informatica

The Chief Analytics Officer's Guide to Getting Analytics Right.

Three Pillars of Data Management for Positive Business Results.

Contents

Tip: Click on parts to jump to
the particular section you want

Introduction

Building a Foundation to Deliver Faster and Greater Business Value.

Building a Foundation to Deliver Faster and Greater Business Value.

The most successful businesses aren't only using data to compete. They're using it to drive transformative shifts by investing in advanced analytics to identify new business opportunities and deeper operational efficiencies. More specifically, they are driving the enterprise-wide use of data and analytics to increase revenue and profits.

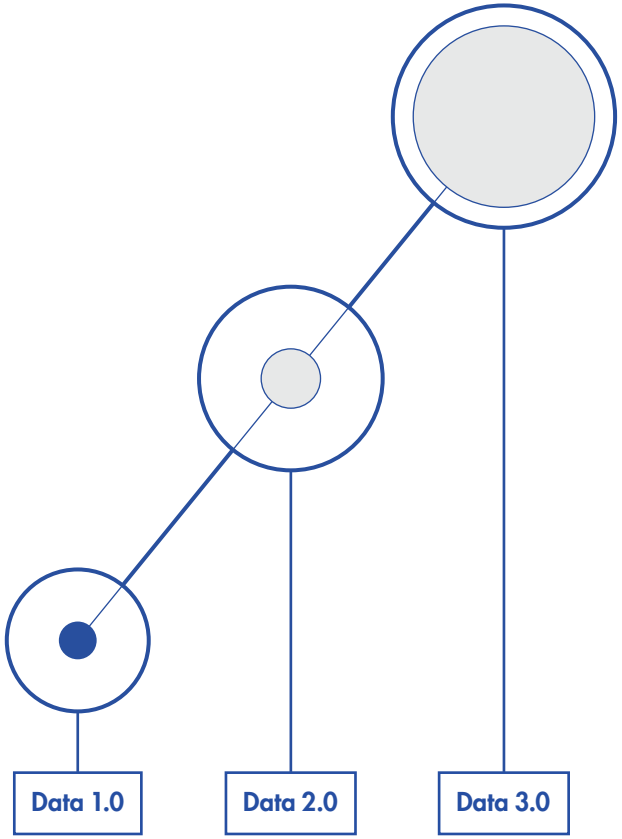
"Notching small analytics victories may not be enough. For leaders with their eyes on the prize, it's all about connecting analytics capabilities across the enterprise," reports Deloitte in its "Analytics Trends 2016" research. "Leaders are beginning to take serious steps toward connecting these successes to create something bigger."

In fact, according to Gartner, BI/analytics has been the top priority for CIOs for the past five years[1]. But, just the fact that it has been the top priority for five years tells you that the job is as yet unfinished.

But it's worth noting how businesses have changed the way they use data to compete over the last few decades. The Data 1.0 age saw businesses using data for specific business functions, such as payroll automation until the 2000s. The Data 2.0 generation of the past 20 years used data to support enterprise functions such as the supply chain.

Today, the Data 3.0 generation of businesses is using analytics based on trusted data as a key differentiator. They're using it to compete in game-changing ways, including:

- Improving customer loyalty and wallet share.
- Delivering better healthcare patient outcomes.
- Driving predictive maintenance of equipment and machinery.
- Delivering better fraud detection and faster reaction time.



Building a Foundation to Deliver Faster and Greater Business Value.

But while businesses are applying such strategic importance to analytics, a full 86 percent of executives consider their organizations to be "at best only somewhat effective at meeting the primary objective of their data and analytics programs," according to a recent McKinsey report[2]. More than one-quarter say they've been ineffective.

To ensure businesses get better at using analytics to drive business transformation, many companies are appointing a dedicated, accountable C-suite executive to lead their initiatives: the Chief Analytics Officer (CAO).

(While analytics leaders in some companies may not have that exact title, they may also be known as the Chief Data Officer, VP of Business Intelligence, or VP/ Director of Analytics. For simplicity's sake, in this eBook we'll refer to these individuals as CAOs.)

The details of the CAO's role are still being fully defined at many organizations. But by and large, CAOs are charged with aligning the analytics strategy across the enterprise with the business strategy. They're responsible for making sure priorities are matched and the analytics initiatives directly deliver business value. They do this by:

- Leading multidisciplinary teams with senior executive peers across business functions and IT to make data and analytics a competitive advantage for the organization.
- Leveraging new analytics technologies (Hadoop, predictive, machine learning, etc.) to accelerate business value delivery.

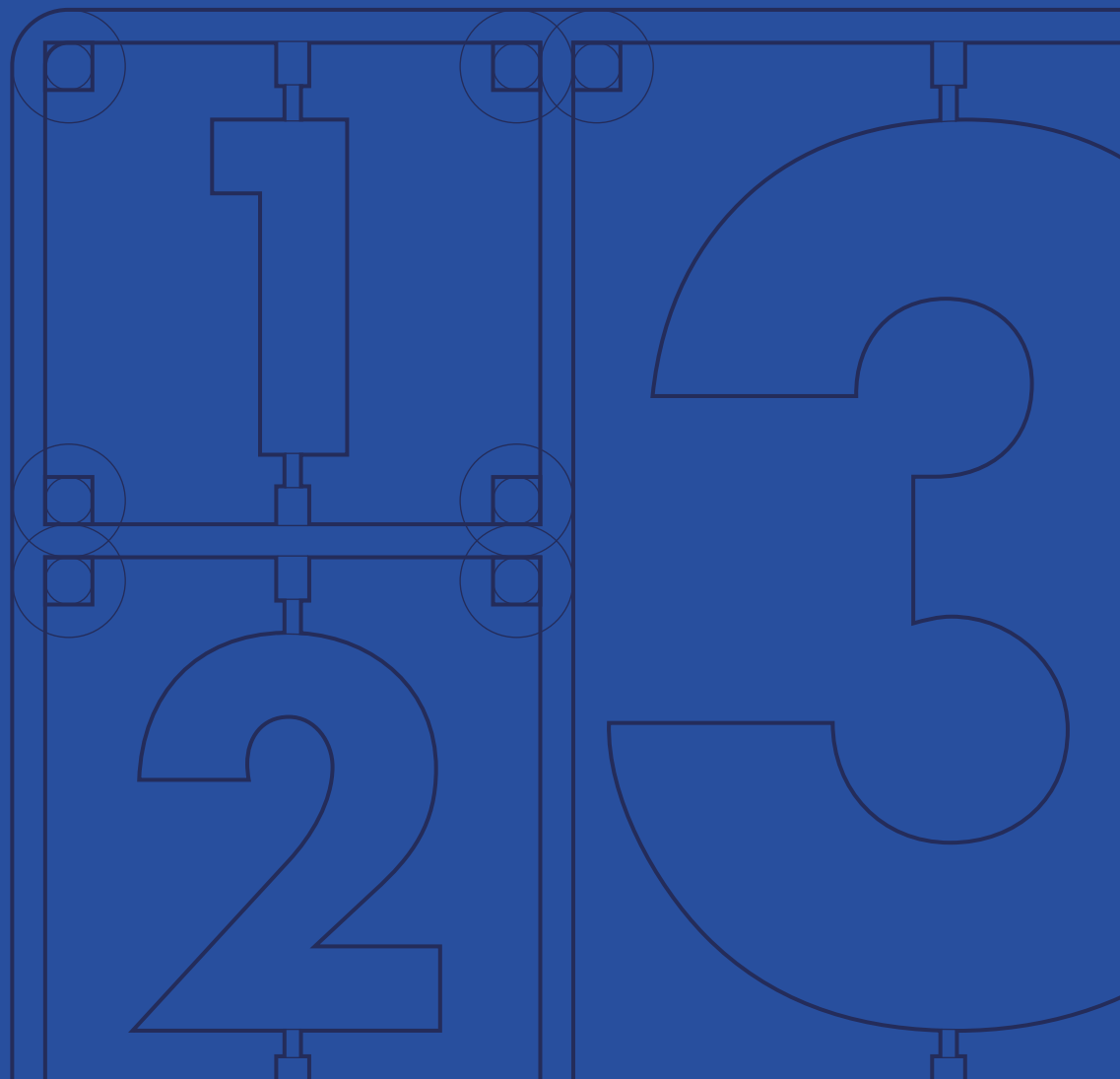
Key to any analytics success (and ultimately driving successful business transformation) is the deployment of an enterprise-wide data management strategy. This encompasses technology and processes to ensure business decision makers, data scientists, and analysts have access to shared data that is drawn from internal and external sources, while also making the data trustworthy, timely, and easy to discover and access for analytics.

