



Informatica®

# Cloud Data Warehousing with Microsoft Azure

A workbook for creating a modern data  
architecture on Azure



# Contents

<b>Introduction</b>	<b>03</b>		
– Accelerate Your Move to Public Clouds	04		
– Benefits of Data Warehousing in the Cloud	05		
<b>Part 1: Three Ways to Leverage Microsoft Azure Synapse Analytics (formerly Azure SQL Data Warehouse)</b>	<b>07</b>		
– Exploring Cloud Data Warehouse Use Cases	08		
– Building a New Cloud Data Warehouse on Microsoft Azure	09		
– Extend On-Premises Enterprise Data Warehouse with Azure Synapse	11		
– Migrate On-Premises Data Warehouse to Azure Synapse	14		
<b>Part 2: New Paradigms Powering the Shift to Cloud Data Management</b>	<b>16</b>		
– Exploring Cloud Data Management Drivers	17		
		<b>Part 3: Data Management Best Practices for Azure Synapse</b>	<b>19</b>
		– Accelerate Data Migration with Automated Data Management	20
		<b>Part 4: Cloud Data Warehousing Checklist</b>	<b>25</b>
		– Identify Your Stakeholders and Their Needs	27
		– Migrating and Integrating Your Data	28
		– Getting the Most Out of Azure Synapse	32
		– Reducing Your Risks and Promoting Success	36
		<b>Part 5: Accelerate Data Warehouse Modernization to Azure with Informatica</b>	<b>38</b>
		– Simplify with Intelligent Data Management for Azure	39
		<b>About Informatica®</b>	<b>41</b>
		<b>About Microsoft Azure</b>	<b>42</b>

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



# Introduction



# Accelerating Your Move to Public Clouds

**Intelligent data is a powerful force enabling businesses to disrupt competitive landscapes and to discover new revenue streams. When you give business leaders and decision makers timely access to secure and relevant data, they can quickly discover new insights, and make transformative decisions.**

Data transforms the way we operate, and businesses today consume data from more sources than ever before: big data, Internet of Things devices, business applications, and social media feeds. Transformative businesses take a data-driven approach to competing and operating in the data-rich economy so they can quickly scale resources and adapt investments to capture new opportunities. They do this by connecting on-premises and cloud data across platforms and business functions.

With their inherent agility, public clouds are ideal platforms for processing the vast quantities of data necessary for advanced analytics. You can speed decision making in your organization when you provide business leaders with cloud-based self-service analytics tools.

IT organizations are moving on-premises workloads to public clouds, such as Microsoft Azure at an increasing rate. In its “2017 State of the Cloud” report, InteropITX Research<sup>1</sup> found that since 2012, the rate of businesses shifting workloads to public clouds nearly doubled to 57 percent in 2016. The top three drivers to public clouds, according to the InteropITX report, are scalability, performance, and better/faster access to resources. Clearly, cloud computing is the new normal.

We’ve written this workbook to help you accelerate your data-driven digital transformation and guide you in modernizing your data architecture with Microsoft Azure. We will show you how cloud data management can enable the next generation of agile analytics initiatives. We’ll describe key cloud data management hurdles and how to overcome them to support common usage patterns for cloud data warehousing with Microsoft Azure.

By the end of the workbook, you’ll be prepared to start your journey toward delivering accessible, secure, and relevant data to your business decision makers.

