Think "Data First" to Drive Business Value

Why data should play the starring role in your reference architectures

Table of Contents

chapter 1

Introduction

chapter 2

Defining Success

chapter 3

Reference Architectures Done Right

chapter 4

Conclusion

Chapter 1

Introduction:
Houston, We Have a Problem

IT Isn't Supporting Business Need for Speed

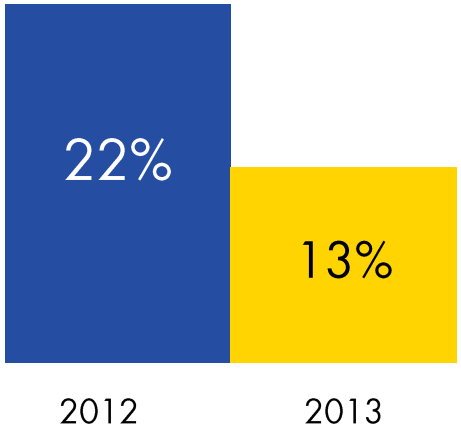
It's time to think differently about enterprise architecture, and enterprise *data* architecture in particular

Let's just call a spade a spade: IT organizations are not delivering value to the business as quickly as the business needs it.

Just as technology is unleashing unprecedented changes in our lives and our society, IT departments are playing a catch-up game. According to a recent McKinsey study[1] on business and technology strategy, overall satisfaction with IT performance was down in 2013 from prior years, among business users and IT management alike. Just 13% of IT executives surveyed said their organizations were effective at introducing new technologies faster compared with competitors, a decrease from 22% in 2012.

Waning Satisfaction with IT Performance

Percent of IT execs who say their organizations were effective at introducing new technologies faster than competition



Source: 2013 Global Annual Survey on Business & Technology Strategy, McKinsey

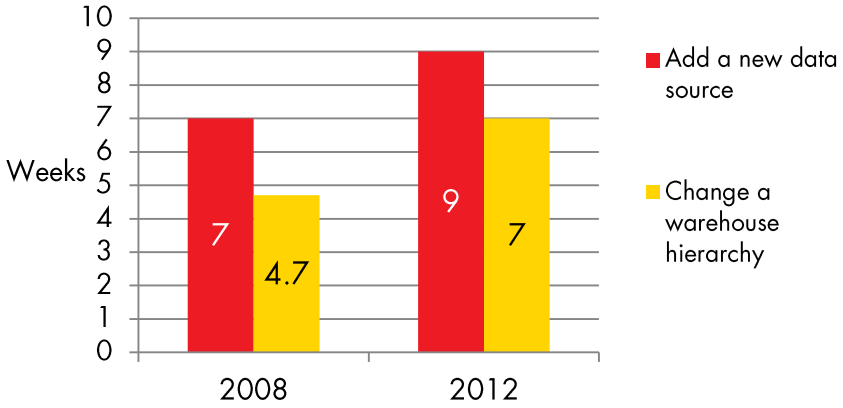
The problem? Data silos tied to applications and increasing data complexity have become a major impediment to delivering value to the business. Architects need to take the lead in three areas:

- Accelerate delivery of business value by fixing the fragmented data architecture where data is tied to applications.
- Deliver immediate access to reliable data for both business and IT users.
- Uphold data security and architectural governance issues.

Some characterize the current state of affairs as a crisis for IT. In its Hype Cycle for Enterprise Architecture, July 2013, Gartner found that organizations, "struggle as they spend too much time focusing on activities (that is, doing EA), such as modeling or capturing artifacts, instead of concentrating their efforts on identifying and enabling the delivery of business outcomes."[2]

Dogged by Data Silos

Information silos and infrastructure complexity are to blame for an increase in time to perform common integration tasks



Source: The Data Warehousing Institute (TDWI) "2012 BI Benchmark Report" and "2008 BI Benchmark Report"

The roots of the problem can be divided into those with which we're all familiar and those that are relatively new: First, the familiar problems:

Data as an afterthought: The traditional, ad hoc approach to building enterprise solutions one system at a time with a primary focus on applications and business processes locks up information in silos with minimal standardization of tools. That may have made sense when companies differentiated on their applications, but more and more those are becoming standardized. The true competitive differentiation now lies with the data and how you use it.

Data complexity. IT is under siege by the sheer volume of available data. And architectural designs are breaking, not just because of data's volume but also because of its complexity. This has been accelerated by the fast pace of M&A activity that introduces new systems, data, and business models to the organization.

Lack of data standards: Too many organizations let data integration developers approach projects in an unplanned fashion, without basic guidelines on processes, standards, or common tooling. This leads to inefficiency, brittleness, and complex data architectures, which increase cost and slow delivery of business value.

"Business-outcome-driven EA [enterprise architecture] is a practical approach to supporting EA that starts and ends with a focus on delivering business and IT leaders with signature-ready recommendations for adjusting policies and projects to achieve target business outcomes, based on the business direction, that capitalize on relevant business disruptions."

–Gartner Inc., Hype Cycle for Enterprise Architecture, July 2013 G00248891 page 3

Then there are problems that are relatively new, such as the following:

Geometric growth in data consumers. Equally problematic is the number of users who expect immediate and easy access to all that data from their tablets, smartphones, sensors, devices, and other applications. Enterprise architectures rarely fully account for uses that go beyond the organizational boundaries of IT into product development, supply chain management, etc.

A hammer-and-chisel approach in an age of robotics. Too much of data integration and management is still done by providing tools to make craftspeople more productive as they do the work of designing and implementing an end result. This model stopped scaling for manufacturing a century ago, and it's stopped scaling in the last few years for delivering data to the business. We need to take a fundamentally different approach to instrumenting our "data factories," where intelligent platforms are doing a lot of the tedious work (much like robots now automate manufacturing lines) and experts are guiding the technology platform in its work.

The importance of analytics for business transformation. Projects to drive insight, serve customers better, and get predictive about the business require data from many sources that are both internal and external to the organization.

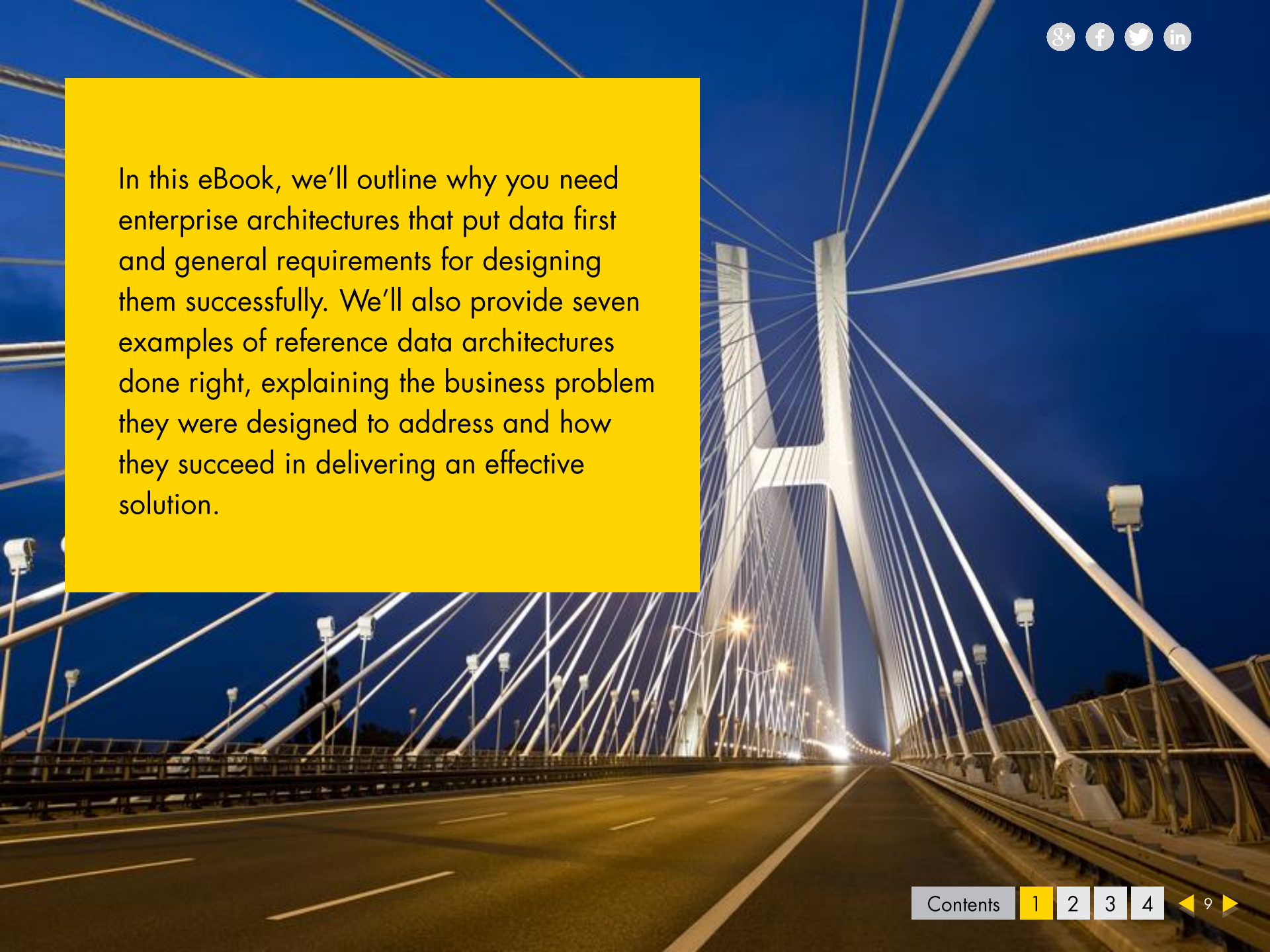
"Infrastructure complexity and lack of coordination among siloed teams is grinding some IT functions to a crawl. … It's not uncommon for more than 50 percent of an IT budget to be devoted to integration."