This eBook is about using these tools and processes to build three foundational pillars of effective, enterprise-wide analytics:

1. **Enterprise data management:** To make all data available to every analytics user—even data that is traditionally locked in application silos. This includes external data, partner data, or data that is in the cloud or new analytics applications. The most interesting analytics insights often come from combining data from widely disparate sources.

2. **Data governance:** To manage data as an asset and ensure that every analytics user has trusted data that is "fit for purpose" for their analytics needs. This includes managing data's meaning, business context, and quality over time.

3. **Data self-service:** New business initiatives require fast time to market and one of the critical requirements is to empower a new class of business users, analysts, and data scientists to access and manage data without IT assistance.

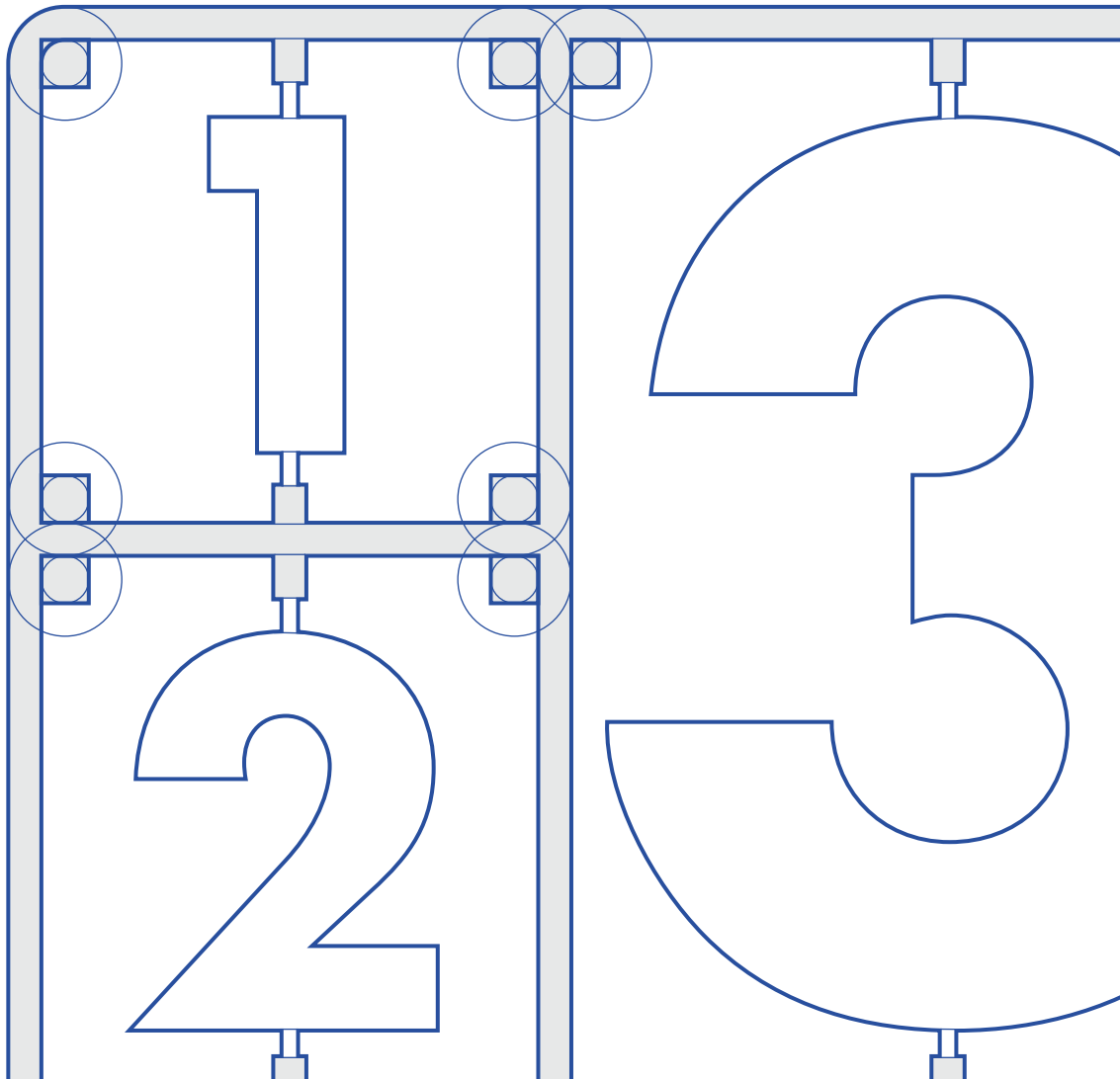


Building a Foundation to Deliver Faster and Greater Business Value.

In this eBook, we won't touch on many of the organizational, political, and cultural challenges that are no doubt on your mind.

But we will show you how a logical, collaborative, and scalable approach to data management can help you overcome some of the most important business analytics challenges and take advantage of new analytical possibilities.

Let's dive in.



Part 1

Enterprise Data Management: Cross-Functional Insight.

Enterprise Data Management: Cross-Functional Insight.

Your organization is practically swimming in data. Much has been written about the growth in data volume, variety (complexity), and velocity. Not only is your data growing, but industry analysts estimate that 50 percent is coming from external sources[3], which means that you may not know the data structure, quality level, or meaning and business context – making data management that much more challenging.

