Managing Data in the Cloud

A Guide to Help You Manage Data Sources, Quality, Processes, Analytics, and More

Author:
Maria C. Villar
April 2021

Table of Contents

What Do You Need to Consider When Managing Data in the Cloud?

As more companies pursue a "cloud first" IT environment, chief data officers (CDOs) will need to reevaluate their data strategy. While managing all critical data is important, managing data in the cloud requires **special consideration**. CDOs can use this shift as an opportunity to simplify and reengineer existing data processes and practices. A successful move to a cloud data warehouse requires a plan for mapping, transforming, and cleaning data. This includes data both at rest and in motion.

When managing data in the cloud, there are three important scenarios:

1. **Managing master data, including hybrid business processes**
   A hybrid business process may well handle customer data in a customer cloud solution, employee data in a human resource cloud solution, and supplier data in a procurement marketplace.

2. **Managing cloud data warehouses**
   A data lake or warehouse is hosted in a cloud infrastructure.

3. **Managing third-party data providers**
   This includes third-party data services such as Dun & Bradstreet Holdings Inc. (D&B), data marketplace providers such as Datarade GmbH, data exchanges such as CDQ GmbH, as well as various social sources.

A recent data strategy series describes both the business and technical capabilities required to manage data effectively.

These high-level capabilities apply to managing all types of critical data, whether on premise or in the cloud. This framework can be used to assess the changes that will be required in each of the three scenarios described above. Once you understand the changes that will be required, what other things should CDOs consider? See Figure 1 for some answers.

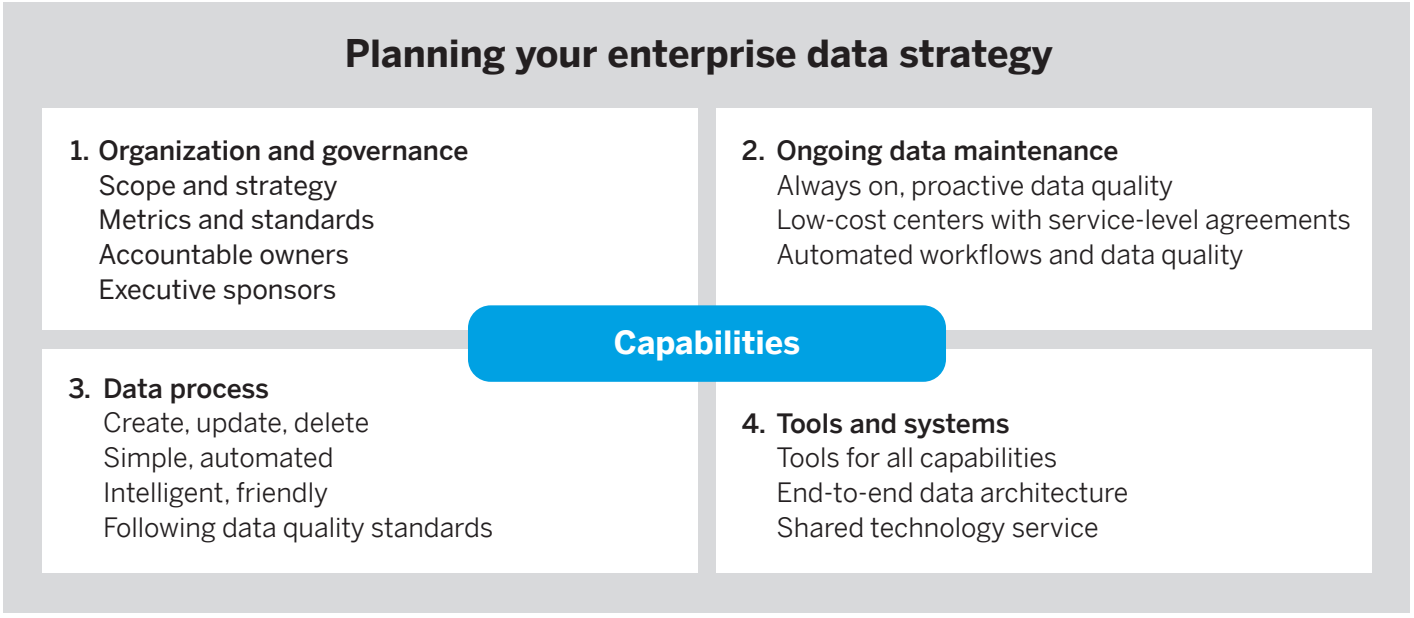


Figure 1: Outcome-Driven Enterprise Data Strategy

Managing Master Data in the Cloud

Let's assume a company has chosen to run one or more of its business software applications or business processes from a cloud software solution provider, for example, software as a service (SaaS), platform as a service (PaaS), or business process as a service (BPaaS). Popular examples include CRM applications for sales management and marketing processes, workforce applications for recruiting and hiring, and supply chain applications for inventory management. SaaS examples include productivity apps such as Microsoft Office, but from a CDO's perspective, these apps will not be managing master data.

Running a cloud software solution involves standard business processes and infrastructure. Whether to use an underlying cloud infrastructure that is private, hybrid, or public is a decision the CIO team must make. Performance and responsiveness will be key criteria of this decision because delighting business users can quickly fail when they have to wait. And wait.

In some cases, CIOs choose to retain some systems on premise while moving workloads to the cloud to meet the needs for responsiveness and performance.

While technical, the decisions about where that data lives and how it is replicated or federated across the business processes has a profound impact on managing master data. All of the above will have implications for the master data. For that reason, the CDO should be involved in the analysis and the decision.

When business processes are run in a cloud software solution, master data will be created and updated in that solution. The cloud solution can also generate new analytical insights about the master data. For a CDO, these data strategy capabilities should be evaluated.

AUTHORITATIVE DATA SOURCES AND LEADING SYSTEMS

Deciding which critical master data fields will be managed (and where they are to be created, maintained, and replicated) is a significant strategic decision for the CDO. These decisions must be made early in your cloud software solution deployment strategy. The decision becomes even more important when multiple cloud software applications use the same master data, which complicates the issue.

