and my IT teams around decision support. So, operations and performance management. You know, the platform is the platform. It’s the Big Data and SQL platforms matched together. Operations & Performance Management is a day-to-day reporting externally, internally, from the executive down to the department head, those KPIs and things. You have to have every single day accurately, efficiently produced. Not that differentiating unless you mess it up and you don’t do it well.

Financial Decision Support for us going forward in the future is more about cost accounting and how do you manage these value-based contracts, these PM/PM-type reimbursement contracts. Clinical Analytics & Decision Support is back to the kind of the closed loops that we talked about earlier. What are you going to do to get to that last loop of decision-making, that 15x opportunity to help physicians do their job and deliver personalized precise medicine to patients. That’s the loop that this is all about.
Population Health is about the upper loop – how do we deal with and understand and strategize around the management of hundreds of thousands and millions of lives. In particular, how do we get into socio-economic data, which drives population health more than anything else and how do we develop intervention strategies into the community to improve socio-economic factors. And then the last is Comparative Analytics and Collective Intelligence and that is how do you aggregate data beyond the walls of your organization so that you understand how you're performing relative to national benchmarks.

**Acknowledgments to the Learning Health System [40:15]**

This is a rather lengthy statement but again this is one that I would advocate that you all think about adopting. This is our aspirational product development statement, and you can see the influence of Big Data behind each of the words in here. We actively parse this statement and we're trying to figure out in a machine learning sense the data, the hypothesis, the performance, that loop, what does it mean to the statement, and the acquisition of large data sets to train those algorithms, which is critical.

So, let me just go through this, and I want to acknowledge the Learning Health System. I borrowed this from that consortium and made my own changes to it but they were the source
of the inspiration. So I can make a health optimization recommendation for you. So there's predictive analytics really right there. That's what that statement implies. Informed not only by your latest clinical trial — so that suggests that we have to figure out a way to incorporate clinical trials data into those training sets at the top of the diagram. Also, local and regional data about patients like you. So now the training data includes local and regional data. So your local data warehouse, plus the regional data warehouse of some kind, to inform that machine learning algorithm, and it has to be around patient registries, precise patient registry – that's what patients like you suggest.

Then I can also compare that to the real-world health outcomes over time of every patient like you who has ever had your illness. What that says is we have to start collecting patient outcomes data that we don't right now. That's a huge missing piece of the ecosystem. So into that training set of the Big Data machine learning algorithm world is outcomes data. So as an industry, it's one of the most glaring missing pieces of data that we have in our ecosystem. No outcomes data. Michael Porter from Harvard is one of the leading thinkers and has lashed on to this in the last few years. I encourage you go out and read some of his work. He published a great editorial of sorts in the New England Journal of Medicine last week. Michael Porter, patient outcomes.

And then I can also tailor your care according to your level of interest and ability to engage in your care. So what does that say? That says we have to start collecting patient activation measures and socio-economic data about patients' geographic, cognitive language, economic ability to participate in their own care, as well as their patient activation measures, their willingness to engage in their own care. That has to feed these training algorithms.

And then I can tell you within a specified range of confidence, there is predictive analytics again, which treatment has the greatest chance of success for a patient like you, and that not only gets in the predictive analytics but it starts touching on prescriptive analytics. Now, you're starting to prescribe treatments that are best suited for this patient based upon the millions and millions of records that you've analyzed in your learning algorithms about patients like this. And then finally after I do all of that, I can tell you how much that treatment will cost, which again, big missing piece of the ecosystem, especially the dialogue between physicians and their patients right now. So we're actively parsing this statement and we're building it into products.
And here's an example of that. We have this product called the Patient Flight Path Profiler that does a lot of what I just mentioned in that previous statement. I will not go into the details of this. We'll put this out on the web and you can look at it for yourselves, but we're actively mapping each statement in that aspirational development statement back to functionality in our applications. And the more things in that statement that we can support with our applications, the more valuable and important they will be to outcomes and our clients.

And so again, this is not something that I'm saying you should do or buy from Health Catalyst®. This is something that you need to do somewhere. Whether it's with Health Catalyst® as a vendor, you're an internal IT shop, or some other vendor, I would advocate that this is the direction you have to go.
If we do this diagram a little differently, I would call this the House of Health Catalyst® with population health at the top, the Health Catalyst® platform as the foundation, and the low-bearing walls listed as they are here – Operations and performance management. You have to run your organization, internal and external reporting, in an efficient accurate way to achieve population health. You have to have accurate financial decision support, especially activity-based costing, so that you understand what your margins are in this PM/PM reimbursements. So cost accounting is a big focus for us this year. We're getting into the data collection business there with our relationships with UPMC and Indiana University Health. You have to have closed loop analytics at the patient level. That's what clinical analytics and decision support are all about. You have to have a care management app that the patients can interact with and the care team can interact with. And finally, you have to have collective analytics from your local data warehouse, your partners, your region, and ideally nationally. That's a product that we're building out, called CAFÉ™, that will aggregate all of the data from all of our clients' data warehouses.
I want to mention one thing here that is kind of interesting. We use a term called Late-Binding ™ about our data engineering patterns and principles and I have been using that in lectures now for over 10 years. In the Big Data world, the exact same concept is known as schema on read. So when you hear Late-Binding ™ or schema on read, they are exactly the same concepts and what it really means is a single data model or schema around your data will not meet all use cases. So there's an unfortunate fallacy among some folks right now that you need to purchase a healthcare data model. When you're building a big data solution or data warehouse, the first thing a lot of folks will ask is where is your data model? Well, that's a wrong question to ask, friends. That's not the question you want to be asking your vendors because a good data warehouse, a good analytics platform will have all sorts of different models and schemas within it, depending on the use cases. So I need to let go of this notion that a single data model will support all of the use cases. It's a fallacy and we've seen lots and lots of failures because of that.

