

Dale Sanders: Welcome everyone that participated in the first webinar, and thanks for joining us on this second. We're going to jump right to the technology portion of the slides. For those of you that participated in the first session, we talked about people, and processes, today we're going to jump into technology. Also we were grateful for the feedback that we received in the first webinar, which encouraged us to be a little more specific, and move right to the topics of interest, and a little bit less chatter between us. So, we'll try to do that as well today.

Dale Sanders: So, people, processes, and technology. On the technology front lots of things to talk about here, because over the last 20 years just a phenomenal changes in the technology that's available in analytics and data warehousing. I think one of the things that I did right in that context, we did right, was ignoring the Enterprise Data Model in favor of late findings. So, if you were a follower of Bill Inmon and Ralph Kimball and those folks, they would generally advocate the production of large Enterprise Data Models, this of course was in the mid to late 90s. Star schema is the center piece of all data models.

Dale Sanders: I learned the hard way in my military, and space and defense background that those Enterprise Data Models, there's no such thing. Because in a fluid environment, an Enterprise Data Model is outdated the second that you think it's complete. Then the other problem with Enterprise Data Models is that you essentially do nothing but model into, and mapped into those models. So, you never really finish the modeling. You never really finish mapping. Every time there's a change in the environment, you have to go back and change the model, and the ETL, and the downstream analytics, and you spin yourself into the ground.

Dale Sanders: I learned that the hard way, and through my mistakes, and watching mistakes of others when we attempted those approaches early on in my career. That's when I came up with this concept that is known as Late Binding, and brought that into healthcare. More commonly now I think in Silicon Valley it's referred to as Schema on Read. One of the interesting benefits of taking this Late Binding approach is that we didn't need expensive ETL tools, because most of your ETL is more object oriented further downstream, smaller grains, not the big massive ETL that was advocated in the mid 90s. There is some downside to Late Binding, we'll talk about that in the next section.

Dale Sanders: I took good advantage of tried and true SMPR architectures. The technologist, the pure technologist would always push towards massively parallel architectures, but the reality is the Teradata's of the world being a good example of that. Those are very expensive architectures, they're hard to maintain. And I always got, and I continue to get a lot of value out of SMPR architectures without the complexity and cost of parallel architectures. So, I think that was something that was generally against the grain of advice through most of my career, but happy that I stuck with the pragmatic approach that worked well for us.

Dale Sanders: We ignored the early love affair with Dupe, eight to 10 years ago everyone was in love with the Dupe. I think we're going to see the same, we're seeing the same sort of thing with blockchain. Which, in my opinion, blockchain is not well suited for analytics. It has some incredibly disruptive potential in transaction systems, especially those that require a distributed application development, and distributed trust. It's going to be incredibly disruptive in that context. But even if you look at blockchain today, the throughput, and the transaction rate that blockchain can support is just phenomenally low. Not even close to being capable of handling the transaction rates of a large health care system with an Enterprise EMR for example.

Dale Sanders: So, in order to achieve those high throughput rates for blockchain you have to throw all sorts of hardware at it, and it's just cost prohibitive. But it'll be interesting to watch blockchain. We're not jumping on blockchain, and we didn't jump on Hadoop, and Hadoop is still not the end all, be all that everyone thought it would. The big data ecosystem is very cool, and I will talk about how we migrated towards that in Health Catalyst. One of the things that we did, especially at North Western from the very beginning was the notion of blending text with discreet data in a data warehouse when literally no one else was doing that. In any setting, but especially so in healthcare, and I'd be interested to hear how that's evolved when Shakeeb talks about current state of affairs of the data warehouse at Northwestern. But now we're seeing that as a pretty common practice, but at the time it wasn't common.

Dale Sanders: And the other thing was in the early 2000s when I went to Northwestern, having come from a very deep IBM and Oracle background in data warehousing, I could see a trajectory of value from Microsoft that I thought was pretty positive. Now it was very early, it was very risky. When I landed at Northwestern we had a very deep relationship with Oracle. We had a very affordable Enterprise license, so my advocacy for Microsoft didn't go over very well for a while. But my opinion at the time, which I think has been reasonably well validated is that Oracle was not showing the evolution, and the pace of evolution that I thought it should. And Oracle's code base was pretty antiquated, or to maintain Oracle was very, very high cost labor, and very intense. Microsoft not having that legacy came long and they offered smarter things about tuning, and indexing, and taking a lot of the DBA functions, and making those easier to manage, and more automated.

Dale Sanders: I'm happy to say that that's been a good decision. We are really happy with Microsoft at Health Catalyst, it worked out well for us at Northwestern. And Oracle, of course, has not done very well, they've struggled I think because they've never really reinvented the code base of the Oracle data base engine. Whereas, Microsoft has done it a couple times. So, those worked out.

Dale Sanders: What could I have done differently here? The reality is in the world of Data Modeling too much Late Binding is also a problem. So, an Enterprise Data Model I think, generally speaking, we all recognize those of us who've been in the profession, that you can't crack open a data warehouse today, a successful one,

and look at a single Data Model at the heart of this system, and expect that to be successful. It's funny how I, as a vendor now, I occasionally still get those questions, "Well let me see your Data Model.". Well that's a pretty outdated question, because the reality is when you crack open a modern data warehouse today what you should see are a lot of Data Models. Various kinds, you might have a star schema, you might have something that looks third normal form, first normal form. More and more Data Modeling is going to become less and less important.

Dale Sanders: If you think about Data Modeling, it's really just caches of data and logic, what we would call bindings of data and logic. But going forward those bindings, and those logic rules will be associated through APIs, and my cache, and a data model. So, your data stream will come in, you'll bind, and you'll call up the APIs that you need for the binding that you're interested in. But I see a future where persistent data models in a data warehouse are going to become less and less important.

Dale Sanders: Now, Late Binding can have, just almost as many problems as an Early Binding Enterprise Data Model. In that if there's no reuse, there's no consistency, if all you practice is late binding, you get this proliferation of data objects in the database that are hard to manage. Performance issues become a problem. Load time, load management becomes a problem, so late binding feels very good for a while, because you're very agile, you're very adaptive to these cases that show up. That's essentially what late binding is for those who don't understand the term, you're binding according to the use case basically at the last minute. I borrowed that term from software engineering. Run time versus compile time binding the libraries, and things. This is essentially run time compile of data is the concept.

