

2020 DATA TRENDS AND PREDICTIONS

to Help You Kickstart
the New Decade

2020 Data Trends and Predictions

As the second decade of this millennium draws to a close, it's only natural to wonder what the future has in store. The world of data has grown exponentially over the last ten years. In 2019 alone, we saw the role data and analytics play within companies continue to expand across departments. More and more teams are now looking to data to help drive successful decision making and assist with critical job functions.

TABLE OF CONTENTS

- 1 [Data regulation is sweeping the nation](#)
- 2 [CDWs and SQL have emerged victorious](#)
- 3 [ELT > ETL](#)
- 4 [JSON and semi-structured data is now mainstream](#)
- 5 [Augmented analytics hold promise](#)
- 6 [Want to unlock the power of data? Keep humans in the loop](#)
- 7 [Self-service is no longer "nice to have." It's essential](#)
- 8 [The re-emergence of the spreadsheet](#)
- 9 [The future will be built on data insights](#)

A new generation of data pioneers

While we'd expect to see financial institutions, e-commerce websites, and telecommunications companies using analytics to govern their strategies and operations, [libraries](#), [cruise lines](#), [video game studios](#), [firefighters](#), and even [tennis players](#) are beginning to harness the power of data to plan for a better future.

[Forrester](#) reports that data-driven companies grow more than 30% annually on average. Leading companies are capitalizing on data and analytics to widen the gap between them and their competition.

Mind the data skills gap

But despite universal agreement that data has the power to improve decision making across all parts of an organization, many companies are struggling to reap its benefits. According to a study by [NewVantage Partners](#), only 31% of executives feel they have been able to build a “data-driven culture.” And according to [Gartner](#), 87% of companies have low BI and analytics maturity.

87%

of companies have low BI and analytics maturity

With the benefits so clear, why are companies having such a rough time realizing the promise of a data-driven future? There are a few factors at play, but the most common reason is a [widening skills gap](#). Few people within organizations have the technical expertise necessary to access, analyze, and draw insights from data directly.

Instead, those with domain expertise across departments must rely on and compete for the time and resources of a limited pool of data engineers, scientists, and analysts. It's an inefficient system that only grows more complex as the volume and scope of data continue to increase.

The future of data is within reach

But there is hope. A new generation of tools and technologies is on the horizon. They hold the potential to transform the speed and ease at which companies can extract value from their data and form impactful insights—all while maintaining evolving security and compliance standards.

We've uncovered eight data trends that will shape the way businesses use data in 2020 and beyond:

1 Data regulation is sweeping the nation

A wave of data-privacy legislation is sweeping the country, and companies will need to pay close attention if they want to remain compliant and avoid legal vulnerabilities.

A perfect storm of scandals, security breaches, and whistle-blower revelations thrust data security and privacy into the spotlight over the last decade. From [Edward Snowden to the Cambridge Analytica exposé](#), the issue of data privacy became part of the public consciousness. For the first time, people were made aware of just how much personal data gets collected and how it's used. The era of tech companies regulating themselves drew swiftly to a close.

Then there was a series of very high profile breaches that affected hundreds of millions of people. In 2018 alone, [over 1 billion people](#) had their data exposed from security breaches. Facebook, T-Mobile, Quora, Google, Orbitz, and dozens of other companies all had sensitive customer data compromised.



Over 1 billion people

had their personal data exposed thanks to security breaches.

But the defining moment was the enactment of [GDPR](#) (General Data Protection Regulation) in May 2018. This set of European regulations had far-reaching effects that went well beyond the continent. Although the laws are designed to protect the information of EU residents, they affect not only European companies, but any organization processing the data of those EU residents. Any company doing business involving European citizens has to comply with GDPR or risk significant fines and penalties.

Data privacy laws are in force at the state level, creating a lattice of overlapping—and at times—incompatible laws and regulations.