And the Big Data folks cut on to that relatively soon. They started off with the notion that you had to have a single schema but they quickly found out that the volume, velocity, and variety of data that they were intaking, they couldn't keep up with schema changes in that high V environment. And so they said, "You know, we can't just keep trying to model all this to a
common schema. We've got to stop doing that." And so, Google and eHarmony and LinkedIn, a few other taught leaders, started talking about the schema on read instead of schema on write a few years ago, about five or six years ago.

So what are we doing? And again, I think this kind of mirrors to what I would suggest is a strategy for a lot of organizations when it comes to Big Data and trying to achieve the things that I have been talking about in healthcare. So, I want to get us out of a data center business as fast as I can. If I were a healthcare CIO, I would be doing the exact same thing. Right now, we have a Health Catalyst® Data Warehouse platform and applications where a number of our clients host their data. We want to get out of that business and we want to essentially photocopy what we have in our existing data warehouse into the Amazon or Microsoft Azure Cloud. Those are our final two choices. I don't think we're going to go with Google.

So the first step towards Big Data is actually getting into the Cloud because once you get into the Cloud, you have the flexibility that you've never had before in your local data warehouse and the Cloud is starting to abstract that complexity, so that now you're dealing with the service, not so much the technology. You don't have to worry so much about the Hadoop
technology, the configuration, the management of that system. What you're interested in is the data that's returned from and the data that goes in and out of the service associated with it. Not the technology stack.

Selection Criteria

1. Total Cost of Ownership
2. Flexibility
3. Scalability
4. Security
5. Reliability

I'll share with you our selection criteria in this decision-making between Health Catalyst® and Azure and Health Catalyst® and Amazon. So we're comparing ourselves, our data center and our offering against what Azure and AWS can offer. And it boils down to these criteria – total cost of ownership has to be affordable, it has to be flexible, it has to feel like that data center is right down our hall, it can't feel like someone else's data or someone else's system, it has to feel like ours, it has to be scalable both at a micro-level, as well as at a microscopic level, and it has to be secure and it has to be reliable. It has to have high uptime. Pretty standard criteria.
And let me give you an example and this is where I will probably make Shawn quite nervous for Microsoft. This is the spreadsheet that we put together yesterday with Microsoft and shared with them. This is how we grade them according to this criteria. So on the left column here, Health Catalyst®, you'll see how we rank ourselves – and this by the way is a 10 point must system. So the winner in each of these categories gets 10 points and then the second place in Azure gets a score that's relative to that. So, you can see here, and I X out the numbers to protect Microsoft's proprietary information, but you can see our total cost of ownership with our labor for the clients that we host is about a $1.75 Million per year, and Microsoft came in, I will not tell you the specific numbers, but they won this category. They are now more affordable than we are. Numbers were close obviously. They got a score of 10 and we scored a 9. The numbers are very close but the cool thing here is that the Cloud is now affordable. And earlier in my career when the Cloud was taking up as a CIO, I can never figure out how to take advantage of it because it just wasn't affordable. Now, the affordability threshold has been broken.

There is no comparison between the scalability. I mean giving us a 5 in relative comparison to Azure is probably being generous. There is no comparison between the scalability that
Microsoft can provide and Amazon will be in the same league. So, no comparison from a scalability perspective.

From a flexibility perspective, it's kind of interesting because the way we can configure our servers and machine actually gives us a little more flexibility at a micro-level than what Microsoft can provide us in their environment. They tell us that that's going to change but right now we're a little more flexible with our configurations than they are from a micro-level, being able to just storage, CPUs and memory, kind of independent of one another according to our clients' needs.

Security – you know, as some of you did not know, but I've been in security for a long time. I grew up in the top secret world of the Air Force as an information systems officer. I worked for the National Security Agency. I think I know my security pretty well. I've done research and development in risk analysis. You know, one of my teams in NSA, we were responsible for hacking nuclear command and control systems. I mean big time stuff, right? There is no way that I can protect our data as well as Amazon and Microsoft can. The level of sophistication that they have around security now is stunning. So if there are any CIOs out there listening today who think that you can protect your data better than Amazon or Microsoft, you're really missing the boat. There is no way that you can keep up with what they are doing. I mean look at the attacks like we witness almost everyday in healthcare, but the Hollywood Presbyterian attack last week, I mean it's really scary, the sophistication of these attacks, and you have to be able to match that. This is not an amateur's game anymore, not at all. And so, security is becoming a just compelling reason to go into the Cloud.

And then finally on reliability, because of that infrastructure that Microsoft and Amazon have and they're financially penalized through their SLA, they can achieve a higher SLA than we can.