Dale Sanders: But if you don't have an effective way of managing all those objects, and if you don't reuse some of those objects when you should, it creates data inconsistency, and it creates data governance problems. It's interesting for us to learn that at Health Catalyst, where the reality is we over applied late binding in a lot of our client sites, and if we have any clients on board I'd love to hear your thoughts. We're now going through re-engineering of our platform, and we're creating what I would call Intermediate Data Models. So, if you look at the far extreme of an Enterprise Data Model, it assumes that you know everything about a use case a priori. If you look at the other extreme, which is late binding, it assumes you know nothing a priori, and you're going to bind the data at the very last minute.

Dale Sanders: While the reality is we do know some things. And healthcare is, kind of, a volatile industry, but we do know some things. For example, CMS Measures are, for the most part, pretty low volatility. So, you can bind those data structures in that intermediate space between enterprise and late binding, and you can achieve greater efficiency for your data analyst. Greater reuse and support, and you can also increase the efficiency of your data governance function, because you get more consistent results. So, there is this intermediate space, and the

rule that I suggest everyone applies when you see comprehensive, and persistent agreement about a data binding, go ahead and lock that down in a data structure. Eventually we'll lock those down in APIs. But look for comprehensive, and persistent agreement, and again almost nothing is permanent in healthcare. We can't even decide what the definition of a hypertensive patient is. But there are some things that have a lower volatility rate than others, some concepts. So, you might as well go ahead and bind that in an intermediate data structure, and increase your consistency of reuse.

Dale Sanders: You know, back to the SMP architectures, I let the technologist talk me into massively parallel processing system at Intermountain. And my gosh, it created all sorts of headaches for us. It almost undermined the complete success of the data warehouse, and it was incredibly expensive for no good reason. So, I vetoed that decision finally, and we fell back to an SMP architecture. It was one of the best decisions that we made at Intermountain, in terms of the success of the data warehouse. Because, prior to that, that SMP architecture was too hard to keep up, and maintain, and just too costly. We were having downtime all the time.

Dale Sanders: The other mistake that I made early was too much faith in an Enterprise Standard Business Intelligence tool. I don't see this as commonly as I used to, but a lot of times CIOs, in particular in the IT staff, they want to drive standards for ease of maintenance, and all those other reason. So, they want to drive towards a common BI tool, that everybody uses, and nothing else would be allowed. Well that doesn't work, and you have to allow people the tool ... That it's essentially bring your own device, well this is bring your own analytics. You have to allow for that. It's interesting when just before Intermountain, I worked at Intel as a consultant, and designed their first Enterprise Data Warehouse. One of the tools that we tried to put on every desktop was Cognos.

Dale Sanders: So, I tried to do the same thing when I came into Intermountain, a couple of times. Crystal Run, Business Objects, Cognos, and I can remember saying that I wanted those business intelligence tools to be as common as Microsoft Word on every desktop. That was my goal. That was a very arrogant, and short sighted goal. Because what assumes is that everyone wants to interact with data within the constraints, and the features of those tools, and it's just not the case. It's not just human nature. So, for those of you that are perusing this ubiquitous single tool that does everything, I would say adjust your decisions, and adjust your strategy.

Dale Sanders: Okay, what are we thinking about in the future? Read-Only batch related warehouses are already outdated. So, I could have put in the what should we have done differently sooner. Should have gotten out of the batch load mentality much sooner. To some degree we did that at Northwestern, I saw the downsides of the batch oriented mentality at Intermountain starting to show up. We moved towards more real time capability at Northwestern, what I call trickle feeds, instead of batch feeds. So, in general that's where we should all go as a profession, is you want slow trickles, constant trickles of data into the

enterprise warehouse platform. Start getting away from these batch loads, there're all sorts of problems with batch loads. The windows become unmanageable. You get these huge performance spikes on board the source system, as well as the data warehouse, and it leads to slow decision making. Your temporal dimension of decision making suffers if you're only uploading data every day, or every week. So, get away from mentality, and get away from that architecture as soon as you can.

Dale Sanders: So, where we're headed now is towards these hybrid transaction, and analytic architectures. Borrowing patterns from Landon CAP architectures in Silicon Valley. That's, let me go into the details here a little bit. So, this is what the Health Catalyst Data Operating System Architectures looks like right now. By the way, it's all in Azure, if you're not moving to the public cloud you need to. There's just no way that you can provide the same kind of agility, and dial up capability that the public could can, or as well as security. We're very, very happy with Azure, and Microsoft trajectory with Azure. I can't advocate strongly enough, if you're not in the public cloud get there as soon as you can.

Dale Sanders: This is a big typical diagram, you've got data sources showing up on the left. All sorts of different file structures, things like that feeding into the platform. We break the data ingestion part up into three parts now, which is Data Integration. What we call the Engine, which is a proprietary tools set that our team developed. And also the Engine now is nationally bound to artificial intelligences, so it's not something that you do as a separate thing in the data pipeline anymore. So, you can invoke from our tools, and from our pipeline, you can invoke AI Algorithms as a natural part of it. Again, moving away from those persistent data structures that are most common right now in healthcare, and all these really are calls out to the EPIs that bind to the rules, and the logic that are important.

Dale Sanders: In the lower levels you see the compute and AI clustering, the transaction data storage. That's what Azure allows us to do as a utility, and you can see then, we're taking advantage of the Dupe, Spark, Elastic search. These modern tools, Spark, are a pipeline in the AI cluster, as well as Sequel. Sequel's going to be around for a long time. Relational structures are still very valuable, but they're not capable of doing everything that we want to do now in a modern environment.