Multiple countries have GDPR-like laws in the works, including China and India. This web of growing international legislation is making it incredibly difficult for companies that collect data on any level to do business. In response, many industry leaders are calling for global data privacy regulation to make it easier to do business. Some, like Satya Nadella, CEO of Microsoft, are even advocating for data privacy to be treated as [a human right](#).

In the United States, there is no single federal data privacy law or central data protection authority enforcing compliance. Data privacy laws are in force at the state level, creating a lattice of overlapping—and at times—incompatible laws and regulations.

Leading the way has been California with the CCPA [California Consumer Privacy Act](#), which goes into effect Jan. 1, 2020. While it won't be enforced until July 1, 2020, it will apply to any data collected with twelve months of its enactment, so it's something businesses need to address now. The number of states enacting similar laws will explode throughout 2020.

How to prepare for this trend:



Audit your processes:

Review each data collection point to ensure it complies with data privacy legislation. Consider training your workforce in data security best practices and procedures. Analyze how data is stored and secured and upgrade/update any systems that are not up to code.



Evaluate your partners:

Because customer data often gets shared between partners and vendors, organizations must ensure their data partners are compliant with data privacy law. Your company may be fully compliant, but it could still face legal accountability if a third-party vendor is not.



Consider updating your tools and services:

Your tool and service providers should ease the burden of compliance, not add to it. Cloud data warehouses (CDWs) can help your organization stay compliant and up to date with data security certifications. For example, [Snowflake](#) offers a wide range of features that make it easier to honor data subject rights and manage data security.



2

CDWs and SQL have emerged victorious

After decades of rapid innovations, cutting-edge cloud data warehouses and the robust, four-decade-old SQL database language have emerged as the dynamic duo of the modern data stack.

Cloud data warehouses (CDWs) are replacing on-premises and hybrid data warehouse solutions at a swift pace, and it's not hard to understand why. Rapid scalability, increased flexibility, lower costs, and greater connectivity are just a few reasons companies are moving their data to the cloud. Snowflake changed the game by decoupling storage from compute, a move other CDWs have started to imitate.

Even when it comes to the issue of security—an area on-premises solutions have historically excelled in—cloud capabilities give today's modern CDWs the upper hand. Today's CDW providers must meet the highest security standards and certifications, including SOC2, ISO27001, HIPAA, PCI, and others. Their offerings are built entirely on data security and encryption, and they invest heavily in these technologies.

\$19k - 25k

According to [Amazon](#), running a data warehouse on your own costs between \$19k - 25k per terabyte per year.

They also employ an army of security and technical experts tasked with maintaining and improving standards as well as responding to incoming threats. Many experts feel the public cloud is [more secure](#) than the majority of on-premises data centers.

As mentioned earlier, CDWs ease the burden of compliance. By storing all company data in one place, organizations don't have to deal with the complexity of searching various discrete data stores to locate individual records. This makes complying with updates, changes, or deletions required under GDPR's data subject rights much easier to manage.

Then there are significant cost savings. According to [Amazon](#), running a data warehouse on yourself costs between \$19k – 25k per terabyte per year. On average, companies see a 96% savings switching to a modern CDW.

By storing all company data in one place, organizations don't have to deal with the complexity of searching various discrete data stores to locate individual records.

Lastly, CDWs enable organizations to draw value from their data faster than ever before. Data expires, grows irrelevant, or gets replaced at an astonishing rate. It's estimated that [60% of corporate data](#) has lost some or even all of its business, legal, or regulatory value.

Modern CDWs allow raw data to be loaded directly into the system and converted directly in the data warehouse (a process known as ELT, or Extract, Load, Transform) instead of needing to be transformed into a staging database first and then loaded into the data warehouse (a process known as ETL, or Extract, Transform, Load). This [greatly reduces the time data is in transit and speeds the time in which data can be accessed and analyzed](#).