So now you're seeing the Cloud is now becoming a compelling story and the Cloud, that first step towards the Cloud, photocopying what you have is the path towards Big Data. That's the point I am trying to make and that's the strategy that we're following.
Step 2: Add Hadoop to the front end of SQL Server for staging text/NLP [53:51]

So, getting our data center into the Cloud is step one. Step two will be adding Hadoop to the front end of SQL Server for staging and also for text and NLP. That's where we're adding our capabilities in that arena this year and Hadoop is just a lot better suited not only from a pure technology platform standpoint but from the tools above it, and Cellar and other tools for text analysis, text processing are just much easier, much better than SQL Server and the relational world. And then we'll blend, that's the little guy down here at the bottom, this cartoon, we'll blend the two environments together, the natural language processing from Hadoop and the discrete area from SQL Server and we'll serve that up in those web applications and mobile applications that I depicted earlier.

And again, I just keep emphasizing, friends, this is not a strategy that should be unique to Health Catalyst®. This is probably a strategy that would fit a lot of organizations, I think, today in healthcare and vendors.
Step 2.1
Deploy CAFÉ™ in a Greenfield Big Data Environment [55:04]

And then step 2.1, in parallel to all this, is we'll deploy CAFÉ™ which is our Collective Analytics For Excellence. That's our data warehouses. And we will deploy that in a Greenfield Big Data Environment. So all of our clients' data warehouses right now are either on premise or in our data center in Health Catalyst®. They are all in SQL Server. We will pull all of their data, de-identify it and load it into a single instance in either the Amazon or Microsoft Cloud and if we could turn this on tomorrow, we would have over 65 million patients with an unprecedented granularity of data and breadth of data for research and outcomes analysis. And that will be updated in near real-time. So this solution, now we've crossed the threshold of volume, velocity, and variety and also certainly crossed the threshold of relational database management performance as well as cost. There is no way that we could possibly deploy CAFÉ™ in any other environment and hope to be successful other than the Cloud and other than in the Big Data environment.
Step 3: Add New Content and Applications to Hadoop/NoSQL.
Atrophy SQL Server Apps over five years
[56:19]

And then step 3, long term, we're going to continue to add new data content and applications to the Big Data platform and we will slowly atrophy SQL Server Apps over the next five years. So we're going to grow this, we're going to shrink this on a case-by-case basis. Again, this, pretty straightforward strategy that probably applies to much more than just us.
Real Life Architectural Example [56:42]

Here's a nice, very cool real life architecture. I borrowed it from Gartner. Again, thank Gartner and Vi Shaffer and Richard Gibson and Laura Davis and everybody there. Here's an example and I purposely borrowed from some place other than healthcare because healthcare does not have a whole lot of great examples right now other than kind of these experiments.

So you can see in this trucking management environment, we've got sensors from the trucks flowing into Kafka, that's the open source tool for streaming data and capturing data in a very quick and easy way, no transformation, loading that into Storm. You've got BI users, typical kind of folks interacting from a SQL perspective and Hive. You've got all the traditional transaction data systems. They are being loaded in through Hive and accessible through Hive and Spark. By the way, the Machine Learning Libraries around Spark and MLlib are what make the Hadoop platform and the ecosystem so compelling as well. So down here you've got these analytics users, kind of the data scientists. These are traditional BI users. These are high-end algorithm guys here. They are interacting with all these data through Spark and MLlib. Then you can drive alerts back to the point of care. You can drive alerts back to the truck drivers. You can drive alerts back to headquarters. And then you've got the BI users with predictive analytics and things here using HBase and what's called the Low-latency NoSQL. So data running in here very quickly being exposed to those predictive algorithms, low-latency back out.
to the BI users and then eventually folding this all back to the point of driving and the point of decision-making as well.

So, this, in some fashion, we needed to start thinking about what this architecture looks like in healthcare, and that's what our team is working on right now at Health Catalyst®, and I would encourage all of you to take this thing and say, okay. What are the metaphors and the analogies in healthcare that we can apply here?

Q&A [58:59]

Alright. We have now entered the Q&A phase of the slides. We're at the top of the hour. We're going to turn Shawn Weisfeld on, get his mic on, and I'm also going to start off by addressing a couple of questions that were submitted prior to the webinar.
I’m hoping you are planning to (or can) discuss the appropriateness of a big data or Hadoop data store for different types of healthcare data. One thing I hear from IT people at healthcare organizations is that they feel like the business/clinical folks hear “big data” or Hadoop and think that is the panacea and everything in their data warehouse should move to Hadoop. My understanding and experience with this is that there are times when this data structure is helpful and times when there won’t really be much net gain from moving from RDBMS to Hadoop.

That's kind of my whole point in a lot of those slides. So I hope I gave you a framework and kind of my thinking about how to do that using that algorithm around volume, velocity, and variety and the pain points of relational database performance and costs. You know, it's not a panacea. It's cool, but from a pure "big data" perspective and from a skills perspective, we're not quite ready to go all on Hadoop and that ecosystem yet in healthcare but it's becoming more and more compelling. A move to the Cloud is going to allow more organizations to adopt Hadoop than ever before and faster because the complexity of managing it, the complexity of interacting with it are being obstructed by the Cloud vendors. So that's going to be the path, I think, of greatest value, least resistance for most organizations. It is not standing these systems up in their own data center but standing them up in the Cloud, especially now as you've seen that the price point now is very competitive with what you can do locally.
Where do you see big data having the greatest positive and negative impacts for healthcare organizations?