– Informatica's VP of Global Integration Services John Schmidt in "Can you architect your way from sluggishness to speed?", Potential at Work, June 2013

Data architects and IT leaders appreciate the problem. They recognize that data has become the central nervous system of the enterprise and are all too aware of the need for agility in responding to the business. However, short-term project schedules tend to trump enterprise data architecture concerns, and architects aren't given the time, people, and budgets necessary to do an effective job. Adding insult to injury, traditional enterprise data architecture has gotten tagged as a bureaucratic process likely to slow already sluggish project schedules—a perception that does little to rally the rest of the organization around the concept.

It's time to think differently about enterprise architecture, and enterprise data architecture in particular. Too much focus on architecting around specific enterprise applications and processes has led us to a point where it is impossible to meet business's time requirements. The cost, complexity, and security issues of ad hoc data architectures are greater than the intended benefits. Architects need to put data first.

Think about it this way: Your enterprise applications will come and go. The one thing that will truly provide competitive differentiation for your organization is your data.



In this eBook, we'll outline why you need enterprise architectures that put data first and general requirements for designing them successfully. We'll also provide seven examples of reference data architectures done right, explaining the business problem they were designed to address and how they succeed in delivering an effective solution.

Chapter 2

Defining Success

Three Design Points for Architectural Success

Gain speed, re-use, and efficiency by standardizing on a common approach to managing data architectures, processes, and technologies

Architectural success is defined as delivering business value faster, better, and cheaper. To accomplish that, you need to standardize on a common approach to managing enterprise data not only from an architectural perspective, but also from a process and technology perspective. And this approach also must design in the constant change and data growth that is our new normal. Among the benefits are speed, re-use, self-service, and efficiency.

1. Design for speed first (without compromising quality or cost)
2. Architect to know everything about your data
3. Architect for systematic self-learning and self-organizing

Design Point #1:

Design for speed first (without compromising quality or cost)

The point of this architectural shift is to enable the business to move faster than their competition through the use of data. There are two key design goals of the new data architecture:

- Accelerate the delivery of business value.
- Empower business stakeholders to act faster in response to market and customer demands.

Action Item:

As an architect, you need to build and manage data as a strategic system in a holistic way, starting design from the viewpoint of the data and data flows and leveraging the entire system as a shared resource. Standardization on a single intelligent data platform will eliminate custom, ad hoc approaches to data integration, which is the primary cause of project delays and complications.

Design Point #2:

Architect to know everything about your data

To know how to apply the right data to a given business scenario, you need to know more about the data, including its context, rather than only its schema or other technical attributes. You need to gain intelligence about how, when, where, and why your data assets are created and used.

In many current enterprise architectures, it's hard to find and access data. That scenario will only worsen as data from third parties, cloud, mobile, social, big data, and the Internet of Things is added into the mix and causes further fragmentation and complexity.

Action Item:

It's not enough to have a shared data repository. The requirement is to build a dynamic, shared metadata layer to collect and manage technical metadata; data relationships; data models; business metadata; semantic meaning; usage patterns; security mappings and privileges; and rules governing how, where, and when data is used to ensure the data is truly understood. This data is essential to creating an environment where tasks can be automated and intelligence added for greater productivity.

Look for an intelligent data platform that is able to learn usage patterns, leverage metadata, and make intelligent recommendations. Such a platform will, over time, learn from its own actions, enhancing its intelligence level. This will accelerate the delivery of clean, safe, and connected data for your business initiatives.

Design Point #3:

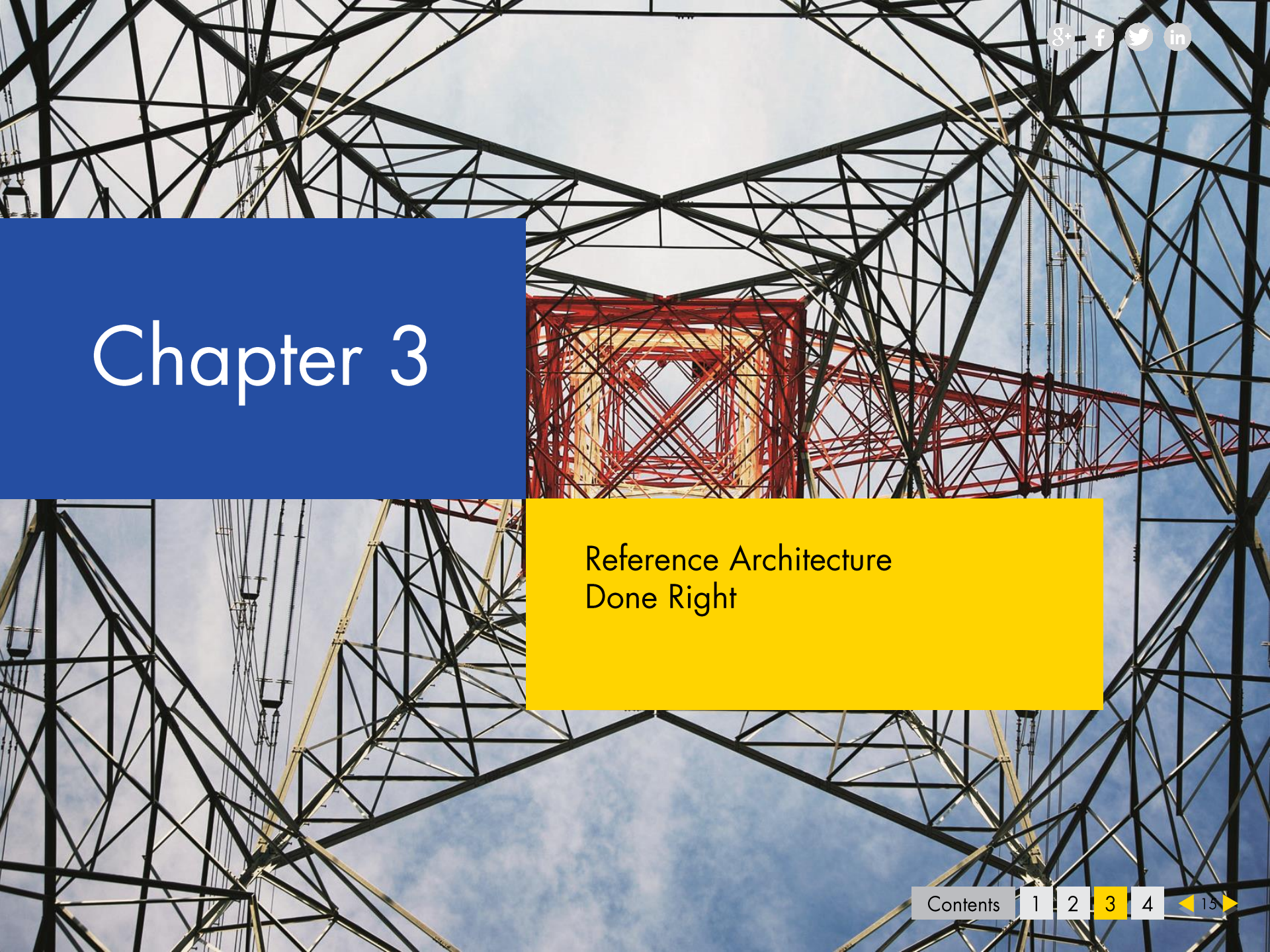
Architect for systematic self-learning and self-organizing

Let's face it: You are not going to get the headcount necessary to organize your current data architecture, let alone deal with growing data complexity. The only way to survive is to increase the level of automation in the system for both IT and business users.