Then there's the incredible story of SQL; a database language developed nearly 45 years ago. There have been many efforts to replace or reinvent SQL over the years. Some experts thought of it as a relic that couldn't scale. This led to the rise of NoSQL databases: Bigtable, Cassandra, MongoDB, and more.



60%

of corporate data has lost some or even all of its business, legal, or regulatory value.

But as data volume and velocity increased exponentially, the problems with these non-SQL databases began to manifest. To address these issues, some database vendors added proprietary “SQL-like” query languages. But there’s been a major return to SQL, and it’s more relevant than ever before.

It’s still the best language for interacting and performing on databases. It’s superior in structure and compatibility and is more robust than its alternatives due to a massive community that has spent decades focusing on security and efficiency for SQL systems. According to [Stack Overflow](#), SQL is the 3rd most popular coding technology. From where we sit, CDWs have emerged victorious and SQL is only growing in importance and popularity.

How to prepare for this trend:



Consider a move to a modern CDW:

If you haven’t already, consider a move to a cloud-based data infrastructure. You can learn more about building a cloud-native data infrastructure in our free eBook, [Building a Cloud Analytics Stack](#), and get some tips about [choosing the right CDW in this article](#).



Unlock the power of SQL for your entire organization:

While investing in a team of data experts who are wizards in SQL is certainly a viable strategy, the ability of your organization to draw and benefit from data insights will be limited as long as power and access remain in the hands of these select few. The solution? Deploy a [tool that blends the ease and flexibility of a spreadsheet with the power of SQL](#). Sigma's uses a spreadsheet-like experience, an interface the vast majority of business users know and love, to explore of data. All the complex SQL is written under the hood *automagically*.

3

ELT > ETL

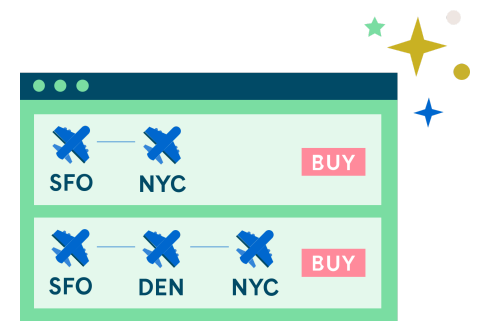
ELT enables data to be transformed in place as needed—limiting the time data is in transit and increasing the speed of analysis and action.

The traditional approach to data integration is known as extract–transform–load (ETL) and has been predominant since the 1970s.

An ETL workflow performs the following steps:

- 1 Data gets extracted using connectors.
- 2 Through a series of transformations, the data is rearranged into models as needed by analysts and end–users.
- 3 Data gets loaded into a data warehouse.
- 4 The data is summarized and visualized through a business intelligence tool.

Orchestration and transformation before loading impose a critical vulnerability on the ETL process. Transformations must be specifically tailored to the unique configurations of both the original data and the destination data. This means that upstream changes to data schemas, as well as downstream changes to business requirements and data models, can break the software that performs the transformations.



Overall, the traditional ETL process has three serious and related downsides:

- 1 It is complex because data pipelines run on custom code dictated by the specific needs of specific transformations. This means the data engineering team develops highly specialized, sometimes non-transferrable skills for managing its codebase.
- 2 It is brittle. For the reasons as mentioned earlier, a combination of brittleness and complexity makes quick adjustments costly or impossible. Parts of the codebase can become nonfunctional with little warning, and new business requirements and use cases require extensive revisions of the code.
- 3 More importantly, ETL is all but inaccessible to smaller organizations without dedicated data engineers. On-premise ETL imposes even further infrastructure costs. Smaller organizations may be forced to sample data or conduct manual, ad hoc reporting.

These shortcomings are a direct result of its origins from a time of scarce and expensive computation, storage, and bandwidth resources. By limiting the volume of data that is processed and stored, ETL preserves computation, storage, and bandwidth at the expense of labor.

