



Whitepaper

BIG DATA 2.0 Cataclysm or Catalyst?

A Leader's Guide to Navigating the Shift in Big Data

Four undeniable trends shape the way we think about data – big or small. While managing big data is ripe with challenges, it continues to represent more than \$15 trillion in untapped value. New analytics and next-generation platforms hold the key to unlocking that value. Understanding these four trends brings light to the current shift taking place from Big Data 1.0 to Big Data 2.0, a cataclysmal shift for those who fail to take action, a catalyst for those who embrace the new opportunity.

The world has changed.

We now live in the digital age. Everything is made up of bytes and pixels. Billions of devices, sensors, and apps create a constant flow of data. All things digital are now connected by the Internet of Things. People now use devices with hundreds of apps utilizing all of the sensor technology embedded in the device and creating new data and new patterns just waiting to be discovered.

The world will keep changing.

The only thing we know to be constant is change. Two elements drive unprecedented change. First, the entire world is in flux. We are in uncharted territory and no one really knows what might happen next. Second, technology innovation is exploding on four vectors all at the same time. The internet connects everything, everywhere. Mobile device adoption outpaces traditional desktop and laptop computing. The cloud is taking over the corporate data center. And big data continues to explode. Change will no longer be measured by seasons or eras. It will happen constantly at a much faster pace than ever before.

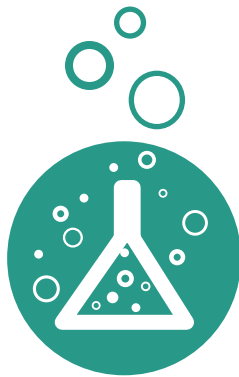
Time is the new gold standard.

As the speed of change continues to accelerate, time becomes the most valuable commodity. You can't make more of it. You can only move faster or use it more wisely. Everyone expects instant access and immediate response. Speed becomes an unfair advantage to those who possess it. Anyone who wants to survive in the new age must combine speed with intelligence in order to come out a winner.

Data is the new currency.

You've heard it said that data is the new oil, but we say that data is the new currency. He who has the data wins in the end. With the dropping cost of storage and new technology designed to capture, store, and analyze data at ever decreasing costs, companies are holding on to more data than ever before. Hidden in the data are the secrets to attaining more customers and keeping them, shaping their behavior, and minimizing enterprise risk at levels never before possible. The combination of massive amounts of data and new analytic algorithms creates a new global currency, more stable than the dollar or the pound or the yen.





The growth and adoption of Apache™ Hadoop continues to astound the market-place. From its first release in December of 2007 until now the widespread distribution of the new data platform has exploded thanks to the open source community and companies like Hortonworks, Cloudera, and MapR. This first phase of big data brings us 5 significant achievements that change the way organizations all over the world approach data and analytics:

1. Enormous, affordable scale compared to old storage paradigms
2. Capture and store all data without first determining its value
3. New types of data present new opportunities for analytics
4. Data discovery and data provisioning now common in most organizations
5. Analytic applications emerge as the number-one user of big data

Big Data 1.0 – A Laboratory Science Experiment

It's in the midst of these four trends that we have exploded into the age of big data. After years of big data hype everyone wants to know the state of the union. Big Data 1.0 has been all about introducing new technologies to take advantage of all the new data being created. One of the frontrunners that has emerged with great promise is Apache™ Hadoop.

From a 50,000 foot view, Hadoop has gone from being an open source phenomenon like no other to becoming a common household name among technical and business people alike. Big data has evolved into something beyond what we would call a market; it's a movement, a global movement.

5 Achievements of Big Data 1.0

Enormous, affordable scale compared to old storage paradigms

The traditional storage market has been under attack from above and below. From above, cloud vendors like Amazon, Google, and Microsoft use cloud storage volumes to drive down the overall cost of simple storage. From below, HDFS (Hadoop Distributed File System) now offers a simpler way of storing data than the higher cost options provided by NAS, SAN, and databases.

Capture and store all data without first determining its value

With the cost of storage at bargain prices, everyone from the phone company to the government decided to keep data that had always been thrown away. Energy companies keep their historian data. Log files mount up at never before experienced rates. And new data from sensors, social media, and the Internet of Things has found its way onto Hadoop storage systems everywhere.

New types of data present new opportunities for analytics

All of the new data being amassed in Hadoop represent new opportunities for analytics. Mobile apps produce huge amounts of behavioral data that can be used to drive hyper-segmentation. Social data provides never before captured context to what is happening around markets, brands, and segments. And sensors create micro-snapshots that can be pieced together into valuable predictive models.

Data discovery and data provisioning now common in most organizations

New storehouses of data pouring into file system storage have spawned new practices around data discovery and data provisioning. The first order of business around big data has been to discover what is actually in the data. The next order has been to provision that data to the different people and applications that need the data.

Analytic applications emerge as the number-one use for big data

For everyone involved in the big data movement, there seems to be consensus around what to do with the oceans of data that have been collected and categorized. Everyone understands that analytics unlocks the value within the data.