Action Item:

Design a modern architecture based on an intelligent data platform that employs metadata and machine learning to self-organize, greatly accelerating project delivery. The system should be able to automate routine tasks, make intelligent recommendations, and provide computer assistance and collaboration for more complex tasks.

Coupled with an intelligent data platform, data-centric architectures will ensure data is self-organizing, self-healing, and intelligently provisioned, freeing IT from a lot of the routine work involved in managing, modeling, and cataloging data. Moreover, an architecture that learns usage patterns and pairs them with analytics will enable a new level of automation that will speed value delivery. Increasing automation will help IT scale to meet the challenges of data complexity while delivering value in a timeframe that meets business needs.



Chapter 3

Reference Architecture Done Right

Seven Enterprise Architectures that Put Data First

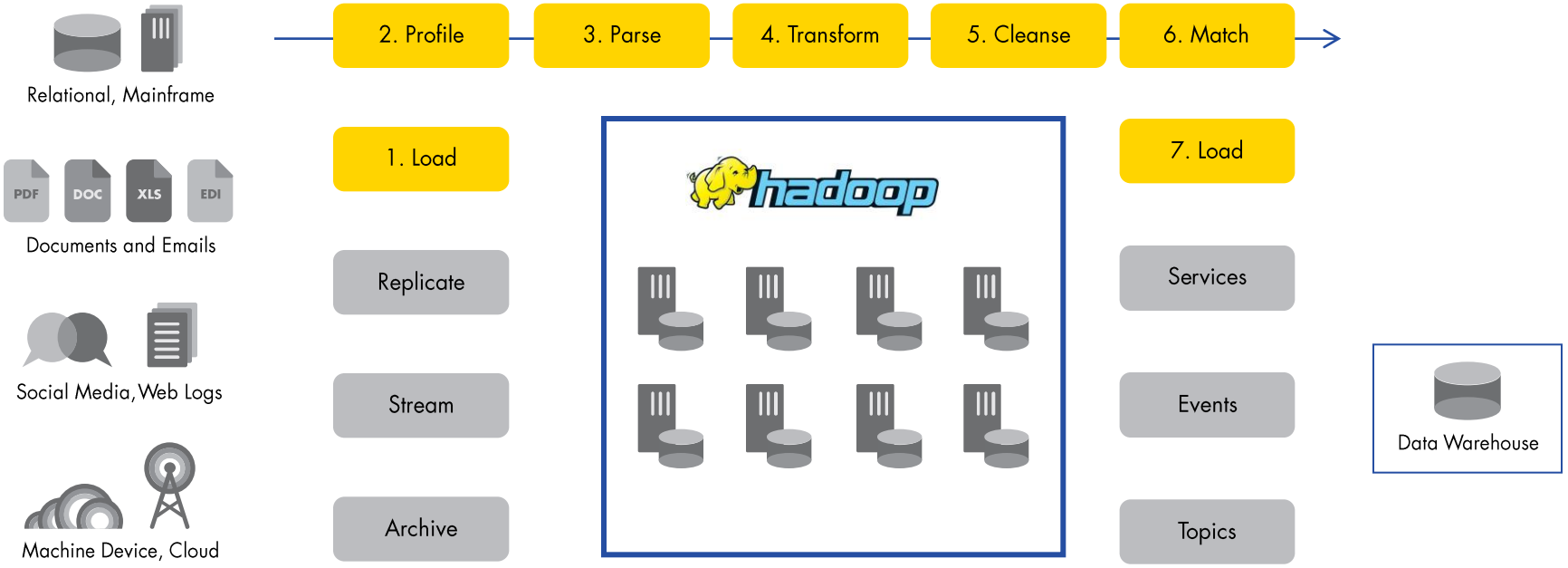
From data warehouse optimization to data security, giving data a starring role in your reference architectures delivers increase speed for business benefit

Shifting to an approach that puts data first can help organizations better address a variety of common business problems and move the overall enterprise architecture forward to meet the demands of the modern age. We offer seven examples of common business processes and applications that are more effective and provide new opportunities when you put data first when planning their architectures.

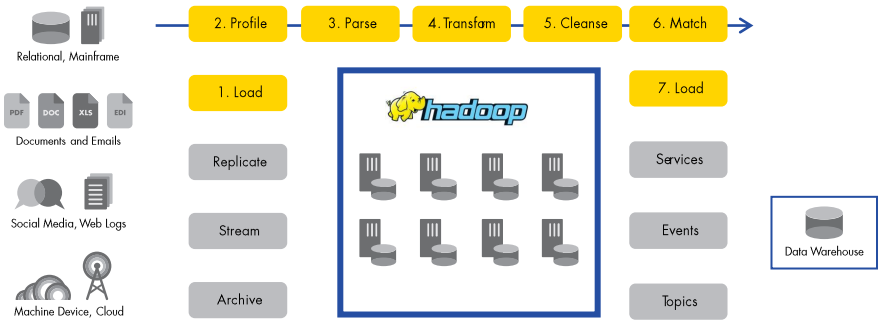
1. Optimized Data Warehouse
2. Modern Data Warehouse
3. 360-Degree View of the Customer
4. Modernizing and Retiring Applications
5. Hybrid IT
6. Next-Generation Shared Services/Data Integration Hub
7. Data-Centric Security

#1: Optimized Data Warehouse

Data warehouse optimized by offloading and preprocessing on Hadoop



#1: Optimized Data Warehouse



Traditional process: Traditional data warehouses and specialized data warehouse appliances are mature and optimized to store and analyze structured data typically found in business.

The challenge: Data warehouses and appliances can be very expensive to operate on a cost-per-terabyte-of-data basis. Managing the data preparation in a commodity environment like Hadoop can result in an order of magnitude in cost savings.

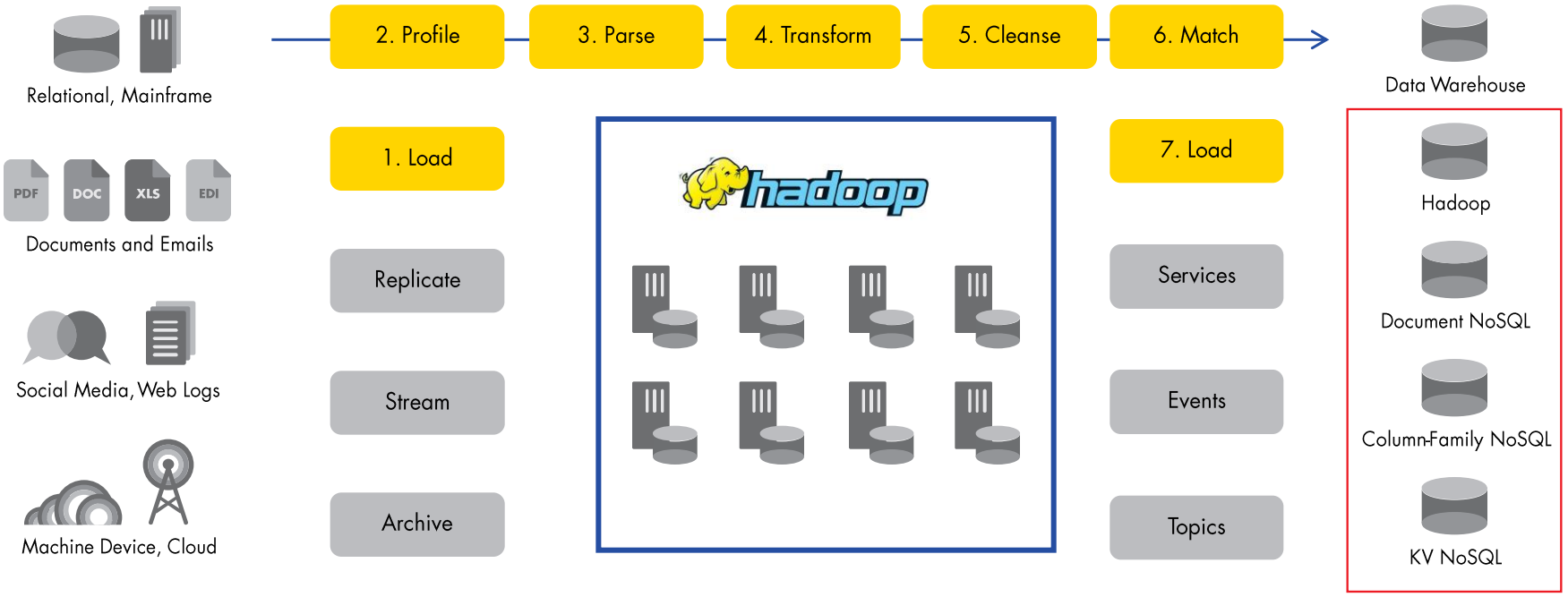
Preparing data for the data warehouse or appliance can consume more than 60 percent of the development time. Using data preparation tools for cataloging, cleansing, transforming, and moving the data in—along with tools that can run across all types of environments (SQL, data warehouse, Hadoop for example)—can result in significant time and cost savings. That way, expensive data scientists are free to focus on what they should be doing—discovering data insights.