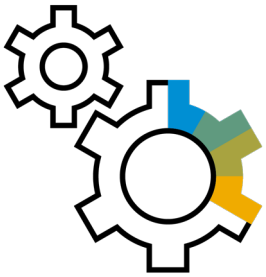
As you approach these decisions, there are a number of related questions you should ask.

For example:

Will the cloud software application's data repository be the authoritative source? Or will the company maintain a separate authoritative master data system, and if so, how will the two stay in sync?

If similar data fields are maintained in both the cloud software application and the trusted source, which is to be the leading system for creating and updating the field?

When running an end-to-end business process in the cloud, multiple subprocesses are often implemented in different SaaS applications hosted by different cloud software vendors.



For executing an **end-to-end** lead management business process, over 8,000 SaaS vendors provide point solutions today. For sales processes, there are over 900 cloud point solutions, for finance processes, over 1,300 cloud point solutions.

The number of point cloud software applications will continue to grow. Often one cloud software solution provider will partner with these point software application providers to provide some form of process and application integration. The integration of data here is not seamless.

For a CDO, the more cloud software applications your company has, the more important the decision becomes of which leading systems to use and which trusted authoritative sources to designate.

Some options to consider:

An on-premise, trusted master data management (MDM) system should be maintained.
The cloud software applications may still capture some critical master data fields and then transfer those fields to the one MDM trusted source system. Other fields are captured only by the MDM system. An example is credit card information. Because of the inherent risks and sensitivity of this information, many companies maintain and store this information only in their in-house MDM system.

Contact and address information are typically in both systems.
You will need to decide on what leading system to use to create and update master data fields. If the fields are dispersed, you will also need an operational process to identify and resolve inconsistencies. In particular, you must determine the rules for handling personal contact data that can be updated by an individual in a cloud software application versus the same "contact data" in an MDM account record. Enterprise contact data should be considered critical enterprise data and should not be overridden without the appropriate checks and balances.

For those companies with no separate MDM solution, but with multiple cloud applications all able to update the same master data, the CDO will have some important considerations to address. Let's take a look at them.

Choose one of the cloud software applications to create and update all master data fields.
That one cloud application will then distribute and publish to all the applications that require the data, whether the application is on premise or in the cloud. Rules for distribution are also needed (for example, to which system and how often). The more widely distributed the cloud processes are outsourced, the more advantageous one solution will be.

Federate the data governance across multiple SaaS solutions using a cloud MDM platform.
That platform will integrate and manage the tasks while leveraging a single data model per master data domain that all cloud software solutions comply with. Without an enabling technology solution, federating master data governance will be highly complex and error-prone.

DATA QUALITY MANAGEMENT

Typically, the primary responsibility and operation for data quality management will still be with the CDO's data management team and not the cloud software solution provider. To assist the data quality efforts, the cloud software solution providers may provide data quality management services or add-ons, for example, D&B customer data lookup services or data marketplaces for data enrichment. Executing data quality tasks within the cloud software business process has its advantages. However, a CDO will still want to be able to integrate the cloud software data quality tasks, results, and monitoring within the overall CDO data quality dashboard.

A data quality checklist includes the following tasks:

- **Define the accountable roles and responsibilities** for ongoing data quality maintenance between the cloud software solution provider and the company's data management team. Additional checks and balances are recommended to ensure the cloud provider performs the tasks effectively.
- **Establish data archive schedules and rules.** While cloud storage is economical, the risks of accumulating too much data is higher. Therefore, establish an archive-and-delete schedule to be performed automatically by the cloud provider.
- **Define or reuse other data governance rules, standards, and policies** for the master data fields that reside in the cloud software infrastructure. Certainly, rules for data access will be reused as will identity management rules.
- **Evaluate the data quality services** provided by the cloud software solution provider. Usually these are add-on components or partner solutions. Criteria for selection should include the functionality, ease of use, and ability to integrate with other CDO data quality tools.
- **Integrate the results in the overall CDO data quality dashboard** if you use cloud software data quality monitoring or reporting services. This ensures a comprehensive view of all MDM data fields, whether on premise or cloud-provided.

DATA QUALITY
CONSIDERATIONS FOR THE CDO

The enterprise will see benefits in performing data quality tasks within the cloud software business process.

The CDO's overall data quality dashboard will still need to integrate important cloud software data quality metrics, tasks, results, and monitoring.

DATA PROCESSES AND LIFECYCLE MANAGEMENT

When deciding which cloud software or MDM solution and repository to use as the leading system and trusted source, you will need to consider which data lifecycle processes will be executed in the cloud software solution. Most cloud software solutions can execute these tasks as part of their standard functionality, but it will be up to the CDO team to decide if that capability is used. If you do not use the "out-of-the-box" cloud capability, then a central create-and-update data process and software application (such as an MDM system) must be developed by the internal CDO team and called from within the cloud software solution.

- **Decide which master data lifecycle** tasks will be executed in the cloud software solution, such as create, update, delete, and archive.
- **Redesign existing legacy processes** with new, simple, user-friendly processes if you are designing a new master data cloud process. Automate the field entry process to speed data entry and maintain consistent data quality. Use known, trusted sources or look-ups where possible
- **For those master data processes not executed in the cloud**, determine how the cloud master data process will be disabled to make sure it is not inadvertently used.
- **Create or reuse business rules and standards** for creating, updating, and archiving master data fields in the cloud software solutions. Often the rules and standards used on premise can be reused in your cloud environment.
- **Define the checks and balances** used to ensure data is current and accurate in cases where your customers, suppliers, and employees can maintain their own contact data in the cloud.

  Enabling employees, customers, and suppliers to maintain their own company and contact data is not without risk. In our experience, self-service does not guarantee high data quality accuracy or timeliness of updates. There will still need to be a data quality review process monitored by the central data team for self-service data.

MASTER DATA ANALYTICS

Most SaaS, PaaS, and BPaaS solutions provide analytics and insights for the master data in their cloud software solution, such as:

- Customers and employees at risk
- Customer and employee churn
- Customer propensity to buy

Decide which insights and analytics generated in the cloud software solution are new master data–critical fields and how they should be stored and managed external to the cloud software solution.