Dale Sanders: Over on the far right you'll see that the arrows now go back into the platform, whereas in a traditional data warehouse it all goes from left to right. Now we have the ability to write application, word flow applications back into this environment. So, that's where we're borrowing Landon CAP consign principles. It's working out quite well. So, we can peel the data out of these source systems, we can put them in the engine, this more modern architecture, with modern tools you can see is microservices based. We're using Angular D3, .Net, Java, Kubernetes, JSON, everything that's modern in Silicon Valley is now reflected in this architecture. The applications that we're writing can now support work flow. So, you got this hybrid combination of analytics, and work

flow in the same user experience just like Amazon would offer. Two thirds of the screen space Amazon occupies is about the analytics of the transaction, not the transaction itself.

Dale Sanders: Then in the upper right quadrant, even if you're not involved with a vendor like Health Catalyst, you should have the ability to develop third party applications against your data warehouse platform in the next generation version of that. So, for us it's a commercial play, but we're also allowing clients to do that, but you should be thinking about that internally. Going back to people for just a second, your data analyst and your software engineers, you need to be coalesced into a single team in a modern data warehouse platform; because you should be writing applications against that platform that go beyond sequel. So, if I were still a practicing CIO, I would mash those two teams together that have typically been separated. I'd have software engineers, and data engineers on the same team, and they'd be building these consolidated work flow analytic applications together.

Dale Sanders: Okay, now let's get onto ... There we go. I think this is Lee?

Lee Pierce: This is Lee.

Dale Sanders: Lee's next?

Lee Pierce: Yep.

Dale Sanders: Okay, no transition slide. Lee you're up friend, thank you.

Lee Pierce: Alright, thank you Dale. So, what did we do right? At Intermountain as the years continued to go forward one of the things that we did was we resisted, as Dale mentioned early but it continued, the ongoing pressure to move data to this Enterprise, or canonical healthcare data models that continued to creep up. One example of that is as Intermountain started doing more work in the genomics space Oracle had a Genomics Data Model for healthcare that is combined with their health care data model. What they were pitching was, "Well, why don't you just transition your whole data warehouse, and all of your data into this? Then look at all these wonderful tools that you can stick on top of it.". We figured out other ways to address the needs, and I think that's something that we did right.

Lee Pierce: The next one is, one of the biggest value add decisions that we made was a purchase, and implement tableau. And I know other organizations use Click, and other visualization tools, but this really was a game changer for the delivery of information at Intermountain. Was to purchase a tool that's in that visualization space, it was very different from the Crystal reports, or Cognos scheduled reports, or just parameter driven reporting that we had done previously. That was really something that has continued to pay dividends at Intermountain.

Lee Pierce: Next is when it came to technology we didn't make decisions in isolation. We collaborated with others. We asked a lot of questions. We talked to vendors, and vendor partners, and even more important is we talked to other healthcare systems. Started to also talk to groups outside of healthcare, other industries. The early work that Dale did, with Susan McFarland, to stand up the Healthcare Data Warehouse Association, now the Healthcare Data and Analytics Association, that continues to be a wonderful forum for collaboration for healthcare system. Over the years, related to technology, I believe we benefited a lot from the ongoing collaboration.

Lee Pierce: The Health Management Academy is another collaboration that Intermountain has participated in with some other health systems. That has been valuable. You know, there's been some regional events. I participated in one in Virginia yesterday, there was tremendous collaboration in sharing, doesn't just relate to technology, but people in process just as much. But related to the technology, there was wonderful collaboration and discussion. So, something that I hope to see continue for sure is the opportunity to learn from each other related to technology so that we can help each other find the value add use cases, and success we are having. Figure out how to assimilate that across healthcare, not just success within our own organizations.

Lee Pierce: Not just success within our own organizations.

Lee Pierce: Next is ... I think we, as the years went on, we used a thoughtful approach to make technology decisions and how we spent money. We didn't put a goal out there around spending X million dollars within a certain number of years. There have been some health systems that have publicly said, this is what we're going to do over a certain period of time, and make a big investment and then, go buy a bunch of technology.

Lee Pierce: We were very careful to say, "How is this going to help us improve the value that data and analytics can provide in healthcare?", and ask that question and, "Does it augment current capabilities?" Or, "Does it replace current capabilities?" And, "Is it worth the investment to try and replace current capabilities with the different tool and approach?" Or, "Where does it fit?" And I'll share something else on the next bullet also on that. Ultimately though, does it enhance our capability to achieve better healthcare outcomes? Is the value, in the end, greater than the investment that it's worth?

Lee Pierce: So, I think we've done a good job. Maybe, at times, have been a little too hesitant to pull the trigger on some technology decisions when we were able to answer yes to a lot of these questions, where it would add value, but in the end I think we've made some very good decisions.

Lee Pierce: Alright, next is that we built an analytics reference architecture, especially these last couple of years, and on the next slide, I'll share with you what that looks like, and really, we used it to rationalize the data and analytic tools and technology decisions that we were trying to make.

Lee Pierce: If you can advance to the next slide then, this ... and the source of this is the data and analytics team at Intermountain and in particular, this last year, I and other team members work closely with Kurt Peterson, who is the ultimate author of what we're looking at here, is he pulled all our different ideas together.

Lee Pierce: And, there's many similarities, actually, in this analytics framework to what Dale shared as part of the data operating system for Health Catalyst and it's organized from left to right, data source, acquiring, storing, additional processing, and then storing of that enriched data, then you have your analytics, and then you have, how are you going to get the data into the hands of people that need it? And, but it's also the closing the loop from distributing data back then, into the source systems and into the workflows and what underpins all of that is data governance and data quality and master data and security requirements that we all need.

Lee Pierce: This has been really helpful as a tool and I think conceptually, this will continue to be, where you can look and say, what's the current state and what's our future states in each of these paths, in each of these steps in the process of trying to get insight and value from data. The other thing that this was used for, that I used it for and why I started down the path to build this, is because you can also lay alongside your internal tools and technologies and vended tools and technologies that are brought to bear, so that you can show if it's an and, or if it's an or. Is it a replacement or is it an augmentation, the current tools and technologies, to have a comprehensive enterprise view of the technologies and the footprint that you have within your organization.

Lee Pierce: So, the actual ... this is kind of the logical view of what it looks like, and then you apply the actual tools that you have and the gaps that you might have and investments that you may need to make and this has just been helpful and I think is something gone right that we had done at Intermountain and hope it'll continue to add value.