But if business users need to find answers to questions like "which customer should a sales rep call on and what should they say"—questions that touch multiple lines of business and teams—they're going to need to access data sources scattered across silos within your company, as well as data sources from outside. This is arguably the defining challenge of the CAO role—to look beyond existing organizational boundaries and "integrate" the enterprise and outside data.

It's why it makes sense for your analytics initiative to approach data and data management from an enterprise-wide perspective.

To be sure, this doesn't mean implementing processes to manage all your data at once. Instead it's about starting with the right analytics projects and logically building toward a state where you have a consistent, repeatable, and trusted way to deliver data for all of your analytics initiatives and users.

Turning Integration Competencies Into Information Competencies

Many organizations have created an Integration Competency Center (ICC) which is focused on creating common standards and practices for data management.

The Next-Generation Integration Competency Center.

To better support the increasingly complex needs of the business, the traditional Integration Competency Center must shift to become an Enterprise Data Competency Center. The key is to connect technical data competency to business value. By broadening "Integration" to "Information"— the data that drives business —and focusing on true business transformation competency, you can better position and manage the business context and value of information for competitive advantage.

Read our blog series about the Next-Generation ICC to find out more.

Enterprise Data Management: Cross-Functional Insight.

Your current ICC, if you have one, is most likely focused on best practices for IT. The challenge going forward will be to extend this to solve data-related issues that affect the whole business – and engaging business experts in the management of the data.

Take the following steps to get started:

- Inventory your data assets to find out what data you have today.
- Conduct a data maturity assessment to find out how standardized and effective your data management is across the organization.
- Prioritize your analytics initiatives so you start with the ones likeliest to have the greatest business impact.
- Find out what data is required but missing to accomplish these initiatives.
- Find out what data management capabilities you need to accomplish these initiatives.

Inventorying Your Data Assets

In large organizations that have many application and analytics systems, it's often hard to know what data you have today and where it all resides, let alone how it's used. In such cases, the first challenge is to do a comprehensive data discovery, using emerging tools.

By standardizing the management of business and technical metadata on a single data management platform, you can then start to see how data moves between systems and what that means for different lines of business.

Aiming for Hybrid Data Management

As you plan your analytics and data management strategy, one very important option to consider is cloud deployment. The advantages include:

- **Faster time to value:** For example, you can get results from new analytics initiatives faster when you stand up a new cloud data warehouse in hours rather than months.
- **Flexibility:** The cloud offers the ability to quickly scale up and down as your business requires.
- **Cost:** Cloud options typically provide the advantages of pay-as-you-go and OPEX versus CAPEX.

Enterprise Data Management: Cross-Functional Insight.

As you consider cloud options, you will also have data management challenges. And it will pay to plan for them early in the process. You'll need to plan for migrating data to new cloud analytics systems and applications, while also maintaining the quality of data in the new system.

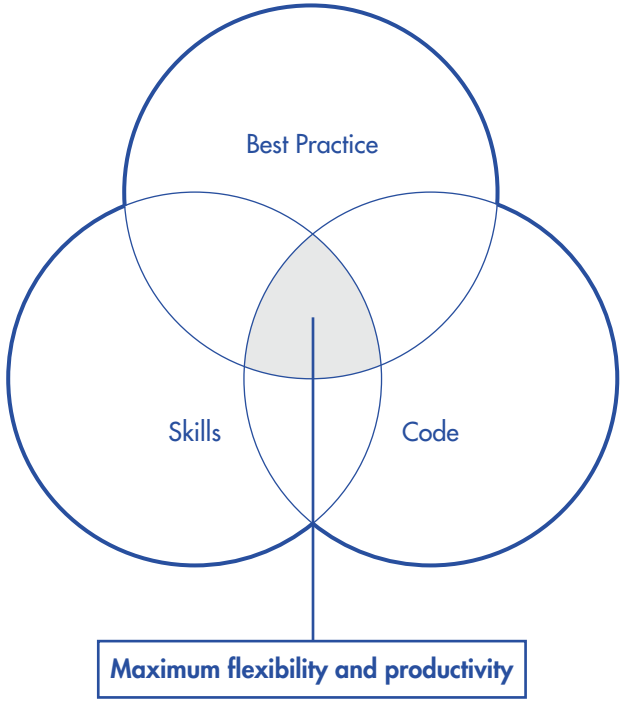
You'll also need to ensure the new system's data is synchronized with other on-premise and cloud systems so that all users have access to the best and most current data.

Most larger organizations have a significant investment in on-premise applications, data warehouses, and analytics that will not be moved to the cloud all at once. So when it comes to selecting a data management platform, it will be important to choose a platform that works across on-premise and hybrid environments.

There are three important options to consider:

- On-premise data management capabilities with connectivity to cloud data sources.
- On-premise data management capabilities that are hosted in the cloud. This provides an "exact-same" experience for IT developers while delivering some of the advantages of cloud for data management.
- An Integration Platform as a Service (iPaaS). This provides integrated, cloud-based data management services with an easy-to-use interface for "citizen integrators."

Whichever options you choose, you should look for an environment where skills, code, and best practices can be shared across the environments for maximum flexibility and productivity.



What is Integration Platform as a Service (iPaaS)?

iPaaS is a cloud service that gives IT a single platform to manage data integration, application integration, and process integration with a very user-friendly user interface.

It powers development, execution, and governance of **any integration pattern** between on-premise, public cloud, and private cloud applications, databases, and other data sources.

It delivers cloud integration services (including data integration and application integration services for batch and real-time scenarios), native connectivity, a robust API integration framework, data management services (including master data management, data quality, test data management, and data security) to manage **any data.**

And crucially, it's built to serve the diverse needs of **any user**. It provides simple governance for IT, reusable logic for line of business developers and mobile application development teams, and ease of use for business users.

Put simply, iPaaS is a cloud-based platform that allows enterprises to rapidly execute any integration pattern, logically manage any data, and easily serve any user in need of integration.

As Gartner notes[4]: "By 2019, iPaaS will be the integration platform of choice for new integration projects, overtaking the annual revenue growth of traditional application integration suites on the way."

To learn more about iPaaS, read "The Cloud Architect's Guide to iPaaS".

Enterprise Data Management: Cross-Functional Insight.

Technological Imperatives for Enterprise Data Management
As you scale up processes for data management to meet the requirements of enterprise information management, you and your team need to be able to get more done with minimal budget increases. So it is vital that your platform has intelligence to automate tasks as much as possible.