Benefits of data-centric solution: You can increase productivity and gain substantial savings by using Hadoop to stage the preprocessing work (transforming raw data into an understandable format) in addition to pre-built connectors and transformation libraries for ETL, data quality, parsing, and profiling. Estimated costs for data stored on specialized hardware ranges from $20,000 to $40,000 a terabyte, while data saved in Hadoop is only around $1,500 per terabyte. In addition to cost savings, this approach doesn't require Hadoop expertise, although it exposes IT to that and other big data technologies. The money and cycle time saved on data preparation work can then be redirected to business-outcome-oriented initiatives.

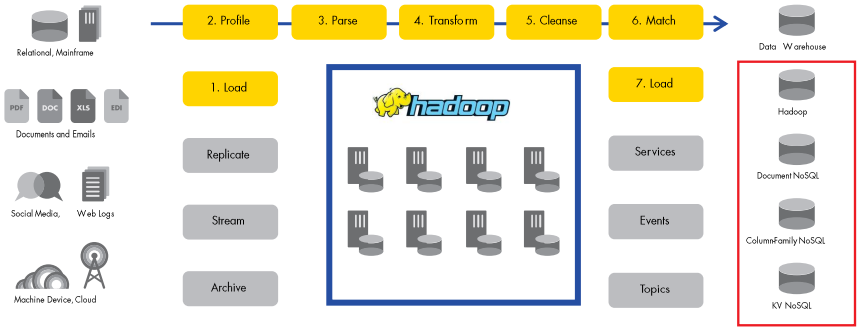
Key takeaways:
If you are looking to get started with big data, this is a good place to get experience.
Look for tools that work across all environments, as they'll save you time and promote re-usability of people, skills, and code.

#2: Modern Data Warehouse

Add next-generation NoSQL and Hadoop storage and analytics



#2: Modern Data Warehouse



Benefits of data-centric approach: Putting data first will enable a single toolset requiring a single set of skills to prepare and load data into the different systems. This enables organizations to leverage the technology that makes the most sense in a cost-effective manner.

Traditional process: While still relatively new, companies are starting to broaden out from the traditional, structured data warehouse to add a variety of data types, particularly unstructured data like social media, web logs, and machine information.

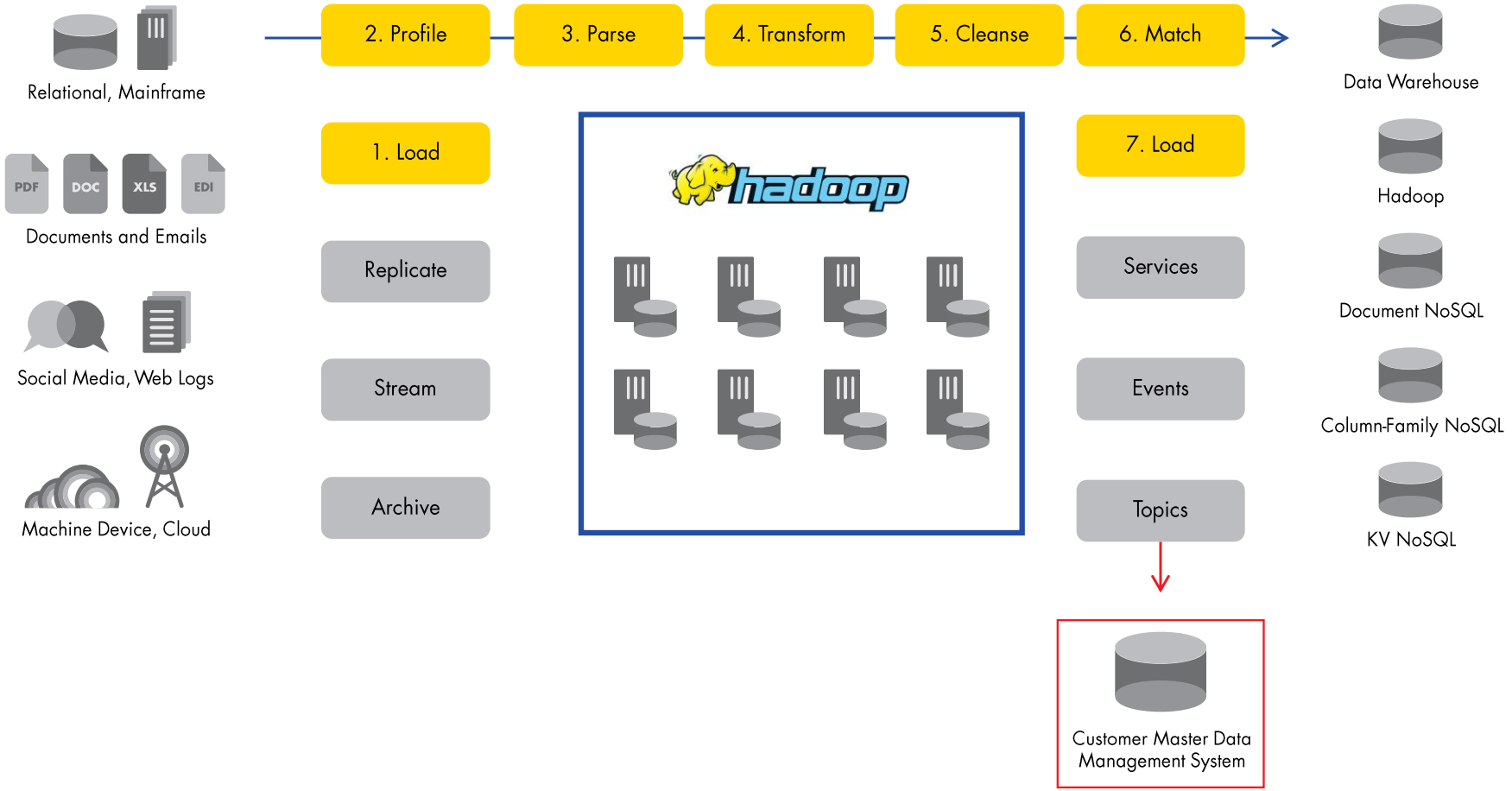
The challenge: Relational databases aren't capable of storing the full range of eclectic data types, so this data hasn't been stored or analyzed historically. Companies adding new technologies like Hadoop and NoSQL to handle the influx of new data may not have the skill sets to take on the task and be effective.