Lee Pierce: All right, so what could we have done differently? One of the decisions, as I look back, is we migrated from Crystal Reports to Cognos and this is not necessarily about the tools and the technology itself, it was more about the effort to migrate from one to the other and versus the additional value that we actually realized. I'm not sure, and some of my Intermountain friends that were involved in this may think differently, but, that was a lot of work and I'm not sure, in the end ... I think we were hoping that Cognos, at the time, would have some of the capabilities that we ended up having with our Tableau purchase with better data visualization and didn't end up having that, at least, in a timely manner, for what we were needing to do. So just a caution there, I guess, for considering the full cost, the migration cost and efforts, is it really worth it?

Lee Pierce: The next one really is similar to what Dale just talked about related to the late-binding, early-binding and really, what happens in the middle? Where's the balance? I wish that we would've started earlier to build what we referred to as

a shared infrastructure, shared or common data marks that are those things that are repeatable, that we should instantiate and then re-use in order to limit the duplication of ETL and data modeling that ended up happening; and along with that, then, is then, to enforce internal data modeling standards or instantiating ETL standards as data marks, downstream data marks were created.

Lee Pierce: I am a big proponent and fan for having a shared or common infrastructure where it's that middle-binding, if I can use that term, where that's what gets instantiated in the middle and then re-used over and over and over again and is a model that Intermountain continues to move forward with and I think, is going to continue to really add value.

Lee Pierce: I would've also built a more meaningful working relationship between our EMR vendor that came in as a replacement for our legacy EMR, which is Cerner and our analytics program and teams and try and really focus on opportunities to deliver value through analytics together. This became something I wish we would've done differently is just have that better working relationship where it wasn't us versus them, which was not something that we started out for it to be that way, but it was a difficult, and continues to be, a difficult discussion.

Lee Pierce: And, part of the relationship that I think could have been handled better. I could have handled it better as the leader of the group and I think ... and just going forward, could've seen signs for and could've led through some of the decisions that were being made to really focus on the value that we're trying to achieve and what's the best way to do that and combine that with the architecture, the framework that we just looked at and look for opportunities for, and, to deliver value and stay focused on improving patient care. That's something that I really wish that I, personally, would've done differently in helping lead the team.

Lee Pierce: All right, and then the last is future. I think machine learning and AI ... the impact is real today and I have a list that I've kept and other of my team members that I've worked with and been keeping around success stories of machine learning and AI in healthcare and it's adding real value today. I think there's no doubt that, that's just going to be part of how we do data management and analytics going forward, but it also needs to be embedded in the business and clinical workflows and which we all know, still requires, in order to get value from machine learning and AI, you still have to do good data and data management, but this doesn't replace the need to do that.

Lee Pierce: Personalized analytics, I think, is something that I hope, for myself and for my family and loved ones and for all of healthcare that instead of focusing as much on populations that the analytics can be applied to each of us individually and make personalized recommendations based on data on us and help us manage our own health. I think that's something that is a big deal for our future. And then cloud, not just the data warehouse, but our ETL tools, data governance tools, reporting tools, I think leveraging the public cloud is going to be a big deal.

Lee Pierce: And, the next slide is a slide that my boss at Sirius has published and it's just seven reasons for using the cloud that I think is a great list of reasons. Elasticity, you think of the high concurrency querying and you can do it when you need to scale up, but you only pay for the actual usage. Experimenting, if you can spin things up quicker to do proof of value kinds of things, I think that's going to add a lot of value.

Lee Pierce: Agility, it's ... you don't have to wait for procurement to acquire the servers and wait for the SA teams to spin that up. That's part of the ... helps with the experimentation, also. Gravity, the data can stay closer to customers where it's created. You don't have to move it all around and I think more and more data is going to natively be created and stored in the cloud and so having, I think, our analytic solutions in the cloud, it just makes sense, close to where it is.

Lee Pierce: Focus on the business. The software as a service model really can commoditize a lot of those lower-level tasks that our teams get bogged down with and the less of that, those maintenance tasks, system administration kind of tasks required, I think, more time we can spend on the business. Continuity, lower costs, scalable backup system, there's no doubt there, because I ... man, we've spent a lot of time over the years having discussions about backing up of the data warehouse and how it's done and how frequent. We just need to let that happen, and I think cloud can support that; and then workload balancing. Excess capacity as needed, just spinning that up and down. So, I thought that was a good list just to share and hopefully it's helpful for some of those dialed in and listening today, just around reasons for using the cloud and I think those all apply towards data and analytics.

Lee Pierce: All right, I think that's it for me.

Dale Sanders: Okay, thank you Lee. Shakeeb, you're up.

Shakeeb Akhter: Great, thanks Dale.

Shakeeb Akhter: It's good to see that a lot of things that we need to cover have already been talked about so I can be short on time.

Shakeeb Akhter: I think, what did we do right at Northwestern? I think a lot of that's related to what Dale talked about and I feel the same way, as well. We utilized Microsoft for database, SSIS, reporting, and analytics, incorporated that entire BI stack for those tools and that helped us quite a bit with not chasing the next shiny object or tools that's on the market and really helped us lower costs by using the native functionality within those Microsoft tools that could meet 80 or 90 percent of our use cases, which is the majority of what we were doing on a day-to-day basis.

Shakeeb Akhter: So, I think that helped us a lot there. I think, tangentially related, too, from a people perspective, that also helps because you don't have a proliferation of

different ETL tools, analytic tools, reporting tools that you need people with different skills to be trained up on, or know a variety of these tools, so I think that was quite helpful.

Shakeeb Akhter: I think we are continuing to do more and more of that and echo some of the comments that we made around Microsoft stack and I think ... I'll talk a little-bit about this later from our cloud strategy, but we've seen a lot of growth with Microsoft. We continue to be very invested in their platform; we continue to kind of partner with them for our cloud strategy, as well, and I'll talk a little-bit more about that.