From a data management point of view, consider the following technologies:

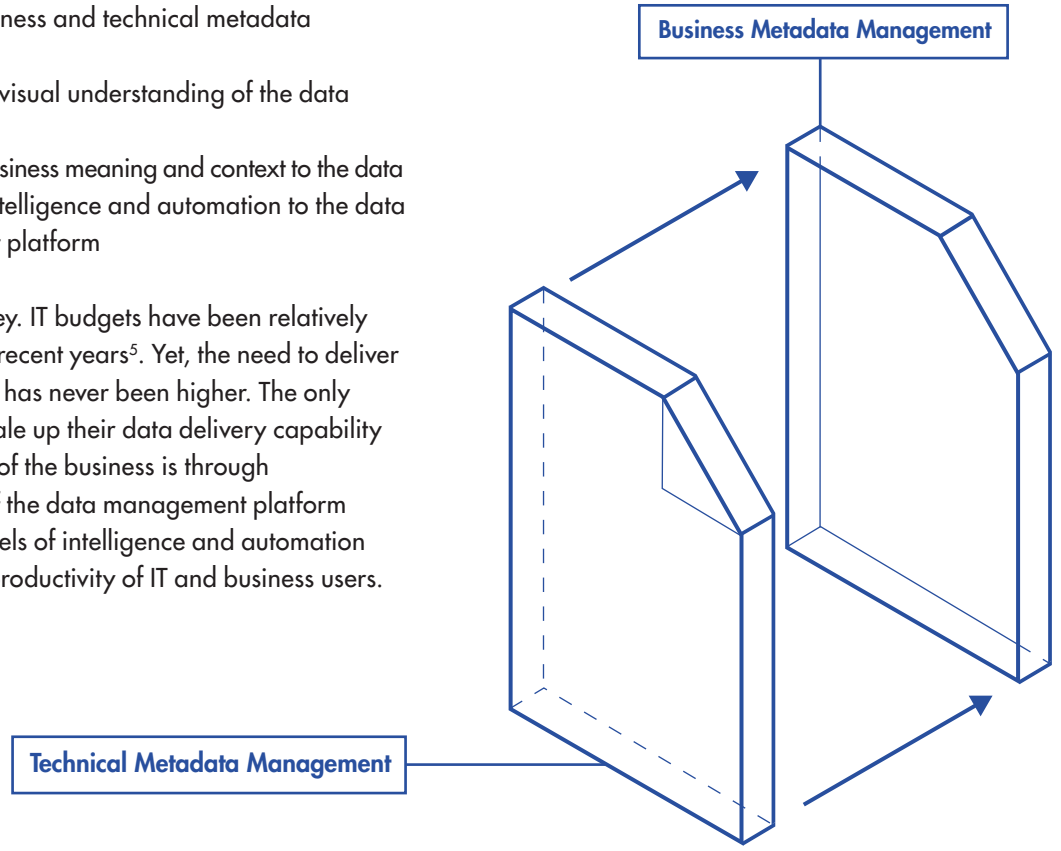
- **Data profiling:** So IT can discover and quickly understand data, while also assessing the state of data as it's being managed.
- **Data integration:** So you can connect disparate data sources with faster batch loads or in real time.
- **Data quality:** So you can automate rules for data quality at scale and deal with exceptions and anomalies as they arise.
- **Master data management:** So that data from across many systems and data sources can be combined and managed to provide a single, trusted, 360-degree view of a business entity, such as a customer, or part number.
- **Metadata management:** So IT has visibility into where the data came from (lineage), how it moved, and how it was transformed. And so that business users know what the data means with a centralized business glossary.

Enterprise Data Management: Cross-Functional Insight.

What to Look for in a Data Management Platform:

- An end-to-end data management solution
  – for all types of data across the full data life cycle.
- The flexibility to add functionality as your needs grow, with minimal disruption.
- Flexibility to run on-premise, in the cloud, or hybrid.
- The flexibility to work with any data, any analytics use case.
- Support for batch, real time or streaming data delivery.
- Business-IT collaboration built in.
- Productivity improvements with:
  – Role-appropriate tools
  – The ability to re-use code and skills
- Enables the business to self-serve data.
- Integrated business and technical metadata management
  – To provide a visual understanding of the data environment
  – To provide business meaning and context to the data
  – To provide intelligence and automation to the data management platform

That last point is key. IT budgets have been relatively flat or declined in recent years[5]. Yet, the need to deliver trusted data faster has never been higher. The only way that IT can scale up their data delivery capability to meet the needs of the business is through standardization of the data management platform and increasing levels of intelligence and automation to accelerate the productivity of IT and business users.



Part 2

Data Governance: Collaborating for Trusted Data.

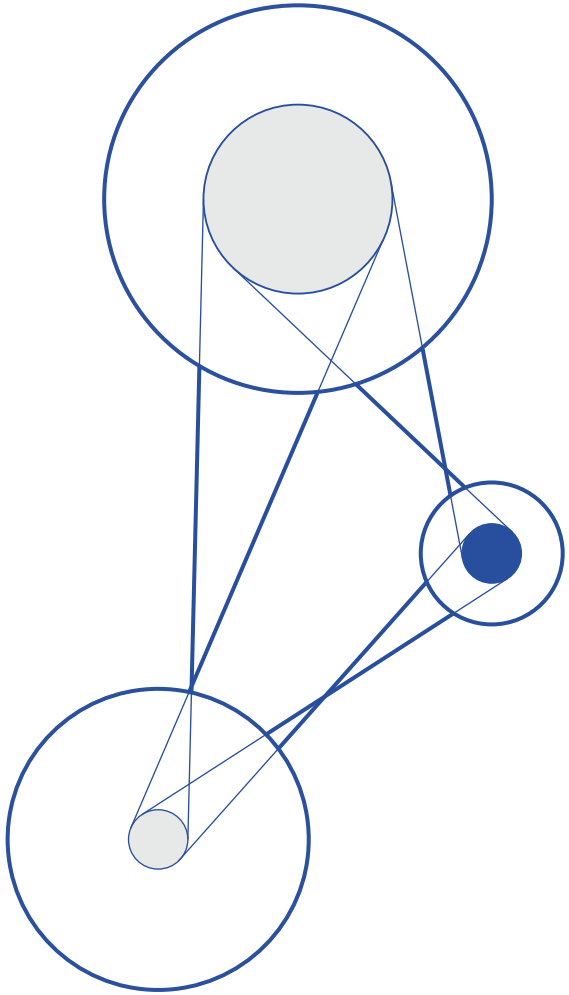
Data Governance: Collaborating for Trusted Data.

We've all been in meetings where people show dashboards with conflicting data. Or, even worse, a compelling analysis is delivered but management does not trust the data enough to act on it. That is where the importance of data governance comes in.

Data governance is all about managing data as an asset. IT is part of the solution, but it can't do it alone. In fact, the business side needs to take the lead in data governance because only you know the critical meaning of data, the business context, the relative priority of data to be managed, and how to define data quality metrics to determine if the data is trustworthy or not.

Driving business meaning and context: In order for data to be usable for any analytics user it needs business meaning and context. This includes metadata such as terms, definitions, data stewards/owners, data domain, data policies, etc.

Engaging different departments: In order for a CAO to be successful, you need to operate across many departments in both business and IT. One of the critical interfaces is with the data governance body within the organization. This process is fundamental to making data useful for analytics purposes.



Data Governance: Collaborating for Trusted Data.

We're often asked how to go about implementing data governance for analytics projects that are well underway. A great example of an organization that has done this is the Cleveland Clinic, featured in the sidebar.

Maintaining procedural hygiene and data quality: Data quality erodes, on average, between 1 and 1.5 percent every month[6] when it isn't actively managed. Data governance includes assigning data steward owners, processes, and policies needed to ensure that data is ready for any analytics use case. This does not mean that the data always has to be perfect. For rapid exploration and faster innovation data may be more important than perfect data. But for critical business processes and decisions, is it important that the data is trustworthy.

