# Data Analytics 2020 And Beyond

blazeclan

A Blazeclan Whitepaper

# Introduction

One of the fundamental elements of the Fourth Industrial Revolution is Data. It has descended upon us like a hurricane that changes the landscape, it passes through. The much-touted maxim ‘Data is the New Oil’ has caught on, not because it sounds fancy but it summarizes every organization’s dream – to have control over forecasting demand, optimizing supply, and more importantly, to stay in the limelight before the competition does.

While no amount of Big Data can act as a crystal ball to accurately predict the future, it can certainly help as a magnifying glass that can help read the lines in our palms. Having organized data is preparation, processing it meaningfully is progress, but leveraging it to guide decisions is priceless.

Even today, the aggressive analysis continues to haunt organizations, thus, slowing down their decision-making process at every step from whiteboard to go-to-market. The reasons for this can be attributed to barriers in information flow that are crucial for managerial decision making. But surprisingly, the data was available, but never processed in a way to distill needle-moving insights.

'How can organizations break the barriers to refine information from packets of data siloed within various divisions of an organization?' is the broadest consistent goal for every CXO.

# Current Trends of Data Management

While increasing adoption to Cloud has helped organizations to achieve agility, large companies still have challenges and queries around data security, data analytics, making sense of structured and unstructured data in unity, data governance, and more.

While it is already a mammoth task to deal with internal challenges to adopting a robust data management process, the external factors such as the introduction of GDPR in 2018 and changing the global mood on data privacy makes the process much more complex.

For many organizations, the adoption of Cloud is a huge step. But a mere migration to Cloud is just the first step in a very long strategic data journey.

Key trends for the organization to adopt a robust data management process must answer some of these fundamental questions.

- What is the role of clouds in collecting, organizing, analyzing, and understanding data?
- What common denominator do structured and unstructured data have in common that will yield a single telescopic view?
- What is the role of maturing data quality and its impact on the managerial decision-making process?
- How to manage external complexity and ease of adhering to compliances in a changing world?

Managing data complexities, both internal and external, has been addressed by cloud service providers such as AWS since the beginning. But what has now changed is the pace, scale, and the speed at which things move. The changing behavior of end consumer who wants everything in a digitally connected world has put organizations on the edge.

While AWS has unveiled several features that allow organizations to handle data, irrespective of size and complexity at ease, the changing dynamics in technology powered by Machine Learning and Artificial Intelligence warrants a much-advanced data management strategy, proven processes, and smart set of tools to implement them.

## *The Emergence of Data Lake*

The big data wave has hit firms in the past to help store and structure data in proper data sets. This technology is in existence for the companies to take business decisions. With Amazon constantly developing products that can add to business intelligence and ease towards decision making, it is expected that data lake is the future that is likely to revolutionize the field of data management.

At AWS reInvent, Amazon also announced 'Lake Formation' which takes the process of setting up a Data Lake from several months to just a few days while also offering prescriptive security policies and access controls.

It is estimated that there are more than 10,000 data lakes built on Amazon Simple Storage Service (S3) indicating that organizations value in consolidating data.

# Technological Challenges

Traditionally, managing large amounts of data by itself was a challenge. The complexities accentuated further as a direct function of the Volume, Veracity, Velocity, and Variety of data leading to a set of technological challenges.

## Computing Power

With unstructured data volumes exceeding structured data, the sheer computing power required is tremendous to move data from disparate places and make sense of it. The cost of running data centers with such advanced computing power is also prohibitive. Estimates indicate that a 25,000 square foot data center will use about $4 million in energy bills.

## Data Governance

With changing legal and policy frameworks, IT systems and processes needed to ensure Data Governance that can turn out to be Achilles' heel for the organization.

## BYOD, Interoperability and Data Lineage

With increased 'Bring Your Own Devices' (BYOD) culture in organizations, data is created across devices – laptops issued by IT, personal mobile and tablets used by employees, apps  that have access to enterprise databases, possess technological challenges such as how to ensure that the data lineage is not disrupted, is made interoperable.