**Let’s start by building the foundation.**

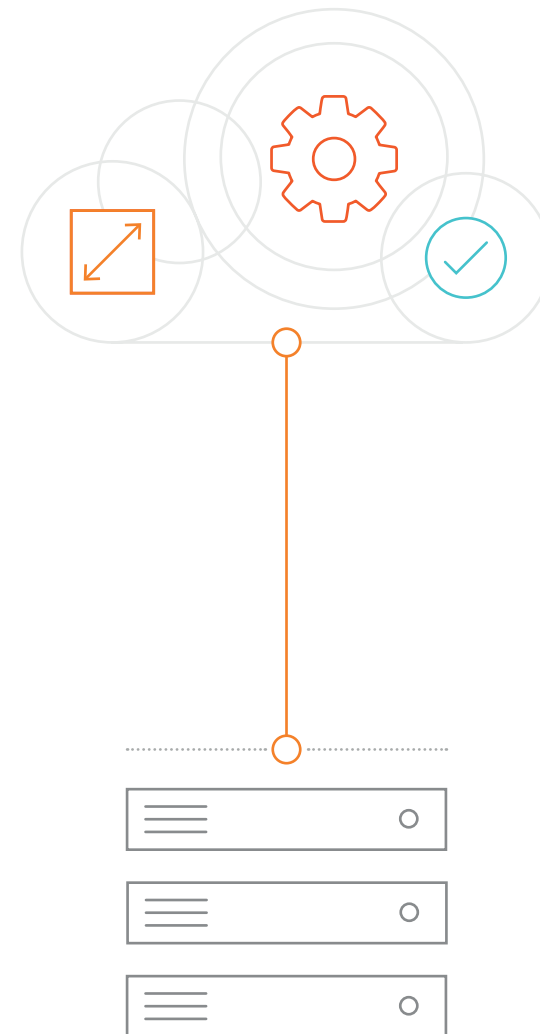
<sup>1</sup> InteropITX Research Reports, “2017 State of the Cloud.” January 2017.



# Benefits of Data Warehousing in the Cloud

Public cloud data warehouses, such as Azure Synapse, provide robust functionality, require virtually no upfront investment, and offer the added benefits of the cloud's elasticity, infinite scale, and agility. With cloud data warehouses, you pay for what you need, when you need it, eliminating the need for complex capacity planning and fixed capital investment.

While cloud data warehousing offers many benefits around agility, scalability, performance, and cost savings, there are several additional benefits of using Azure Synapse. Those include:



- **Speed:** Azure Synapse can query unstructured and semi-structured data stored in Azure Blob Storage or Azure Data Lake Store using familiar T-SQL skills. This means it's easy to combine data sets no matter where they are stored. While traditional data warehouse models require data to be moved into the instance to be accessible, Azure Synapse allows the data to stay where it is and combine the results with relational data via common T-SQL constructs. This keeps your data costs low and lets you choose the query speed that you need.
- **Scale:** Whether your data is on-premises, in the cloud, or in a combination of both, Azure Synapse helps you meet many use cases. You can burst into a public cloud environment to handle extra load when your existing on-premises environment needs a temporary increase in processing or data capacity. You can shrink the environment back to on-premises later. Azure Synapse allows you to scale quickly to your needs, meaning you only pay for what you use. This means you save on your compute power, whether you are deploying new virtual machines (VMs), moving a few workloads, or even migrating your data centers.
- **Elasticity:** Azure Synapse independently scales compute and storage so you only pay for the query performance they need. This means you can grow or shrink query compute in seconds. And since compute and storage scale independently, costs can be easy to forecast.
- **Security:** Azure Synapse helps you comply with industry regulations with its auditing, encryption, and threat detection capabilities. You gain insight into discrepancies so you can quickly address security concerns. Azure also lets you perform authentication, authorization, and access control through SQL Server authentication or Azure Active Directory.

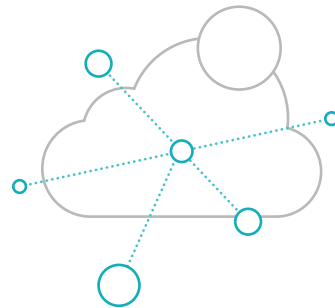
Part 1

# Three Ways to Leverage Microsoft Azure Synapse Analytics (formerly Azure SQL Data Warehouse)

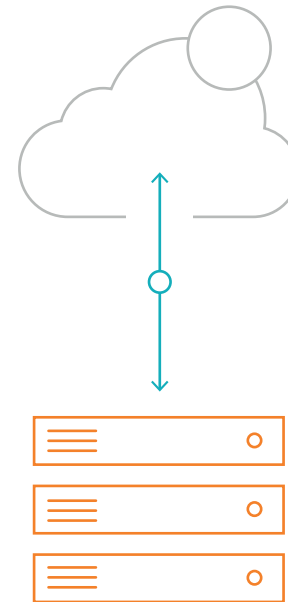
# Exploring Cloud Data Warehouse Use Cases

Organizations typically fall into three common scenarios for cloud data warehousing with Azure Synapse, as seen in Figure 1.1.