Well, I think it's the accumulation of massive amounts of data that will train the machine learning algorithms to a greater level of precision than would otherwise be possible. So volume needs accuracy when it comes to machine learning. One of the problems that Watson has been facing is that it does not have a training set that's equivalent to what impress all of us in jeopardy. There is no equivalent digitized training set of data in healthcare that Watson can consume. If you look at what IBM has been doing, they have been trying to acquire that data volume. I'm trying to remember the name of the radiology company they acquired, and I think it was Merge maybe, with all of the ViTel reports. They acquired ViTel. They acquired Exploris. They acquired Truven. And you know, I don't know the inner workings or what's happening there but if I were them, I'm sure they're probably trying to train Watson's algorithms around that data. And so, the
<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>biggest impact that Big Data is going to have is the ability to process those enormous data sets once we've ramped up the data collection system in the country and then layer over the top of that, all of these cool algorithms and tools for machine learning that then we have to fold all back to the point of care for better decision support, not just the point of care but the point of decision making. They could be at the population level in a board room, it could be in a conference room where you're tweaking clinical protocols for a cohort of patients, or it could be a patient's mobile device or a physician's workstation, but we have to bend those algorithms back to the point of decision making.</td>
<td>I would think so, yeah. Again, we have to be cognizant of our current state of affairs. That one slide that I showed in the case that we only have about 8 percent of the data that we really need for population and precision medicine. So, throwing technology at a problem like Hadoop is not going to solve the problem if we don't have the data to feed the engine. So data is the fuel that feeds and operates the Hadoop system but we don't have that much Big Data in healthcare. So, all of us have to start ramping up the notion of the data that we're going to collect, and going back to that earlier slide, we've got to round out that ecosystem of data so that we can leverage the volumes and the velocity and the variety of data that Hadoop can manage.</td>
</tr>
<tr>
<td>Do you anticipate individual clinics and hospitals will be motivated by the potential benefits of combining big data with machine learning types of applications?</td>
<td>The challenges that healthcare organizations will face, it's skills, I think, fundamentally. And then there's going to be this issue of just kind of a lack of data. We just don't have as much data as what we need to really fuel this engine yet.</td>
</tr>
<tr>
<td>What challenges do healthcare organizations face when considering adoption of techniques like machine learning and big data?</td>
<td>Internet of Things will definitely be a factor in the future. You know, we're collecting data right now from patient devices in that context, all the usual suspects that did and jawbone and all those things. We are collecting all that data right now. It's not entirely clear what we're going to do with it and what value it will have clinically, but that's the nature of data. You don't always know what you're going to do with the data but once you get your hands on it, you'll figure something out. So, expanding the ecosystems to the Internet of Things and personal wearable devices and 7x24 biometrics clearly on the horizon and that's going to drive that volume, velocity, and variety algorithm.</td>
</tr>
<tr>
<td>I can imagine how the Internet of Things can be a factor in the future. Not only will we have the regular data sources to contend with but also many intelligent devices will be collecting a wide variety of types of data. Potentially, useful data. Do you think as the volume of data being collected continues to grow, potentially dramatically, is going to motivate adoption of big data technologies in healthcare organizations?</td>
<td></td>
</tr>
</tbody>
</table>
[Dale Sanders]
Okay. I think that might be – yeah, those are the end of the questions that were submitted ahead of the webinar. And Tyler, do we have other questions that are queued up here, friend?

[Tyler Morgan]
We certainly do, Dale. We are making those available for you right now.

Quick Poll
How interested are you in someone from Health Catalyst® reaching out to you to schedule a demonstration of our solutions? [65:14]

And while we’re doing that, we do have a poll for everyone. We would like to remind everyone that our webinars are meant to be educational about the various aspects. I think we saw a lot of great educational content from Dale today. We have had so many requests for more information about Health Catalyst®, where or what we do. I think Dale was able to give some additional information there as well. But if you are interested in having someone from Health
Catalyst reach out to you to schedule a demonstration of any of our solutions, please answer this poll question.

We'll leave this poll up for a bit. And Dale, can you see those questions now in your control panel?

[Dale Sanders]
I can. Let me pull this out so I can read it. They've changed the user interface a little here. Let's see...Yup. There's a question about can we download the slides. Yeah. Tyler will make the slides available. By the way, they're not proprietary. They're all labeled proprietary but they'll just be creative common copyright protected, not proprietary. Several questions along those lines.

