IMMUTA

2022

Data Access & Analytics Trendbook

Expert perspectives on where data use is heading

Table of Contents

In a world of information, nothing is more important than **access.**

Organizations across industries are accessing and analyzing data to discover essential information. From developing life-saving vaccines to personalizing customer experiences, there is no one-size-fits-all use case for data.

Similarly, there's no definitive answer on how data use will continue to evolve. As organizations of all sizes continue to push into new horizons, it is necessary to share best practices and understand just how much is possible with data.

In the spirit of accessibility, we tapped into the minds of some of our most knowledgeable colleagues to compile their thoughts on the current and future state of data use. From data science and analytics, to organizational structure and emerging technologies, this collection of outlooks and concepts draws a roadmap of where data usage is and where it's poised to go.

This Trendbook aims to make valuable ideas accessible to all who can benefit from it. Our peers' perspectives are wide-ranging and grounded in decades of firsthand experience.

As you explore these ideas, consider how they might apply to your own data strategies. Where are you now, and where would you like to go next? Consider your own perspectives and those included here, and keep the conversation going within your networks. By making our ideas accessible, we can take action and drive data further together.

Automate & Innovate Your Data Stack

Data is a dynamic resource that is constantly being created, collected, and examined. As such, the approaches to its use and analysis are constantly evolving. New tools and technologies are developed, tested, and implemented in line with developing methodologies and approaches. Recent years have seen the integration of Artificial Intelligence (AI) and Machine Learning (ML) into data analysis, as well as the surge of automated software added into cloud environments.

How does this apply to the future of organizations' data usage? How will these modernizing technologies continue to evolve going forward?

> **"Consumers are going to expect ethical, explainable and socially responsible AI."**
>
> – Ajay Sahu, Director, Enterprise Data Management

Chris Atkinson
Partner Global CTO

snowflake

We are seeing significant adoption of data governance and catalog technologies to drive internal data product delivery, not just mitigate compliance demands. In line with internal adoption of data products, there is also significant increase in 3rd party data usage to enrich internal data. This also creates compliance and privacy demands around PII protection, which in turn is driving a wave of tokenization technology adoption.

Certainly in line with automating data requests across organizations, we are seeing wide scale adoption of data access control automation to manage role access to data services and products. You can't really have one without the other.

Chintan Sanghavi
Sr. Partner
Solutions Architect

aws

Data volumes are exploding and organizations are enhancing software to deliver new and compelling customer experiences.

They are improving existing software and building new, modern applications that enable customers to do more – look up information, buy products and services, communicate with each other – which in turn creates more and more data.

Hence, modernization of data analytics platforms and real-time data analytics were trends last year and will continue to be this year as well. In addition, organizations want to gain more insights by applying various data science and machine learning technologies.

Organizations are adopting cloud technologies, such as AWS Lake Formation, and machine learning technologies, such as predictive and prescriptive analytics technologies, to drive their business with data. They are also using purpose-built databases and modernizing their data warehousing solutions and creating organization-wide data lakes.

Feature Engineering will become an extension to machine learning.

With over 300 datastores in the market today, data science engineers are finding it difficult to create features from raw data sets. There's just too much noise and complexity.

This trend will also cause the concept of "data pods" to flourish. Data pods refer to sets of virtual data environments and controls that may reference multiple source data sets but have their own governance and access controls. Users can search for data and define the model on the data, regardless of which underlying database the data is stored in.

Mohamad Thahir
Data Architect

DataStax

Make Your Data **Accessible**

Sounds pretty straightforward, right? It should be! All the data in the world wouldn't be of use to anyone if they couldn't facilitate access to it.

That's why organizations are putting a greater emphasis on democratizing access to data. Whether providing data access for internal stakeholders, external data sharing, or for consumers themselves, there is a drive to get data safely and efficiently into the hands of those who need it.

"Organizations are building intelligent ecosystems with customers, partners, suppliers, and employees. This creates the need for an ecosystem driven–data strategy that transcends the normal data boundaries."

– Shekaran Sury, Sr. Partner, Data & Analytics Consulting

Deepak Rajendran
Senior Manager,
Data & Analytics

EY

Organizations these days are sitting on goldmines of data, which they have acquired over time, from their different lines of business and operations.

These organizations are increasingly looking for opportunities to augment their data assets, with more contextual and meaningful external data. A data marketplace or exchange facilitates this transaction of data, and offers a whole ecosystem of services and products built around it.