As part of the CIO's service-level agreement and the performance monitoring of the cloud software provider's IT services, the CDO should add appropriate master data KPI metrics for the accountable data tasks.

DATA LIABILITY AND RISKS

While cloud software environments often have stronger security and data protection than what is found on premise, there will still be risks the CDO must evaluate. In addition to defining the data lifecycle tasks to be performed by the cloud software provider, the CDO must also understand and clearly document the liability risk for data exposure and data risk mitigation obligations.

What must be in the contract?

- **Privacy compliance responsibility:** Within the context of the General Data Protection Regulation (GDPR) and other similar privacy regulations, the cloud software solution provider usually has the role of the "data processor" as defined by GDPR. How the cloud software provider executes that role should be clearly understood by the CDO and the chief privacy officer.
- **Data location assessment:** In some industries and in some countries, legal and often regulatory requirements specify where critical data is to reside physically. Assess where sensitive data (such as customer, employee, and patient information) can be hosted.
- **Security and data transfer risk management:** Assess the encryption and anonymization provided by the cloud provider.
- **A contingency plan:** Establishing a contingency plan allows your company to switch vendors and ensure all company data in the cloud software solution is provided. Many customers establish a refresh plan and maintain a complete copy of all the data they need. Having a complete copy supports continuation of service in cases when the cloud provider service is not operational.
- **Identity management:** Identity management becomes even more important when the company has multiple SaaS, PasS, or BPaaS vendors. Key mapping of critical fields across SaaS is also important.

TECHNICAL CONSIDERATIONS

Implementing one business process in the cloud may require multiple SaaS, PaaS, and BPaaS solutions, each with its own data model. For a CDO, this may introduce complicated data integration and data definition issues as well as new data silos.

Resolution of these issues requires synchronization across hybrid business processes that spans on-premise and cloud solutions. You must also consider key capabilities, such as common data models, data governance, key mapping, identity management, and data cataloging.

As mentioned, the cloud software provider may be able to provide data quality tools and processes, which can supplement the CDO's own data quality tools and processes. An example of this includes SAP® Ariba® solutions, which support supplier industry data standardization and product hierarchy.

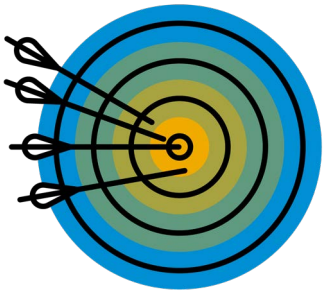
If the CDO team already uses third-party data sources for data quality enrichment, important technical considerations include the usability and ease of accessing and integrating third-party data to supplement the cloud master data process and analytics.

Most cloud solutions are designed for low-volume master data updates. Consider a staging area or data sidecar for processing volume master data in the cloud.

KEY TAKEAWAYS

To federate or to consolidate master data governance management is a key strategic decision for the CDO.

CDOs must to consider which data lifecycle processes to execute in the cloud software solution. Implementing one business process in the cloud could require multiple SaaS, PaaS, and BPaaS solutions, introducing new data silos as well as complicated data integration and data definition issues the CDO must resolve.



Implementing **one business process** in the cloud could require multiple SaaS, PaaS, and BPaaS solutions, introducing new data silos as well as complicated data integration and data definition issues the CDO must resolve.

Managing Cloud Data Warehouses

In the cloud data management scenario, a data repository is hosted in a cloud infrastructure environment, such as a data lake or a data warehouse as a service (DWaaS). Cloud repositories are often created to perform analytics or collect and store diverse information from numerous sources in a data warehouse. New, fast cloud warehouses are rising in popularity and are replacing in-house warehouses for "fast path" business-led initiatives.

A modern, cloud data storage architecture has clear advantages in its elasticity, efficiency, addressability, and query optimization. In addition to its technical architectural advantages, you can also realize business benefits.

Deployment speed is always a concern. With data warehousing in the cloud, an individual department can acquire nearly unlimited computing power and data storage with just a few clicks.

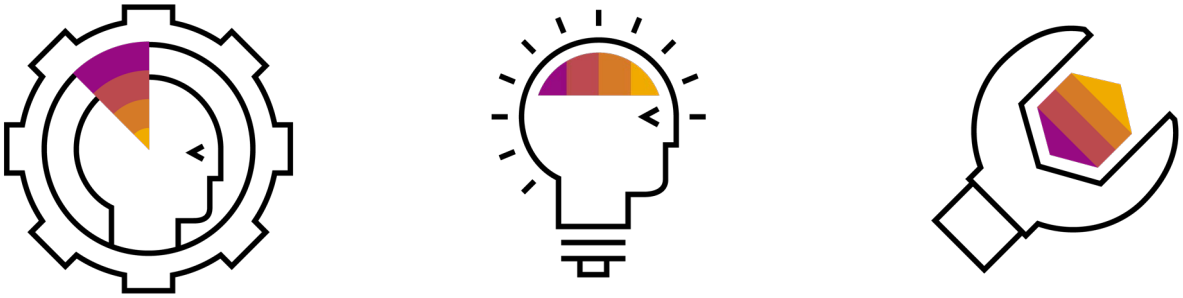
Total cost of ownership (TCO) is also key. The DWaaS pricing model is set up to charge only for resources used. This means no forecasting or paying for long-term needs or up-front costs, such as expensive hardware, server rooms, or maintenance staff. Separating storage pricing from computing pricing has the added advantage of driving down costs.

What about flexibility and scalability? A cloud data warehouse can scale up or down dynamically, as needed. The cloud gives us a virtualized, highly distributed environment that can manage huge volumes of data in a highly scalable way.

Now more than ever, security and disaster recovery are on a CDO's mind. In many cases, cloud data warehouse providers offer stronger data security and encryption than that available in many in-house, on-premise data warehouses. You may also configure your cloud system to ensure data is automatically duplicated and backed up to minimize the risk of lost data. Remember to verify that your service-level agreements cover these requirements.