# Scale Paradox Drivers

| Driver | Impact |
|---|---|
| Talent Shortfall | Forcing new thinking about how to leverage current talent and how to acquire more.<br>Driving new talent models of collaboration and crowdsourced problem-solving. |
| Democratization of Data | Businesses will likely have growing access to internal data sources and may need to link that data with external structured and unstructured sources to gain more relevant business insight.<br>The conversation should move away from "big data to "smart data"—focusing resources on identifying the right data and information needed to solve the problem at hand. |
| Ecosystem Collaboration | Historical divisions in the value chain are being blurred by the advent of more and more third-party data—and the need for all stakeholders to get closer to the consumers.<br>Many organizations are increasingly willing to share their information assets in order to get a holistic value-chain picture. |
| Strategic Imperative | Widespread recognition of analytics as a strategic imperative at the top of many organizations creates new opportunities to act upon data in new ways.<br>Organizations are beginning to become more open to challenging ideas that were once off-limits for analytics. |
| Analytics Tools | Analytical tools are becoming less expensive and more easily available to companies of all sizes.<br>In addition, analytics tools are more user-friendly and intuitive, allowing individuals within the business to perform a level of analysis and insight that was previously reserved for a select few. |
| Disruptive Technologies | Mobility, cloud, and social are disruptive forces that have built the expectation of "insight on demand"—getting actionable information into the hands of decision makers who need it, when they need it, no matter where they happen to be. |

# Data Modernization Push

Businesses face constant push to become data-driven to succeed. To become data-driven, a need for technology must acknowledge the existing system, process, and business architecture and manipulate data accordingly to provide actionable insights.

Conventional databases and on-premise infrastructure were regarded as strong IT. However, the increasing percentage of unstructured data with no real-time integration, low scalability, lack of DevOps, and high cost of storage have demanded a data warehouse or a better technology that will assess all types of data and offer reliable analytics. All the technologies which apply conventional algorithms no longer work.

These challenges are the reasons for the data modernization to occur in the year to come. The intention is to automate equipment function and allow algorithmic trading that will permit businesses to apply any kind of device and generate real-time on-demand data decisions.

## *Future of data Management*

From the market observations of Blazeclan, an award-winning audited partner of Amazon Web Services helping enterprises transform their business using the power of Cloud, it is evident that big data analytics is likely to receive an upgrade.

The ability for companies to organize a large amount of data at a central location and make it available to cross-functional teams, as desired, is now possible with as Cloud computing evolutions. With advanced analytics along with pre-built machine learning models, there's a room for agile analytics.

"Data Modernization supported by DevOps and Day Zero Security will help organizations foster the analytics-driven culture and empower them with smart decision taking competency to achieve the maximum profits in the long run."

***Baxish Mission,***
***Head of Analytics at Blazeclan***

With the world's changing behaviour on data privacy, companies need to re-risk and evaluate how it handles PII data. With emerging tools on the Cloud, it is now within the reach to collect, gather, sort, and analyze data at any scale without compromising on data governance norms.

**Speed and accuracy are two other factors that influence data management processes. Businesses look forward to converting streaming data into actionable insights for accelerating their decision making.**

## *Data Management with Amazon*

Amazon has maximized big data solution with the introduction of the data lake. It does not demand the user to pre-structure the data and run analytics. Instead, it accepts data (raw or processed, structured or unstructured) and allows users to perform analytics right from dashboards. The big data processing typically happens with the help of visualizations. The integration of machine learning algorithm to this data lake promises to boost business performance.

As we progress with the data usage and big data processing, the data lake assures to make data management simpler by creating schema on the go, during analysis.