It's important to note that a completely manual approach to data quality is costly to scale. So it pays to invest in automating data quality rules that are shareable and repeatable with minimal human intervention.

Cleveland Clinic: Data Governance as a Foundation for Predictive Analytics

Cleveland Clinic is a non-profit healthcare leader that specializes in heart and brain healthcare. Cleveland Clinic wanted to make the transition from traditional business intelligence reporting to predictive analytics.

To accomplish this, Cleveland Clinic began an Enterprise Information Management and Analytics (EIM&A) initiative focusing not just on technology but also around four pillars: data, people, process, and technology. A key component of the data domain is data governance.

The focus on governance led to the establishment of a council comprised of executive leaders, senior stakeholders, clinical representation, and a newly created Senior Director of Data Governance position.

In addition, they formed an advisory council that collected input, feedback, and concerns from a large cross-section of the organization. Insight gained from the multidisciplinary advisory council was incorporated into the data governance council, where decision-making occurred. This ensured quicker decision making while maintaining close alignment to the opinions and voice of the customers. Win-win.

The result: EIM&A laid the foundation of data governance which is increasingly enabling the delivery of timely, trusted data, Cleveland Clinic can now expand on its advanced analytics, such as forecasting operating room activity eight weeks out for over 100 operating rooms, increasing efficiency and enabling better resource planning.

Read all about Cleveland Clinic's impressive journey toward predictive analytics.

Data Governance: Collaborating for Trusted Data.

Managing Technical Metadata: Technical metadata management is a key tool for IT productivity. Data lineage diagrams can quickly show how data flows end-to-end through the data management environment. That's incredibly useful because it means it takes less time to understand your data management environment. In turn, that means you're less likely to make errors when making changes – because everyone has a clear view of how data flows through the organization.

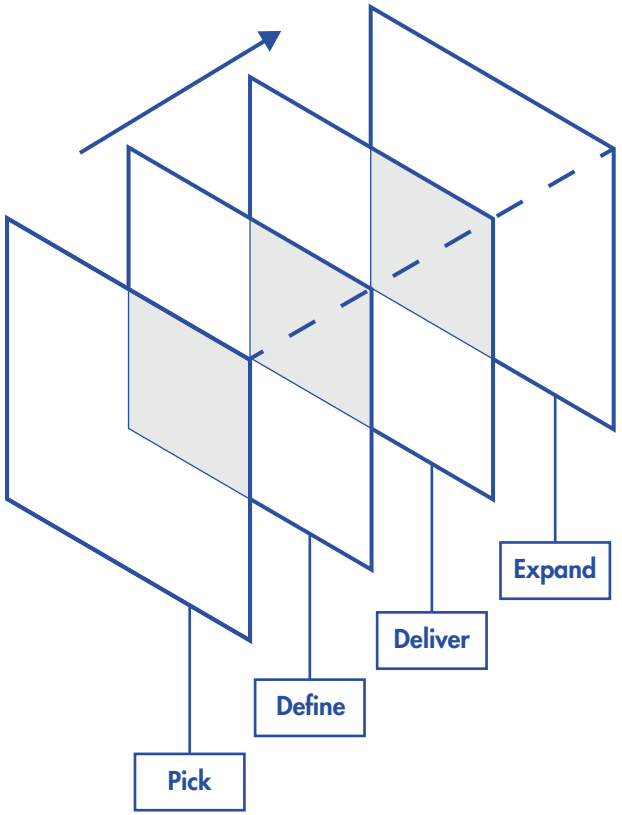
Impact analysis shows the impact of a proposed change to the environment before it is implemented. In fact, one financial services company and a healthcare organization have reported that impact analysis saves them up to 90 percent of the effort of making a change without introducing errors. Metadata management is therefore critical for IT agility when it comes to managing complex data management environment.

Depending on your priorities, technical metadata can also enable you to maintain audit trails so you can report on who did what and when they did it.

"Just enough" Data Governance
Many organizations have experienced the failure of starting too big or taking a top-down approach to data governance. It pays to take a more agile, business-oriented approach to governance.

We're big believers in "just enough" data governance —an approach that emphasizes no more or no less effort than is needed to make your governance program successful for the intended business initiative. It allows you to follow a process that looks like this:

- Start small: Pick a specific business initiative, with budget and clear business outcomes defined.
- Define metrics so you can measure your program's impact on this business initiative.
- Deliver quick wins and then communicate why they matter.
- Expand to additional business initiatives.



Data Governance: Collaborating for Trusted Data.

In the same way that analytics initiatives fail when they try to encompass more than they were designed for, data governance should also follow a "start small, prove value, scale fast" approach. Depending on the data you need to kick-start your analytics program, focus on one area that needs governance and then prove its value before you scale it.

We've written about "just enough" data governance in more detail. Read our eBook to find out how to get it right.

Technological Imperatives for Data Governance
A critical starting point for choosing the technology for data governance is that it must include built-in business-IT collaboration as a fundamental capability. Some other capabilities include:

- **Data quality:** So your governance team can set metrics and track data quality trends on an ongoing basis, remediating anomalies and data quality problems as they arise.
- **Metadata management:** So IT stakeholders can have an understanding of the data integration environment, the data flows, and how data is being transformed.
- **Business glossary:** So business subject matter experts can create, manage, and share the business meaning and context of data.
- **Data security:** So you can ensure a policy-based approach to data access with data masking to obfuscate sensitive information, application security to proliferate policies within applications, and encryption to protect data where it lives.

Part 3

Self-Service: Democratizing Data to Empower New Users.

Part 3

Self-Service: Democratizing Data to Empower New Users.

Providing business analysts with self-service business data is one of the newer requirements for CAOs. However, due to the data complexities that we've described in this eBook, IT has struggled to deliver data at the quality level and in the time frame required by the business.

As a result, businesses are increasingly moving toward a data-driven self-service model where IT makes data available to business analysts and data scientists who can do their own data discovery and data prep. That's led to far less IT involvement (and cost) and much greater business agility, particularly in the area of analytics.

But self-service isn't dumping a load of raw data into a repository such as Hadoop and hoping for the best. There needs to be a plan for providing data that is appropriate for the business use case, making that data easily available for rapid iteration and innovation, and ensuring the data itself is reliable enough given the types of business decisions it will be used for.

In some cases, raw data might be just what your scientists and analysts need for exploration. But more often than not, you'll need some data management to make sure users spend less time in data prep—a process that can take as much as 80 percent of their time[7].

For instance, everybody will need some data cleansing, but they will also need ways to join data sets to provide interesting and useful insights. That will take some pre-planning in terms of tags and keys.

You will probably also need to do some data security work so you can identify sensitive data. Also, you may be creating new sensitive data by joining different data types. In which case, you may need data masking to obscure personally identifiable information.

Self-Service: Democratizing Data to Empower New Users.