Integrating new technology is also a factor. Cloud data warehouse providers often include new technologies, such as machine learning, artificial intelligence, and tools for data analysis. You can easily integrate these technologies into the cloud data warehouse solution (see Figure 2).

Additional cloud data warehouse capabilities



Machine learning **Artificial intelligence** **Analytical tools**

Figure 2: Cloud Data Warehouse

From a data management perspective, the cloud data warehouse can handle high data volumes, data types, data complexity, and data velocity, but may also introduce more data risks.

In our experience, many CDOs identify "moving to the cloud" as their most pressing challenge.

TRUSTED DATA SOURCES

The cloud warehouse integrates and persists many sources of data, including internal company data and external third-party data. With multiple possible data sources for the same data type and field, a CDO must provide guidance on the authoritative trusted source for each of the critical fields in the cloud repository. A critical field could be master data, transactional data, or reference data.

The decision of which source to use for each field's authoritative trusted source becomes even more important when a broader cloud repository is to be used and shared across the enterprise. The broader the use, the more critical the decision becomes. In addition to advising on the internal authoritative sources in the cloud repository, the CDO should also drive (or participate in) the acquisition and management of the third-party data. This will help ensure the data source is reliable, representative, and as risk-free as possible. Further details are provided in "Managing Third-Party Data Providers in the Cloud".

DATA QUALITY MANAGEMENT

There are no inherent data quality benefits with a cloud data warehouse. Data quality must still be proactively managed by the CDO team. Many of the same master data quality practices and on-premise warehouse data quality practices apply.

Similar to on-premise data warehouses, cloud data warehouses require ongoing data quality management to ensure the data quality fits the purpose intended for the repository. For the cloud data warehouse, the usual data quality parameters are still relevant, such as data consistency, validity, accuracy, completeness, timeliness, and uniqueness (see Figure 3).

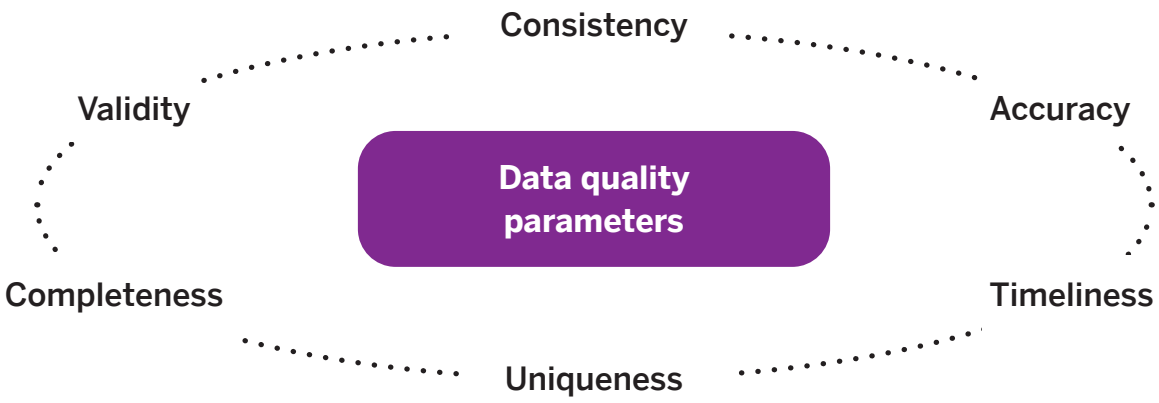


Figure 3: Data Quality Parameters

A cloud data warehouse for research purposes requires a different level of data quality and different data quality parameters than a repository for financial analytics. In partnership with the business users, the CDO must determine the dimensions of quality that will materially impact the cloud data warehouse's business purpose and the scope of the resulting data quality program.

As with an on-premise data warehouse, once the critical data quality parameters are identified, the CDO can establish the corresponding data quality rules, tasks, and accountable roles. Another important task is to evaluate the reuse of existing business rules and standards from the on-premise data quality program.

There are many parallels between on-premise and cloud warehouse data quality management. However, a few tasks involved in cloud warehouse data quality management require special consideration given the large volume, quantity, and variety of data integrated in that scenario.

Cloud data warehouse quality management tasks include:

- Developing an automated, proactive archive-and-delete calendar and rules for "end of purpose" checks is important. As with master data quality management, while cloud storage is inexpensive, keeping data after its use has been exhausted presents data risks. Additionally, stale data is problematic for predictive analytics and can lead to poor business decisions.

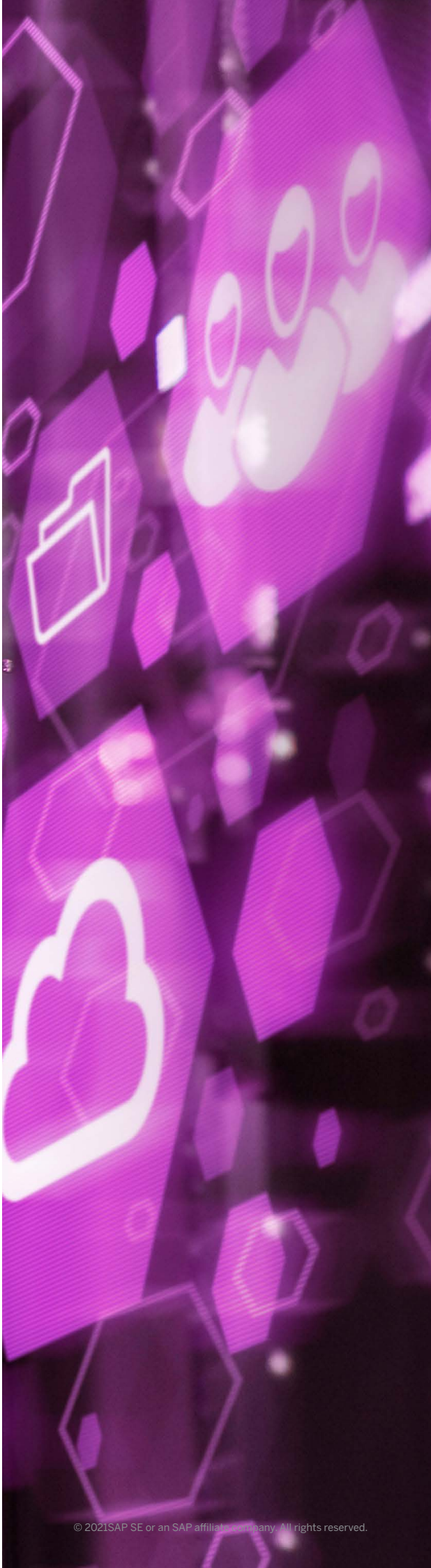
- Data quality rules and tasks for duplicate matching, elimination, and best-record consolidation are especially significant and complex.

