



Modern Data Management

How next-generation data tools
eliminate data maintenance



Introduction

As more organizations integrate data into every level of their business, the volume of data generated even by small companies has exploded.

Data on customer behavior, market characteristics, product inventories, and more can all provide critical insights. This has the potential to make companies incredibly agile, able to identify signals in their data as they come in and to make quick decisions about how to respond.

But that data also comes with new challenges, chiefly in the form of access and management. As companies increase the amount and complexity of their data, processes that work on a small scale may introduce friction that slows the time from data collection to insight.

The current state of data

While just capturing data used to be tough for organizations outside the Fortune 500, in recent years the need for and availability of data has exploded. That proliferation has resulted in a variety of challenges that cut across industries, company sizes, and roles.

Big Data Meets Too-Big-for-Excel Data:

“Big data” has been a buzzword in business for nearly a decade, and for good reason. But while most companies probably aren’t generating terabytes of data on a regular basis, many organizations are generating enough data to overwhelm tools like spreadsheets. Whether the issue is files that are too large to load or complex VLOOKUPS and pivots that bog down software, overloading a spreadsheet is surprisingly easy to do.

The Need for a 360 View:

As data becomes more readily available, users’ appetite for sophisticated analysis has grown. For example, following a customer’s journey from the first digital ad they click to the point where they become a power user—or loyal buyer—is both possible and illuminating. But mapping that journey requires connecting marketing, product, and help desk data, a process that requires serious interdepartmental coordination

Strategic Automation:

From workflow automation tools to data syncing platforms, teams of all sizes are in search of ways to make their data flow smoothly...and without unnecessary manual effort. Organizations aren’t just doing more with their data, they’re also looking to inject efficiency into their processes so hands-on work can be reserved for when it’s really needed.

Closing the Loop:

For a long time, data was mostly a read-only affair, with the implications of analysis divorced from “upstream” data stack components. More recently, organizations have begun looking for ways to close the loop by pushing data and insights back into applications and operational tools where they can be surfaced to users in a relevant context, not just via BI.

No/Low-Code Data Tools:

Self-service BI probably kickstarted this revolution, but it’s certainly not the only innovation in the space. From drag-and-drop data pipelines to visual database design platforms, data tools are becoming increasingly accessible to less technical users.

Together, these shifts are prompting organizations to invest in data earlier and to do more complex analyses, often despite limited staffing and resources. Analytics has become table stakes in today’s business environment, but getting data analysis-ready can be a lot harder than it seems.

Common Data Problems

As their data operations evolve, many companies face similar issues. Early on, figuring out processes for managing data is a common problem. But as organizations mature in their handling of data, they're likely to become frustrated by both unsustainable lag around analysis and opportunity costs caused by inadequate data infrastructure.

Ad hoc analytics workflows lack “flow”

When first starting out with data, many businesses use a kind of ad hoc analytics workflow built on a collection of apps:

Custom code for periodic emails or occasional CSV dumps

Microsoft Excel or Google Sheets for collecting, organizing, and visualizing data

Dropbox, Google Drive, or Sharepoint for storing and sharing data

Google Analytics, Salesforce, or Shopify for tracking customer and commercial data

The benefit of this approach is that end users can jump into a familiar app and start working with data right away. Whether they're using in-app reporting to pull basic metrics or they're exporting the data to a spreadsheet for further analysis, this setup provides immediate access and has little to no learning curve, so it's incredibly pervasive.

But ad hoc workflows aren't a good way to store or manipulate data over the long run. Spreadsheets make collaboration difficult and aren't built for version control. Likewise, without fanatical organization, storage solutions like Dropbox or Google Drive can quickly become garbage heaps filled with outdated files of uncertain origin and accuracy.

Put simply, these tools just aren't built to work as a [single source of truth \(SSOT\)](#), and businesses that attempt to use them that way are likely to become frustrated as they scale or mature.

Analysis takes time, but not all of that time is well spent

Unless you're using canned insights, you're always going to have to dedicate some time to making data accessible and query-ready. And while organizations can rely on stock insights from tools like Google Analytics and Salesforce for a while, pre-built reporting is never going to provide the same level of analysis—or the same competitive edge—that a data analyst can deliver.