<table>
<thead>
<tr>
<th>QUESTIONS</th>
<th>ANSWERS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hadoop is open source but far from free.</td>
<td>Yeah, I agree with that but it's also a lot cheaper than the total cost of ownership for relational database management systems right now. Orders in magnitude are less expensive. So, not free but relatively speaking, almost.</td>
</tr>
<tr>
<td>Is the speaker Dale Sanders?</td>
<td>Yup. You must have missed the intro slide.</td>
</tr>
<tr>
<td>What is your view of IBM's Watson and share some thoughts there.</td>
<td>You know, the Loop A decision support, I think Tim is asking will Watson be valuable in that Loop A at the patient level, and I think that's ultimately where we all have to go. So, I absolutely believe that it could be but the challenge will be the training set and training Watson to the level of what we saw in jeopardy is going to be very hard because we just don't have that kind of equivalent training set in healthcare. And I would also say that there's huge missing pieces of data that will allow the Watson's learning algorithms to evolve, same thing that we're going to face in Health Catalyst®, and that is if you don't have patient reported outcomes data and if you don't have socio-economic data, it's going to be very very hard to tune those algorithms in a very precise patient-specific way. So we have to close out that ecosystem around patient reported outcomes and socio-economic factors for us to really achieve what we should in Loop A.</td>
</tr>
<tr>
<td>What do you think the timeline is for a skilled and knowledgeable healthcare work force to catch up with Big Data technologies and what should companies and health systems do in the meantime if they want to take advantage of Big Data, say for example clinical decision support?</td>
<td>I think that the best path forward for the most rapid value, time to value for healthcare and Big Data, will be taking advantage of the Cloud. I just don't think every organization – I don't think you want to spend the limited resources you do have on the administration, the set up and configuration of Hadoop, in that ecosystem. I think you want to take advantage of the offerings that are in the Cloud. And</td>
</tr>
</tbody>
</table>
not only distance yourself from the kind of barebones administration of the systems, but also the tools and services for data management and data analytics are evolving in the Cloud, so you don’t even have to know anything about Big Data underneath it. You can go out right now to Amazon and Microsoft Azure and you can submit data sets for machine learning processing and algorithms and in the background it’s being processed in Hadoop but you don’t know it and you don’t need to know it. So you can take advantage of Big Data right now and know absolutely nothing about the backend and the skills.

You talked about Hadoop but Google BigQuery seems to have some advantages over it. Are you looking at it as potential tech platform for data management?

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not right now. We believe that the overall strategic trajectory of Cloud services from Amazon and Microsoft are better than Google. Google has some really cool very narrow areas of technology like BigQuery, they’re impressive. But when you look at the total ecosystem of Cloud services that are being offered from a business perspective, Amazon and Microsoft are quite always ahead.</td>
<td></td>
</tr>
</tbody>
</table>

Will a non-confidential version be available?

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes.</td>
<td></td>
</tr>
</tbody>
</table>

Great topic and excellent presentation. Data analytics, insight-driven population and personal health management is based on the premise that variety of person level data, financial, clinical, socio-economic, device generated is either centralized, as is the case with Facebook and Amazon are really available for real-time analytics to enable the workflows. What are your thoughts on data sharing across organizational boundaries and silos? What role vendors like Health Catalyst™ could play in enabling real-time data sharing?

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Well, it’s a problem because now the way the economic model works in healthcare, there is no incentive for anybody to share data truthfully. I mean there are pockets of data sharing that are working kind of economically but not much. I think the only way to really change this data sharing interoperability problem, and I know it’s a bit of a dream, but would be for the patient to own all of the data about their health ecosystem and then they decide how it’s going to be shared and put them in control of it. The situation right now, and I know that’s their dream because I just don’t see this changing very soon. I think that five years from now I think we are still going to be basically where we are right now with interoperability. I think we have some tools like FHIR that would make it easier technically but we still haven’t evolved the economic model that forces and encourages sharing of data and it’s still the sticky relationship to healthcare providers and insurance companies and others. So, they want to hang on to that data. So I don’t know. It’s a weird thing. Not one that I’m happy about as a member of the industry or a patient.</td>
<td></td>
</tr>
</tbody>
</table>

What visualization tool do you suggest for 130-provider group and approximately 70,000 patients?

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power BI from Microsoft is a great suite of products and a lot of times organizations already own the licenses to that. So that’s a great choice if you don’t have much money, especially for sort of small to mid-</td>
<td></td>
</tr>
</tbody>
</table>
range group like you folks have and most provider groups don’t want to spend much money on IT. So I would probably go in with that recommendation first. If you can afford it, Qlik is a phenomenally good product. That’s the one that we had selected and we believe is the best overall visualization tool on the market today. It’s a little bit pricey but you might think about a situation in which you buy a handful of licenses for really high-end sophisticated users that are Qlik-based and deliver the rest back to the masses through Power BI. Tableau is also a very very good product, similar to QlikView. The edge and I would say the thumbs up kind of goes to Qlik in that context but Tableau is still very good as well.

**What is the impact of the federal interoperability in direct technology?**

I would say it has not been very effective. I still think that we need to be harder federally about forcing interoperability and making it a law, not making it an incentive. And I draw parallels to HIPAA. I came into healthcare from space and defense and intelligence communities where information security was really important. Came in to healthcare, there was zero security. I mean it was like the most open unprotected information I had ever seen. And there wasn’t a lot of move towards improving that. There was no incentive to improve that until HIPAA came along. HIPAA has been kind of brutal, and like all federal interventions, it has all sorts of problems and it still has problems. It’s better than it used to be. But I think we have to have some sort of federal legislation around interoperability. We cannot make this as optional and incentive-based as it has been before.

Direct, I’m kind of on direct. I mean it’s good but there are other modern new ways to exchange data that I think are more effective than direct frankly. So I don’t have a very positive opinion about where we stand with interoperability and I think it’s about time we legislate it, and I’m not necessarily the guy that encourages federal intervention on very many things.

The majority of healthcare providers I speak with do not have a basic exposure experience to data integration or data quality tools. Isn't that the place to start before Big Data? Keep It Simple, Stupid or KISS.

Well I could say that actually Big Data makes integration and data quality a little easier to manage than the relational database world. Like I mentioned, these young smart guys, coupled with some older senior guys have been evolving the Hadoop NoSQL environment. They learned from all the shortcomings of the relational world. And so, there are some very easy-to-use, very effective data integration, data quality tools in the Hadoop world, the Big Data world, but frankly still don't exist in the relational world. So I
think you can actually get into these issues probably easier and more simply in the Big Data world than you can the relational world.

