Modern Data Strategy

cloudera

Modern Data Strategy - It's Not Really About Big Data.

Yes, "big data" is already a cliché. But it's also reality.

If you're not already modernizing your data management infrastructure, you are laying the groundwork to be perennially behind—struggling to interpret this newer way of thinking about, and using, data to drive your business. Simply stated, a modern data strategy involves changing the way you think about data, and then plotting a course to effectively take advantage of it in many ways.

Every IT vendor today is telling you the same things: "Big data is critical. Big data is the most important trend of the decade. We are the experts in big data." In fact, every vendor will go on to say that they'll not only help you with a big data strategy, but that they offer a suite of products and services to help you implement it, as well.

But what is it, exactly? How does big data relate to your business today? Is it really just about volume, variety, and velocity? And, do you need to act now? This paper proposes a different perspective on big data and asserts that it's not the "what" of data but, rather, the "how" that really matters. It also argues that if you don't have a well-thought strategy, you're not going to get very far and will find yourself at a competitive disadvantage.

Data is a huge challenge. But it's also an enormous opportunity.

Computing as we know it has changed—major advancements in processing, networking, and storage have opened the door to a whole new set of possibilities for how to leverage data. Whether it's self-driving cars, automated financial planning, legal help, factory automation, medical diagnoses, or any number of other applications, data and advanced analytics are working their way into virtually every corner of our personal and professional lives. Machine learning, a practice previously reserved for highly technical applications, is making its way into solutions across industry thanks to these advancements in technology—and the possibilities are endless.

However, to take advantage of these new capabilities, you must first devise a modern data strategy. A modern data strategy is more than just acquiring data and the technology needed to store, process, and analyze it. In reality, it's a strategy inclusive of all the people, process, and technology changes you need to consider when transforming your business to become data driven. A modern data architecture is part of the solution, but so too is a clear grasp on the business problems you're looking to address, as well as a thoughtful plan for how you're going to drive transformation from start to finish, and beyond. Dismiss this fact and you're likely to struggle, if not fail outright.

Today's world runs on data. What does that mean to you and your business?

Think about your personal life for a moment. Whether it's obvious or not, data drives nearly everything you do, where you go, how you get there, and how you pay for it. How long could you survive without your smartphone? When was the last time you went inside a bank? Used a map? Talked with a travel agent? Waited outside in the rain for a taxi? Went more than a few minutes with a nagging question you wanted answered? All of these capabilities we take for granted are powered by data. In fact, data drives nearly everything in the modern world.

As individuals, we've come to expect these things. As a business, we strive to mimic and implement them. What if we could completely change the way we interact with customers? Completely change how we track and manage our supply chain? Completely change how we mitigate business risk? Leading organizations have figured this out. They've realized that through advanced analytics, they can change how they solve key business problems and, in doing so, are separating themselves from the pack. This is the difference a modern data strategy can make.

A Modern Data Strategy – The Key to Success with Big Data

To truly unlock the potential within data, a new approach is required. By now, it should be obvious that the dynamics of data has changed. The so-called 3V's of data—volume, variety, velocity—are driving organizations to not only rethink how they architect data management solutions, but also how they can leverage all that's locked up in that data to potentially solve problems in new and innovative ways. If you peel back the layers, a modern data strategy is nothing more than a well-defined "crawl, walk, run" approach to success with big data. It's easy to get excited about all the things you can

accomplish, but if you don't thoroughly think through how you'll get there, you'll struggle to make progress. Or worse, fail outright. A modern data strategy helps ensure success by breaking the problem down into manageable pieces—so you know precisely where you're starting, where you're trying to go, and all the steps along the way to getting there.

Think business first

IT teams don't just embark on technology projects for the sake of it. There's always a purpose behind it. The first, and perhaps most important, step in the journey to a modern data strategy is to identify the key business impact that will be realized from the end result. There are hundreds of use cases to which big data can be applied, so often times it's difficult to pinpoint just one. Fortunately most, if not all, businesses have a finite set of concerns, so identifying the business priority isn't as difficult as it seems.

Drive audience insight

Many organizations are focused mainly on one thing: making money. As markets evolve and become increasingly competitive, the business that wins is the one that uses data to its advantage when it comes to understanding customer behaviors—so as to build long-lasting relationships that lead to a generous revenue stream. This "segment of one" view not only allows you to understand past customer behaviors, but, more importantly, enables future ones—keeping you one step ahead of their next action.

