How to Choose a **Data Analytics Platform**

Open for Innovation
KNIME

Table of Contents

What's the Next Corporate Balance Sheet?

While many say that data is the new oil of the economy, McKinsey suggests data usage is the next corporate balance sheet. In a new report, **McKinsey** states that by 2025, most employees will use data to optimize nearly every aspect of their work. This will not be limited to the daily reporting currently expected of most teams; instead they will "naturally and regularly" apply data analytics to resolve problems in hours (not weeks) and efficiently collaborate with internal and external stakeholders on projects.

While this might seem hard to believe after decades of noise around digital transformation, the last few years have seen increased momentum for data usage. The pandemic **accelerated digitization across practically every sector** by as many as four years. Businesses were pushed to prioritize efficiency and agility in the face of unstable economic conditions, disrupted supply chains, and labor shortages. And the competitive environment has added fuel to the fire. Studies have shown that organizations which fail to effectively leverage data are being sidelined by those who do, particularly in data-rich industries like banking and manufacturing.

> **"Smart workflows and seamless interactions** among humans and machines will likely be as standard as the corporate balance sheet"
>
> McKinsey, 2022

Bridging the Analytics Skills Gap

While many enterprises have made headway in digitizing their processes and building data strategies, a significant obstacle remains: A limited number of people can effectively make sense of data. **With a labor shortage of over 250,000 in the U.S. alone,** there are not enough data scientists to solve all of a company's data problems, let alone upskill an organization to handle even a fraction of their work.

87% of executives report that they currently face or anticipate facing skill shortages, **according to a 2020 survey of 1,216 participants** across the finance, retail, manufacturing and telecom industries. In the same report, data and analytics was reported as the biggest area of concern.

The problem is that data experts are often siloed from the groups they need to support and upskill, sitting in different departments and using tools and scripting languages few can understand.

Dividing these working groups has two negative impacts:

1. Business and domain experts cannot efficiently collaborate with data experts. Because any analysis has to be communicated through briefs or made available through custom-built frontend interfaces, the feedback loop is slow. Data problems can take months to solve, irrespective of urgency or manpower.
2. Business and domain experts are deterred from self-sufficient advanced analysis because the only tools available to them (e.g. Python or R) have a significant learning curve.

For this reason, leaders should consider no-code/low-code analytics platforms, which offer complete end-to-end support of any data project and more analytic depth than spreadsheets and BI tools, all while providing an intuitive drag-and-drop interface for fast onboarding.

Must-Have Capabilities of a Low-Code/No-Code Platform

When leaders look at the saturated market of no-code/low-code tools, it's difficult for them to know how to evaluate vendors. To expedite your search, we've gathered research from major analysts like McKinsey and Forrester, as well as peer review sites such as G2, Capterra, and Gartner.

We can summarize their criteria under three broad categories:

1 Analytic Depth

The platform should go beyond what's possible in spreadsheets and business intelligence tools, putting analytical techniques in the hands of business users and unlocking upskilling in an intuitive, easy-to-use platform. The platform must also support advanced use cases, leveraging the latest AI/ML techniques on the market to prove useful to data experts.

2 Complete Data Science Capabilities

The platform needs to support not just one-off analyses, but also enterprise-scale productionization. Teams need a way to deploy custom data products to end users, whether in the form of reports, dashboards, or APIs, with all the necessary training, validation, and monitoring to support those models in a production environment. Users should also be able to bring insights across the organization to 10, 50, or 1,000 through their deployments.

3 Enterprise-Scale

The platform needs to support large-scale onboarding and collaboration, supporting knowledge-sharing across an organization. At the same time, the platform needs to support data governance, centralized user authentication, and access controls. It also needs to be easily maintainable to reduce the burden and reliance on IT.

Vendors take different approaches toward these categories. Here is a full breakdown of what is necessary for a complete data science platform.

Analytic Breadth & Depth

Data Accessibility

Data accessibility is a prerequisite capability for any advanced analytics work. Many teams encounter too much friction when trying to get data out of their tools or move large amounts of data. If your platform doesn't easily connect all your data sources, it will require complex technical workarounds or expertise that isn't available on most business or data teams.