A hundred Megabytes per patient per year, not counting imaging info, what is the source of the numbers? What is the future implying for future growth?

That comes from my experience as a CIO when I had to sit down and calculate these numbers for storage, engineering storage, management storage, acquisitions, and purchase. So that comes from me. That’s first hand empirical observation and measurement from my experience as a CIO. And the last time – now, granted, the last time I did that was in 2013. But if you look at the accumulation of data on a per-patient-basis per year, it has not changed much. I would argue it has not changed at all in the last three or four years. So, I think that number is still pretty valid. If there are others out there, other CIOs that are listening that have a different view on that, I would love to hear and maybe share that in the question.

And the future plan for growth is, you know, it depends on how fast we move out in the other areas of the data ecosystem for patients and how willing patients are to share that data and how we make it easy for them to share that data. So, strategically the future of healthcare, we have to start thinking, and our leadership has to start thinking, that the game is not the acquisition of bricks and mortar. The game is the acquisition of data that rounds out our understanding of the human healthcare data ecosystem and EHRs only account for about 8 percent of that. So we have to have as a nation very deliberate data acquisition strategy around healthcare to go way beyond EHR. That’s a national level strategy. We spent $30 Billion in what I would consider largely misspent money on the EHR world and we have done nothing to think about strategically the acquisition of data beyond those EHRs that for the most part represent a billing environment, not a clinical environment.

So, I believe if ONC were going to step up and be more useful and HHS and CMS, they would start pushing a strategic data plan for the acquisition of more data in healthcare on a very deliberate basis, as if they believe that health information exchanges and EHRs are the end of the story, and they are just the beginning. So very short-sided view of that on a national level right now. And I would argue that a lot of healthcare CEOs have no clue about this, right? There’s generational gap right now among a lot of healthcare CEOs that do not understand that their whole business runs on software and data now. It’s not about great facilities.
It’s not even about great physicians and nurses and staff anymore. If you surround them with poor data and poor software, they cannot achieve their full potential. So, as I mentioned in that earlier slide, the cultural problems we have in healthcare are not culture, it’s software and data that’s constraining people.

<table>
<thead>
<tr>
<th>Does the Patient Flight Path Application from Health Catalyst® have machine learning built in?</th>
<th>Yes, that does. It’s in the background. David Crockett, our Chief Data Scientist, is the primary brain tower behind that along with this team.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Is it possible to copy...</td>
<td>Oh yes, it is. Copies will be made available.</td>
</tr>
</tbody>
</table>
| Healthcare does have a data measure management monitor. Their cost of providing care seems like a great place to start through. | Yeah, again, it’s a great, it’s a huge missing piece of the ecosystem. We’re talking about becoming more economically driven on our decisions with healthcare, moving from fee for service to fee for quality and outcomes-based treatment, but we don’t know what our costs are. So, if you are expecting to take on a capitated PM/PM contract and manage your finances to the margins underneath that and you don’t know your costs, you’re flying without instruments in a thunderstorm.

So, where it’s taking it upon ourselves at Health Catalyst, we’re building out a cost accounting system to start closing this gap of understanding through costs in healthcare. It’s another area where we’re changing and pivoting as a company to fill out these missing pieces of ecosystem. So cost accounting being one of them, patient reported outcomes being one, socio-economic data being one, genomics data, and the biometrics 7x24 data. Those are all in active development because we have to round out that ecosystem and cost accounting is critically important.

| There is not enough data to feed new machine learning engines. Do you mean that the data is out there, yet not formatted to be consumed? | Jacob Barhak is a name that you can Google. Jacob is doing some very cool things with Big Data and modeling the progression of disease. It’s pretty fascinating. Jacob Barhak. Big Data as it relates to the progression of disease.

I mean that we’re not going to be able to feed these machine learning engines. They’re not going to achieve their full potential until we start driving more volume into them. So there’s clearly things and we’re doing it. You could say the Patient Flight Path Profiler. We are doing the best we can with the data that we have and we are seeing some pretty cool results, but we are not going to see breakthrough results until we round out the ecosystem of data that can train those
| **What is the learning health system?** | It's a consortium of vendors and providers, all kind of aspire to constant learning and real-time decision making and constant improvement. It's largely just a philosophy of thought. It's a club right now. I would argue not a whole lot of meaning coming from it. Not to disparage them and not to dismiss them, but it's more just everybody getting together and say, yeah, we should do this. We are taking what we had learned from the learning health system, that aspirational statement, and that is now a part of our development environment here. We are parsing that statement that I mentioned earlier and we're building everything that we can into our products to support every dimension of that aspirational statement. |
| **Can you also use our studio platform as it pertain to healthcare?** | Yes, you bet, and we are doing that right now. And R is quite popular amongst some more sophisticated data scientist and organizations in healthcare and we are definitely taking advantage of it here in Health Catalyst®. |

**[Dale Sanders]**
I'm going to pause for a second and I would love to hear from Shawn. Shawn, why don't you comment about what you see for R in the SQL world, as well as the Cloud world from Microsoft?