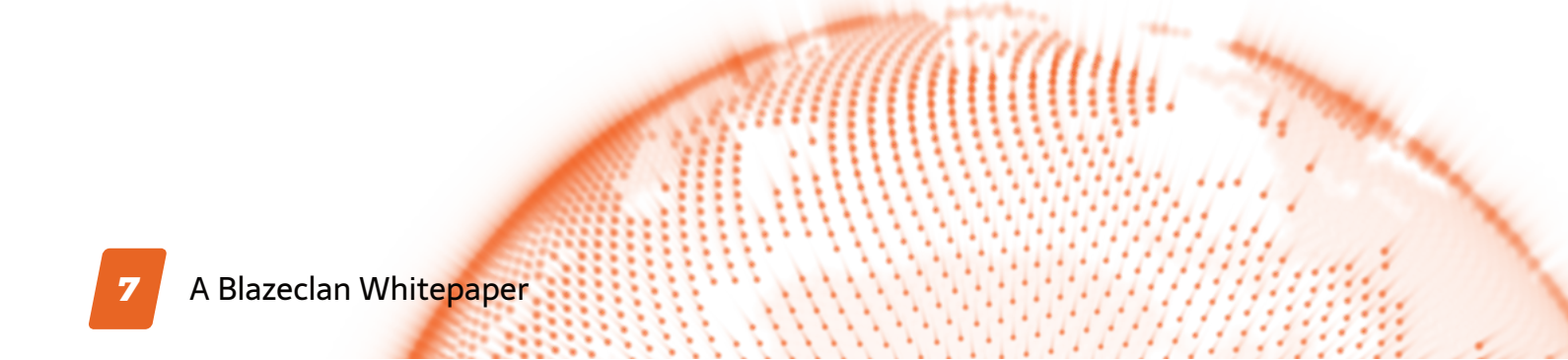
AWS users can save more money due to the low-cost storage feature available. On the other hand, conventional data management applications used business intelligence and batch reporting to analyze and inform results. Data lake takes a bigger leap with the integration of predictive analytics, machine learning, and profiling.

## *DataOps - Data Analytics*

The changes taking place with data analytics are not only transforming its use cases but its methodology as well. DataOps is an agile operations methodology that is being used to combine DevOps teams and data engineers. The term DataOps was inspired by DevOps in software engineering which combined software development and information technology to build and deploy software products faster.

Organizations are using DataOps to build data analytics platforms more efficiently. In 2018, Nexla found that 73% of companies were investing in DataOps. The increase in investment has come as a response to the challenges of developing and maintaining data analytics pipelines.

A collective definition of DataOps is hard to come by because this term acts as an umbrella term for the development and deployment of data analytics pipelines.



# Data Lake Industry Analysis

The industry insights provided by O'Reilly and other researchers indicate that the global data lake market is expected to attain 14.01 billion USD value by 2023. Major industries that are looking to switch over to data lakes include retail, media, healthcare, manufacturing, entertainment and IT. As per Blazeclan's experts, it is also identified that new industries are switching to data lakes in recent times.

One of the clients of Blazeclan, Chartered Accountants Australia and New Zealand, faced problems with data management and required a real-time solution that will fetch desired results in a few seconds. The strong expertise of Blazeclan has enabled the client to receive a customized AWS platform that performed all the essential data management services. Upon a few days of implementation, the client has reported satisfaction in terms of efficiency, traffic load balance, improved security, quick insights, and zero downtime.

On the other hand, the group company from Malaysia looked out to grab a solution that could capture data from over 121 data sources and analyze and provide results in a few seconds.

They focused more on scalability, application services, and fault tolerance. Blazeclan experts deployed data lake that allowed ingestion of data from multiple sources providing self-service, ad hoc analysis, and guaranteed real-time capability. As a result, a single customer view was enhanced and the client maximized their results.

The experience of working for a leading energy solution provider has further enhanced our expertise. The customer was specific about the creation of an independent data lake that could reduce dependency over EICT database. This use case had data from several external sources. We made use of data ingestion layer to accept landing data and the ESP teams analyzed and reconciled data with data marts along with the tools like S3, Lambda, Python, Redshift, Matillion and SQS, and various other technologies.

Hence, these use cases are clear indicators pointing out the fact that data lakes simplify data management without compromising on security.