Shakeeb Akhter: Very similar to kind of what Lee and Dale mentioned as well, I think one of the really good decisions that we made is not buying into the hype of new architectures and products early. Particularly around two things that have already been mentioned, which is MPP and Hadoop. I was also presented with this, we at Northwestern, conducted a kick off a EDW modernization exercise of modernizing our entire EDW infrastructure a couple years ago. We were going through a new EMR implementation and took stock of moving from Cerner to Epic, took stock of how are we delivering analytically the enterprise and what are things we need to do and took on the initiative of revamping a bigger warehouse as well as all of our analytical tools, reports, rewriting all of those and systems over the last couple of years, so we're just emerging out of that.

Shakeeb Akhter: As part of that, what was presented with a lot of options was just to the Massive Parallel Processing architecture as well as Hadoop potentially being your Enterprise Data Platform. I think that there was a lot of pressure for that, because they were the pricey thing on the market, particularly around moving ... was mentioned as an appliance. Moving ... I was caught in the middle because we're in appliances today, but a lot of that workload is moving the Cloud. So, do we invest in appliance, or do we invest in where the market and vendors are going, which is the Cloud? I'm very happy to say that we chose the latter and that we've stuck with our SMT architecture on REM.

Shakeeb Akhter: We upgraded people from SQL 2012 to SQL 2016 which itself provides a ton of additional benefits from indexing, optimization and performance perspective. But, then, we're raising Cloud for our advanced analytics workloads. I think we're very happy with that decision.

Shakeeb Akhter: Then, in modernizing the platform once market was mature, so again related very much into that and learning from others that have invested in appliance and others early, that really wasn't a huge return on value and waiting until market was a little-bit more mature where we could differentiate between the different lenders that are out there and pick the right strategy with something that's can be a big benefit for us.

Shakeeb Akhter: What would we have done differently? I mean, I think one of the things I think about is modernizing the platform slightly earlier. I think this is tangentially related to a lot of the comments that were made earlier around late binding,

middle binding, early binding. One of the things I think that, as Dale mentioned, that at Northwestern really took the ELP approach, really late binding and loading as soon as possible to begin with and then doing the transformation as the new stages came in.

Shakeeb Akhter: That was great at the start. I think over the years, when I particularly had the data warehouse, we had so much data in there and so many source systems that are taking up a lot of the prime real estate of storage and our sequel server, that performance was really taking a hit from that perspective. And, so, what we would like to do is modernize the platform slightly earlier and really put a lot of the data management tools in place and governance in place and say that if the data is going into the data warehouse, then there is a ready use case for us. And, that's why we're investing time in modeling and designing something that's going to be a scalable solution that's going to be used multiple times and continue to add value instead of binding late and then getting a lot of data into the data warehouse.

Shakeeb Akhter: I think that's definitely something we learn as we continue to embrace in the future. And, I think, in the second point we really want to focus on this because them more and more we think about leveraging data historically, and particularly in healthcare, we've been stuck in this leveraging a lot of our data for a ton of things; there's diagnostics, analytics, there's descriptive analytics as what it's done has been done in the past. As we move more into a lot of the trends that we're seeing, which I think are predictive analytic, tell me what's going to happen; prescriptive, dual networks, and particularly around the real data propositions and things that we're seeing is embedding the analytical insight from the data into the work wall.

Shakeeb Akhter: In my opinion, in order for physicians to trust that data, we really have to spend a lot of time doing data governance and data quality and make sure that the decisions that are being made off of that data are being made off of credible data, which we can prove. Evidence based medicine and also evidenced based data quality. So, I think, we're going to be spending a lot of time there and doing that a little-bit earlier which helps mitigate a lot of data that goes in the DW, but then also continue to monitor the credibility of the analytics products that we're sending out to physicians.

Shakeeb Akhter: What are we thinking about in the future? I think one of the things that we did with the modernization of the data warehouse, not just the infrastructure, but also, we thought about the data model, which Dale and me talked a lot about. I agree with a lot of things that were said. We spent a lot of time over the last couple of years to really think about what we call integrated data structures and if you look at our health systems, or healthcare data. These integrated data structures that are very well known on what you need from an analytic perspective, at least the foundational ones: patients, providers, diagnoses, labs, procedures, bio, those types of things.

Shakeeb Akhter: We spent a lot of time over the last couple years really building foundational data structures that integrate data across our current EMR as well as our historical EMRs. I mentioned this a little-bit, previously we had one EDW for research and for clinical and operational analytics and those are very different use cases, so we have to be able to model the data that's historically will make that and present that in our integrated data system as well.

Shakeeb Akhter: So, what we're doing now is, now that we have the foundation laid out, we are now building on top of those integrative structures to really enable self-service analytics. My vision for that here is really to build subject area specific data marts on top of the integrated data charts. So, a data mart for ED where we are integrating ED data and generating a cube on top of that, that would give them the ability to do self-service analytics. I still think that's in its infancy in health care by having self-serve analytics, so really a large focus throughout all our business units have self-service focus unit schools that then plug them into data that's been very nicely curated for that. So, investing a lot in that.

Shakeeb Akhter: The other thing is, extracting value form NLP. There've been a lot of statistics, and 80% of the value in healthcare data is in the text, in the notes, etc. And, at least here and thinking about the community abroad, I'm really not seeing very sophisticated use systems for extracting value from the notes, yet, that have been permeated and adopted into the workflows, clinically. And, so, we're really spending a lot of time thinking about how we're going to do that, because there seems to be a lot of value there.

Shakeeb Akhter: And, also, really thinking that some of the predictive models that we've created around readmissions and others, and presenting them in a way in workflows that increase adoption. The question we are always asking ourselves is, "That's a great predictive model, it's very accurate, what are the changes in outcomes from the model?" And, that really gets to answer the question on how much it's been adopted and how widely used is it? So, we're really focusing a lot on that.

Shakeeb Akhter: Everything that has been said for the cloud in the conversation, I completely agree with. Our strategy has been, we are sticking with Microsoft, we have invested heavily in, we're going to be using that for big data workloads, if you're on the cloud yet, you have some way to get there, out of there, some of the reason you just mentioned on the previous slide, It really is sealable, flexible, I would say more secure and we're embracing that for a lot of our big data workloads, as well as, really using that as our disaster recovery.