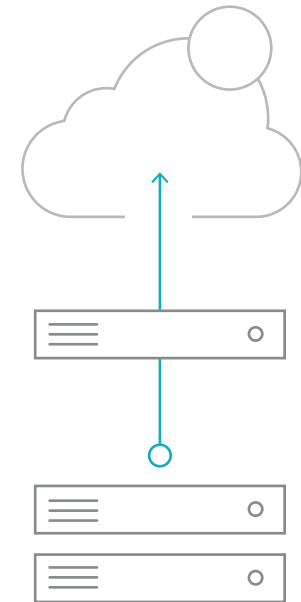
**Note:** The role of data management and integration as you move to the cloud. No longer are your data management and integration platforms limited only to on-premises locations; now they can also exist in the cloud. Your data management and data integration platforms should complement and simplify both on-premises and cloud computing environments, and integration between those environments.



New cloud data warehouse  
(born in the cloud)



Extend into  
cloud data warehouse



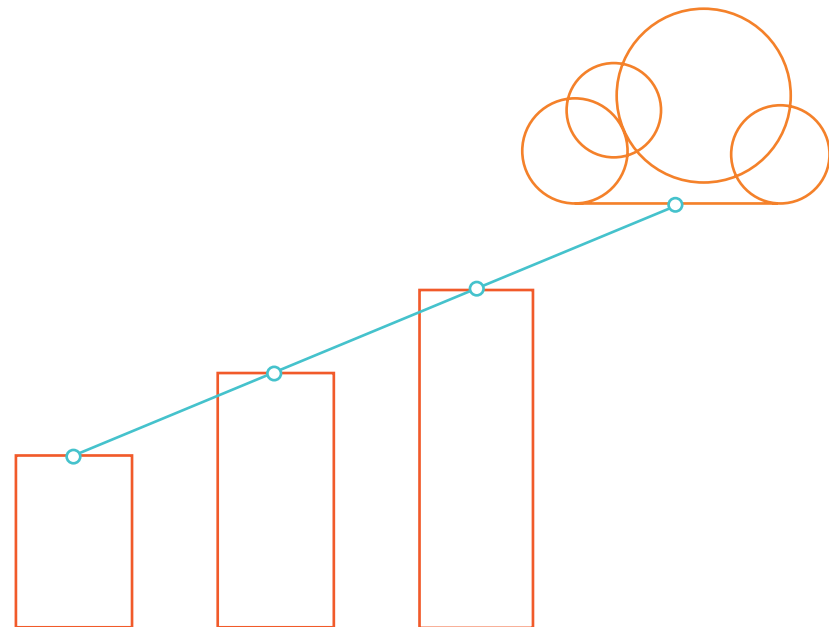
Migrate into  
cloud data warehouse



# 1. Building a New Cloud Data Warehouse on Microsoft Azure

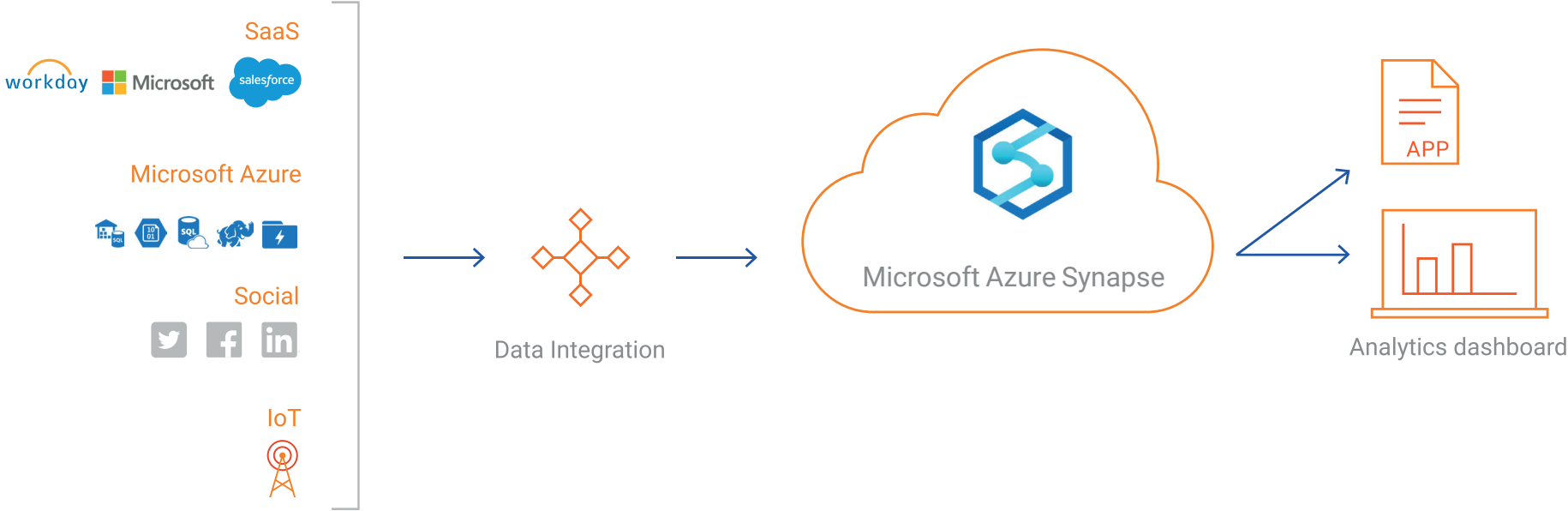
There are many reasons why you might create a net-new data warehouse in the cloud (“born in the cloud”) using Azure Synapse. For example, this could be appropriate when planning a new analytics initiative project for a business unit, or for an area that previously had not benefited from a strong analytics effort. A new Azure Synapse is highly complementary to an agile development effort, which is often part and parcel with new business ventures or initiatives.