5 Challenges of Big Data 1.0

Even with all the strength of a global movement, most big data projects are stuck in the laboratory. It is still a science experiment in need of production. Millions have been invested and the executives behind the funding are still waiting to see a return on their investment. So, what is holding back the advancement of big data? Big Data 1.0 unearthed 5 big challenges that need to be overcome in the next wave of the movement:

1. Extremely complex to integrate, deploy, operate, and manage
2. Requires specialty programmer skillsets to deploy the system and process data
3. Lacks enterprise-class capabilities like security and high availability
4. Shortage of data science skills
5. Performance lags behind the need for immediate response times

Extremely complex to integrate, deploy, operate, and manage

The Hadoop open source organization has spawned a myriad of component projects being run by diverse groups all over the world. It is difficult, at any given time, to determine which of the projects will graduate into the code base and which will fail. The complexity of the pieces parts combined with the diversity and explosion of data make integrating, deploying, operating, and managing massive Hadoop environments difficult. There is no such thing as an out-of-the-box big data solution.

Requires specialty programmer skillsets to deploy the system and process data

Many companies who initially stood up their Hadoop clusters thinking they were going to run predictive analytics on their new platform have abandoned that pursuit. When they saw the amount of MapReduce, Java, or Python programming that it was going to take to deploy the system, they scaled back their effort, focused Hadoop on data provisioning, and brought in analytic platforms to do the heavy lifting analytics. There are not enough resources available in the world to fulfill everyone's big data dreams and that is not going to change any time soon.

Lacks enterprise-class capabilities like security and high availability

Even though big data platforms have sprung up in most companies, they have stayed out of production because of the difficulty of protecting the data and keeping the systems up and running. Even big companies like Google and Facebook, companies that run their operations on Hadoop, won't let anyone or anything near the clusters for fear they will bring the entire system down. While both of these issues are being addressed by the open source community and other vendors, they remain a barrier to production.

Shortage of data science skills

The top performer in the world of big data is the data scientist – the person who mines all the data to find the valuable nuggets to solve specific business problems. However, it is such a unique skillset that very few people actually possess what it takes to succeed in this role. Success takes a combination of mathematical expertise, business acumen, understanding of the new data types, out of the box thinking, and storytelling. Multiple studies, including a recent study done by McKinsey, have concluded that there will continue to be a shortage of data scientists.

Performance lags behind the need for immediate response times

Because Hadoop is a file system built on batch processing, wait times continue to be too long for the operational needs of most businesses. To compound matters, MapReduce was designed to search and find data, not to run the sophisticated analytics that companies expect from their big data implementation.

Big Data 2.0 – Transformational Value

Big Data 2.0 is all about releasing the \$15 trillion of value from below the water line of the big data iceberg. The race is on to provide affordable access to the 88% of the data that has never been touched. To move quickly to the transition to Big Data 2.0, you can expect six major shifts on the technology front to give companies the ability to move quickly, turn on a dime, and stay ahead of the competition, even in the face of relentless change:

1. Cooperative processing will deliver faster time to value and better price performance
2. Analytic building blocks will provide accessibility for non-skilled and less-skilled workers
3. Moving processing to the data will operationalize big data and push toward real-time
4. Combining non-relational and relational data will enable a richer set of analytics
5. Services layers will abstract away the complexity of underlying infrastructure
6. Unified platforms will provide modular approaches covering the entire analytic process



Cooperative processing will deliver faster time to value and better price performance.

The concept of cooperative analytical processing began to emerge in Big Data 1.0. In Big Data 2.0 it will become the norm. Everyone agrees that there is no one single platform that can handle all the workloads necessary to run today's complex organizations. Running workloads on platforms that weren't designed with that genre in mind will always cost more money. When companies implement Hadoop for data management and provisioning, traditional databases for operational workloads, high performance analytical databases for predictive and prescriptive workloads, they will naturally drive down the cost of computing and accelerate time to value. When these new systems support cooperative processing across the different platforms they will extract even more value from the data.

Analytic building blocks will provide accessibility for non-skilled and less-skilled workers.

With the shortage of data scientists written in the books, Big Data 2.0 opens up the world of big data analytics to non-skilled and less skilled workers. Analytic functions delivered through visual interfaces or embedded in databases become building blocks accessible to anyone via simple SQL calls or a drag and drop work bench. Users no longer have to know how to write sophisticated algorithms; they just need to know how they work and how to use them.

Moving processing to the data will operationalize big data and push toward real-time.

It is the concept of massively parallel processing that drives the operationalization of big data. Up until recently, only massively parallel analytic databases were able to move the processing to where the data lives, right on each node. Platforms like Actian Analytics Platform™ optimize and then compile the query to run on hundreds of compute nodes. With the introduction of YARN (Yet Another Resource Negotiator) introduced as part of Hadoop 2.0, products like Actian DataFlow™ now run sophisticated data transformations and complex analytics directly on the HDFS (Hadoop Distributed File System) node. This major breakthrough puts hundreds of Hadoop nodes to work and greatly accelerates the speed of computing.

Combining non-relational and relational data will enable a richer set of analytics.