Shakeeb Akhter: Cold storage, we're taking a lot of data out of our production sequels driven data bases and putting it into cold storage. Really, the 80/20% rule: 80 percent of the data in the EDW may not be being used, so how do you get that all to a lower cost for storage. Really thinking about that, and Azure is going to play in for us in advanced analytics and these other areas in a big way.

Shakeeb Akhter: The last thing is the formal data quality program. We're definitely launching that and thinking about that. I think what we're seeing is that we really want to

move the needle on machine learning and getting an AI, getting some of those things embedded in a data proven culture at Northwestern, and really be a part of the way we think. But, in order to do that we really need to have good data quality checks and balances in place. So, we're launching a data quality program plan during the next fiscal year to really focus on that area.

Shakeeb Akhter: And, the last couple slides, the traditional data warehousing approach, is I think most of you have talked about is that data sources are being transformed into very stark views of use that are very locked down and then you're writing reports and dashboards all for that. There usually 24 hour to 48 hour delay. I think that's where a lot of the traditional data warehouses have been.

Shakeeb Akhter: I think, if you go on to the next slide, with the modernization, this has really been ... what we've been thinking about is what's important to us. And it's really having our structured data and our non-relational data sources that we know are already here, but we are not utilizing; Fitbits, apple devices, IOT and all those other sensors, social media. I really think getting those data persons to sit next to each other in a variety of different data management platforms. But, I think one of things we see on the top right of the slide is the Single-Query Model. We don't expect our end-users to know Sequel, and then a bunch of other syntax to be able to query non-relational and relational data. We want that to be seamless for them and I think that's where for Microsoft, Polybase and Sequel has really helped us. And so that was a big driver with us in adopting that.

Shakeeb Akhter: As our platforms for the future and really having to do the data sources side by side, and easy to read, we talked about that. Not modeling everything that goes in EDW and then leveraging cloud driven performance for a lot of the analytics, we won't have a ton of overhead infrastructure costs. Plus, you're not wasting time on a lot of the maintenance on the cloud form, which were focusing in on the business value that you can generate from the data.

Shakeeb Akhter: I think that really sets us up well for data science, which is ...

Shakeeb Akhter: I think that really sets us up well for data science which is does a couple things right now in the towns we have that. In the traditional model, you're moving ... if you ask your data scientist, you're moving data to a lot of different places in order to conduct the data science and a lot of time has been spent on that versus just making the tools available where the data already is, and I think that where I think the cloud space will be a huge tool for us with all the software applications are readily available where your data already is. And, so that's kind of my thoughts.

Chris Keller: Great. Thank you, Shakeeb. Great update. Thank you.

Dale Sanders: Thank you to all three of you. Chris, who's got a poll for us.

Chris Keller: That's right.

Dale Sanders: Oh, this is the marketing poll. This is where-

Chris Keller: This is the one you love.

Dale Sanders: This is the one I don't love.

Chris Keller: Excuse me. The reality is our webinars are focused on education.

Dale Sanders: Yeah.

Chris Keller: However, we still have some people who want to know more about our products and services. If that's you, please take a moment, answer this poll question. We're going to go ahead and move on to two other things. We want to let you know about our Healthcare Analytics Summit that happens in September. We have a great line up of speakers. I'll show that slide in just a moment. And we also have an upcoming webinar on value-based care and contracting, bundled contracting specifically.

Chris Keller: Dale, why don't you go ahead and move into-

Dale Sanders: Q and A?

Chris Keller: Some of these questions?

Dale Sanders: Sure. You bet. We have some questions from the first webinar and I think some of these are still pertinent so let's take a shot at some of these really quick. This is from Johnny Bui a former colleague at Northwestern actually, a teammate there. What are your thoughts on getting the analysts on the EMR side of things more exposure to what analytics can provide? They most likely are closer to provider users on a daily basis.

Dale Sanders: Yeah, and I think Johnny's right. I think what he's suggesting here is that you have the folks who take care of the EMR are typically a separate team from the folks who take care of analytics who are a separate team and, in fact, many times there aren't any teams for software engineering and application development in most healthcare systems. So, yeah, if I were still a practicing CIO, there'd be a lot of cross fertilization, bringing those teams and those functions together so that analytics gets driven to the front end of care and is ... decision support is more helpful to the front end of care than what it is right now.

Dale Sanders: Let's see here. Wes Grimes asks, do the central model analytics report to IS? So, Wes is asking organizationally, do the central model analytics report to information systems? That was always the case for me because I was a very data driven CIO. I would say across our clients, it's not common. I'd say maybe 30% of

the time analytics reports to the CIO and two thirds of the time it doesn't. Lee and Shakeeb, what's your observation there?

Shakeeb Akhter: That's an interesting question. I think in Northwestern we would see that analytics reports up to the senior VP and the senior VP of analytics, so that's not within IT. And then the EDW team rolls up to the CIL. So we have a little bit of a different structure then I'd say most places do.

Dale Sanders: Yeah. Lee?

Lee Pierce: So, at Intermountain we forged, say, most recently a data and analytics services function where you had data and analytics together and less important was the direct line reporting relationship. More important was the definition of the data and analytics program and how everybody works together. Although, we also have been working towards having the data analyst community also change their reporting relationship. But that was going to be outside of IT. So, it has traditionally been with an IT, but in IS, but is now outside of IS at Intermountain Healthcare.

Dale Sanders: Got it. Thank you guys. We have a question from Mark Morrison who's on the call today. It says, I believe he and Shakeeb both mentioned in the first that they have about 10% monthly users market penetration in adopters of the platform at Intermountain and Northwestern. So this is going to be a tough question to answer, but we'll take a shot at it. What are the main clinical indications that those users are going after? Lee, do you want to take a shot at that first? What were the primary clinical-

Lee Pierce: Sure.

Dale Sanders: Drivers, you know, the users there at Intermountain?