Azure Synapse, when supported by a robust cloud data management platform, allows you to rapidly experiment, test, and evolve new ideas as they come in, without long development lead times and heavy upfront investment in systems and resources. As your new data warehouse project grows and data increases in volume and complexity, you can rapidly scale and enhance your analytics environment with the elasticity and performance capabilities inherent to Azure Synapse coupled with a robust data integration solution.



**About this architecture**

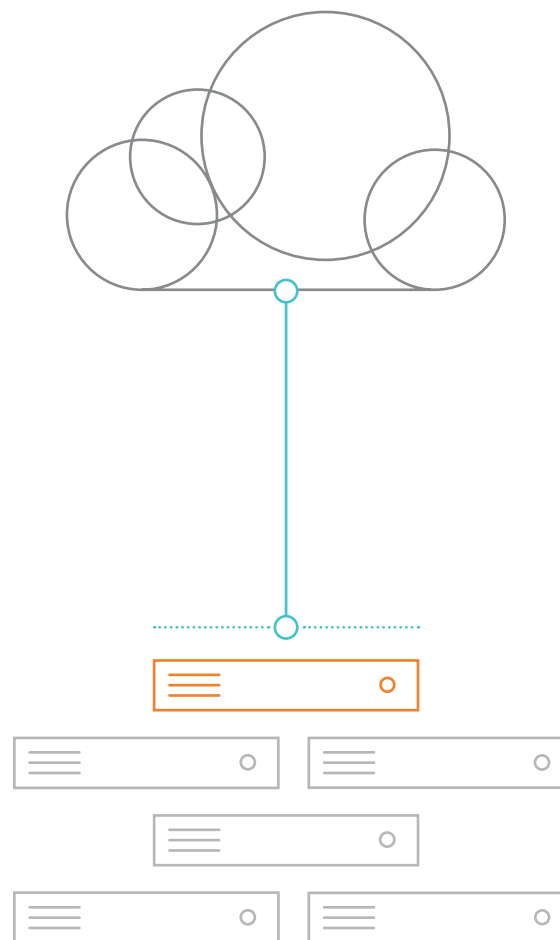
Building a new cloud data warehouse helps you reduce the time and money required to support low-risk experimentation and analytics projects through Microsoft Azure's elasticity.



## 2. Extend On-Premises Enterprise Data Warehouse with Azure Synapse

You may want to keep assets (data and applications) on-premises for a number of reasons, such as for regulatory compliance. But for agility, scalability, and flexibility, it makes sense to create a cloud data warehouse environment by extending your existing enterprise data warehouse into the cloud with Azure Synapse.

Extended cloud environments integrate on-premises resources with cloud-based computing. This is an easy way to start using the cloud without a major change to your current IT methodology. When you move newer workloads or those experiencing performance or reliability problems from on-premises to the cloud, you can take advantage of cloud-based applications and more easily integrate with new data sources.



