

Data warehouse platforms, demystified



How to choose the right platform
for your analytics workload



Contents

1

Choosing an analytics platform

Consider the pros and cons of each data warehouse platform option to optimize analytics results.

2

Cloud-based data warehouse

When agility and scalability are paramount, a cloud environment may be just what you need.

3

Integrated analytics solution

An integrated data warehouse appliance can give you the agility of cloud, right in your data center.

4

On-premises data warehouse

If your data already exists on-premises and regulations make it difficult to move, analyzing it in place can be the most effective solution.

5

Hybrid environments

The key to capitalizing on a multifaceted data warehouse approach is a common underlying platform.

6

IBM data warehousing

Support any and all of your data warehouse deployment options with IBM solutions.

04 Choosing an analytics platform

Today's complex analytics workloads require more than strong performance. They involve a diverse array of data sources and types—from structured transactional data residing on premises to unstructured, born-on-the-cloud data flowing in from Internet of Things (IoT) sensors and mobile devices.

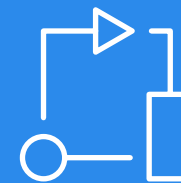
IT architects, business analysts and data scientists must integrate this data to generate the most impactful insights. By bringing together a wide variety of data, they can run more sophisticated queries and more complex analyses that ultimately yield deeper insights and fuel better data-driven decisions.

How do you choose the right data warehouse platform for your analytics workloads? Whether you are looking to improve analytics performance, incorporate new data sources, accommodate growing data volumes or better connect data scientists with the right data, you'll need to consider the pros and cons of each platform to optimize results.

Some organizations will need a holistic combination of multiple data warehouse solutions to create a cost-efficient, flexible foundation that can satisfy all users and needs across the business—be it the scalability of cloud, the security of on-premises data warehousing or appliances optimized for data science.



A cloud-based data warehouse might be the right choice for analyzing data that is born in the cloud.



An integrated analytics solution can offer high performance while minimizing the management burdens of operating a traditional “software-defined” data warehouse.



An on-premises data warehouse might be the best choice when data already exists on premises or when government regulations restrict moving data across state or country lines.

02

Is a cloud-based data warehouse the right choice for you?

The agility, scalability and elasticity of cloud solutions are spurring many organizations to consider the cloud for new IT initiatives. Here are several factors that can help you determine if the cloud is the right place for your workloads.

Analyze born-on-the-cloud data

Is your data “born on the cloud?” If you’re collecting IoT data from sensors or mobile devices, you might decide to analyze that data in the cloud as well. You can avoid the costs, complexity and risks of transferring tremendous volumes of cloud-generated data to your on-premises environment. Analyzing data where it resides helps accelerate results and reduce complexity.

Streamline budgeting and speed deployment

How fast do you need it? The budgeting and planning processes alone for a new on-premises data warehouse environment can be time-consuming. You might need to pull together people and information from multiple departments. Installing, configuring and testing the new data warehouse takes additional time. With a cloud-based solution, you can be up and running in minutes, not days or weeks.



33:

Average number of
unique data sources
that an organization
uses for analysis.

Source: Aberdeen Group, 2017 Big Data Survey, October 2017

Conserve IT resources

Do you want to reserve your IT staff resources for something other than managing an in-house data warehouse? A fully managed cloud-based data warehouse enables you to keep your staff focused on more strategic tasks. Offloading analytics to the cloud also helps you conserve your system resources.

Scale rapidly

Cloud-based data warehouse solutions can offer faster, easier scalability than on-premises environments. You can start with a short-term proof of concept and then scale up to match the requirements of the project when it goes live. For your in-market data warehouses, you can use elastic scaling to handle both planned and unplanned spikes in data warehouse requests or demand for computing power. Elastic computing helps you accommodate these short-term needs while continuing to meet service-level agreements (SLAs) for the applications. Elastic computing also enables elastic pricing: you can scale up—and down—while paying for only what you use.

Because a cloud-based data warehouse can be brought up and taken down with minimal investment in time and resources, it offers an ideal solution for augmenting your on-premises environment with data marts. For example, you can add cloud-based data warehousing resources for those few days per quarter when your regional offices need to run financial analyses.