- Managing high data volumes and real-time data makes automation and real-time testing even more vital. Most cloud data warehouse providers offer cloud data quality management add-on services or tools.

- Management of data quality for purchased third-party data may not be possible. For that reason, it is especially important for the CDO team to profile the data prior to acquisition to ensure the quality level fits the purpose. Third-party data providers want feedback on data quality issues.

DATA PROCESSES AND LIFECYCLE MANAGEMENT

Unlike SaaS and BPaaS cloud solutions, cloud data warehouses do not typically create or update master data. The typical data lifecycle process in a DWaaS consists of data transfers, data integration and transformation, data quality management, data ingestion, data protection, data backup, and data deletion. The growing trend is to automate these data processes, reducing overall TCO. Cloud data management platforms provide integrated sets of tools to manage these tasks, often augmenting the tools with advanced AI and machine learning.



For each of the applicable data lifecycle processes, the data management roles and responsibilities across all stakeholders should be agreed on and part of the overall contract. This agreement should include the DWaaS provider, your company's internal IT teams, and the CDO team.

If multiple cloud providers are involved, or a combination of cloud and on-premise systems are involved, the CDO must design data policies for security, compliance, and data lifecycle management that span all environments.

Finally, the CDO must establish specific data process KPIs to track the execution of the DWaaS provider`s against the provider's primary data tasks and service-level agreement.

ANALYTICS

The COVID-19 pandemic has forced companies to be more nimble, flexible, and fast. Analytics is at the heart of this transformation. It enables speed through insight-based decision-making and faster course corrections to ultimately become faster and more agile.

In modernizing the analytics environment, the CDO has to consider the multiple advantages of implementing analytics in the cloud. Benefits include speed of innovation and operation, greater cost savings, secure access to strategic insights, real-time analysis of data regardless of its location (on premise or in the cloud), increased mobility, and the potential for more accurate and timely forecasting. These make cloud analytics a high-impact driver of innovation and value.

Recent market trends show that analytics and data management for analytics are the top two use cases accelerating the move to the cloud due to the pandemic.

Cloud analytics enables more insight-driven decision-making but is also becoming more complex. Before you decide to implement cloud analytics, you must first address some key questions in order to ensure success.

Everyone in the organization makes decisions (executives, managers, analysts, and front-line workers). Some decisions are strategic, some are operational, and some are tactical. How do you support the diverse decision-making needs of your entire organization?

Decision-making is a continuous and collaborative process that involves a number of steps. You may start with a question that requires an answer, or you may have received an alert that prompts you to investigate. Then, you analyze what's happening and predict what's likely to happen next so you can plan for it accordingly. That plan may lead to actions, such as making an incremental investment, launching a new product, or hiring a new employee. Ultimately you will report on the outcomes of each action. How do you ensure that the complete process is supported in the most efficient way with no disconnect between all these activities?

Successful analytics pulls insights from multiple and various applications and data sources. Data gravity is a key influencer for your analytics strategy.

Cloud analytics is the most effective way to access cloud data. However, this option may affect your ability to access all of your company data, with a large part of your data remaining on premise for some time. What hybrid architecture will be best suited to ensure timely access to all your data?

Looking at all these questions and considering organizational structure and requirements, the CDO can now begin to plan for moving forward with cloud analytics.

Cloud analytics is your best way to access cloud data but may affect your ability to access all company data, requiring some of your company's data to remain on premise. What **hybrid architecture** is best suited to ensure timely access to all your data?

TECHNICAL CONSIDERATIONS

Cloud data management platforms that manage all the data tasks in a DWaaS are becoming more available, including those from traditional data management software vendors. These cloud data management platforms are comprehensive and include integrated cloud data management services, such as data discovery and cataloging, data ingestion and integration, data quality, and data governance. Often the platform includes engines for AI and machine learning as well as analytic platforms. Integrated platforms have advantages over "best in class" point solutions for each of these capabilities, such as ease of use and ease of integration across the various data services.

Data cataloging and robust metadata are especially important in a DWaaS for effective data discovery, search, and reuse of third-party data and nontraditional data sources. These cloud data management platforms will include these capabilities.

Additional data management technical tasks (see Figure 4) will still be performed by your company's IT team:

- Data protection and security review
- Key management
- Creation of a virtual network to separate the object storage from users (assuming zero trust for network and user access, with no single sign-on and where sign-on must be reverified in the cloud)