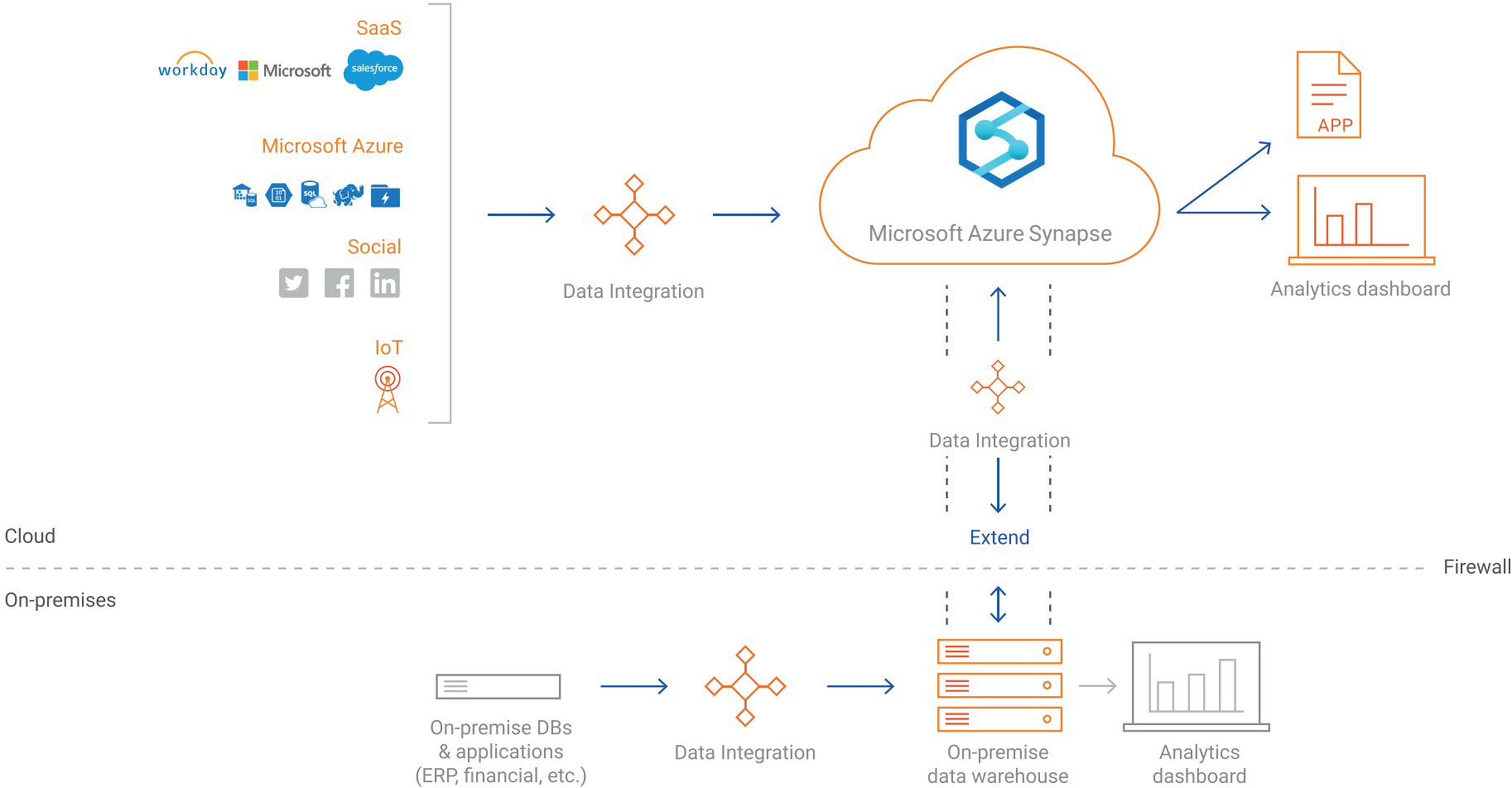
To create an extended cloud data warehousing environment, you create a new cloud data warehouse with Azure Synapse and extend your existing on-premises enterprise data warehouse architecture into it. You can then leverage either cloud-based or on-premises data integration solutions to move new data to the cloud and leave less frequently used data on-premises (or data that must remain on-premises). Having the flexibility to recommend your data integration deployment where most of your data—and future data—is expecting to live allows you to easily move between cloud and on-premises data warehouses. This approach solves many challenges for organizations that want the benefits of cloud computing but need to move at a slower or more controlled pace.