Computation, storage, and bandwidth have, in recent years, plummeted in cost and became radically accessible to organizations of any size or means. This means the sequence of transformation and loading can be reversed. Delaying the modeling and transformation steps of the analytics workflow allows teams to preserve the labor of engineers and directly provide analysts with a comprehensive replica of all the organization's data to model at their discretion.

How to prepare for this trend:



Compare ETL and ELT solutions:

Choose a vendor that manages multiple data sources, including support for structured and unstructured data—even if you don't need that support today. This may come into play down the road, and if it does, you won't need to change providers. Be sure the vendor works well with your data warehouse of choice.

4

JSON and semi-structured data is now mainstream

The wealth of semi-structured data pouring in from apps, websites, mobile devices, etc. combined with the capabilities to search, manage, and analyze this data in CDWs have led to a data breakthrough.

JSON, or Javascript Object Notation, is a method that was [discovered by Douglas Crockford](#) in 1996. Since then, it has become the defacto format for transferring data on the web. Lightweight, human and machine-readable, and parseable in every programming language, JSON has skyrocketed in popularity. According to [Stack Overflow](#), more questions get asked about JSON than any other data interchange format.

Data that has business value no longer gets restricted to structured data. [Unstructured and semi-structured data now makes up more than 80% of enterprise data](#) and is growing at a rate of 55% and 65% per year. This data, in the form of emails, documents, text messages, chats, videos, photos, mp3s (and more), is pouring in through apps, websites, mobile devices, IoT devices, and sensors. It gets generated at ever-increasing volumes by both humans and machines.

Semi-structured data used to be more difficult to search, manage, and analyze. Housing It required multiple on-premises storage systems, which added significant complexity and expense. But now, structured and semi-structured data can be managed in the same system. Innovations like [Snowflake's variant data type](#) allow semi-structured data to be loaded into a column in a table, and then accessed natively with some minor SQL extensions.



This eliminates the need to parse out and ETL the data into traditional tables and columns — not mention makes it all cloud accessible. The result has made JSON and other semi-structured data easier to store, analyze, consume, and even create analytics for, so organizations can reap the full value.

How to prepare for this trend:



Use a modern CDW and business intelligence tool to get value out of your unstructured data:

[Sigma unlocks the benefits of semi-structured data](#) – expedience, transparency of data lineage, flexibility to add new fields, depth, breadth – without the classic trade-offs of cost, confusion, and technical training. By empowering business experts to explore and comprehend raw data, Sigma allows companies to get more out of their data.

5

Augmented analytics hold promise

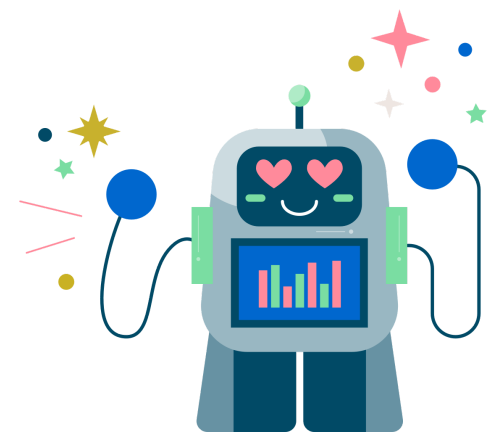
The use of artificial intelligence and machine learning to augment data analytics has the potential to change the way analytical data is shared, generated, and processed.

Augmented analytics topped [Gartner's list of data and analytics trends](#) this year and has many in the industry excited at the possibilities. [Gartner](#) defines augmented analytics this way:

“Augmented analytics is the use of enabling technologies such as machine learning and AI to assist with data preparation, insight generation and insight explanation to augment how people explore and analyze data in analytics and BI platforms. It also augments the expert and citizen data scientists by automating many aspects of data science, machine learning, and AI model development, management and deployment.”

Simply put, it's a process that uses AI and ML protocols to change the way analytical data is shared, generated, and processed and it has the potential to transform the entire industry.