Key takeaways:

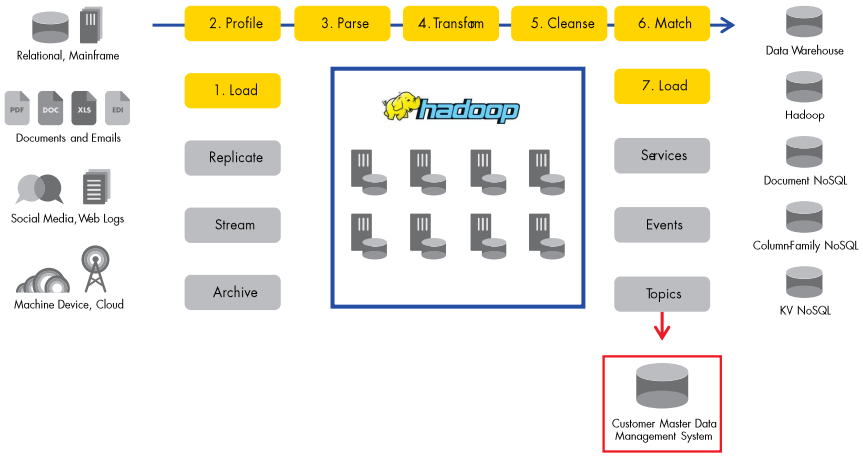
- You can easily grow incrementally from the Optimized Data Warehouse architecture to this one.
- With tools that run across multiple environments, you can adopt the best new repository technologies available with minimal re-training.

#3: 360-Degree View of the Customer

Connect events and transactions with customer master data



#3: 360-Degree View of the Customer



Traditional process: With one-off customizations and integration projects the norm, customer, product, and transaction data is traditionally fragmented and spread across different siloed systems.

The challenge: Your customer (or other business entity) data is typically scattered across dozens of systems. Without a single 360-degree view, companies miss opportunities for cross-selling, upselling, and other marketing initiatives. Not to mention, they run the risk of customer service issues.

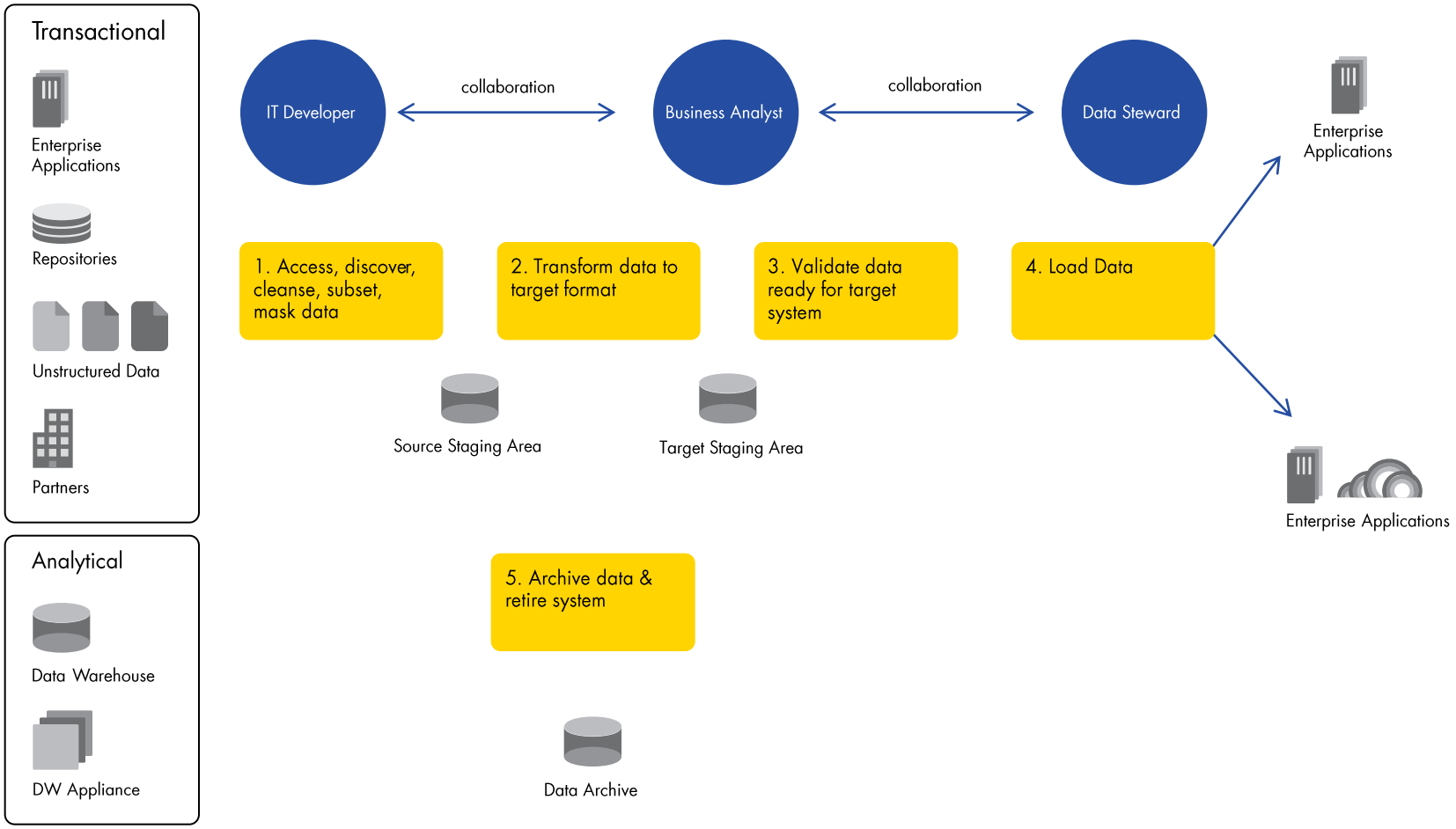
Benefits of a data-centric approach: By leveraging master data capabilities, automation, and intelligence, a data-centric approach can deliver the broader view of a customer that can be leveraged to help spawn new revenue opportunities, improve customer service, and create efficiencies.

Key takeaways:

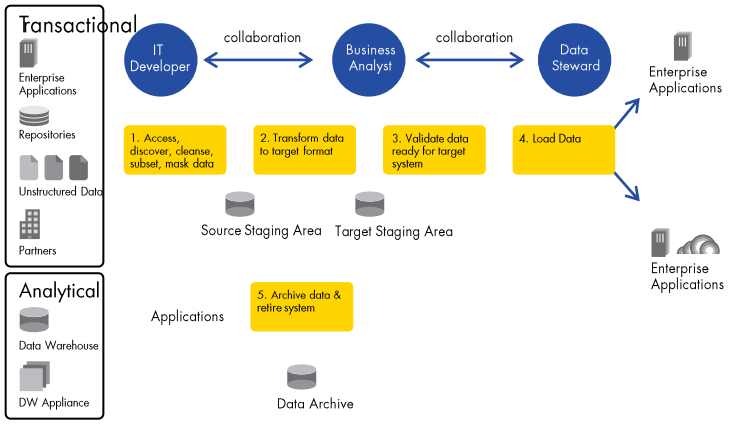
- Implementing a master data initiative around a business entity such as customer, partner, or part number will provide a solid basis of trusted data for your current business initiative that can be re-used in any other application, analysis, or initiative that needs data about that business entity. This will speed delivery of new projects that use that data.
- Adding this to the first and second reference architectures will provide access to a broad variety of internal and external data.

#4: Modernizing and Retiring Applications

Move data between enterprise systems with low risk



#4: Modernizing and Retiring Applications



Traditional process: Companies accumulate systems over time that need to be consolidated, upgraded, or modernized. This is particularly true in M&A situations. The problem is that while all companies need to do this on a regular basis, few develop a competence or repeatable processes in this area, leading to significant risk of project overruns or failure.

The challenge: By attempting these migrations as one-off projects as opposed to following a standardized set of best practices, companies often encounter data quality and data conversion problems that don't surface until late in the game. (See our white paper "10 Things You Need to Know Before Modernizing Your Applications" for a list of best practices.) The fallout is usually project delays or total project failures, depending on the severity of the data quality issues.