True Corporation creates a true 360 degree view of customers with Cloudera

Build better products and services with data

Often times organizations are merely trying to drive excessive costs out of their business. Many are using data to develop entirely new products or services (like selling your customers comparative data on their own use of your products, or selling completely new services based on combining live, historic, and public data in novel ways). Predictive analytics, using machine learning, is becoming the norm, and it helping leading enterprises proactively notify customers of a problem that neither the customer nor the vendor had previously been aware of (at Cloudera we do that ourselves and with our own products).

Lower business risk

Perhaps one of the most powerful byproducts that comes from analyzing huge amounts of data is the ability to spot trends and anomalies that were not visible before. By looking at huge populations and vast numbers of events over very long periods of time, things come into focus in ways that weren't possible when you were only looking at a few hours, or a few days, worth of data. That new capability can be directly applied to problems like defending your network against cyber attacks, understanding the behavior of your customers more deeply, detecting insider trading, improving the efficiency of supply chains, and allocating financial risk in new and more sophisticated ways.

Modernize architecture to meet the changing dynamics of data

We previously asserted that it is not the 3V's of big data that matter, but that it is how you use that data that really counts. Well, that's not entirely true. There's no denying that the dynamics of data have changed. It's now easier than ever to get your hands on as much data (regardless of type) as you so choose, and do so as frequently as you like. Most companies still rely on systems that were conceived several decades ago—systems that are deeply entrenched, or systems that are highly reliable but hard to scale economically (or in a way that can cope with the modern world of data). Enter Apache Hadoop.

Hadoop has emerged as the de facto solution for managing the challenges associated with the 3V's of data. Hadoop's distributed processing and storage nature makes it cost effective and technically feasible to store, process, and analyze as much data as you so choose. Cloudera offers the fastest, easiest, and most secure Hadoop platform—Cloudera Enterprise—for modern data management and analytics.

cloudera®

But the 3V's aren't the only reason we modernize architecture. More profound is what we do with the data, and what architecture modernization enables—namely a completely new way to think about analytics. Whereas data and analytics teams were once mostly limited to looking backwards using historical snapshots of enterprise data, today we can not only look backwards to see what happened, but we can also understand why we're seeing those particular results in the first place. Machine learning is increasingly allowing us to take analytics one step further: predicting, or actually causing, outcomes to occur. Previously limited to just uber-technical domains and hindered by high-cost compute requirements, machine learning is rapidly coming to the forefront—thanks to rapid advances in networking, storage, and memory architectures. The end result is machine learning quite conceivably being applied to a whole host of problem sets we never thought before. With machine learning, not only can we predict what will happen next, we can actually be one step ahead of what's going to occur, and start to steer towards a desired outcome.

All of this begins by building an enterprise data hub with Cloudera Enterprise—allowing you to handle any data, anywhere, and do it while scaling analytics and data science to the masses.

Anything

In the era of big data, there's no reason why you should not be leveraging as much data as possible. This applies not only to the aforementioned 3V's, but also to how we interact with data. With Hadoop, it's possible to manage massive amounts of structured and unstructured data in a single environment that can be then be transformed by data analytics. Furthermore, this data can be at rest or in motion. Once thought impossible, you can now analyze data as it's generated, in order to make critical, near real-time decisions that can provide a business with a significant competitive edge.

Anywhere

We've long debated the merits of cloud, with many organizations interested in adopting, yet limited in their ability due to security or compliance constraints. In response, many are adopting a hybrid approach and easing their way into cloud by deploying only select workloads to minimize business risk. So it should come as no surprise that, when it comes to Hadoop, portability must be a priority. Most organizations don't want to be locked into one cloud vendor or another, and furthermore, want the flexibility to move workloads on premises as needed. Furthermore, security, governance and administration must be consistent across all environments so as to ensure corporate requirements for each is met. Hadoop makes this all possible.

Anyone

Data science is key to this entire equation. Unfortunately, there aren't enough of them to go around. Moreover, it's often hard to put their work into production in systems like Hadoop without the need for highly specialized skills. When data science resources are scarce how can you make sure that they are working on the most valuable projects and their work makes its way into enterprise systems? CrowdFlower 2016 Data Science Report found that 83 percent of respondents said there was a pronounced shortage of data scientists, an increase from 79 percent reported a year earlier. (http://www.datanami.com/2016/03/25/tracking-data-science-talent-gap/). What if we could simplify data science such that more could do it?

Putting it all together: mapping your journey to success with big data

Imagine you're going to run the first marathon of your life six months from today. What would you do? Would you do nothing, go about your routine with no additional preparation and show up race day hoping for the best? Probably not. Instead, you'd create a plan—starting from today—that would help you get ready for race day. Your plan would assess your current fitness level and dietary needs and involve a slow progression for getting you to the full distance when the big day arrives. The plan makes no assumptions—and Big data, in a lot of ways, is a similar challenge. Assume you know everything and bite off more than you can choose, and you're likely to struggle.