Facilitate test and development

The flexibility of a cloud-based data warehouse makes it an ideal solution for many test and development projects. You can spin up resources quickly, expand your project as its scope changes and then spin resources down when you're done.

03 Integrated analytics solution

For some organizations and some workloads, an integrated data warehouse appliance, which combines hardware and software in a single solution, is the best approach.

Keep sensitive data in-house

Like implementing an on-premises data warehouse, deploying an integrated analytics solution on premises lets you keep your sensitive data in-house, where it's easier to comply with rigorous privacy regulations.

Speed deployment

Choosing an integrated solution helps accelerate deployment. You can avoid the time-consuming processes of procuring equipment, installing software and configuring the environment—the solution components are designed to work together, as they should, right out of the box. It's like having cloud agility but in your data center, where you can plug into your network and start loading data the same day.

Reduce management complexity

Selecting an integrated solution simplifies your data warehouse environment and reduces management burdens. You can save a significant amount of staff time by minimizing, and in some cases eliminating, administrative tasks such as tuning, indexing and aggregating tables.



Support a variety of data sources and types

The right integrated solution can support a full range of data sources and types—including structured, unstructured and semi-structured data—from a variety of on-premises and cloud sources. It will also provide flexible deployment options. An integrated solution that employs the same SQL engine as on-premises and cloud-based solutions lets you run analytics on premises today and in the cloud tomorrow.



Access robust performance

With the right integrated solution, you can tap into a high-performance, massively parallel system to analyze large volumes of data and generate real-time insights. By analyzing data where it resides, you can streamline queries and reduce the latency of results while also avoiding the cost and complexity of data movement.



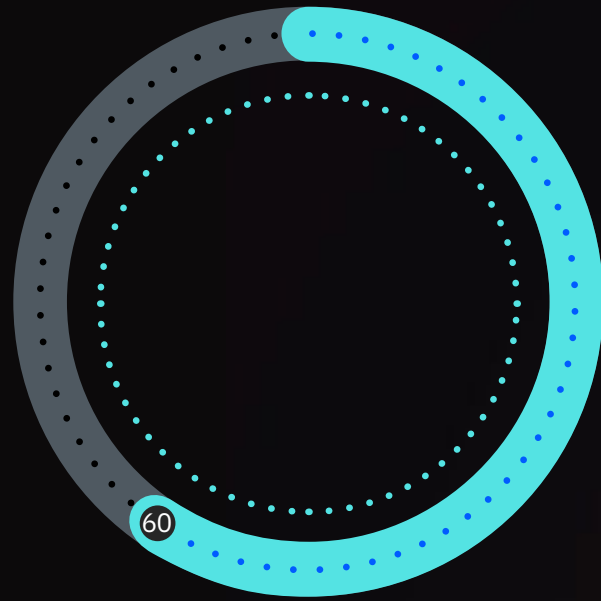
Support fast-growing data volumes

The right solution can also help you scale to support fast-growing data volumes. You can accommodate petabytes of data in a single environment and scale up performance in step to maintain a balanced system.



Capitalize on data science technologies

Some solutions enable you to support advanced analytics initiatives by leveraging the latest data science technologies to make better and faster decisions. For example, you can employ machine learning or artificial intelligence to improve stock recommendations, produce targeted advertising or enhance fraud detection.



60%
of respondents named
COMPLEXITY
as their main data
warehouse complaint.

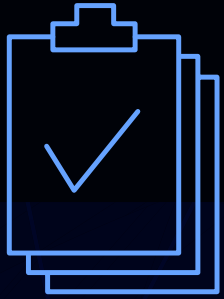
Source: Panoply, "Data warehouse trends report 2017,"
<https://learn.panoply.io/data-warehousing-trends-report-2017>

04 On-premises data warehouse

Does your data already exist on premises in a data mart or data warehouse? Then analyzing the data where it already resides might be your most efficient and effective option.

Employing an on-premises data warehouse can minimize analytics latency and reduce the risk and complexity of moving data while helping you maintain tight control over your data.

It also helps you avoid the potentially high costs of moving large amounts of data to another environment. Those costs could include expenses for high-speed network lines—rates can be especially high in regions around the world that lack a robust network infrastructure.