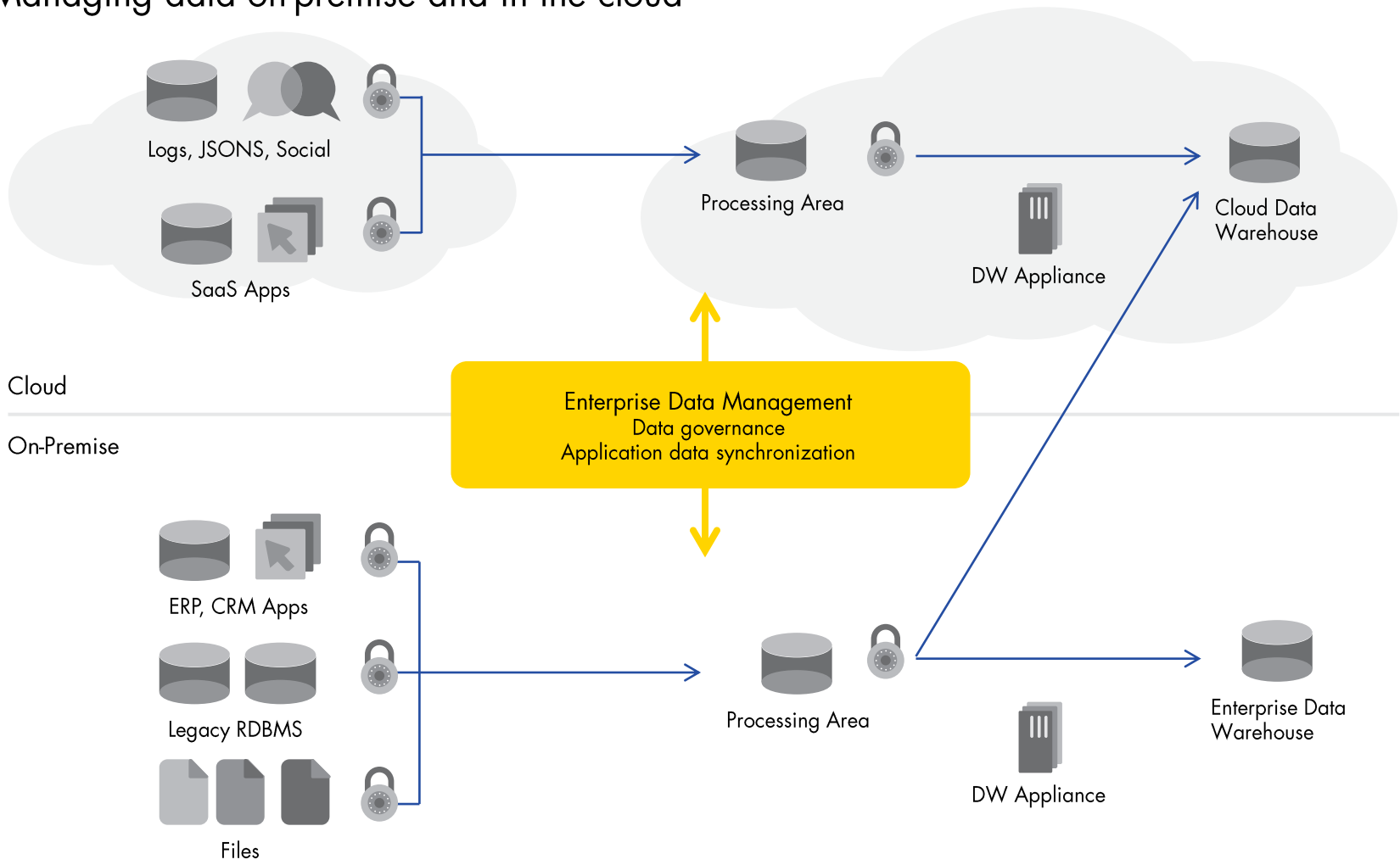
Benefits of a data-centric approach: Establishing repeatable processes that address the problem in a systematic way while examining data as it flows from the source system to source staging area can help avoid these issues. Business, working in concert with IT, can provide context for the data and ensure that it follows specific business rules. Employ a standard sequence of tools for validation, cleansing, and transformation so potential problems can be identified on the front end, allowing them to be fixed early. Compared with traditional ad hoc methods that lead to "code, load, and explode" scenarios, the approach that puts data first minimizes costs and risks while speeding migrations.

Key takeaways:

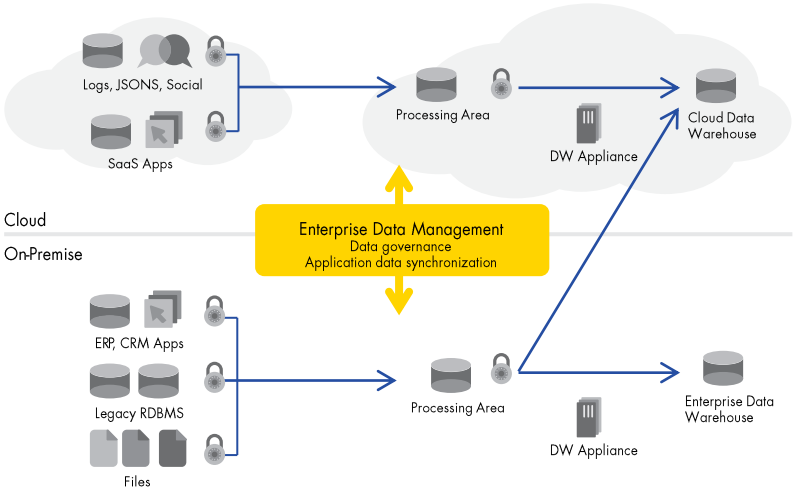
- Standard processes, practices, and approaches for these projects can be useful for other similar projects across the organization. This is an easy way to establish these practices.
- An architecture like this one can be an excellent starting point to a data governance initiative for managing data as an asset across the organization.

#5: Hybrid IT

Managing data on-premise and in the cloud



#5: Hybrid IT



Traditional process: Cloud applications have reached a tipping point—enterprise deployments are accelerating rapidly. All too often, these projects are ad hoc and disconnected from the rest of the enterprise, creating another silo of data.

The challenge: These systems are brought on-line without regard to enterprise architectures and data governance controls, exacerbating data fragmentation problems. Architects need to ensure that data in cloud applications is managed, secured, and integrated with other key systems that require this data, both in the cloud and on premise.

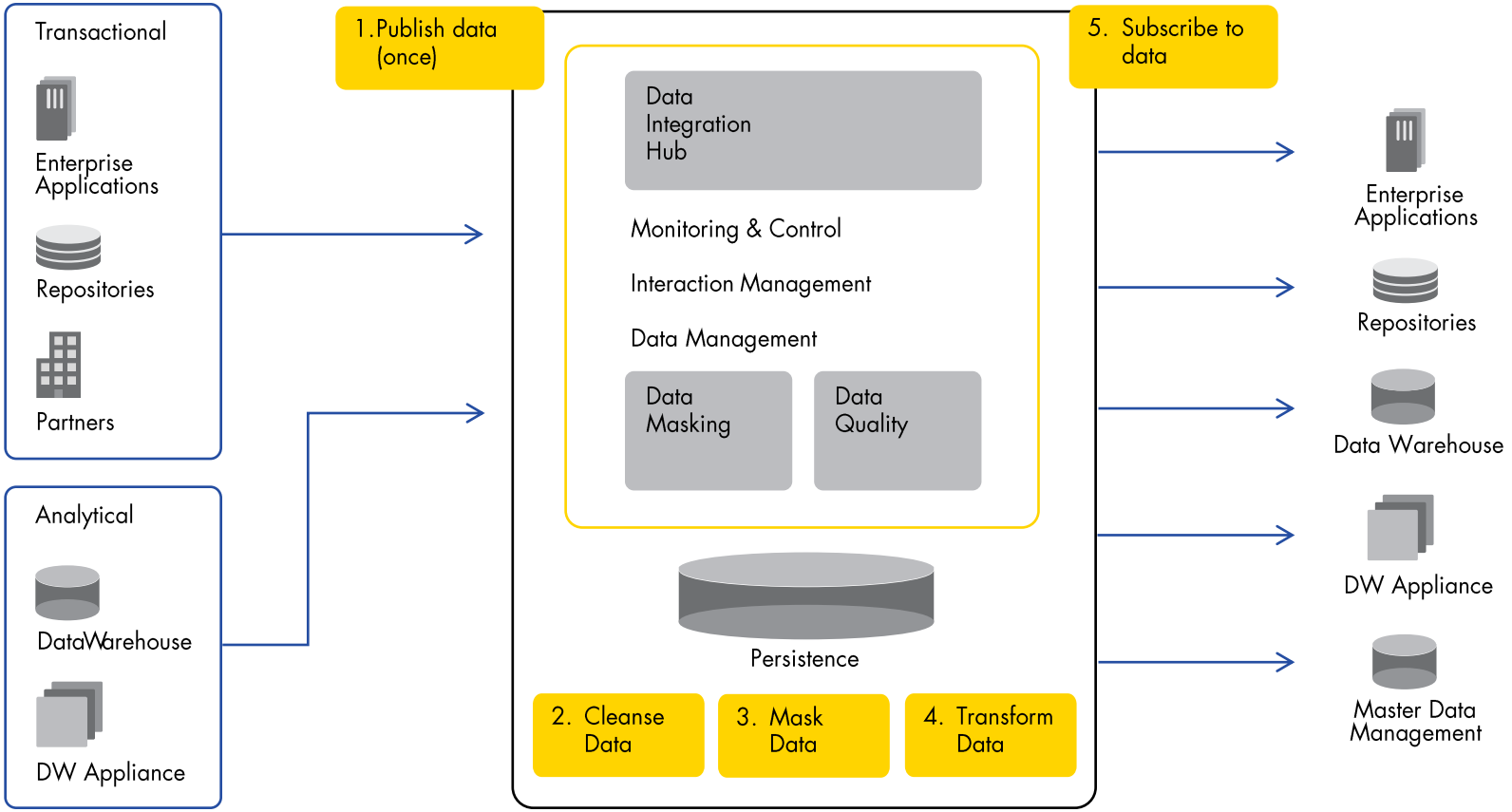
Benefits of a data-centric approach: Architect for a single set of data management tools that deliver a holistic view of on-premise, cloud, big data, relational data, and real-time data. This ensures a central repository complete with the proper architectural constructs and governance controls. Cloud applications can take advantage of the elasticity and pay-as-you-go advantages of the cloud, but must be managed, governed, secured, and synchronized with data in the current enterprise architecture.