This isn't to say you need to make sure everyone has access to fully conformed, penny-perfect data. In fact, when it comes to self-service, the priority should be on delivering data that is "fit for purpose."

- For innovation, perfect data is often not required. Often the data only has to be good enough to see if the question is worth exploring in more detail. The priority here is to ask questions quickly, iterate, and find the useful questions as quickly as possible.

- For critical business decisions and business processes the data will typically have to be more trustworthy—in direct relation to the business impact of not having good data.

But it isn't enough to just create a centralized data hub for self-service experimentation—you've also got to ensure your users have the right tools for self-service data prep and analysis.

The key is to deliver and prepare data quickly enough to fuel self-service, but still maintain enough data quality depending on the use case—for instance, light data preparation for experiments and strong data preparation for operationalized queries.

What Self-Service Users Need

A TDWI report[8] found that business-side users prioritized the following four tasks as things they need to be able to do on their own:

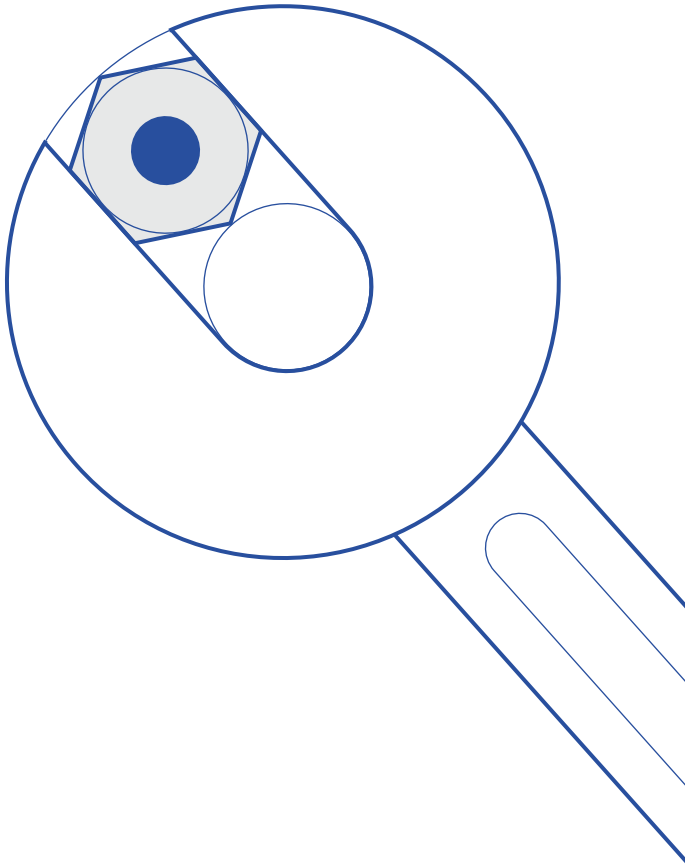
1. **Discover data:** To quickly and easily access "good enough" data to test their hypotheses and run different queries.

2. **Prepare data:** To make changes to the data without permanently changing IT's core data assets.

3. **Visualize data:** To present their findings to business stakeholders using a range of visualizations.

4. **Author dashboards:** To provision interactive views of the key performance indicators their lines of business care about.

This may require some data preparation work by IT, but the idea is to do the minimal and appropriate level required for the intended use. This will reduce IT costs, and speed up analytics delivery.



Self-Service: Democratizing Data to Empower New Users.

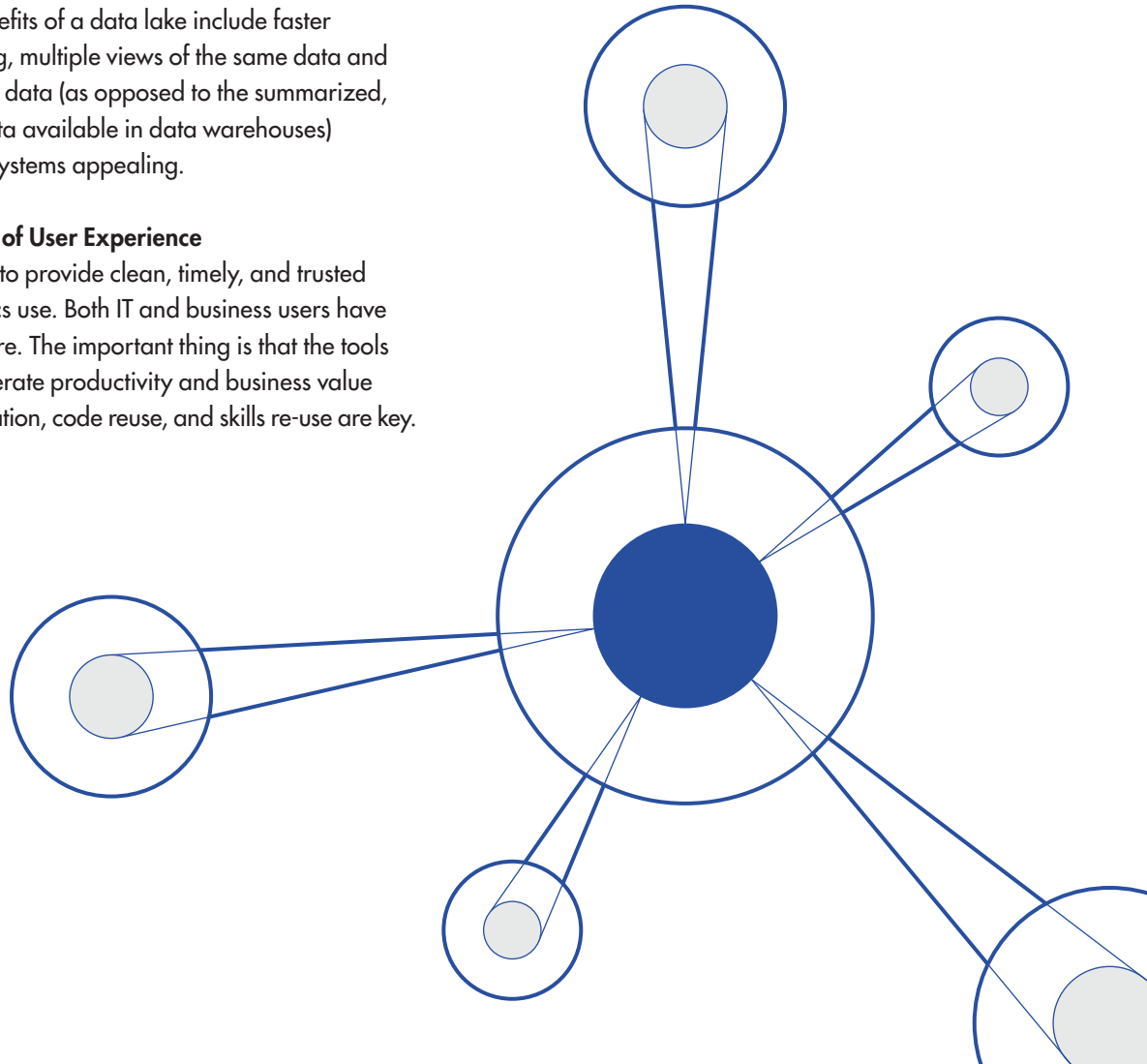
Creating a Self-Service Environment
For self-service to work, you need to be able to feed data from various sources into a secure, centralized location where business users can experiment and interrogate data sources on their own.