**[Shawn Weisfeld]**
Well, I mean we're pretty heavy on that topic. I mean we have huge investments in R both for on and off-premise workloads and we really see the platform as an enabler, regardless of it when you choose R, Hadoop or whatever technology you choose. It's the idea that we have, the right platform in order to enable you to focus on making the right decisions for your business and focus less on installing and maintaining servers and figuring out how to set up clusters and things of that nature.

**[Dale Sanders]**
Yeah. And you guys acquired them – I'm trying to think of the company you acquired that had deep background in our tools. I can't remember the names.

**[Shawn Weisfeld]**
Revolution (85:35).
[Dale Sanders]

[Shawn Weisfeld]
Yeah. And we're big proponents and supporters of the open source world. I think I blow most people away when I tell them 25 percent of Azure is running Linux and some of our Hadoop workloads are running on Linux. So, a Microsoft service that's running on Linux, not on Windows. It's a new world, it's a new Microsoft.

[Dale Sanders]
Yes. I mean it's an amazing new company under Satya Nadella and it's pretty crazy what you guys are doing. I read the other day that Microsoft last year was the single greatest contributor on a lines of code basis to the open source world. Microsoft contributed more lines of code to open source community last year than any other single vendor which is just crazy to think. I mean Steve Ballmer, if he were dead, would be rolling over in his grave.

[Shawn Weisfeld]
He is playing basketball now.

[Dale Sanders]
Yeah, actually. Yeah, good for him.

<table>
<thead>
<tr>
<th>QUESTION</th>
<th>ANSWER</th>
</tr>
</thead>
</table>
| The strategy that you are present starting with data warehouse scales upwards very nicely. However, not all organizations have that kind of foresight. How do you feel about predictive analytics applications? E.g. a model for readmissions deployed on the edges of the data ecosystem, be a third party such as prediction or data queue, bypassing all the steps in between. | That's a great, very insightful thought there. I don’t know exactly. I mean these kind of brings edge products, like prediction, have to have data volumes to be accurate. And so, there's almost no way around the requirement and need to aggregate data. There's just no way to get around it right now. There are HIEs that will say, well, we've got data. Well, it's not deep enough and it's not broad enough to affect population health on a broader strategy. So, I think the only way that those edge third party products are going to work is if they can interact with a platform like Health Catalyst®, and this is something that we have to do. I mentioned that the four modalities, we have to allow people to bring their own tool. So we've got Qlik, Tableau kind of as a preferred visualization environment. We are going to be building web applications and HTML5. We are building mobile applications in a variety of platforms and in angular primarily.

Then, Bring Your Own Tool has to be the fourth mode of access to that data, and that's where I see prediction and tools like that interacting with aggregators of data like Health Catalyst® – because,
you know, it's kind of interesting. We evaluated in the development cycle that we went through for our care management app. We looked at about 20 different applications that you could call kind of fringe applications. We thought about either firing them or partnering with them. And in the end, we decided to build our own because we think we have got some ideas that all of those vendors were missing. But one of the things that they're all consistently missing is they don't have any data infrastructure underneath them to inform their algorithms and their registries. So if you don't have that data infrastructure underneath, your applications have almost no value. So, there is no way to get around the need to aggregate data and you can't do it virtually. The virtual data warehouses don't work. HIE's don't have the data that you need for population health. They have a portion of it. So those are my thoughts there. I hope that it helps.

[Dale Sanders]
I can stay for probably another 10 minutes, friend.

[Shawn]
Well the first use case that we have for us, and I think it's pretty reasonable for other organizations, is to use Hadoop as kind of a front-end place where you can stream and stage data on the front-end of the data warehouse and then it's a great place and it's easy to get started around NLP and text to analytics. So those are the two areas that I think we are going to start our strategy for product delivery in that realm. Let me think if there's any other.

Shawn, are there any other patterns? An initial first step that you would suggest?

<table>
<thead>
<tr>
<th>QUESTIONS</th>
<th>ANSWERS</th>
</tr>
</thead>
</table>
| You mentioned the compelling business case of Hadoop for data warehouse augmentation. Can you expand on some use cases that can serve as a starting point? | [Dale Sanders]  
Well the first use case that we have for us, and I think it's pretty reasonable for other organizations, is to use Hadoop as kind of a front-end place where you can stream and stage data on the front-end of the data warehouse and then it's a great place and it's easy to get started around NLP and text to analytics. So those are the two areas that I think we are going to start our strategy for product delivery in that realm. Let me think if there's any other. |

**You mentioned the compelling business case of Hadoop for data warehouse augmentation. Can you expand on some use cases that can serve as a starting point?**

[Dale Sanders]  
Well the first use case that we have for us, and I think it's pretty reasonable for other organizations, is to use Hadoop as kind of a front-end place where you can stream and stage data on the front-end of the data warehouse and then it's a great place and it's easy to get started around NLP and text to analytics. So those are the two areas that I think we are going to start our strategy for product delivery in that realm. Let me think if there's any other.
**Shawn Weisfeld**

Yeah, I think **[91:05]** on ahead. Collecting the data is the hard thing. It is the first thing that I would start with. We have a service called Azure Data Lake, which is a bottom list repository for HDFS, Hadoop File System Base storage of however much data you want. And that’s where I see a lot of folks starting by getting all of that data, collecting it all into one place, and then figuring out what are the neat interesting things that we can do with this data now that we have it, and we have it collected into a central location where we can scale and get to clean some insights from this repositories of information of data.