It's true that introducing new technology can cause some disruption and lead to projects taking longer than desired. But in the case of big data, more times than not, it is the people and process factors that, if not thought through carefully, will result in more issues. What you need is a plan for how your organization is going to fill any gaps, how they're going to start small with meaningful impact, and still do so with a vision for long term success. In this sense, big data is very much a marathon.

Think globally, act locally

It's important to tie exactly what you're trying to do with data to what the business actually cares about. Maybe your organization is trying to grow revenue, eliminate costs, or minimize risks. Data has an important role to play in helping solve all of those problems. You need to be able to articulate the "why" of big data in these terms, otherwise, no one in the business is going to care. We talked about this earlier in this document.

Key to getting started, though, is making sure you're realistic about what you can accomplish short term. A use case discovery exercise is a great way to get going. Sit down with your broader team and talk about what it is you're trying to accomplish with your big data project. This will not only help you understand what the key business priorities are, but also what you're going to need to meet your goals.

Second, it's key to acknowledge that there are cross-functional concerns that need to be taken into consideration, such as data governance or data security. Oftentimes, decisions regarding these things involve multiple stakeholders, and with competing interests, so the idea of change can be a sensitive subject. Approached incorrectly, barriers are likely to arise, limiting your ability to make progress. Find an acute pain point in one part of the company and discuss how big data can positively impact that business function. Scoped correctly, success will be fast and real, making it easy to scale your efforts across the rest of the company when you're ready.

Assess your strengths (and weaknesses)

It's also valuable to assess the skills and maturity of the individuals on your team. Data science, for example, is a new domain that requires a unique and complex skill set—those of which are typically not available in most organizations. Identifying this early on and building a mitigation plan is critical. Are these skills you can grow organically, or do you need to hire in order to acquire them? What's the ramp time and when will the team be truly productive? Better to know the answers to these questions ahead of time, rather than finding out after expectations have been set.

How can Cloudera bring you into the modern era for data management and analytics?

Simply put, no one knows Hadoop like Cloudera. We can help you assess your business goals, help you figure out which of those goals can be achieved by using the power of data, and then build a system that actually works. And we'll do it in a reasonable period of time.

We're not claiming this is simple—implementing a modern data strategy with Hadoop, no matter what your vendor tells you, is not a trivial undertaking. But Cloudera has the broadest and deepest knowledge of the technology, the most comprehensive technical support and professional services organization, and the largest ecosystem of partners (well over 2,000 as of this writing). We also train more people than anyone on Hadoop development, cluster implementation, data science, and even on newer projects like Apache Spark. In short, our focus is on making you successful and cutting through all the hype. We know the power of big data, and we want to help your company take advantage of it.

The first step is to arrange for an initial meeting with our team to understand what you're trying to accomplish, and how we can help. After that, most of our customers want to do some form of proof-of-concept to get a sense of the power of our solutions.

Sometimes we'll want to engage with one of our systems integrator partners, or perhaps you'd rather engage directly with Cloudera's professional services team—either is fine with us. And, whether you want to deploy on-premise or in the cloud, we will support you 100%. Generally, our customers can get an initial solution up and running quickly, and a proof-of-concept project can often be live in a matter of a couple of months.

 Lastly, in order to improve the likelihood of success, we strongly recommend that you put your people through training.. In fact, Cloudera trains more people on Hadoop than anyone else in the world. In addition to making sure you get the the most out of your Cloudera deployment, training will help you with both the recruiting and retention of the dev/ops people essential to Hadoop.

cloudera

A modern data strategy is within your reach

This is not a science fiction story. While you're probably skeptical of big data due to so much hype being around it, the fact is that you already rely upon it constantly in your personal life. Now it's time to bring your company into the 21st century when it comes to data management and analytics.

Visit Cloudera.com for more information.

About Cloudera

Cloudera delivers the modern platform for data management and analytics. Public sector organizations trust Cloudera to help them apply data to the center of their missions with Cloudera Enterprise—the fastest, easiest, and most secure platform built on Apache Hadoop and the latest open source technologies. Agencies can efficiently capture, store, process, and analyze vast amounts of data—empowering them to use advanced analytics to drive business decisions quickly, flexibly, and at lower cost than has been possible before. Focused on customer success, Cloudera offers comprehensive support, training, and professional services. Learn more at cloudera.com.