The goal is to surface critical insights that take less time, less skill, and have less bias. As was brought out in our introduction, data scientists spend over 80% of their time doing simple mechanical things such as labeling and cleaning their data. Augmented analytics would use AI to do most of the heavy lifting and could even lead to completely autonomous business analytics systems in the future.



A word of caution: As with most technology, there's a drive to remove humans, or at least abstract them, as much as possible from decision making. But AI isn't and will never be perfect. Many of the problems inherent in human-led decision making have the potential to be amplified exponentially by AI, [including bias](#).

The ever-growing complexity, volume, and velocity of enterprise data have made the promise of augmented analytics especially appealing to many in the industry, making this a key trend to watch over the next several years.

How to prepare for this trend:



Build a data-driven culture:

As data continues to become the basis for strategic decision making across departments, it's critical to build data literacy amongst all teams. Augmented analytics has the power to surface valuable insights, but will require data literate humans to verify and appraise these insights before taking action on them. Building data literacy now will prepare your organization to make the most of augmented analytics and other new technologies on the horizon.

6 Want to unlock the power of data? Keep humans in the loop

Despite rapid advances in artificial intelligence and machine learning, humans possess unique, tacit knowledge that can't be replicated.

There's a simple concept in the machine learning world called [human in the loop](#). When training AI, companies will inevitably encounter edge cases or situations that are too close to call. In these moments, they bring in the “human in the loop” to make the final judgment call, which helps the AI system improve its decision-making ability.

The idea of AI systems automatically surfacing insights seem attractive. But to exclude humans from the decision-making process leads to missed opportunities at best and utter catastrophes at worst.

Humans are smart. As a species, we can perceive things that are not easily measured or quantified. We can draw connections between events that are not obvious. We have tacit knowledge. And we shouldn't underestimate the human experience or the human brain.



Good AI in the context of business analytics is about recommendations. It's about how the system can help you find the data you need. It balances the system with human signals.



Rob Woollen
CEO of Sigma



Dr. Lance Eliot, CEO, Techbrium Inc and Executive Director of the Cybernetic AI Self-Driving Car Institute, highlighted the example of the [Viking Sky cruise ship accident](#) as an example of the importance role of human governance over automated systems. The rough seas caused the oil-level sensors to detect that the amount of oil was dangerously low — nearly non-existent. If there was no oil, the best thing would be to shut down the engines, and that's exactly what the automated system did.

But there was oil, the sea was just extremely choppy. By cutting the engines prematurely, the ship was turned into a bobbing cork. No humans were alerted to this before it happened, and therefore did not intervene. Hundreds of passengers and crew had to be airlifted in a very dangerous operation.

Humans could have reasoned the rough seas were responsible for the low oil reading, or steered the ship to a safer place before cutting the engine. It's true humans also make mistakes sometimes, but humans ultimately add unique qualities and considerations to the decision-making process machines simply cannot:

- 1 Intelligence
- 2 Ethics
- 3 Emotion and compassion
- 4 The ability to quickly recognize a course of action is incorrect
- 5 The ability to mitigate the impact of wrong decisions

The best decision-making systems in the future will be close partnerships between humans and smart technology. AI systems will be able to take over the manual and tedious parts of data exploration and bring humans closer to the most relevant parts of it. Far from removing them from the equation, the best AI data and analytics tools of the future will amplify human strengths and allow them to focus on the most impactful work.

How to prepare for this trend:



Promote natural human curiosity:

Encourage everyone on your team — non-coders included — to go beyond the obvious, and empower them to find the root cause of fluctuations, spikes, and dips in your data. Connect data sources and join data sets to help domain experts more easily identify commonalities and trends. Promote [divergent thinking](#) and encourage everyone to ask ‘why?’



Limit human-out-of-loop systems:

Fully automated systems are best suited for mechanical or low impact tasks. Keep humans in the loop for high impact or strategic decision making. According to a study by [The Economist](#): “Although technical limits are constantly being overcome, the increasing demand for accountability—especially following the financial crisis—means that important business decisions must ultimately rest with a human, not a machine.”