It is evident to note that the future of businesses will rely on data lakes that can allow the decision-makers to harness values from data and also achieve better customer interactions. With the proliferation of IoT devices and sensors that chug-out a huge amount of data every minute, a data lake is a perfect solution to run analytics and automate processes in an organization.

For companies running their Analytics on Cloud, deploying data lake is a matter of few weeks.

By analyzing the trend of data in the year 2020 and beyond, there is a better chance for businesses to achieve data governance and data integration with easy to use tools. AWS is not only data-driven but highly productive and promising. It has an answer to incompetence, expensive infrastructure, and chaotic data management. The future is data-driven and getting started with Data Lake is a step in the right direction.

# Data Analytics In 2020 & Beyond

According to Gartner Inc., the top trends in data analytics technology lies in augmented analytics, continuous intelligence, and artificial intelligence (AI). As per the research analysts of Gartner, these three elements have disruptive potential in the next 3 to 5 years.

"The story of data and analytics keeps evolving, from supporting internal decision making to continuous intelligence, information products and appointing chief data officers. It's critical to gain a deeper understanding of the technology trends fueling that evolving story and prioritize them based on business value."

***Rita Sallam,***
***Research Vice President at Gartner***

In addition to Rita Sallam's opinion, Donald Feinberg, Vice President & Distinguished Analyst at Gartner said, 'The size, complexity, distributed nature of data, speed of action and the continuous intelligence required by a digital business means that rigid and centralized architectures and tools break down. The continued survival of any business will depend upon an agile, data-centric architecture that responds to the constant rate of change.'

Therefore, following are top trends of data and analytics technology that enabled the data analysts to predict the outcomes of 2020.

*Trend 1*

## *Augmented Analytics*

Augmented analytics can be rightly termed as the disruption pioneer in the data and analytics segment. With the help of Machine Learning (ML) and Artificial Intelligence, it remodels the content development, consumption, and sharing of analytics.

By 2020, it is anticipated that augmented analytics will be among the most dominant drivers of Analytics and Business Intelligence, Data Science, Machine Learning platforms, and Embedded Analytics. Hence, it is highly recommended to the leaders of Data and Analytics that they must adopt Augmented Analytics in 2020.

*Trend 2*

## *Augmented Data Management*

If Augmented Analytics is the new wave of disruption, Augmented Data Management is the need of the day. By leveraging Machine Learning and Artificial Intelligence techniques, Augmented Data Management develops enterprise information management which encompasses data quality, metadata and master data management, and data integration. It also involved self-configuring of database management systems. One of the key elements of Augmented Data Management is that it automates various manual tasks allowing less technically skilled users to use data. The technical resources can now focus on technically critical tasks that require expert technical intervention.

Moreover, Augmented data management empowers dynamic systems by converting metadata that is used for audit, lineage, and reporting. Hence, it is right to say that metadata is becoming one of the primary drivers in Augmented Data Management.

As per the experts and various data analysts, the manual tasks of data management will be reduced by 45% with the help of machine learning and automated service-level management, by the end of 2022.

### *Trend 3*
## *Continuous Intelligence*

Continuous intelligence refers to a design pattern of real-time analytics incorporated within the business operation. It helps in processing historical as well as current data that prescribes actions in response to the events.

Leveraging technologies like augmented analytics, event stream processing, optimization, business rule management, and ML, it bestows empowers data analysts with decision automation and decision support.

According to data experts, continuous intelligence will be incorporated in more than half of new business systems by 2022. The core reason for this is that it uses real-time context data that helps in making improvised decisions.

### *Trend 4*
## *Explainable AI*

The ever-evolving demand for accurate and seamless data management, new and increasing AI models is a great relief to decision-makers. Although some times, the decision-makers have to justify the usage of these models. It is, thus, important for the application experts to make these models more explainable to the users.

For example, explainable AI in data science and machine learning platforms generates model explanations for accuracy, statistics, attributes, etc.