The first slide of the Facebook presentation at the 2013 Strata conference in New York City stated, "Big Data Does Not Equal Hadoop." The second slide stated, "Big Data Equals Hadoop Plus Relational." When a top contributor to open source Hadoop put a stake in the ground around the need for both of these two platforms together, the world listens. Companies that want to excel in their industries and utilize their data in the most profound ways possible will implement new technologies that work together across both relational and non-relational data.

Services layers will abstract away the complexity of underlying infrastructure.

One of the greatest challenges of Big Data 1.0 was the complexity of data types and platforms required to handle all of the different workloads processed by most organizations. When you add in the complexity of running analytics using extreme SQL, Java programming, MapReduce, Python and other scripting or NoSQL languages, the complexity can be overwhelming. In Big Data 2.0 the complexity will be abstracted away using new technology like extended optimizers, common metadata, and visual frameworks that make big data analytics accessible to the masses.

Unified platforms will provide modular approaches covering the entire analytic process.

New unified platforms designed for Big Data 2.0 and beyond will cover the entire analytic process from connecting, preparing, and enriching data all the way to applying and combining sophisticated analytical functions. To support the continual expansion of big data projects and the growth of data over time, these new platforms will deploy in modules, growing by function, node, or data storage volume. In addition, expect modern data platforms to deploy on-premise, in the cloud, or in hybrid environments, all on commodity hardware. This modular approach will give way to adaptive infrastructure, where nodes can be deployed on one platform and quickly spun down and spun up to support a workload increase on another platform.

Big Data 2.0 Means Big Value Times Ten

So, what does all this mean for executives and companies who have invested in big data technology or who have an investment in their 2014 plans? The combination of capabilities being deployed in Big Data 2.0 should yield four points of distinct value for those who are willing to lead the charge:

1. Better price performance
2. Faster time to value
3. Continuous transformation
4. Unfair advantage

Better price performance

Expect newer technology to drive down the cost of computing. Older, traditional technology was not designed to handle diverse data or to crunch through huge data analytics. Because modern platforms run on commodity hardware and were designed for the data deluge they will cost you less and perform better. Get out your calculator or spreadsheet. If you don't see better price performance, don't buy it.

Faster time to value

By abstracting away complexity and speeding the processing of data preparation and analytics, you will accelerate your time to value. You'll cut data processing times in half and speed analytics iterations by magnitudes of more than 100. Not only will you get value out of your data faster, you'll be able to deploy new data services and analytic applications faster than ever before.

Continuous transformation

To keep up with the change around us, it's no longer enough to be able to transform your company. The speed of Big Data 2.0 will give you the innovation acceleration you need to transform your company over and over again. You will be able to use analytics different than your competition, disrupt your market, and lead your industry.

Unfair advantage

All of this creates an unfair advantage for your organization. When you combine better price performance and faster time to value with continuous transformation, no one will be able to catch you. You will become the next Amazon, the next Google.

Transformational Value

So, what makes next-generation technology transformational? It's the extreme speed, extreme scale, and extreme agility that technological advances bring to the marketplace. Truly innovative companies are moving beyond the confines of old technology and first generation big data platforms to embrace Big Data 2.0.

Think of technology like Actian Analytics Platform as the exoskeleton for Hadoop. Hadoop captures all the data and becomes the data operating system. It prepares and provisions data for anyone, anywhere. It enriches data sets with fresh perspectives turning ordinary analytics into a strategic weapon. Actian surrounds big data open source components with extreme performance, extreme scale, and extreme agility. It provides access to less skilled and non-skilled contributors. It opens the door to the \$15 trillion of value hidden below the water line.

In order to succeed in the new world that is continuously emerging, we need to do three things: accelerate everything, innovate constantly, and transform continuously.

Accelerate everything.

Every step of data's journey through next-generation platforms must be accelerated. We must leave no stone unturned when it comes to faster processing, better optimization, and quicker human interaction with data and machines.

Innovate constantly.

Since the only constant is change, we must constantly innovate. A single innovation will only set us apart or put us in the lead for a short period of time. We need more innovation on the tail of every successful innovation we take to market.

Transform continuously.

The old paradigm of enterprise transformation has given way to the new model of continuous transformation, where there is always some part of the organization being transformed. In many cases new transformation cannibalizes old parts of the business, but this is necessary to stay alive and to stay ahead.



It's Business Time!

With next-generation technology emerging to match the explosion of big data, it's business time. It's time for the business to reap the benefit of their big data investments. Visual frameworks combined with drag-and-drop operators are ready to plug into any big data reservoir and produce immediate value. High-performance analytics platforms are poised to accelerate the proliferation of analytics across entire ecosystems. Big Data 2.0 will be the catalyst for the next wave of Amazons, Googles, and Facebooks. It's business time for Big Data 2.0.

Action transforms big data into business value for any organization – not just the privileged few. Our next-generation Action Analytics Platform software delivers extreme performance, scalability, and agility on off-the-shelf hardware, overcoming key technical and economic barriers to broaden adoption of big data, delivering Big Data for the Rest of Us™. Action also makes Hadoop enterprise-grade by providing high-performance ELT, visual design and SQL analytics on Hadoop without the need for MapReduce skills.