Key takeaways:

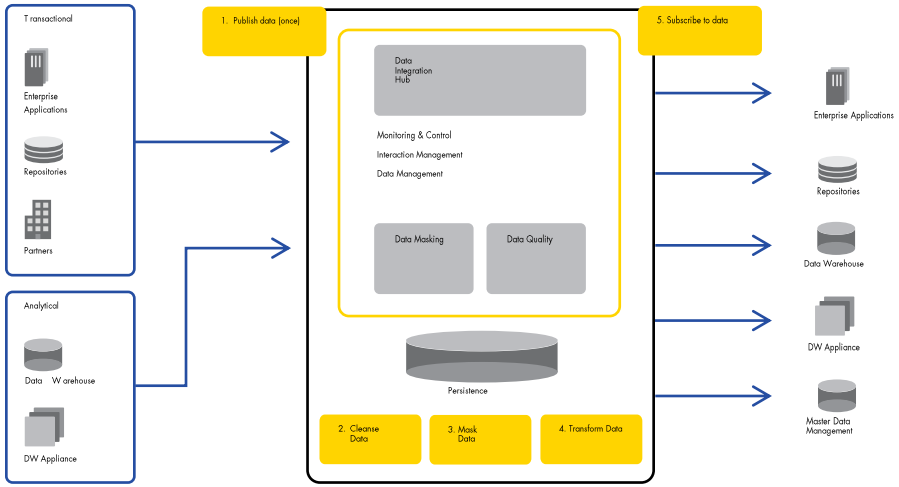
- Architect to manage data regardless of its location.
- Think about security and the need to share and synchronize data up front.

#6: Next-Generation Shared Services/Data Integration Hub

Publish and subscribe to create shared data services and rationalize the data integration environment



#6: Next-Generation Shared Services/Data Integration Hub



Traditional process: Most organizations have a classic "spaghetti diagram" of all the different data integrations and technologies used to move data between applications.

The challenge: Not only is the many-to-many style of integration costly and inflexible to change, it's also very hard to understand and modify. In addition, each use of data is a new integration.

The benefits of a data-centric approach: With a data integration hub, application owners publish data once to a central location, which then makes it available to any and all subscribers requesting that specific data topic. The hub enforces consistent data quality, transformation, and validation rules while also delivering real-time visibility to the data and SLAs. The concept of persistence means the data can be published once and be made available to many subscribers upon request without requiring additional integration work. This aids in reuse and can result in an 80% reduction in application on-boarding costs and effort.

Key takeaways:

- A data integration approach to data architecture can significantly simplify your environment, saving costs and accelerating data delivery.
- As you think about change management, this is a key tool to simplify the process and reduce the risk involved in changes.

#7: Data-Centric Security

Make data secure at the source

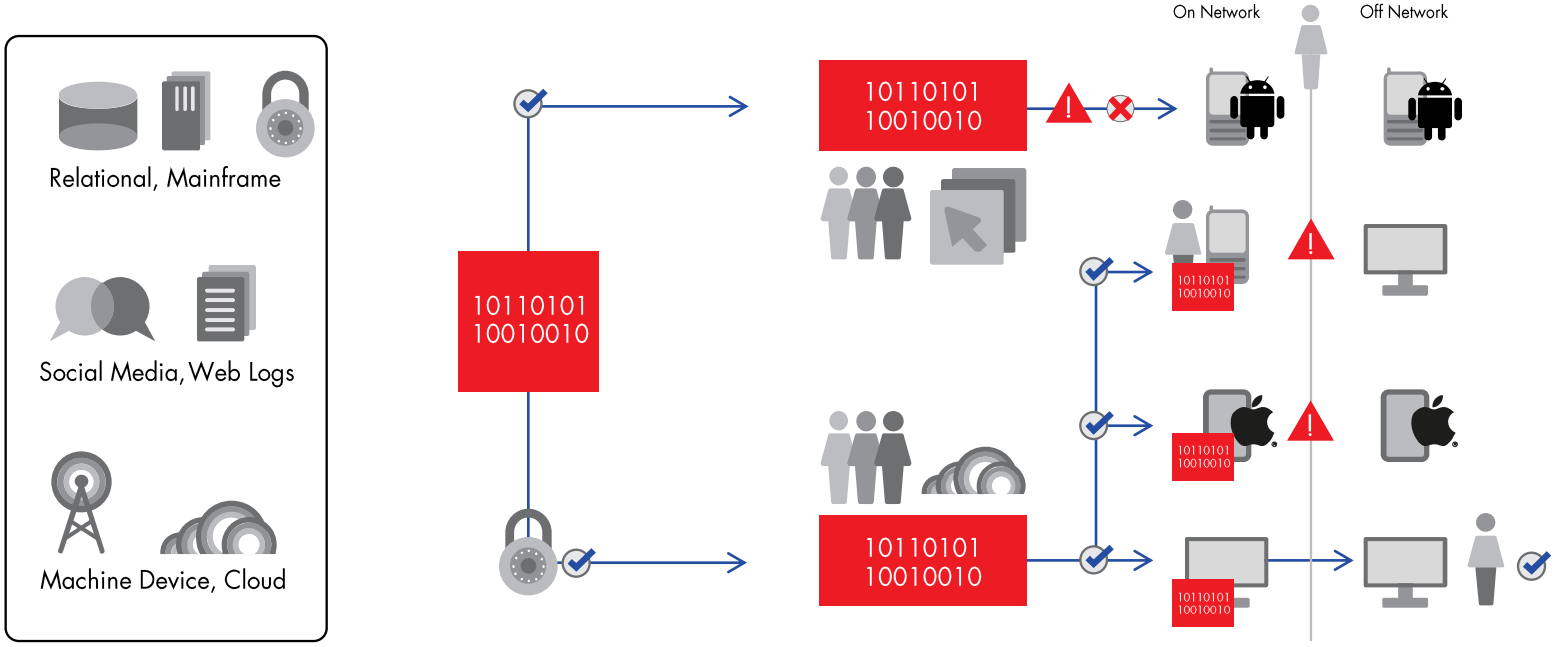
1. Connect to all data sources

2. Profile and locate sensitive data and its proliferation

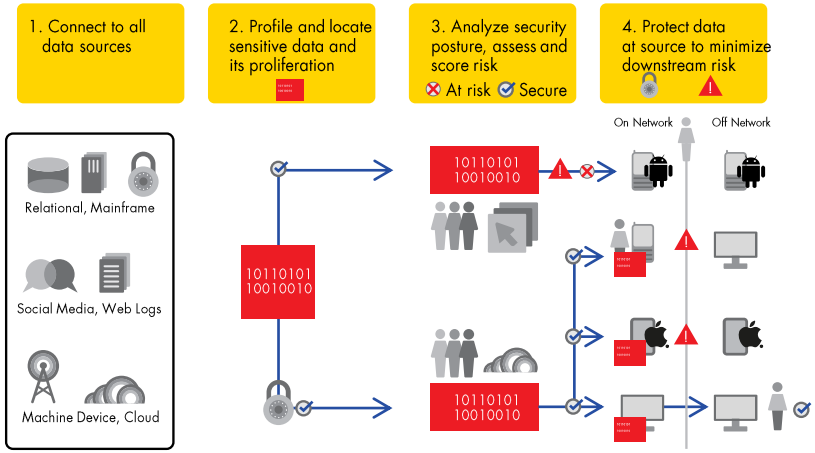
3. Analyze security posture, assess and score risk