**[Dale Sanders]**

Yeah. I mean Hadoop is just a really cheap simple place to dump huge volumes of data. So you're not consuming disk and CPU and memory and things like that on an extensive relational database system. So it’s a great place to pre-process data. Long-term, that notion of pre-processing data before it hits the data warehouse is probably not a good strategy because it increases the latency of your data. Every step you put between the source of the data and the consumption of the data and use of the data adds to the latency of it. And so, we’re headed towards real-time health systems where we’re delivering data to the right person at the right time in the right modality.

And so, having Hadoop on the front-end of a data warehouse long-term is not going to be a good strategy, which is why I indicated in our long-term strategy, this low atrophy of relational database technology, so that's eventually gone. And we'll be streaming data into Hadoop and delivering it out in a low-latency form almost immediately. But for now it's okay to start off with that on the front end of a data warehouse and no problem with doing that as long as you recognize that long term it's not the architecture you want to stay on.

What is your view on how AWS and Azure provide not just the infrastructure as a service but the monitoring and management of client's environments?

**[Dale Sanders]**

Well that's one of the big attractors. Not only is it the infrastructure but it's just incredible surrounding management and monitoring environment as well that my team does not have to do and we don't have to buy necessarily the tools for that.

Shawn, what are your thoughts on that, friend? You want to expound on that a little bit?
| **[Shawn Weisfeld]** | Yeah and that's one of the big drivers I see for folks moving to a Cloud platform, is getting out of the business of the housekeeping, getting out of the business of taking servers out of boxes and worrying about do I have enough hard drive space to satisfy my needs over the next X time period and just getting into the business of adding value to your end customer, get the storage you need when you need it by flipping a switch and a portal or via powershell or a command line command as opposed to a 90-day purchase procurement cycle to a manufacturer of hardware. |
| **[Dale Sanders]** | Yup. Yup. The Cloud is here, friends. You really need to see it. It's an amazing evolution in this business. I have been in IT now since 1983 and it is one of the coolest things that's happened in my career. Well I'd also add that if you are managing a data center, you can expect the assistant administrator to manage. They usually have a ratio of somewhere around one systems administrator to around something like 20 servers. That's kind of stretching it actually, depending on the activity on the servers of course, but somewhere around 1 to 20, the ratio. Well, the Cloud, the amazing efficiency of that environment now is you've got systems administrators that can effectively manage thousands of servers. I mean it's just orders of magnitude more efficient to manage in the Cloud than it is in a local data center now. |

| I missed the presenter's name and his Twitter handle. | It's Dale Sanders and the slides will be available. My Twitter handle is @drsanders. A lot of people think that’s Doctor Sanders. That’s not. I'm not a doctor. My middle name is Royce. So it's @drsanders. |
| What do you think about the use of block chain technology for EMR-derived data? | I have no clue how to answer that. I know a little bit about block chain technology, not even close to being an expert, and I've never thought about it in an EMR context. So you'll give me something to go read and think about there because I can't give you any good answer. |
| You mentioned the importance of social determinants for the personalization of healthcare but I've noticed it is rarely incorporated in healthcare analytics. | I totally agree. |
| Do you have any recommendations on how to get business buy-in to incorporate social determinants into healthcare analytics? | Well we are building in our application the ability for patients to submit that data through our care management app. And if you look at the IOM study a few years ago, two years ago, that finally put this topic in the headlines. There's only about 10 to 15 data |
elements that if you collect those have a gigantically valuable effect on the analytics in the algorithms of machine learning for personalized care, as well as population health management. So these socio-economic data attributes are important at the personal level but they are also hugely important at the population level. And it’s not Big Data. It’s small data. I mean it’s literally like 10 to 15 data elements. If you look at the value of those to the algorithms, it’s gigantic. So, I don’t know where most healthcare systems are with this but we are most definitely going to fill this gap with our application so we can collect that data and it’s relatively easy to do.

What do you mean that EMRs only account for 8 percent of the data?

We, in this precision medicine secondary use data project in Alberta, we stepped back and we put together the strategic data acquisition plan for that initiative and we estimated and we put together – you know, this is all the data that we think we need to be fully informed at the population level, as well as at the patient level for our precision medicine goals. And when we came back from that analysis, John, we can concluded that the EHR is just a tiny sliver of the total data that we need.

Now, the interesting thing in that analysis is the genomics kind of skewed the results a bit because genomics is so voluminous and can be, but it was essentially on a – we use volume of data, not number of data attributes to come up with that 8 percent number. Volume of data, not attributes of data. Different ways you can count it. You can have very low attribution data that is hugely voluminous or you can have very small volumes of data with very large number of attributes. So it’s a little tricky. So I wouldn’t get too hung up on the 8 percent number. I would just acknowledge, I think, what we should probably face, and that is we have just begun digitizing healthcare. And so, EHR is just representing a small piece.
Thank You [99:33]

[Dale Sanders]
Okay. We better set down, I guess. Thanks everyone. I really enjoyed the dialogue and question. Thank you so much.

[Tyler Morgan]
Thank you so much, Dale. Thank you, Shawn. Shortly after this webinar, everyone will receive an email with links to the recording of the webinar, the presentation slides and also please look forward to the transcript notification we will send you once it is ready. So on behalf of Dale Sanders, as well as the rest of us here at Health Catalyst®, thank you so much for joining us today. This webinar is now concluded.

[END OF TRANSCRIPT]