Not all no-code/low-code platforms are open. While many can handle standard databases, they struggle with accessing other data sources, like noSQL databases or custom data lakes. Open platforms offer the advantage of connectivity into any data source, whether it's a database, a data lake, an app, or simply a desktop folder. The better the connectivity, the more efficient and agile the users. And for those looking to work with more than strings and integers, it's also important to pick a tool that can support a wide array of data types, such as images, sounds, or spatial data, as well as domain-specific file formats like networks and molecules.

Because data analytics is still a rapidly evolving landscape, it's also important to evaluate how quickly your vendor of choice innovates. In this area, open-source platforms with a strong community offer a way to future-proof your investment, since they will consistently integrate new data sources.

Data Preparation

45% of any data worker's time is spent on data prep, according to a recent Anaconda survey. This process of extracting, transforming, validating, standardizing, and sometimes anonymizing is the most time-consuming and demotivating part of data work. Thus a platform which streamlines and automates parts of this process is a prerequisite for any team looking to work with a wide range of analytic techniques.

No-code/low-code platforms abstract all the technical complexity of automating getting data out of legacy sources, cleaning, transforming and finally loading it into a target system. This can be especially helpful with legacy tools which few engineers and data experts know how to work with. These platforms also typically let users save, reuse, and automate their data prep processes (among others) to reduce the time spent on repetitive manual work.

> **45% of any data worker's time is spent on data prep,** according to a recent Anaconda survey

It's also helpful to examine how no-code/low-code tools self-document — a feature which ensures easy handover, standardization, and knowledge sharing, which is really not possible with a scripting language.

Range of Analytical Techniques

No-code/low-code analytics platforms widen access to advanced analysis that goes beyond what a user could do in Excel or a business intelligence tool. However, various vendors of such platforms support different techniques for this. If a platform's range of analytical techniques is too narrow, it's less likely that experts can leverage it.

The most sophisticated no-code/low-code platforms enable users to do analyses of similar complexity to what can be done using a scripting language like Python, including classification, regression, time series analysis, deep learning, ensemble learning, clustering, and dimensionality reduction. Users can choose to apply any of these techniques without writing a line of code (hence "no-code"), or integrate their favorite scripting language (hence "low-code").

Another important aspect to consider is the granularity and flexibility with which one can apply these techniques. Many platforms purpose-built for business or spreadsheet users do not offer functions for a wide variety of other use cases, limiting advanced users' ability to execute ideas for solving a problem.

Access to AI/ML Libraries

Most AI/ML libraries let data experts leverage the latest innovations in data science. Because these libraries are sometimes built by thousands of contributors, it's difficult for any platform to match such capabilities on their own. Therefore, a no-code/low-code platform's integrations with all the common AI/ML libraries, such as TensorFlow or H2O, are essential for any team that plans on taking advantage of the most cutting-edge techniques on the market.

While many platforms have built-in capabilities or extensions that allow for access to AI/ML libraries, open platforms typically enable the widest access to the latest and greatest analytical techniques, and are typically the first to adapt new integrations.

Scripting Support

Many data scientists have extensive experience and training with building models in Python or R, and new libraries and techniques are often first introduced in these scripting communities. Extremely custom algorithms, or data prep methods that solve an edge case, can often only be performed with a scripting language. It's critical to choose a vendor that lets advanced users integrate code into their analytics workflow.

While nearly all vendors support scripting to some degree, it's important to evaluate how scripts perform within the platform, as well as its packaging and sharing capabilities. If the script is easy to share, Python scripters' expertise can be leveraged more widely in the organization.

Completeness of Platform

Ease of Deployment

Data science still faces a final-mile problem. Teams often encounter many technical hurdles before getting their models into production — i.e. making those models available for end users to consume through dashboards or data apps, or making them available as APIs for integration with a third-party tool.

Transforming data science into apps and services is non-trivial, and typically requires either dependence on IT or deep technical knowledge around deploying client server architectures. This is the precise reason many data scientists, despite having complete fluency in Python, decide to use a no-code/low-code platform which abstracts the complexity of deploying analytics as data products.