The data usage patterns have evolved radically in the last decade or so and consequently, the access patterns [have] as well. The traditional way of defining a people-centric access model for data management and then assigning roles to operationalize it will not be enough in the era of IoT and streaming data.

The regulatory landscape has also changed drastically and a global organization must design access policies, taking into consideration the various global and regional data privacy regulations. These drivers are leading to a hybrid model of data access management, combining the traits of people-centric, role-based models (RBAC) with more nuanced and contextual, feature/attribute-centric models (ABAC).

Igor Chtivelband
VP Data & CRM

BILLIE

What is common for the tools, which are considered to be part of the modern data stack, is that all of them were developed in startups, not huge companies. Thus, consciously or unconsciously, these startups built products that are a great fit for other startups.

The pricing model of these tools is reasonable and more flexible, compared to the traditional enterprise solutions. The learning curve is not steep, so it allows you to start building an MVP in no time. Also, these tools work well with each other.

So the times when one had to make a commitment to a tech giant, pay an enormous amount ahead, and use only the technology from one vendor, are luckily over.

Steve Touw
Chief Technology Officer

MMUTA

All the data to all the people. That might sound crazy to some, but if you aren't trying to get there, that likely means you will fall behind your competitors.

We are moving from a world where raw data had to be abstracted through visualization tools and apps that pump out pre-computed decisions from ever-changing data, to a world where more advanced data analysts want the data in its rawest form with compute resources available to process it as needed. This opens a whole new world of challenges, ranging from consistent data access control to data platforms with flexibility of compute, to data products (a la Data Mesh).

Vineeth Menon
Head of Data
Lake Engineering

Swedbank

We are going through turbulent times with the pandemic disrupting industries around the world, and enterprises have no option but to adapt to the changes with a quick turnaround time.

Data science and analytics have and will continue to play a big role in this journey. Some of the important trends I see coming in this area are:

1. **Hyperpersonalization:** Leveraging AI and advanced analytics in terms of delivering the most relevant content and product offerings customized to individual users.
2. **Automated Machine Learning (Auto ML):** This is still at an early stage, but AutoML software platforms make machine learning more user-friendly and give organizations without a specialized data scientist or machine learning expert access to work ML Models, which eventually helps in better utilization of the power of AI and Machine Learning at the enterprise level.
3. **AI Democratization:** This will be a key trend going forward, as it will make AI accessible across the business as a whole, helping enterprises in their journey towards AI-driven decision making capabilities. Data accessibility, quality, and simpler UX will play a key role in AI democratization.

Develop & Diversify Your Data Team

For data goals to be met, organizations need the right people driving data initiatives forward. Company structures are adjusting to suit the needs of data's continued evolution, and as a result, roles are regularly being expanded or added.

We've seen the rise of the Chief Data Officer, but organizational transformations go even deeper. Whether building out DataOps teams, adding new qualifications and training, or better enabling data teams through restructuring, it's the people operationalizing the data who deliver results.

> **"CDOs can improve business value by setting goals and strategies, ensuring that organizations are regulatory compliant, and using data and analytics to assess revenue impacts and reduce risks."**
>
> – Chintan Sanghavi,
> Sr. Partner Solutions Architect

aws

Slava Frid
Chief Technology Officer

WORLDQUANT PREDICTIVE

We have been evolving new roles that lie between traditional data engineers and data scientists. These are data scouting and shaping analysts who have a custom stack of skills and tools to quickly evaluate and adapt data sources for use with each research and data science need. Organizationally, they are located at the intersection between product, research, and engineering.

For our engineering organization, they are internal customers for new Data Discovery, Lineage, and Shaping capabilities. At WorldQuant Predictive, we feel these roles are going to be an important contributor to our success.

Any modern organization has to be thinking of what their toolkit for discovering and evaluating ever increasing sources of data would be.

Shekaran Sury
Sr. Partner, Data & Analytics Consulting

wipro

The Chief data office function has begun to play a pivotal role in shaping the future of data driven organizations.

Global and diverse organizations are driving insights at scale through business aligned product teams powered by multi-disciplinary data pods. These self-contained pods are orchestrated through degrees of freedom within a data & insights framework.

Legacy data support models are evolving into data command centers that measure experience rather than outcomes by tapping into Data Ops, ML Ops, and DevSecOps constructs.