Lee Pierce: Well, you know, our analytic staff for one of the early drivers of the need for data management and analytics was clinical quality improvement or what Intermountain referred to as their clinical programs and that has continued over the years. And, so when you think about what are the clinical drivers, the way it was organized is by light clinical conditions. So you had a cardiovascular group that had clinical leadership and also data analysts, statistician, and they work very closely in with the data warehousing team and figuring out what they need to work on. And that was repeated for originally I think it was eight clinical programs. And I'm not sure of the number today, was up to 10 or 11 at one point. But, the clinical questions, you know, are driven by the clinical leadership by cardiovascular oncology, primary care, women and newborn, all those- in those areas and would change from year to year depending on which clinical condition they were focused on.

Lee Pierce: So, it's hard to say here are the top ones. Although, they did focus on those that were the highest, frequent highest cost kind of procedures that had variability that they were trying to drive out. So, that's how I generally answer that.

Dale Sanders: Shakeeb?

Shakeeb Akhter: Yeah, I think from my perspective, very similar, more of an evolution. I think one of the big drives was just for that where are the clinical quality programs and the reporting related to patient outcomes for your regular care oriented programs. It started with clinical use, importing CMS, so that drove a lot of our clinical quality focus area. And then, more so lately, in population health where we've really taken a much more broader approach to develop registries across condition that track patient's time and their outcomes in the system.

Shakeeb Akhter: And so, I think it's been an evolution. It's not, it's hard to pick the top five drivers. But, really has been just the clinical quality improvement and the evolution of that program and analytics tracking side by side to make sure we have the data needed to be able to support reporting related to those programs.

Dale Sanders: Yeah. And what I see across the industry now is more an analytic use cases being driven by reimbursement risk which at Intermountain and Northwestern when I was there, that wasn't as big an issue. But, now ... which I actually kind of worry about because I think in many cases, the industry's being driven by these quality process measures and less about the clinical quality and cost measures internal to the organization, which Intermountain was always very good at.

Dale Sanders: And, so now it's all things that provide sort of, anything that contributes to a never event, for example, is driving a lot of the use cases: CAUTI, CLABSI, sepsis, readmissions, that kind of thing. Patient safety issues are on the rise because nobody's getting reimbursed for those mistakes anymore which is good. That's a good part of the focus, but yeah. There are about 20 ... I mean we could publish these. They're out on our website. There are about 20 clinical conditions that account for most of the patient volume and cost in healthcare. So, it's not a big mystery.

Lee Pierce: Yeah.

Dale Sanders: But, it's funny how much the industry tends to ignore those things. I'm also starting to see more adoption of things like choosing wisely. Like, in general, how do we avoid over treatment for patients where there's no evidence that indicates that treatment is advocated.

Dale Sanders: So, by the way, we're at the top of the hour. I can spill over a few minutes if Lee and Shakeeb, you guys if you have other meeting you can drop. If you want to stay on that would be great. I can spill over a few minutes as we still have 135 people on the line.

Lee Pierce: I probably can for five or 10 minutes, Dale.

Dale Sanders: Okay, great friend. Sounds good.

Dale Sanders: Heather Hohenstein asks, making a career pivot from teaching at a community college with a master's in MIS to hands on data analysis, she'd like to get into healthcare. Where do I get a foothold? Learning SQL, Tableau, Python, presently she is. And this is a broader question, and this is how do we deal with what is the incredible shortage in data analysis skills in healthcare? And, in particular, data science skills?

Dale Sanders: And my short answer to this is go first to Coursera. Coursera has great online, very affordable, in fact sometimes free courses on healthcare related data analysis and machine learning. Johns Hopkins, their entire program is open. So go first there, friends. That's a great way to jump start your knowledge. It's not enough, but it's a great way.

Dale Sanders: Lee and Shakeeb? Lee, why don't you go first, friend?

Dale Sanders: Did we lose you? Lee?

Dale Sanders: No comment from Lee. How about you, Shakeeb?

Shakeeb Akhter: Yeah, I guess my number one thought as well is really going to Coursera and to some of the MOOC that are out there. A lot of this is very open. You know there's also edX and some courses from-

Dale Sanders: edX-

Shakeeb Akhter: MIT and Harvard. Some of these areas that have been a really, really valuable and very low cost. And I think it's a great way to kind of start with your skills. And then I would say, as an add on, maybe pursue some of the certifications which combine a couple of those courses so you can show your mastery of the subject as you gain more skills to perspective employers.

Dale Sanders: Yep. Okay. Next question, what do you think about ontology based data integration? Hasan asks that. I think it's interesting theoretically. You should think about. But, I think it's a bit academic and if you buy into it too deeply, I think you have the chance to bend yourself into the ground. But, it's interesting to keep it in mind and not ignore it.

Dale Sanders: And, by the way, I think we've lost the audio for Lee so I'm going to defer to Shakeeb right now. Shakeeb, go ahead.

Shakeeb Akhter: Yeah, it's a good thing to keep in mind, I think. We haven't invested as much time in doing our data integration based on that, but I think there's certainly a large value in keeping the data aligned to one of those standards and ontologies

In particular areas, education, et cetera that is very large value add for data. So I think we would try to keep that close for medications and some other areas where we know there's easy ways to reference standards ontology that has a lot of values. So, we think about as in pockets where applicable for analytical use cases but, not more broadly for all data integration.

Dale Sanders: Yeah. And I do tend to still advocate a late binding approach to vocabulary binding. So, you know ... you should, when necessary: when you need to report externally, or compare across organizations, that sort of thing, then certainly mapping to all the standard ontologies is important. But, don't assume that you have to do that in an early binding fashion.

Dale Sanders: David Bertoke asks, what are the key factors for self-service analytics to be successful? Oh, and Lee says he's back online. Okay. Lee, what are the key factors for self-service analytics to be successful in your mind?

Lee Pierce: I would say the first key to success is understanding who it is you're trying to provide self-service to because self-service is kind of a loaded term. I think you need to understand if you're trying to provide that to some of your power users and giving them access to tools and data so that they can produce analytics for others to consume in the organization. Are you trying to provide self-service to kind of a mid-level user group that doesn't have the technical skills, maybe the SQL knowledge or understanding how to build things in Tableau or CLiQ, but you need to provide them tools to do some drag and drop. That's a different kind of skill set with different kinds of tools.