### *Trend 5*
## *Graphs*

Graph analytics refers to an analytic technique that explores the relationships between entities of interest. The biggest benefit of graph analytics is that graph data stores can aptly explore and query data with complicated interrelationships.

According to graph analysts, the analytics will grow incredibly in the next few years and graph processing and DBMS's application will grow at 100% annually by the end of 2022. This will empower complex and adaptive data science.

### *Trend 6*
## *Data Fabric*

Data sharing is an extremely crucial task in a distributed data environment. Its security is always of prime concern for the organizations. Data fabric helps in facilitating smooth access and data sharing. It legitimizes a data management framework that enables smooth data access and processing.

It is anticipated by several data experts that by the end of 2022, data fabric designs will be distributed and considered as a static infrastructure, helping organizations to enter the new wave of disruption through dynamic data mesh approaches.

### *Trend 7*
## *Conversational Analytics*

After analyzing the previous insights, analysts have identified that 50% of the analytics queries will be auto-generated through search, natural language processing (NLP) or voice, by 2020. Analyzing the complex data combinations will be of prime concern for the analysts. Moreover, analytics will be made accessible to everyone in the organization which will further help in driving broader adoptions as it will make analytical tools user-friendly and enabled with virtual assistants.

### *Trend 8*
## *Commercial AI and ML*

According to Gartner, '75% of new end-user solutions leveraging AI and ML techniques will be built with commercial solutions rather than open-source platforms, by 2022.'

It is interesting to witness that commercial vendors are collaborating with open-source ecosystems to offer enterprise features that are necessary to scale and democratize artificial intelligence and machine learning which are lacking in an individual open-source environment.

### *Trend 9*
## *Blockchain*

The objective of blockchain is to provide a decentralized trust to the network of untrusted participants. Hence, ramifications in such cases are quite evident for those taking advantage of participant relationships and interactions.

Therefore, until blockchain technology is over-powered by any other enhanced versions of network decentralization technology, end-users must integrate with blockchain technologies. This will include integration with existing data and analytics infrastructure increasing the overall costs of integration. However, it is important to acknowledge that blockchains are not a database but just data sources. Hence, it will never replace data management technology.

### *Trend 10*
## *Persistent Memory Servers*

Persistent Memory is an emerging technology that is expected to contribute to improvising application performance, clustering methods, availability, and security best practices while having a cost-effective approach. It will also help in reducing the complexity of adopting IMC-enabled architecture.

According to Mr. Feinberg from Gartner, 'The amount of data is growing quickly and the urgency of transforming data into value in real-time is growing at an equally rapid pace. New server workloads are demanding not just faster CPU performance, but massive memory and faster storage.'

## *Benefits Of Analytics, Predicted To Be Served In 2020*

1. Crowd-aided analytics will leverage the power of the crowd by feedback process, for a wider reach and accurate results.
2. Faster analytics through robust algorithms will enhance both descriptive and prescriptive analytics.
3. Defining & Measuring Data-driven insights will be accessible to all end-users through analytics. A more flexible approach will be followed.
4. The Data scientist demand will surge as the availability of huge and complicated data sets will aid business decisions.
5. Cloud Analytics will allow seamless interaction with the large **analytical and visualization systems** in the cloud predominantly.
6. Advances in natural language processing (NLP) will enable users to analyze the vast variety of data from different channels.

# Conclusion

As Amazon says, 'AWS-powered data lakes handle scalability, agility, and flexibility required to combine different types of data and analytics approaches to gain deeper insights, in ways that traditional data silos and data warehouses cannot.'

In the era of a connected living where every device and software application with us and around us is creating data, the scale of growth is exponential.  Organizations that practice a culture of using data to arrive at decisions are already on Cloud and data lake is one up on that.

For a new-age technology-driven business that was born in the cloud era, data lake creates a compelling edge to be agile and analytics-ready from day one.

For legacy businesses who are often sitting on the fence between on-premise and cloud, the challenge is manifold if their competition is tech-savvy. Each passing day creates more gap between them and reality that becomes difficult to bridge.




sales@blazeclan.com