7

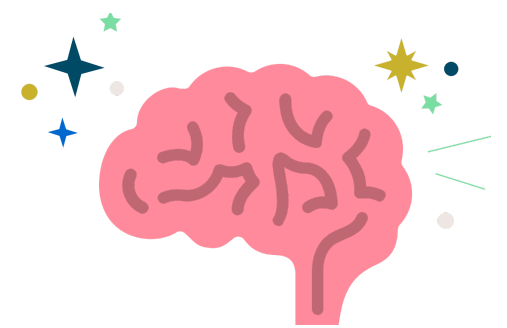
Self-service is no longer “nice to have.” It’s essential

For organizations to get the most value out of their data, they need to enable domain experts to go beyond simple dashboards and directly access their data warehouse.

[Gartner](#) defines self-service analytics as “a form of business intelligence in which line-of-business professionals are enabled and encouraged to perform queries and generate reports on their own, with nominal IT support.”

Self-service is the holy grail of analytics. Many BI providers even claim to have achieved it. But [BI adoption hovers around a mere 30%](#), far from enabling all employees to answer meaningful questions with data. Even when a BI tool is in use, domain experts often require extensive help from IT teams.

For most business users, “self-service analytics” is a limited dashboard built around a single set of metrics or a spreadsheet of expiring data. The promise of completely self-serve analytics has remained elusive... until now.



30%

BI adoption hovers around a mere 30%, far from enabling all employees to answer meaningful questions with data.

The modern cloud analytics stack is making this dream a reality.

Pipeline tools like Fivetran, cloud data warehouses like Snowflake, and intuitive, cloud-native analytics tools like Sigma, allow business users across the organization to explore data without help from IT or the need to know SQL. These capabilities are more important than ever as the ability to quickly act on and draw insights from data becomes a key competitive advantage.

Additionally, these modern tools allow IT to establish robust data governance, keeping report sprawl to a minimum and access to sensitive information restricted. In tools like Sigma, administrators can set permissions by team and namespace, and can even restrict data access directly from the database.

In a recent [MIT Sloan research report](#), 52% of people said they have been unable to access the data they need to perform their job, and 63% are unable to access data in the required timeframe. The modern cloud data analytics stack empowers business users and data teams to work together to quickly draw value from data and use it as the basis for strategic decision making.

Companies that have succeeded at making data-driven decisions are [1.5 times more likely to have reported revenue growth of at least 10%](#) in the past three years. The ability for every employee to access and analyze data is no longer just a nice to have, it's essential for a healthy, growing business.

Why is Data Pipeline Self-Service Essential?

Automated self-service solutions that require minimal configuration and setup can radically reduce the engineering workload of a data project and leverage the expertise of specialists who have stress-tested the solutions against many corner cases. As companies adopt more and more cloud applications, data pipelines will become far more voluminous and complicated, and in-house solutions will become untenable for the vast majority of companies.

How to prepare for this trend:



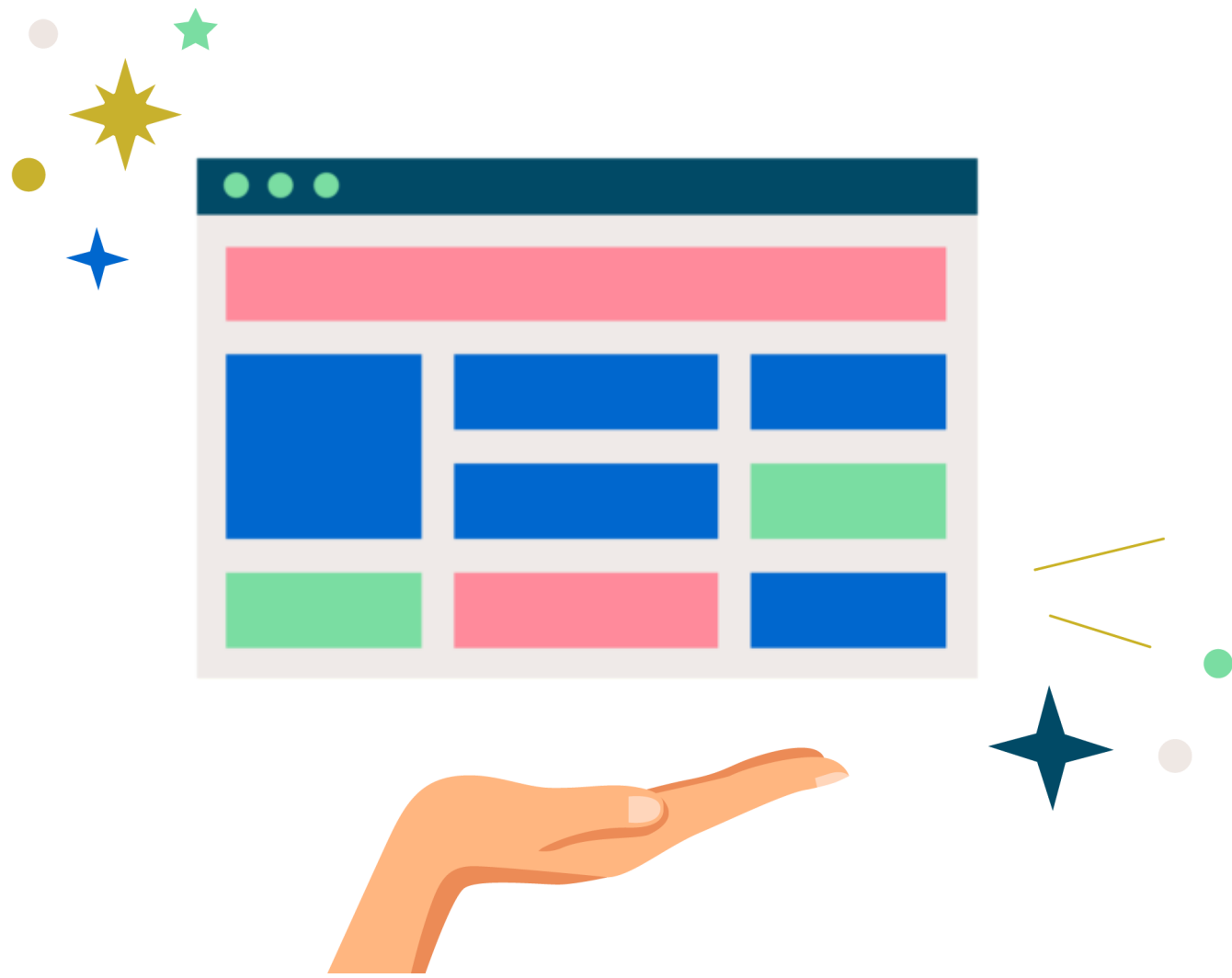
Move beyond simple dashboards:

Built-in dashboards packaged in popular apps provide a narrow slice of data and don't enable users to ask new questions and go beyond a few basic analytics. BI dashboards offer a little more flexibility, but don't empower people to ask deeper questions. More often than not, they're created by data teams and not domain experts. To truly get value out of your data, you need to provide everyone on your team with the ability to explore the data themselves.



Promote data literacy:

Education and greater access to data is the magic combination that can close the gap between business experts and data teams. Add in access to a common toolset and you'll have everyone in your organization on the same page and getting the full value of data insights.



8 The re-emergence of the spreadsheet

Despite efforts to kill it or replace it with alternatives, the spreadsheet remains the easiest, most accessible way for people to analyze and explore large amounts of data.