Lee Pierce: So, I think that's one of the biggest mistakes when it comes to self-service is just trying to ... well, we need self-service let's find a tool. But, not identifying who the audience really is, and what their skills are, and what they are trying to accomplish.

Dale Sanders: Yep. Totally agree. I use a metaphor of a grocery store that is ... you know, first you have to give people an easy data store to shop on their own. Right? So, part of the self service begins with the ability for an analyst or a knowledge worker to shop for data on their own without asking IT.

Dale Sanders: We tend to move immediately towards the visualization tools, but the reality is we want to give them the ability to shop around, identify through metadata tools, and essentially create their own shopping cart of data. Right? So, shop around, expose the content just like you would in a grocery store, give them the tools to create their own shopping cart of data and then, as Lee suggests, work with them individually about their cooking skills. Right? Not everyone needs to be a five star chef. Not everyone wants to be a five star chef. Right?

Dale Sanders: So, work with them individually to give them the tools to allow them to cook and prepare that data in a way that meets their needs and their skill level. And if you do that to boost up their cooking skills a bit then have some classes.

Sponsor those cooking literacy classes. But, starting with the tool first is a ridiculous way to go about this-

Dale Sanders: And I've seen it happen time and time again. It's not about the tool. It's about having multiple tools and it's about providing a shopping cart experience for the end user.

Lee Pierce: I like that analogy.

Dale Sanders: How about you, Shakeeb?

Shakeeb Akhter: Yeah, I think those are great points. I totally agree with them and I think the only thing I would add to that is also understanding the how the data is used has been really an eye opener for us. You know, meeting with the end users ahead of time, giving them access to data so they can discover, and then tell you how they're planning on using that data is really going to decide what kind of architecture and product is best suited to deliver for them. And I think that's going to keep us understanding how they use the data in their day to day so that we can build the right technical solution for them.

Dale Sanders: Yeah.

Lee Pierce: Yeah.

Dale Sanders: Okay, we have time for two more quick questions then I bet we have to roll off. Simon asks here, what are the success stories in healthcare and, interesting, and what is known as discounted added values? I'm not sure what that means, Simon, but ... Let me point you out to success stories. So one of the things we've, incredibly passionate about it at Health Catalyst, is making sure we add value to what we're doing and return on investment. We have somewhere around 150 published success stories out on our website and those success stories are not marketing material. They have to be validated and approved by clients, audited by clients before we're allowed to publish them. That's a cultural belief that we have.

Dale Sanders: So, if you're new to healthcare analytics, you're wondering about use cases and how to apply it, go out there and browse those 150, 160 success stories and I think you'll see a lot of role model activity out there with our clients.

Dale Sanders: Let's see here. What does your data quality program look like? Brian Young is asking that. Structure, resources, position, and org chart, service lines, etc. Shakeeb, you want to take a shot at that?

Shakeeb Akhter: Sure. I think that currently what we're doing, Dale, is we're working through what that program would look like and our thought, making sure kind of what our goal, our vision for that is essentially that we're looking to create a feedback loop, a process that potentially in an automated fashion can check the quality of

data in the data warehouse. And that there are data stores that are assigned to managing and monitoring the quality of the data that they are able to explore. And then from there, once we identify discrepancies and any data quality issues within the data, we want a process that essentially has a feedback loop to the upstream application team. In most cases, to get that data quality issue resolved.

Shakeeb Akhter: And so I think what we're thinking about is how do we create a closed feedback loop process that essentially has, visits users that have stake in the quality of the data and are actively managing it through, you know, text tool and then identifying discrepancies in an automated fashion. And then three, getting those informed to the application team and getting them resolved in the actual sources. I think those are kind of the keys to the program that we're thinking through. We haven't thought through kind of the sources and org structure just quite yet.

Dale Sanders: Lee, any comments on ... data quality?

Lee Pierce: Sure. Yeah, so data quality is for us at Intermountain was just one of the functional areas in data government. And so, we initiated about three years ago a data governance program, data governance office with one of the first areas that we needed to build out and mature was data quality. And so, the data governance director ... we hired a manager who was just accountable for data quality and that individual spent time building the best practices and the concepts using quality improvement kind of concepts that are already inherent at Intermountain and applying that to data quality and improvement in close back loops like you talked about, Shakeeb.

Lee Pierce: And then, that has to sit alongside the development of and enabling of data stewards so they can do their part of the data quality job. We did hire, a few months before the end of the year, one full time data quality analyst that worked for the manager of data quality who was helping with specific projects. So would run the data profiling and help the data stewards with their projects. And I really think that the foundation that was laid is going to be really successful.

Lee Pierce: That all reported up to me as the chief data officer within the organization. And, that's how we approached it. Still a lot to learn as we go forward, but I think this foundation is something ... The manager built a best practice, a kind of a data quality workbook that I think is part of that foundation to help educate on why data quality, and how to do it, and so that was really, really helpful and what I would add.

Dale Sanders: Great. Okay. Well, I think we've reached the end of the webinar. Can I say one more thing about technology that I forgot, friends? I'm going to fold back to technology for just a second.

Dale Sanders: We don't have enough data in healthcare yet to offer personalized and precision medicine. If you look at the data that we have, it's only a sampling of data every three times or three times per year, generally for patients. That's not enough data to really understand at a personal and precise level the patient at the center of care. So from a technical perspective, all of us as data professionals have to start pushing our data collection outside the four walls of healthcare delivery: the data that shows up in supply chain finance, accounting, EMRs. We've got to bathe the patients that we deal with in sensors, passive sensors, and data collection about them for the 362 days of the year that we don't see them.

Dale Sanders: So, I just want to advocate to everyone. I should have mentioned this in our slides. We're creating an initiative in Health Catalyst I call Digitization and the Patient where we're going to build out and we're going to drive the sensor market based upon the analytics that we need to improve outcomes for patients. I would advocate that all of you need to do that. We've got to move outside the four walls of healthcare delivery.

Dale Sanders: Okay, that's the last of my comments. Thank you, Shakeeb and Lee. Thank you, guys.