While many organizations have delivered self-service analytics using traditional relational storage technology, businesses are now beginning to deploy Hadoop-based data lakes. The benefits include fast data ingestion and the flexibility to allow users to ask a wider range of analytics questions.

Hadoop is attractive because of the "schema-on-read" capability that allows you to store data from various sources first and structure it later. Typically, data lake managers will create a data schema with Hive (for example) to solve a specific business problem. But once the data is in the lake, then different data models can be applied to ask new questions or perform different, unanticipated analyses.

Some of the benefits of a data lake include faster data onboarding, multiple views of the same data and access to atomic data (as opposed to the summarized, consolidated data available in data warehouses) make these ecosystems appealing.

The Importance of User Experience
The goal here is to provide clean, timely, and trusted data for analytics use. Both IT and business users have requirements here. The important thing is that the tools used must accelerate productivity and business value delivery. Automation, code reuse, and skills re-use are key.



Self-Service: Democratizing Data to Empower New Users.

For Business Users

A crucial consideration when selecting data visualization, data catalog/discovery, data preparation, and data integration tools for self-service is user experience. In particular, less technical users need intuitive interfaces that make it easier to interact with the data.

For instance, the most popular data prep tool use by business analysts today is Excel. The better new data prep tools offer an Excel-like interface that feel familiar to these users. Also, cloud-based integration tools like "Integration Platform as a Service" or "iPaaS" make it easy enough for moderately-technical, "citizen integrators" to manage data.

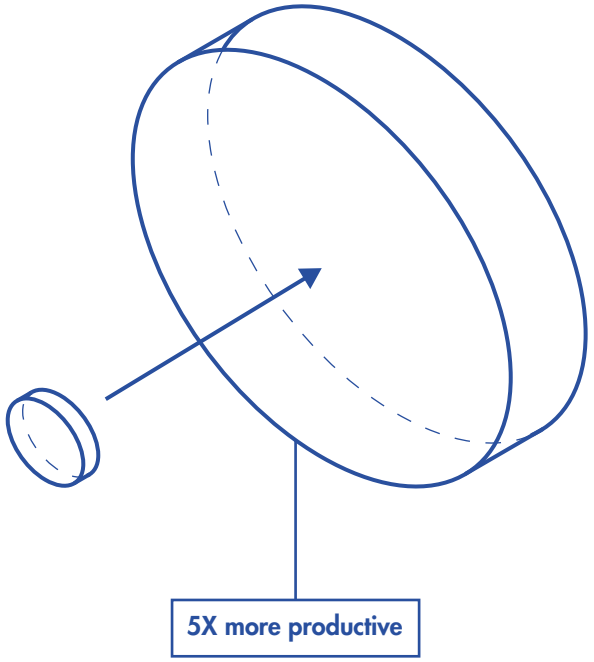
Another important factor to consider is how repeatable your users' experiments and explorations are. Data management platforms should enable users to discover, share, and reuse data prep "recipes" created by others before them.

For IT Developers

Many studies have shown that a good GUI-based, no-coding development environment is 5X or better more productive than hand coding. Look for tools that:

- Are easy to learn and use.
- Have a large base of trained practitioners who can be hired.
- Enable code reuse, skills reuse, and sharing.
- Work across the widest variety of analytics use cases.
- Shield the organization from technology change.

The last point is crucial—data types are changing and so is analytics technology. Your data management tools must protect your developers from these changes by abstracting rules away from the underlying technology. That way, new data and tools won't break your environment and cause expensive and time consuming re-work.



Powering Self-Service Experimentation and Rapid IT Operationalization—with One Platform

A large North American automotive conglomerate with 24 business units needed to overcome its siloed approach to data so different business units could run their own experiments based on a common view of customer relationships.

So it pioneered a big data platform capable of supporting both a "big data lab" and a "big data factory".

Using Hadoop, the team built four separate data staging environments or "layers" for its data based on the amount of data preparation work that had been done. In this way the environment provides "fit for purpose" data: less prepped data for rapid iteration and fully-prepped data for critical business uses.

The four different layers are:

Raw: For data that's pulled in right from the source, unchanged. In many cases, it can actually leverage this raw data as-is, directly from the source.

Published: For lightly modified data that's streamlined for use by the business. The data in this layer is cleansed and where appropriate, masked, while also reflecting the latest changes from source systems.

Core: Here, business units can build new metrics, assets, and business rules to apply to the data. For instance, the business units can build reusable metrics that join customer data with inventory data.

Projects: The most tightly focused of the four layers; data is stored and managed here with specific projects and use cases in mind.

By dividing the platform up into these four layers of data quality, the company has ensured developers and analysts can focus on different users and use cases, while IT has the common foundation it needs to then quickly operationalize successful experiments. It's a great solution to the trade-off between speed of delivery and quality of data.

Read "From Lab to Factory: The Big Data Workbook" to learn how to power experimentation and rapid operationalization with a single platform for big data management.

Self-Service: Democratizing Data to Empower New Users.

Technological Imperatives for Data Self-Service:

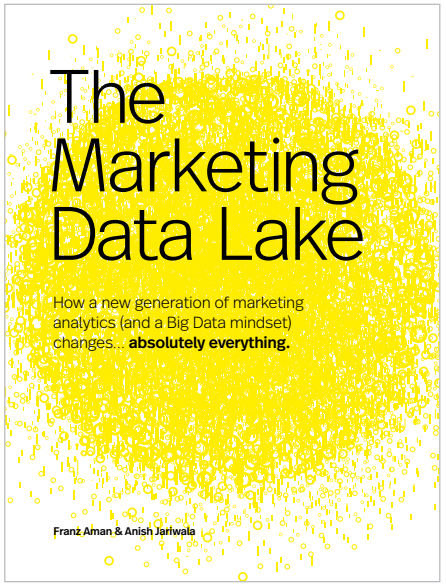
- **Self-service data environment:** So analysts have role-appropriate tools and a secure environment to discover and access a variety of data sources.
- **Data catalog:** So analysts can rapidly discover the data they are looking for.
- **Data preparation tools:** So analysts can easily join, clean, and secure the data prior to use in analysis.
- **Easy-to-use data integration:** Tools that give both IT developers and "citizen integrators" the speed and agility they need to connect different systems when they need to.
- **Reusable transformations:** So all users can leverage a common set of transformation patterns and create "recipes" that make their data prep work more reproducible.

The Marketing Data Lake

The Informatica marketing team recently built a marketing data lake, which gives us the agility needed to move toward an account-based marketing view of leads and customer activity. That means we can shape our marketing campaigns and events accordingly, even ranking which prospects to reach out to next. The result is better marketing productivity and conversion rates because we're reaching out to the right customers with the right content for the maximum effectiveness.