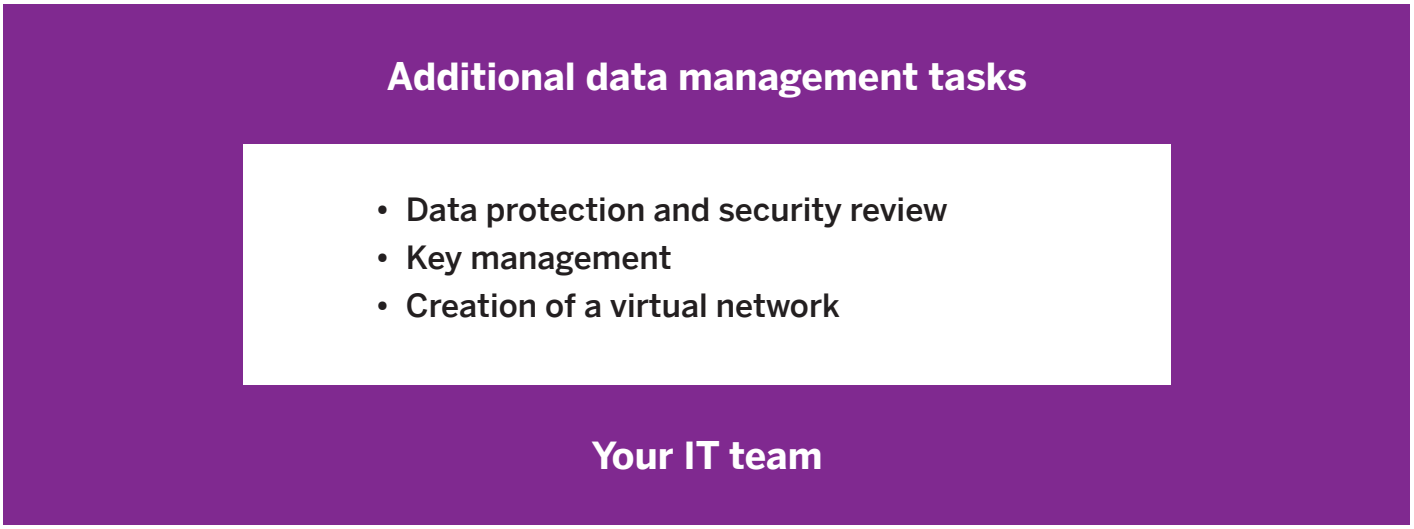
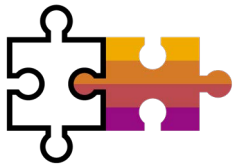


Figure 4: IT Team Tasks Required

New cloud skills may be required by your IT team in areas such as cloud architecture, cloud administration, cloud integration, and automation.

Data architecture skills are essential to keep the data workload in the right place relative to the data being used. This reduces network traffic and costs due to data movement. Otherwise high data transfer and egress fees may result when data is moved out of a cloud environment.



Often the platform includes engines for AI, machine learning, and analytics. These **integrated platforms** have advantages over "best in class" point solutions for ease of use and ease of integration across the various data services.

DATA LIABILITY AND RISKS

Certainly, data liability and risks exist in DWaaS similar to those in on-premise data warehouses. Often the data protection offered by your cloud provider will be superior to what exists in the internal infrastructure.

The CDO team, in partnership with the IT security team and chief privacy officer, should review the cloud provider's encryption, masking, and anonymization methods to ensure they comply with your company's security and privacy standards. This is especially important for protecting personally identifiable information (PII) and data in motion, either from IoT devices or real-time data streams.

CDOs must understand how the data warehouse will be used by the business community, whether on premise or in the cloud. Remember that one of the fast-growing risk types is "data ethics" risks caused by bias in the data and the analytic algorithms. Best-of-breed organizations implement data ethics guidelines and develop an operational framework to manage these risks (see Figure 5).

Developing your data ethics guidelines

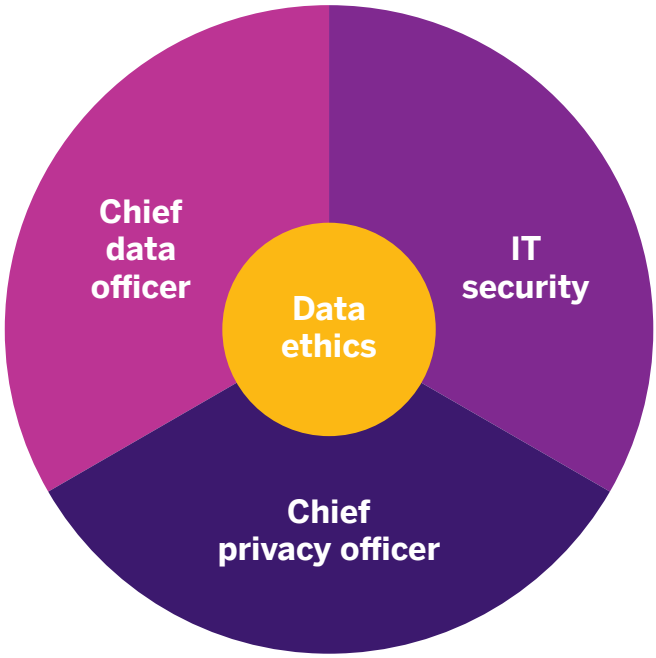


Figure 5: Data Ethics Guidelines

KEY TAKEAWAYS

- Data exchanges and data services, available from the cloud provider, can facilitate data management tasks.
- Data quality management for on-premise and cloud data warehouses is similar in many ways. However, the ease of integrating new data types and volumes in real time in the cloud warehouse increases the need for automation and real-time testing.
- Cloud data management platforms that manage all data tasks in a DWaaS provide comprehensive services that include integrated cloud data management services. This offers clear advantages over best-of-breed point solutions.

MANAGING CLOUD DATA AND DATA ETHICS

CDOs must understand how the business community will use the data warehouse, both on premise or in the cloud.

Keep in mind the risks associated with "data ethics" caused by bias in your data and your analytic algorithms. Data ethics guidelines and an operational framework enable best-of-breed organizations to manage these risks.

Managing Third-Party Data Providers in the Cloud

In managing your third-party data sources in the cloud, what type of things should you consider?

Acquiring, storing, and using third-party data sources have been requirements for CDOs since the 1980s, especially in the financial services and telecommunication industries.

Third-party data has commonly been used to enable and enhance internal master data content for data quality management and to supplement business-led research and analytics.

What have we learned over time?

- **Finding suitable third-party data providers** is time-consuming and often requires using unconventional (and perhaps outdated) sources, such as industry magazines, domain experts, and country residents. Often the same or similar data is purchased separately by different business units.

- **The process to integrate third-party data** into the company's internal infrastructure is often lengthy and complicated, not to mention the need to ensure the data is used legally across the corporate ecosystem.

- **Third-party metadata** is often missing or incomplete, requiring internal data teams to perform metadata remediation manually.