Comply with regulations

Are you prohibited from moving data? In healthcare, financial services and other fields, regulations might require you to keep sensitive data on premises. Even if regulations permit the transfer of data to the cloud, you might be in a country with strict requirements on where your data can reside.

Some regulations permit data to be moved to the cloud, but do not allow organizations to transfer data across state or country lines. With these considerations, you might decide to keep data on premises so you can retain better control of data and provide tighter security than some public cloud providers.



Maintain flexibility

Selecting an on-premises environment does not necessarily restrict your flexibility. For example, you can choose a data warehouse solution that lets you use your preferred hardware in a private cloud or virtual private cloud configuration.

A virtualized environment can also help enhance agility. With the right solution, you can deploy a pre-configured data warehouse on a Docker container in minutes. Automated scalability helps accommodate new analytics demands easily.



Leverage your existing IT environment

Have you made significant investments in your on-premises data warehouse? If you have an advanced infrastructure and strong, established skills for managing your data warehouse, you have good incentives to continue using those resources.



Enhance performance

By choosing an on-premises data warehouse that combines in-memory processing with in-database analytics, you can capitalize on robust performance that enables fast processing of even complex queries. Using an on-premises solution for analyzing on-premises data lets you keep latency to a minimum and helps reduce the complexity and risk of moving data to an analytics cluster.

52% of respondents state that
the enterprise data warehouse
is the most likely way
for them to derive value
from big data.

52

Source: Gartner Research, Inc., “Efficiently Evolving Data from the Data Lake to the Data Warehouse,” Rick Greenwald, Ehtisham Zaidi, November 2017.

Hybrid environments

Many organizations can benefit from combining multiple data warehousing platforms.

For example, you might add a cloud-based data warehouse to support key parts of the business with data marts, or incorporate a cloud-based data warehouse to accommodate a new data type. You could analyze IoT data in the cloud while keeping other transactional data on premises.

- 
- **You might also decide to add an integrated data warehouse solution to your existing infrastructure.** For example, you could implement an integrated system in a regional office to provide on-premises analytics capabilities while avoiding the need to hire IT specialists to run it. An integrated solution could also complement a cloud-based approach, enabling you to keep sensitive data in house without adding administrative complexity.
 - **The key to capitalizing on a multifaceted data warehouse approach is a common underlying platform.** If you choose solutions that share a common SQL engine, embedded Apache Spark analytics capabilities, common tools and unifying data software, you can leverage existing skills while improving efficiency across transaction and analytic workloads. A common SQL engine across data warehouse solutions enables you to move workloads to easily match the workload with the right deployment model. Embedded Spark analytics let you run analytics in each environment where the data resides, without having to move data to a specialized Spark cluster.

IBM data warehousing

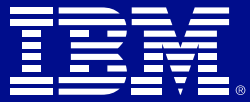


As you look for ways to expand and enhance your data warehousing capabilities, you'll need to evaluate several possible platforms—including on-premises, cloud and integrated solutions—and choose the best match for your workload, technical and business requirements.

Fortunately, you don't have to choose just one platform. The IBM® Db2® portfolio of data warehouse offerings includes on-premises, cloud and integrated solutions—and all support a hybrid approach to data warehousing. A common SQL engine, embedded Spark analytics, common tools and a unifying database platform give you the flexibility to combine more than one platform so you can satisfy all of your data warehouse needs.

Ready to learn more?

ibm.com/analytics/data-management/data-warehouse



© Copyright IBM Corporation 2018

IBM Corporation
New Orchard Road
Armonk, NY 10504

Produced in the United States of America
April 2018

IBM, the IBM logo, ibm.com, and Db2 are trademarks of IBM Corp. registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the web at “Copyright and trademark information” at ibm.com/legal/copytrade.shtml

This document is current as of the initial date of publication and may be changed by IBM at any time. Not all offerings are available in every country in which IBM operates.

THE INFORMATION IN THIS DOCUMENT IS PROVIDED “AS IS” WITHOUT ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT. IBM products are warranted according to the terms and conditions of the agreements under which they are provided.

The client is responsible for ensuring compliance with laws and regulations applicable to it. IBM does not provide legal advice or represent or warrant that its services or products will ensure that the client is in compliance with any law or regulation.

17014717-USEN-01