When evaluating a vendor, it's important to consider how easy it is to take the model or data preparation routine that has been built and apply it to a production environment, and where models can be deployed — whether in the cloud, on-prem, or hybrid.

Sharing Insights with End Users

The experience of the person making decisions based on the analytical model is an important consideration for anyone building data solutions for their organization. No-code/low-code platforms provide varying degrees of flexibility for anyone designing this end user experience.

Ultimately, end users consume analytics in three ways:

- **As a static report shared on an ongoing basis**. All vendors provide the capability to email or upload static reports on a scheduled basis.
- **As a data app or dashboard.** Some vendors enable users to create dynamic and parameterized applications, flexibly and without any code. While some vendors charge extra per end users, others allow for deployment to unlimited end users.
- **Through a third-party application, integrated via API.** While end users don't interact with APIs, the breadth of API support (REST, HTTP, etc.) varies between vendors, and makes the integration with third-party tools more or less feasible.

MLOps & Support of CDDS

MLOps is still a poorly defined function. The basic premise, borrowed from DevOps, is to create a process in which a software (or in our case, an analytical model) can be tested, validated, deployed, and monitored. This process is referred to as the continuous delivery of data science (CDDS). Most of this process needs to happen in an automatic or semi-automatic way to decrease bottlenecks or manual work for the administrator, who ensures all applications work as documented and adhere to guidelines for security or best practices.

While most vendors have some features that support the training, testing, and deployment of models, this is typically not enough to support most MLOps processes. Additionally, most organizations must be able to:

- Completely (and, for some, physically) separate dev, testing and production environments
- Define and customize the testing, validation and deployment stages (and more, if needed)
- Monitor the production process and define quality measures
- Trigger retraining or rollback of a workflow in production

Although most organizations share similar concerns for deployment practices, the deployment process itself can vary greatly. Factors such as the size of the organization, industry, resources, and IT and governance requirements determine what that process needs to look like. For this reason, any features or frameworks provided by the vendor need to be highly flexible and adaptable.

Enterprise Scale

Ease of Use

A no-code/low-code platform needs to make it easy for a user with any level of previous experience, so that they can quickly start building and deploying their analytical models.

A key quality to look out for is how much can truly be done without ever writing a line of code. While some platforms make it easy to build analytical models, they still require code for some aspect of productionization, like making the model available as a parameterized data application. The extent of the “no-code” capability is the true enabler for many in an organization.

Pre-canned Working Examples

For visual workflows, pre-built examples are a great way to help users understand the basics of no-code/low-code, and how to apply it to their business case. Rather than start with a blank canvas, they can download freely available working examples and make small adjustments as they become familiar with the tool.

This empowers technical and enablement teams to accelerate the upskilling of non-technical teams so they can concentrate on more complex tasks.

While some vendors provide a limited number of “getting started” kits, others have an active user community that continually contributes new working examples, freely available online. A large selection of working examples ensures that a wider range of users can get started quickly and onboard self-sufficiently.

Ease of Collaboration

Data science is a collaborative discipline, and no-code/low-code provides just the right environment for collaborating within and between teams.

An important capability to compare between vendors is bundling and sharing expertise across disciplines. It's essential that Python experts, ML engineers, and other technical experts can share their work as features non-experts can later access and reuse.

It's also important that teams of people with similar experience can easily work on the same project, building on each other's developments. Being able to easily share and version-control data solutions unlocks efficient and agile collaboration between team members.

Auditability

Lack of transparency around data is one of the most expensive risks a data team can introduce into an organization. Because no-code/low-code tools model a data process, they are inherently more transparent than scripted models. Even so, when applying machine learning or other more advanced techniques, it can be difficult even for a data analyst to understand what is being done to data.

For both compliance and best practices, businesses need to be able to demonstrate what is being done to data at every step. They also need to know exactly which libraries were used at execution time and what "non-controllable actions" (such as calling external packages) happened, so that further investigative activities can be performed if needed.