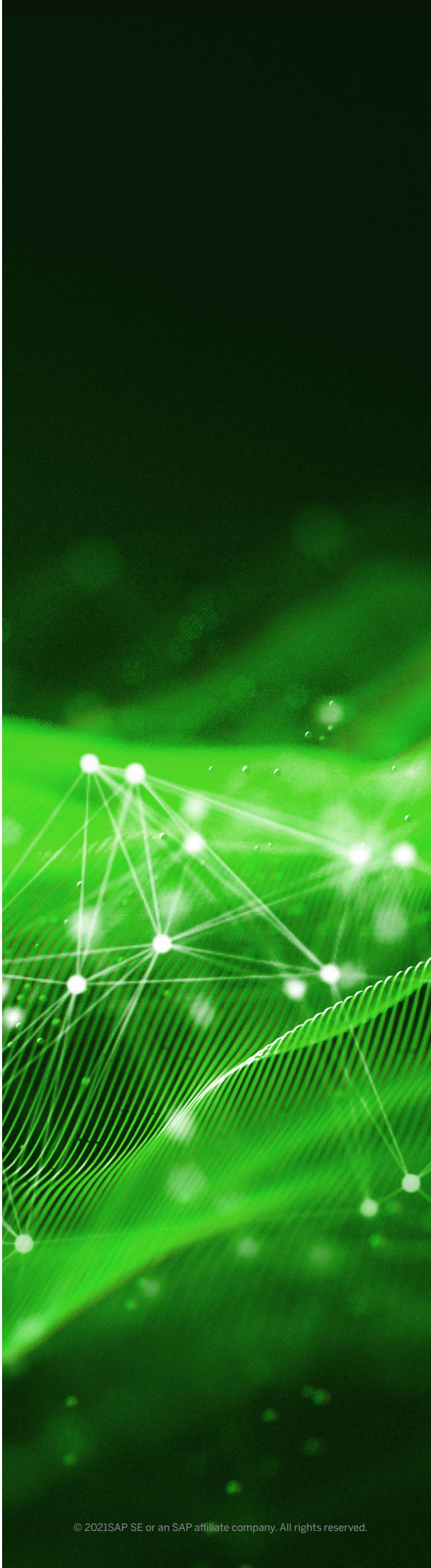
Yet, there is an ever-growing need for more and different third-party data. This is driven by new technology, business trends, and the emerging need of your company to predict when to expect "the next normal," for example, a post-COVID-19 environment. In the case of the COVID-19 pandemic, responses include ongoing, real-time tracking of data from third-parties, such as health institutions, plus federal, state, and local governments.

Responses also involve complying with regulatory requirements like protected health information (PHI) and the California Consumer Privacy Act (CCPA). These factors drive new data sources, new buyers, and new data delivery methods, including third-party data in the cloud.

The growing trend is for third-party data to be provided as add-on features to SaaS, PaaS, and BPaaS cloud solutions, such as the AWS Data Exchange or a data exchange network from SAP like **SAP Asset Intelligence Network** or the data marketplace for the **SAP Data Warehouse Cloud** solution. Also growing are data marketplace solutions such as Datarade, which offers buyers thousands of data source options, as well as data-sharing exchanges such as CDQ, which offers data-sharing and collaboration services. These new delivery methods address many of the challenges that existed in the past with locating, integrating, and using third-party data. We expect to see the use of data marketplaces and exchanges become more prevalent as they become more available and as integration of data marketplaces and exchanges increase within cloud software solutions.

Today, data marketplaces typically connect buyers to sellers, with little data quality assessments provided. The "fit for purpose" evaluation is left to the buyer, with some limited help from select providers. However, the emerging trend is for marketplaces to provide some crowdsourced ratings as a way to qualify and quantify the satisfaction and usage of other buyers.

Another emerging trend is for data marketplaces to match willing buyers with sellers based on specific requirements and target price, similar to what is available in other marketplaces. Sound simple? It is. However, given the abstract nature of evaluating data as an asset, there are numerous areas to think through: ownership rights, end-of-term actions, and pricing models, among other factors that surface with a bespoke approach.



PLANNING YOUR STRATEGY

With the growing demand for third-party data, CDOs (regardless of industry) need to address the requirements for third-party data. It starts with explicitly including third-party data management in the overall enterprise data strategy as a way of doing business.

Designate a senior team member to be your head of data acquisition, reporting to the CDO. No longer a fragmented or part-time assignment, this senior role includes defining and driving the data acquisition strategy and partnering with IT, stakeholders, and legal to facilitate the integration of third-party data.

As part of your acquisition strategy, you must also define the third-party data needed to enable the business outcomes in the enterprise data strategy. This includes sourcing, delivery, security, privacy, and vendor management processes.

Below is a checklist for planning business outcomes in your enterprise data strategy. For each third-party data required, you should ask the following:

- Should the acquisition of each data type be centralized or localized?
- Where will new data come from, for example, from a marketplace or a sole provider? Where will it be integrated?
- Who will use it? For what purpose?
  Who pays for it?
  Are there usage restrictions?
- What data format is required?
- What service-level agreements are needed to ensure solution efficacy?
- If this is PII data, is it opt-in data?
  Does it require anonymization?
- What happens to the data if or when the third-party data contract expires?

Designate a **head of data acquisition**, a senior team member reporting to the CDO. No longer a fragmented part-time role, this person will lead your data acquisition strategy, partnering with IT, stakeholders, and legal to integrate third-party data.

As CDO, you must sanction the use of data marketplaces and exchanges within your company. As these data marketplaces and exchanges become increasingly embedded in PaaS, SaaS, and BPaaS solutions, you must provide guidelines for how to use these offerings and which offerings to avoid. Communicate the value of an enterprise-level acquisition strategy versus a point solution acquisition strategy.

Functional data purposes will not go away. The easier and more complete the marketplace is for finding and using third-party data, the less individuals within the organization will seek other avenues.

In addition to the tasks carried out by the head of data acquisition, marketplace data content must be proactively managed after it has been sourced, similar to other internal critical fields. The task is to ensure that the data is "fit for purpose" (data quality) in an ongoing feedback loop. This is to be performed before the data is accepted, during updates and live testing of the data, and when the data is in use. With input from the data science team, it will also be possible to detect usage issues, such as data bias and source coverage.