At risk Secure

4. Protect data at source to minimize downstream risk



#7: Data-Centric Security Architecture



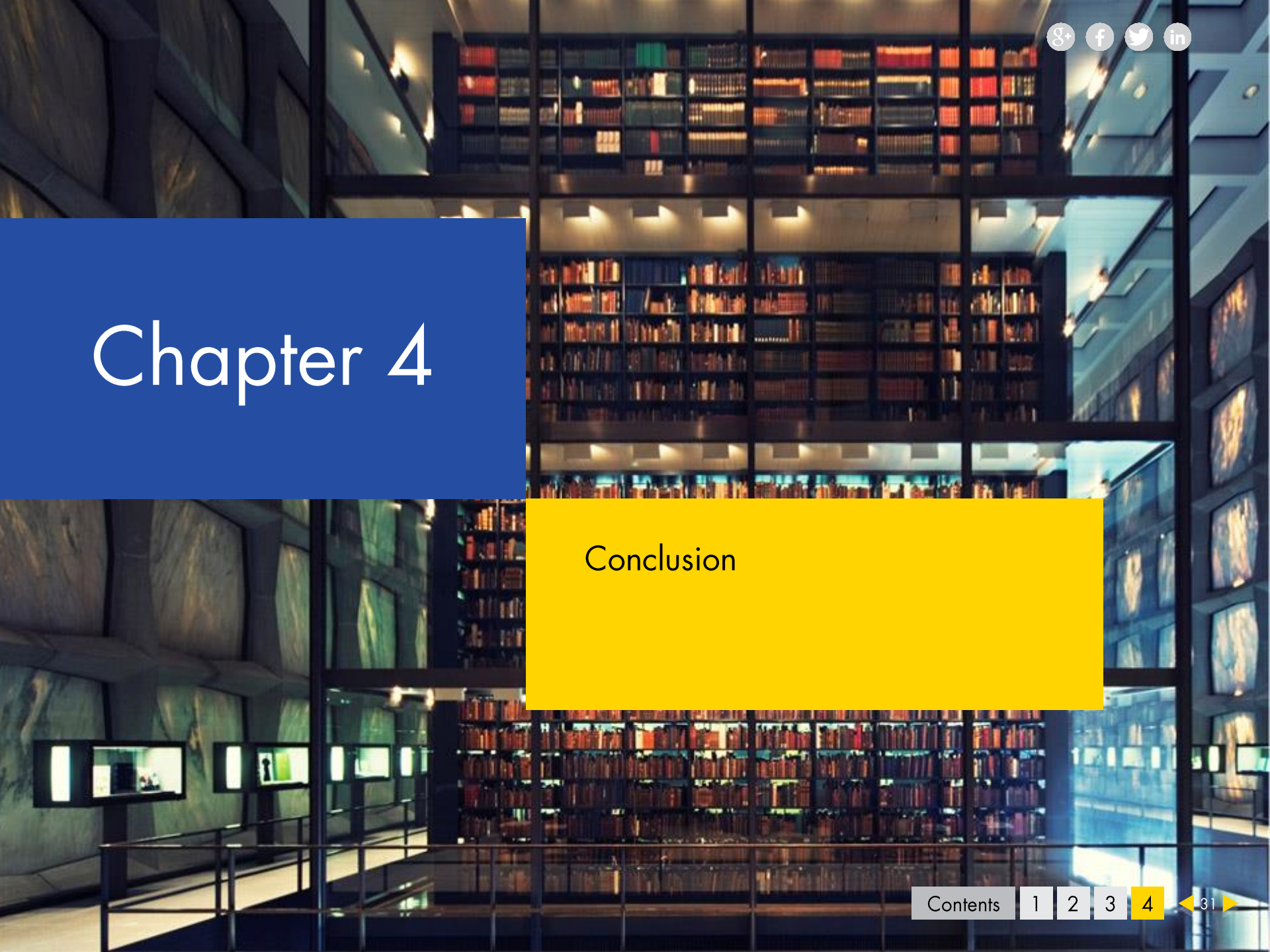
Traditional process: Most organizations address security at the application and data center level, employing firewalls, virtual private networks, antivirus software, and other data loss prevention technologies to safeguard data within the perimeter. While highly effective at that level, these approaches are insufficient for safeguarding data once it proliferates outside the perimeter. (See Ponemon Institute's "The State of Data Centric Security" for more on how organizations are responding to threats today.)

The challenge: Even if data is properly secured within the perimeter today, companies remain at risk for data loss or cyber-attack from employees and from mobile devices outside the network perimeter. This puts pressure on IT to augment traditional perimeter-based security architectures with capabilities that can properly protect data even if it's outside the scope of existing security controls.

Benefits of a data-centric approach: With a security architecture that puts data first, the data is secured at its source. Tags, classifications, and policies are mapped directly to the data and stay with it as it moves throughout the enterprise and beyond. In this manner, security can be applied at a granular level, and enforcement is agnostic to the technology platform, consumption model, or geographic location. This allows companies to easily respond to the changes required by business users without completely locking down data. Moreover, data is fully protected inside and outside of the perimeter, lessening the risk of data breaches or non-compliance.

Key takeaways:

- Think about how much risk you have today with data moving outside your network and traditional protection mechanisms.
- Consider this approach as an augmentation, not a replacement, for your current security systems.



Chapter 4

Conclusion

Right Data to the Right Users in the Right Way, Right Away

Our vision for the intelligent data platform supports enterprise architectures that allow IT to meet business's time requirements

Informatica's vision is for a next-generation "intelligent data platform" that will enable organizations to use their data as the competitive advantage we believe it to be. The platform serves as a cornerstone for new applications and devices, functioning as the "data highway" that connects the right data to the right users and systems while also ensuring data is properly protected and secured with governance throughout its lifecycle.

Used as the basis for enterprise data architecture, the intelligent data platform will facilitate a next-generation approach to data management, leveraging machine learning, automation, and computer assistance to allow both IT and non-technical business users to easily and swiftly integrate, consume, and analyze all types of data. The intelligent data platform will be architected to support today's hybrid world so companies don't have to rip and replace older technology. Plus, it is designed for scalability to serve individual users through the entire enterprise.

Some example use cases include:

Intelligently finding data. An intelligent data platform should have an easy and intelligent way for users to find data. This approach should understand the user's business context, past actions, and common use cases relating to the current search. In this way, the platform can intelligently recommend relevant data to the user.

Intelligently preparing data. Business users often need to find and combine data sets, possibly from outside the organization. An intelligent data platform should be able to recommend ways to join the data, cleanse it, and augment it without the user having to know the technical details of how this is done. The intelligent data platform should also be able to save and recommend data file sets and processes that other people have already built.

Intelligent data ingestion. An intelligent data platform should be able to learn and infer a great deal of information about new data being brought into the organization. That way, the process of categorizing, modeling, and sharing the data can be done much more quickly.

Intelligent data security. An intelligent data platform should be able to automatically detect sensitive data and recommend policies to attach to that data to ensure it is secured and compliant wherever it is consumed.

Data is the lifeblood of the business. To thrive, organizations need to take an approach to enterprise architectures that puts data first. And standardize on an intelligent data platform that can provide the levels of skills and code reuse, computer assistance, collaboration, and automation to increase the productivity of IT and enable business users to serve themselves. Only then will enterprise data architectures deliver value at the speed that business needs.

If you're interested in reading more about how businesses that see data as their lifeblood are gaining competitive advantage over those that don't, download our "Research: Data Drives Profit" ebook. Or click here to contact us. We'd welcome the opportunity to share more about our vision for how an intelligent data platform can be used as the basis for enterprise data architectures that differentiate your company.

Sources

1. 2013 Global Annual Survey on Business & Technology Strategy, McKinsey, http://www.mckinsey.com/insights/business_technology/it_under_pressure_mckinsey_global_survey_results

2. The Hype Cycle for Enterprise Architecture, July 2013, page 4, Gartner, Inc. G00248891

Published July 2014

IN18_0614_2657