Extended environments also provide additional processing power and storage capacity to make your IT operations faster and cheaper via Azure than you could achieve by upgrading on-premises hardware, networking, and storage.

Microsoft Azure provides unlimited capacity and performance during critical spikes in processing when it is needed, providing elasticity and cost benefits.

### About this architecture

Lower costs and improve agility by extending/offloading undifferentiated heavy lifting to Microsoft Azure, while maintaining an on-premises data warehouse footprint.



# 3. Migrate On-Premises Data Warehouse to Azure Synapse

Some organizations modernize their traditional on-premises enterprise data warehouse by moving it into Azure Synapse. For many organizations increased costs, complexity, and the crushing overhead of maintaining on-premises infrastructures is driving the move to cloud. For those organizations, cloud data warehousing is the solution to many of their most pressing challenges.

In this scenario, businesses migrate their on-premises data warehouse into a cloud-only environment. This approach results in an entirely cloud-based data warehouse where your on-premises enterprise data warehouse is ultimately decommissioned. This allows you to eliminate data center maintenance and support costs and immediately take advantage of new cloud data warehouse features. Your data is now in the cloud, so it readily integrates with other cloud-based applications and services.

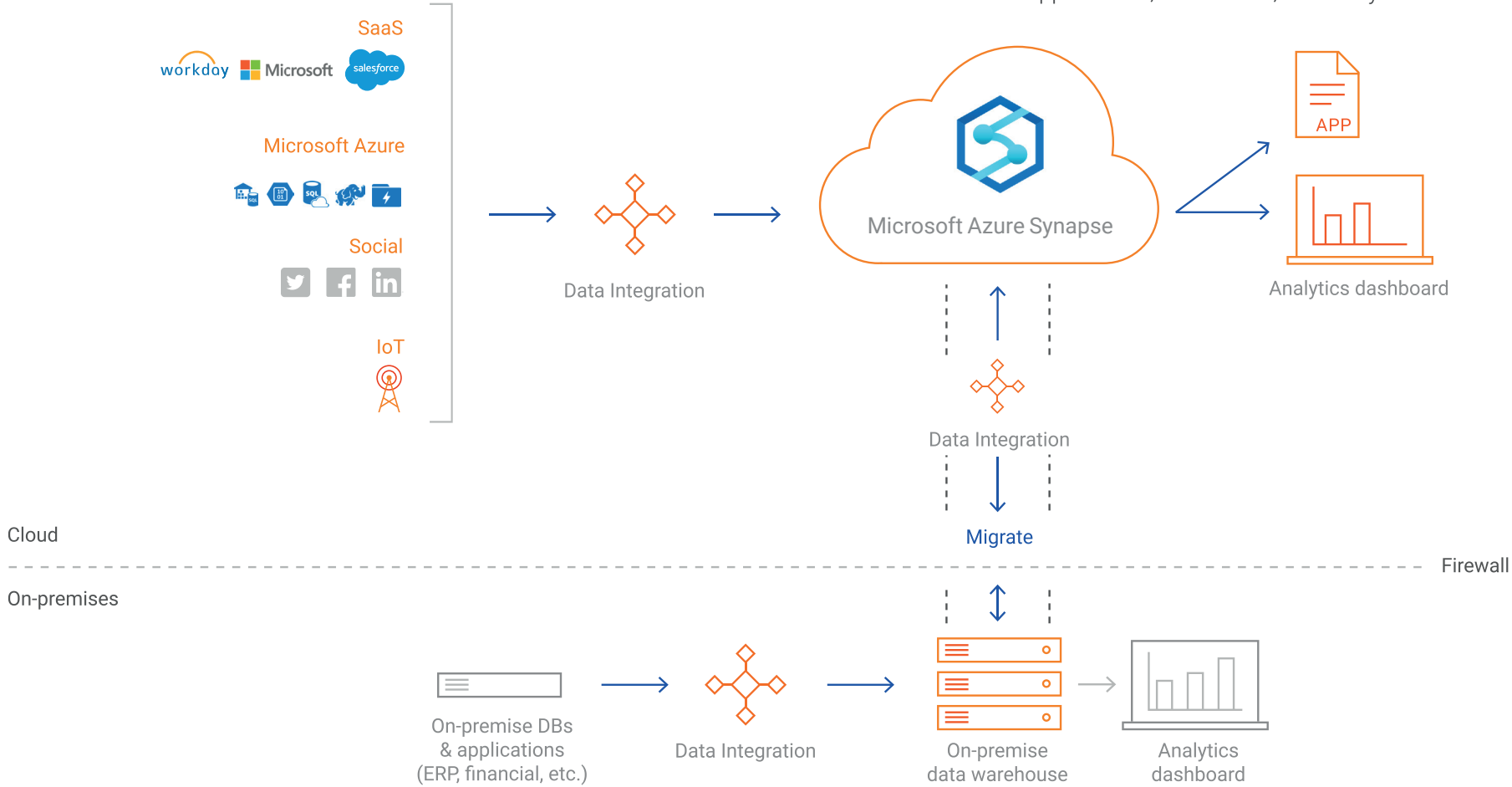
Benefits of this approach include lower upfront costs and the “pay-as-you-go” model of cloud computing. Just as important, this architecture provides elasticity so you can scale your analytics environment as your data sources increase and data volumes expand.