But even with a dedicated data team, a lot of time is required just to access data and make it ready for analysis:



While all of these tasks are necessary, when access and prep account for a third or more of your data team's time, that's a significant opportunity cost. Put another way, your data team is spending hours on busywork rather than being able to focus on the important work of running analyses or deriving insights.

Though handing over operational tasks to Engineering or IT is a common workaround, those teams can become blockers because data operations is only one of their many responsibilities. In fact, a recent survey revealed that 62% of data analysts reported that their work was regularly blocked by a lack of engineering resources¹.

Thankfully, there's a better solution. By developing an effective—and efficient!—data stack, organizations can free up time for analysis, making analysts more independent, data more accessible, and insights more abundant.

1. Source: [2020 State of Data Analytics, Fivetran](#)

2. Source: [2020 State of Data Science, Anaconda](#)

Poor data infrastructure is a barrier to success

Traditional data infrastructure involves on-premise deployments of expensive hardware. That approach often requires extensive in-house IT operations and can be slow to provide value to the organization at large.

The rise of cloud computing has shifted the burden of setup, deployment, and server maintenance away from in-house teams and onto cloud platforms like Amazon Web Services (AWS). While these solutions are often cheaper and more reliable, they still leave plenty of room for improvement.

[In surveys of data warehouse users conducted by Panoply](#), over 60% of respondents using the biggest cloud warehouse vendors (Amazon Redshift, Google BigQuery, and MS Azure SQL Server) still rated their data warehouse solution as “difficult” or “very difficult” to use. When asked why, 47% of respondents pointed to complex user interfaces and setup processes.

Cloud data warehousing has made storing data easier than ever. However, even the most common cloud-based storage still requires specialized knowledge and technical skills that mean data collection and prep are often divorced from analysis.

As a result, data operations and decision-making operations often live in separate silos, ultimately extending the time from data collection to actionable insight.

Fortunately, the data landscape is evolving. Tools are becoming easier to set up, maintain, and use, making advanced data operations a reality for companies of all sizes and freeing up teams to deliver valuable insights rather than dedicating hours to tedious infrastructure management.

Data stack overview

Setting up a data operation usually requires a combination of several services that handle different elements of data collection, storage, processing, and analysis. At its core, an effective data stack makes it possible to perform six basic operations:



Collection:

Data from payment processors, ad platforms, CRMs, CMSs, ecommerce platforms, web and mobile analytics tools, and social media sources can all be gathered and combined. Companies looking to centralize all their business data may find themselves juggling a variety of pre-built ETL tools—not all of which are user-friendly—alongside custom code.



Normalization:

Before being stored, data often needs to be formatted, combined, or normalized. For example, an organization could concatenate separate “first name” and “last name” fields into a single “name” field, impose a standardized date format across all their data, or otherwise unravel messy JSON or CSV files to make them work with a standard SQL database or apply essential business logic so the data is ready for downstream use in a BI tool.



Storage:

Once ingested, data needs to be stored in a place where it’s accessible for analysis. While this might seem like a simple operation, manually managing your storage requires serious technical skill to avoid decisions that will negatively impact cost or performance.



Transformation:

Once stored, raw data is prepared for analysis. Common transformations include joining tables, creating aggregations, and building in key business logic. Although transformation can occur alongside normalization in legacy workflows, post-storage transformation is a key part of a modern data stack.



Optimization:

With the high availability of data on potentially every aspect of a company's business, it's not enough simply to collect, clean, and store data. As datasets grow, further considerations need to be made—can the data be formatted to optimize the space it takes up, or the time it takes to query? Tools that automate these processes enhance performance by making data flow more efficiently.



Analysis:

Data analytics is the top of the data pyramid, the operation that every other part of the data stack is designed to support. Analysis encompasses a range of practices with varying degrees of complexity, from familiar business intelligence approaches such as dashboard construction to more complex machine learning algorithms.

Dialing in your data stack

Any stack that makes these six basic operations possible has the potential to generate useful insights for an analytics-focused organization. But there's a surprising amount of variety in the tools that manage these basic operations. Because tools have different assumptions about how best to manage your data baked into their features and processes, it's worth understanding the differences between their approaches.