The focus is rapidly shifting towards data and business observability.

There is a need for increased data literacy, and upskilling data and analytics teams on new ways of engaging with and creating data products that can be consumed by internal and external stakeholders. This requires a data maturity framework and a methodology to enable effective learning within the work spectrum.

Data Governance must support the diverse personas in the organization – including Data Scientists, Data Stewards, and business users – and enable them to see this as part of their regular business processes.

We are also seeing an openness to exploration, discovering new ways of engaging with data and listening to customer experiences as a means of improving the value realized from the relationship. This requires enabling customers to provide feedback to the development and delivery of products. Thus, transparency and collaboration are key to the connected ecosystem.

While most large organizations still operate separate data and analytics teams,

the more nimble ones are starting to realize the benefits from embedding data and analytics experts within the product teams.

This synergy results in analytics-infused products, as opposed to analytics being an afterthought.

Aaman D. Lamba
Senior Industry Principal

Infosys®

Ajay Sahu
Director, Enterprise Data Management

J.B. HUNT

The Endurance of **Data-Driven Insights**

Collecting, analyzing, and interpreting data are all done with a common goal in mind: gaining valuable insights to inform and drive initiatives. Data does not exist in a vacuum, instead acting as a driving force of business operations and value creation.

New technologies, approaches, and attitudes toward data science are shifting the way we gain data-driven insights, and providing organizations with the ability to ask more of their data. Where might this bring us next?

> **"We measure success by comparing the number of available data sources to the numbers actually used in modeling and customer-facing products."**
>
> – Slava Frid, Chief Technology Officer
>
> WORLDQUANT PREDICTIVE

Scott Stafford
Solutions Architect

databricks

Businesses will adopt any people, processes, and technologies that help answer these questions:

1. **Can you describe the past and present?**
2. **What trends do we expect in the future?**
3. **Can you answer these questions given our policies?**

Besides the overall move to cloud data stores, data teams are adopting much more comprehensive data testing and observability frameworks to stay on top of their processing pipelines and proactively detecting errors. Increasingly, these pipelines are driven through version-controlled code, run in staging and dev environments, and exposed to power consumers through tools like dbt.

How can we reliably scale the impact of data products?

Now that the tooling is in place, and patterns for experts to manage data well (e.g. lakehouse, ELT architectures), companies must come up with larger plans for making quality data products available to more and more consumers.

Stephen Bailey
Data Engineer

whatnot

William McKnight
President

MCKNIGHT CONSULTING GROUP

While the use of a public cloud is mostly "plan A" for new initiatives in an organization, and frequently the target for reengineering legacy applications, it has brought with it a new set of challenges that not every organization is ready for.

An important new trend is fully adapting to the cloud, and all that it means.

It means understanding the new software model, planning development and quality assurance, planning recovery from outage and credit for downtime, understanding safe harbor and cross-border restrictions, dealing with the new ways for capacity planning and growth, security and privacy, understanding query performance and service levels, understanding data interchange in the cloud and formulating a plan for effectively staffing a cloud-first organization.

How do I allow the average user to combine two data sets or more together? In the aggregate of this data, I can get better insights, precision of my models, greater accuracy of the data, all sorts of things. It's democratizing the aggregation of data –and that's super hard to do. If I had to write code to do that across multiple different systems, the barrier of entry to do that is really high.

The first kind of important trend in data science and analytics is this true movement to a consistent data infrastructure environment where data is physically separated, but logically you can unify it with pretty great ease.

How do I move the data into a cloud environment and make it really easy for someone to aggregate data? It's not about ad hoc access to data, it's about the aggregation.

Everyone's making it easier to combine data.

Matt Carroll
Founder and Chief Executive Officer

MMUTA

More About Immuta

Immuta is the leading automated data access platform. Our mission is to make sure that everyone can easily access data in the cloud, so long as they have the right permissions. Organizations that use Immuta can easily achieve scale, efficiency, and agility while being both secure and compliant. By separating policy from platform, Immuta provides a single platform to automate access control for any data, on any cloud service, across all compute infrastructures. Data teams across industries trust Immuta to help control access to all their data, without risking security, privacy, or performance.

To learn more about Immuta, visit www.immuta.com or find us on LinkedIn and Twitter!

IMMUTA™

25 Thomson Place, 4th Floor, Boston, MA 02210 | immuta.com | (800) 655-0982