Consider expanding the data steward role or creating a new one for each data source used in the marketplace. Duties of this role include working closely with the head of data acquisition, business sponsors, and all data users in the company.

Your enterprise data catalog should include third-party data content information and metadata to ease finding and using this data. Remember to include any contract or usage restrictions in the catalog.

Your company's procurement process can support the acquisition strategy by ensuring that any data acquisition exceptions are reviewed by the head of data acquisition or the CDO.

HOW TO SUCCEED IN MOVING TO CLOUD FIRST

You will experience both challenges and opportunities in moving to a "cloud first" IT environment.

To maximize the benefits across the enterprise, CDOs will need to reevaluate their data strategy, simplifying and refining existing data processes and practices.

Look for marketplaces and exchanges that can easily be integrated and used within the business process solution (not as a separate batch job). These solutions have usability and integration advantages over other marketplaces and can potentially provide a more efficient and cost-effective approach.

SUMMARY

Your move to a "cloud first" IT environment presents both challenges and opportunities. To succeed, chief data officers will need to reevaluate their data strategy to simplify and refine existing data processes and practices.

You can run one or more business software applications as a service (SaaS, PaaS, or BPaaS) from a cloud software solution provider, as many companies do for CRM, workforce, supply chain, and other applications. The guidelines discussed in this document will enable you to manage data sources better, including third-party sources. Data quality management, important data processes, and analytics all come into play, along with managing risk and balancing technical considerations.

If you use a data repository hosted in the cloud, such as a data lake or a data warehouse as a service, you can easily take advantage of its analytics and functions for collecting and storing diverse information from numerous sources. Today's DWaaS options are replacing in-house warehouses for "fast path" business-led initiatives. This means clear advantages in elasticity, efficiency, addressability, and query optimization.

Want to learn more?

Contact your SAP representative for more information or visit us at sap.com.

NEXT STEPS

SAP knows that a connected, integrated data landscape offers near limitless potential for improving business outcomes. For that reason, SAP created an open, integrated platform built for business that supports both on-premise and cloud-native environments and unifies data management, analytics, AI, IoT, blockchain, and application development technologies. SAP Business Technology Platform is the platform for the Intelligent Enterprise (see Figure 6). Customers can achieve agility, business value, and continual innovation through integration, data to value, and extensibility of all SAP and third-party applications and data assets.

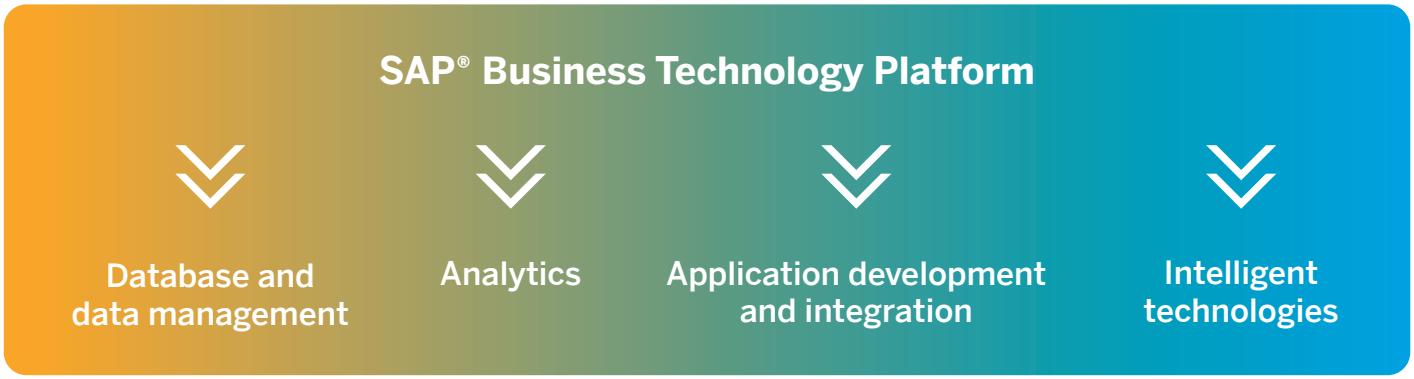


Figure 6: SAP® Business Technology Platform

- SAP Database and Data Management solutions support SAP Business Technology Platform, enabling data-driven decisions with solutions that manage, govern, and integrate your enterprise data to feed analytics and drive confident business decisions.
- An enterprise data strategy is a business-prioritized, comprehensive road map of data-dependent capabilities focused on achieving outcome-driven transformations. Beyond technology, this strategy considers securing sponsorship, establishing a funding model, and introducing organizational change management, success metrics, and engagement strategies.
- Learn how to unravel the complex value creation processes using this simple data value formula. It explains how data value can be determined and identifies three essential value drivers: data volume, data quality, and data use.

ABOUT THE AUTHOR

Maria C. Villar brings over 25 years of experience as a Chief Data Officer responsible for building enterprise data management organizations from the ground up and leading the culture change across the enterprise. She has held this position in both the technology and financial sectors, most recently at SAP where she was the CDO from 2009–2017.

Today, Maria is Head of Enterprise Data Strategy and Transformation for SAP NA. She advises SAP customers on the crucial role of data management in their digital transformation, leveraging her practical, operational experience as CDO.

In 2017, Maria was honored with the "Transformation of Collaboration from Inwards to Outwards" award from the International Society of Chief Data Officers (isCDO) where she serves as a board member. This award recognizes outstanding CDO leadership in driving business outcomes and business collaboration.

In addition to her operational experience, Maria publishes and speaks regularly. Among her many accomplishments, she has:

- Co-authored a book with Theresa Kushner, "Managing Your Business Data from Chaos to Confidence"
- Developed 2 online classes
- Written numerous articles, including "Data Voices Unite: Why One Enterprise Data Strategy Is The Answer"
- Most recently, published a data strategy master class available on YouTube

Follow us

www.sap.com/contactsap

74433enUS (21/04)

THE BEST RUN SAP