Total Cost of Ownership

No-code/low-code platforms can be expensive, so it's important to assess the ultimate cost as your internal data user community grows. Low start-up costs can be essential to getting buy-in across an organization before investing in features that support large-scale data science productionization.

Data science is a collaborative discipline, and no-code/low-code provides just the right environment for collaborating within and between teams.

Scale Data Impact with KNIME

KNIME Software is one of the leading no-code/low-code analytic platforms, unique in the industry for its open-source approach and low total cost of ownership.

Compared to other vendors, KNIME offers:

The most complete range of advanced analytic techniques available.

KNIME is intuitive enough for the business user, but sophisticated enough for the expert data scientist. KNIME's open ecosystem ensures access to dedicated ML libraries like TensorFlow, Keras, and H2O, and enables access to nearly all data types, like text, numbers, images, sound, spatial, molecule, and much more. KNIME can connect to all major databases and data warehouses such as SQL Server, Postgres, MySQL, Snowflake, Redshift, BigQuery, and more.

Data scientists who also want to use Python can run their scripts seamlessly in KNIME workflows with the KNIME Python Integration. The extension supports scripting, model building and prediction, and visualizations. KNIME users can choose to bundle Python scripts as reusable components to make them available to colleagues to use without needing to know Python. Furthermore, scripters can build out new functionality to enable non-experts to fully benefit from the python ecosystem.

Support for every step of the data science life cycle,

past deployment to monitoring models running in production. KNIME uniquely offers deployment through the same no-code/low-code interface as it does for the creation of data science. Through a process called integrated deployment, KNIME enables users to easily capture preprocessing and preparation steps, along with the models and analysis for reuse in the production environment. Users can easily make their analysis available as REST services or as data apps.

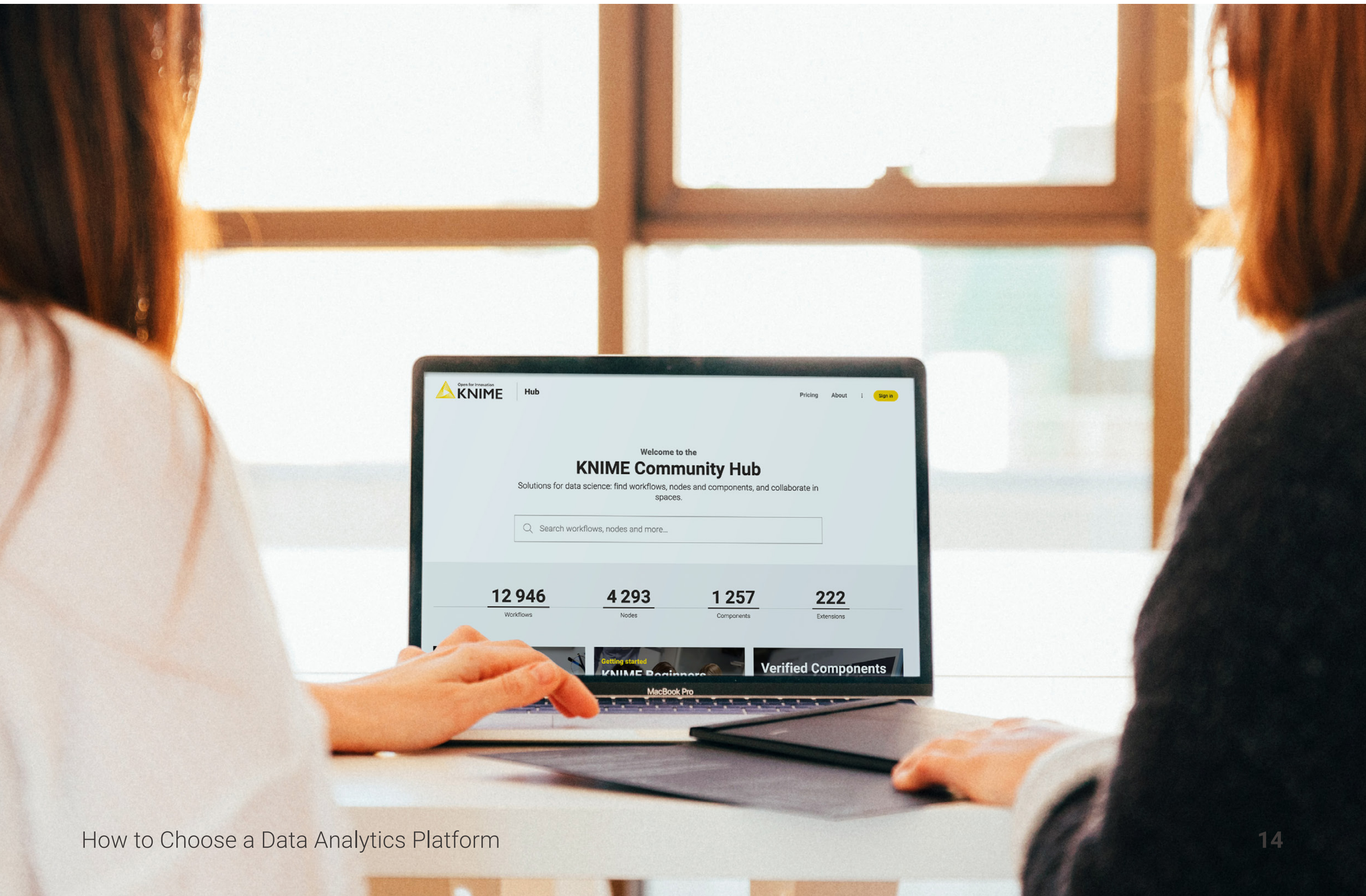
KNIME also supports large-scale productionization of data science by providing all the infrastructure appropriate for the testing, validation, deployment, and monitoring of models.