Spreadsheets, in paper form, have been around for hundreds, even thousands of years. In digital form, they were first introduced to the public in the [1970s with Visicalc](#). Since then, there have been all sorts of attempts to replace them. Hundreds of interfaces, each with their own idiosyncrasies, pros, and cons have been developed to display large amounts of data in a readable way. But despite all of these alternatives, the spreadsheet endures.

In fact, we've seen a re-emergence of the spreadsheet as of late with apps like AirTable, Smartsheets, and Spreadsheet.com taking the interface to new heights. There are reasons for this renewed popularity.

Data in many organizations exists in siloes: walled gardens that only a select few people know how to access and manipulate. But there are lots of smart people in the world who don't work on the IT team. For them, the spreadsheet is access, freedom, and power. A whopping [85% of people use spreadsheets in their work](#), and a surprising 76% of people rate their skill level with them as either good or excellent.

Even those who do know how to write SQL see the value of spreadsheets. [88% people who write SQL still use Excel](#) when exploring data. Unmatched in its power and familiarity, the spreadsheet remains the best interface to ask questions, iterate, and collaborate on data with other people. Providing an intuitive spreadsheet interface enables the most people to have the most power and the most freedom. It's a headstart on the path to data discovery and exploration. We predict an increase in spreadsheet interfaces in the BI world and beyond in 2020.

88%

of people use spreadsheets in their work and a surprising 76% of people rate their skill level with them as either good or excellent.

How to prepare for this trend:



Brush up on your spreadsheet skills:

Spreadsheet formulas and techniques are easy to learn and practice. In the words of [Anne-Marie Charrett](#), CEO of Testing Times, “The great thing about spreadsheets is this: Everybody knows EXCEL and has it on their PC. It's available to everyone, needs no training, maintenance or configuration. People understand and are familiar with spreadsheets.”



Take advantage of the latest generation of spreadsheet tools:

Consider switching to the latest spreadsheet-based tools to give your team a headstart and help increase adoption. Spreadsheets give users a familiar and powerful tool to explore data that can't be matched by existing BI solutions.



The future will be built on data insights

Over the next ten years, data will continue to be generated at an exponential rate. And its role in strategic decision making will grow right along with it. “The continued survival of any business will depend upon an agile, data-centric architecture that responds to the constant rate of change,” explains [Donald Feinberg](#), vice president and distinguished analyst at Gartner.

Organizations that leverage data most effectively will have the advantage and widen the gap between themselves and the competition. It’s critical for companies to recognize these growing trends and put themselves in the best position to capitalize on them if they want to keep pace in the new decade.

Take a future-focused position on business analytics and empower your team to get the full value from your data. Learn how Sigma’s no-code-necessary technology and intuitive spreadsheet interface for your cloud data warehouse can change the way your company approaches data. Sign up for a [free 14-day trial](#) today.

2020 Data Trends and Predictions brought to you by:



About Sigma

Sigma Computing is changing the way organizations use data. As an innovator in analytics and business intelligence (A&BI), Sigma accelerates time to insight for data stored in cloud data warehouses. Sigma is the only cloud-built, no-code A&BI tool enabling business leaders and domain experts to analyze data in real time. Powered by the cloud, a spreadsheet-like interface, and SQL under the hood, Sigma is a single source of truth that keeps data accurate, secure, and in context. Headquartered in San Francisco, Sigma is democratizing data and leading the Collaborative Analytics movement. To learn more, visit www.sigmacomputing.com.



About Fivetran

Fivetran is the leader in automated data integration, delivering ready-to-use connectors, transformations and analytics templates that adapt as schemas and APIs change to ensure reliable data access. Fivetran enables businesses to quickly adopt new SaaS apps, and keep up with continual app changes. Fivetran is ready to go, with prebuilt connectors and customizable templates that accelerate analytics. Data is ready to access with continuous data synchronization from sources to warehouse. Fivetran is ready to adapt by dynamically adjusting to schema and API changes. Learn more about data integration that keeps up with change: www.fivetran.com