Read the 'Naked Marketing' series to find out how we did it.

Or better yet, read our book "The Marketing Data Lake".



Planning Your Architecture

Building an Analytics Architecture to Support Your Business Objectives.

Building an Analytics Architecture to Support Your Business Objectives.

When it comes to building out a data management architecture that can support your business initiative's goals, it's important that you decide where to start based on your company's specific needs and capabilities.

But whatever you decide, it's crucial that you build an architecture that allows you to do the following:

Drive increased data management productivity: The goal here is to make the data management processes more responsive to the needs of the business —for both IT and business users. This means less time doing data prep, and more time sharing, automating, and delivering actionable insights.

And crucially, your data management platform needs the ability to capture and store the steps ("recipes") that users create as they are doing data preparation work. The platform should be able to make these steps discoverable and shareable by other users to accelerate productivity. It should also be able to take these steps back into IT and easily and quickly operationalize them into repeatable processes that IT can run for the business. These types of capabilities will give business users the flexibility and speed they need while at the same time increasing the overall productivity of the organization.

Abstract data management away from underlying technology: Data formats (or the lack thereof), analytics technology, and applications will continue to change at a rapid pace. It is essential that your development environment protects your developers from changes in the underlying technology.

For instance, a change from MapReduce to Spark, should not break your data integration, data quality, or data security logic.

Abstraction allows you to maintain your data mappings and transformations, so you can leverage new analytics platforms and technologies as they emerge—without having to reinvent the wheel. Your data management platform must be able to connect to anything so your architecture can evolve and grow over time without creating new "islands of data".

Building an Analytics Architecture to Support Your Business Objectives.

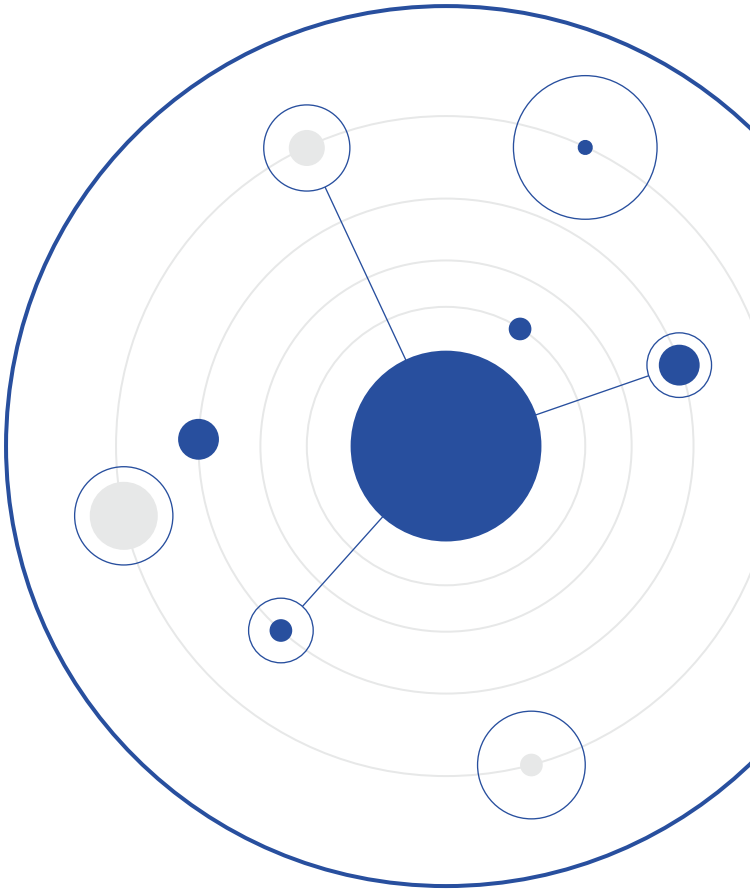
Leverages all available skills: When new platforms like MapReduce and Spark emerge, it can become difficult and expensive to find analysts and data scientists who can use those platforms. By abstracting the data management away from those platforms, you ensure you can always use your existing data management skills and interfaces to keep working on the data.

As a result, you aren't constantly searching for new hard-to-find and expensive talent. And you aren't forcing your analysts to spend too much of their time on data preparation instead of analysis.

Centrally manage metadata: When your data management platform uses a common repository of technical and business metadata, you ensure IT and business analysts know where the data came from and what it means.

Make sure your data management platform can harness the metadata to provide intelligence that improves everyone's productivity. For instance, a good data management platform can make recommendations on how to join data, whatever the data structure is. This means more productivity and faster time to business value.

Think hybrid: You need your data management platform to run across on-premise, cloud, big data and hybrid environments in order to make sure you can work with any data and support any use case.



Conclusion

The Building Blocks for Analytics Success.

"I think we'll look back in five to 10 years and we will laugh at how few people in our companies have access to analytics."

Christian Chabot,
CEO of Tableau at Informatica World 2016

The Building Blocks for Analytics Success.

Advanced analytics will change the way enterprises make decisions—one might justly argue that they already have. Analytics spending remains the top CIO spending priority for the fifth year in a row[9].

But as McKinsey found, most businesses' data and analytics programs are still only "somewhat effective". Clearly there is still a lot of work to be done. And a lot of that work, as the McKinsey study has found, is in the area of data management.

Before the vision of a dashboard on every desk and data-driven decision-making at scale can become a reality, enterprises have to lay the right data management foundations.

It's why the role of the Chief Analytics Officer is so crucial. And it's why the decisions you make around data are critical to the success of your organization. By laying the groundwork for enterprise data management, data governance, and self-service, you're empowering everyone in your organization to take advantage of its institutional knowledge.

Data management plays a crucial role in delivering that vision. So we hope the best practice we've shared here show you how to take advantage of emerging technologies, current enterprise realities, and most importantly—the analytical talent distributed across your enterprise.

Improve the productivity of your people, the reproducibility of their work, and the reliability of the data they use, and there's no telling how much they'll be able to innovate.

Sources

1. "Gartner: BI & Analytics Top Priority for CIOs in 2016," Business Intelligence Solutions Review. March 11, 2016.

2. "The need to lead in data and analytics," McKinsey&Company. April 2016.

3. "CIOs are underprepared for data-driven business, Computer Weekly. March 12, 2015

4. Gartner Magic Quadrant for Enterprise Integration Platform as a Service, Worldwide. Keith Guttridge, Massimo Pezzini, Paolo Malinverno, Kimihiko Iijima, Jess Thompson, Eric Thoo, Elizabeth Golluscio. 24 March 2016.

5. "IT budgets 2016: Surveys, software and services," ZDNet. October 1, 2015.

6. "Data Migration Customer Survey," Bloor. February 24, 2014.

7. "For Big-Data Scientists, 'Janitor Work' Is Key Hurdle to Insights," New York Times. August 17, 2014.

8. "TDWI Best Practices Report: Emerging Technologies For Business Intelligence, Analytics, and Data Warehousing," TDWI. October 1, 2015.

9. "BI & Analytics Top Priority for CIOs in 2016," Business Intelligence Solutions Review. March 11, 2016.