Enterprise scale

With origins in life sciences, KNIME has had enterprise-scale productionization in its product road map since Day 1.

The scalable, cloud-native architecture enables KNIME Business Hub to support an unlimited number of users, running any number of models, which are then deployed to any number of consumers. While IT has centralized control, teams can act independently and manage their own resources within parameters specified by IT. The result is high scalability, with minimal burden on IT.

KNIME has a vibrant open-source community of hundreds of thousands of users who share their data science solutions and extensions as components and workflows on KNIME Community Hub to help users accelerate their time to value. The solutions span across use cases and industries, with dedicated public spaces for Manufacturing, Financial Services, Retail & CPG, FP&A, Marketing Analytics, Customer Analytics, HR Analytics, and much more.



Peer Reviews on **KNIME**

“KNIME, **‘Swiss Army Knife’** For Data Blending, Transformation And Modeling”

Submitted Oct 29, 2020 | **5.0** ★★★★★ **Overall User Rating**

“...we use KNIME Analytics Platform as “Swiss Army Knife” for data blending, transformation and modeling. Major advantages are: friendly graphic environment to develop all kinds of data processes, great performance and, most important, a huge community ready to help users and develop new functionalities. Everything in an Open Source context (key point)”

Read the full review on Gartner.

“A Fantastic Analysis Tool With A Graphical User Interface”

Submitted Jun 8, 2022 | **5.0** ★★★★★ **Overall User Rating**

“Connect to open source: It also works well with a variety of other open source software for image analysis, including Python, R, Spark, and ImageJ. Outstanding functional integration: We never move data between applications/platforms to complete the project. Raw data can be easily ingested into the tool, processed, statistics run, summarized, and exported in a variety of formats.”

Read the full review on Gartner.

“KNIME Is The Paradigm Shift In Data Science With An **Open Analytics Platform** For Innovation”

Submitted Oct 6, 2020 | **5.0** ★★★★★ **Overall User Rating**

“Working with KNIME has been a very productive and pleasant experience. They seem to know the analytics filed (needs and wants of the customers) and their software offerings to prescribe an excellent match for powerful analytics solution. The analytics platforms itself is very intuitive, capable, modular and extremely easy to learn/use. “ **Read the full review on Gartner.**

Open for Innovation
KNIME

KNIME AG
Talacker 50
8001 Zurich, Switzerland

info@knime.com
www.knime.com