Collection and transformation

Traditionally, data transformation was tightly coupled with data transfer because optimizing your data prior to putting it in a warehouse just made sense when storage came at a premium. Today, storage is significantly cheaper, which makes it possible—and arguably smarter—to store raw data and apply transformations afterward. This shift has given rise to two ways of approaching the process:

ETL (extract, transform, load):

An ETL tool manages the extraction of structured and unstructured data from various sources including spreadsheets, databases, or file stores. In this legacy approach, data is extracted from a source, cleaned or otherwise transformed, and then loaded into storage. Because the ETL process was developed in an environment where storage was costly, it applied transformations before making finalized data available for analysis.

ELT (extract, load, transform):

This modern manner of handling data cuts the time from extraction to insight. Because transformations are applied after data is loaded into a warehouse, analysts can get their hands on raw data as soon as it's in the repository. Storing raw data also makes analysis more flexible, as new or enhanced transformations can be instantly applied to stored data instead of requiring the ETL process to be restarted—and all the data reloaded—every time a change is required. For these reasons, many modern data analytics operations have shifted to an ELT framework in order to increase the agility and speed of their data operations.

Read more about [ETL vs ELT](#)

Storage

For many organizations, storing data means housing metrics within the apps that create them (e.g., leaving customer metrics in Zendesk) or exporting metrics to spreadsheets housed in a centralized tool like Google Drive or Dropbox.

While ad hoc storage can work for a while, companies eventually need to invest in a data stack. Organizations often reach the tipping point when they realize that they need to:

Streamline routine reporting that takes hours to assemble manually

Answer sophisticated questions that require data from multiple sources

Make data readily accessible to users throughout the organization

Ensure that all users are working from the most recent and most accurate data

Enable a newly hired analyst or newly assembled data team to deliver meaningful insights

At that point, companies need real data storage in the form of a database. Databases come in many flavors, but SQL-based relational databases have long dominated the field, and for good reason: the codebase is solid and the resulting databases are robust and relatively easy to configure.

At their core, databases and data warehouses are pretty similar in that both store data. But a data warehouse is a specialized type of storage designed to support analytics operations and that, through the use of data modeling and transformation, becomes an SSOT providing both easy access to data and consistent reporting across an organization.

Poor data infrastructure is a barrier to success

“Data analysis” covers a wide range of activities of varying degrees of complexity, but most analytics tools fall into two categories:

Business Intelligence:

The types of questions answered by BI are often necessary but relatively simple things like the current state of your inventory, how many customers you have, tracking incoming and outgoing payments, and so on. The data usually appears in routine reporting, simple data plots, or dashboards available to the entire team. However, more sophisticated questions—such as those that can only be answered by referring to data from multiple sources—quickly outstrip most BI tools’ built-in abilities and work best with a complete modern data stack.

Analytics and Data Science:

Data science uses more complex statistical techniques, machine learning, and potentially huge datasets to identify key performance indicators and generate predictions. Models of future consumer behavior based on past data rely heavily on data warehousing, as it provides both the historical training data that models are based upon and ongoing data that allow them to evolve. Applications of advanced analytics include machine learning algorithms for churn and risk prediction, fraud detection, and product recommendations.

For most organizations, BI makes up the majority of their analytics operation. In contrast, advanced analytics is less prevalent, but is growing in attractiveness as companies seek out answers to more complex, experimental, or predictive questions.

Read more: [BI tools](#) and analytics tools

Conclusion

The world of data has changed dramatically over the past decade. Companies are generating ever-more data, storage costs have plummeted, ELT has all but replaced inflexible ETL, and analysis has become table stakes for companies of all sizes.

Cutting through the jargon to figure out what you really need from a data stack can be challenging. But now that you have a handle on how data stacks have evolved and the most common components in a modern data stack, you can use this information to streamline data operations at your organization.

Ready to take a modern approach to managing your data? Get started with the easiest way to sync, store, and access all your data.

TRY IT NOW



partner
network



Google Cloud
Partner

About Panoply

Panoply is a cloud data platform that makes it easy to sync, store, and access your data. Panoply enables you to:

- Connect all your data sources without complicated code
- Automatically store raw data in the cloud in analysis-ready tables
- Build core business logic into your data to keep metrics consistent
- Seamlessly update dashboards and BI tools, no manual effort required
- Spend more time on analysis and less on managing data

If you'd like to learn more about Panoply and whether we're a good fit for your modern data stack, [book a demo](#) with us! We'd love to show off what Panoply can do and learn more about what your organization is hoping to achieve with data.

REQUEST DEMO