## Integration is Everywhere

The move to cloud is a journey, and even though you have created a new cloud data warehouse, you will likely still have on-premises assets in the form of databases, applications, and infrastructure; this is your extended cloud computing environment. Data integration between your on-premises data sources and your cloud data warehouse must still occur, as must data integration between your Azure Synapse and other cloud-based data sources. A comprehensive data management platform supporting both on-premises and cloud-based data integration plays a critical role in supporting each use case with a single set of skills, and tools.

### About this architecture

Migrating to a cloud data warehouse embraces the full benefits of the cloud and maximizes integration with modern cloud applications, data stores, and analytics.



Part 2

# New Paradigms Powering the Shift to Cloud Data





# Exploring Cloud Data Management Drivers

## Many businesses migrate to cloud data warehouses to help them take advantage of generational market disruption driven by data. Here are some cloud data management drivers and their challenges:

- Increased data sources and volume: Businesses amass a great deal of data from many different sources and in many different formats. This gives businesses a great opportunity to discover new and transformative insights—fast. However, it can be expensive and challenging to maintain the specialist skills required to produce custom applications that smoothly integrate when APIs continually change. Monitoring and managing data in production is key to catching and correcting problems early before they escalate out of control. The increased variety of data sources brings a new set of complexities to this task.
- Increased data silos: The historical growth of on-premises applications to solve key business needs has resulted in today's data silos. But data silos can also exist in cloud environments through multiple SaaS apps. An enterprise integration strategy that supports on-premises and cloud helps you deliver trusted and relevant data for transformative insights.
- Data security continues to challenge. As organizations evolve toward an on-premises and cloud data management model, they need to pay extra attention to securing sensitive data at rest and in motion, whether they are moving data from on-premises to cloud or from cloud to cloud. Consider a cloud data management solution that delivers data security intelligence so you can understand sensitive data risks and vulnerabilities.
- Data migration planning is critical: Migrating workloads to Azure can be a large and complex undertaking and often involve the migration of large volumes of data. This is particularly true if you're moving many different data types, from standard relational structures and predictable batches to huge volumes, massive variety, and real-time streaming data. Having the right plan can make a huge difference in this process.
- Market disruption requires IT to respond to business needs faster: Business decision makers have more new data types to analyze and their analysis needs can evolve rapidly. Often this means IT must develop new processes—quickly—to support this. When new data sources come online, IT must quickly find a way to integrate this data even if it is a new concept for the organization.

- **Context is crucial:** Business analysts can get valuable insights by knowing their data’s “life story”, enriching many analytics initiatives. Data visibility and lineage allows businesses to discover and understand the history of the data: where it came from, the path it took, who uses it, and so on. To do this, you need a unified metadata view that looks at technical metadata, business context, user annotations, relationships, data quality and usage. Data governance—the management and oversight of data by aligning people, process, policy, and technology—plays a critical role in helping you to deliver data that’s complete, contextual, and trustworthy for analytics. A cloud data management platform provides data governance so you can deliver trusted, secure, and governed data across cloud and on-premises.

- **The need for high quality data remains:** “Garbage-in, garbage-out” has never been truer. Data assurance, quality, and integration are critical factors in IT and especially analytics. Ensuring trusted, high quality data is an intensive, complex requirement, magnified by the volume, velocity, and variety of the modern data landscape.

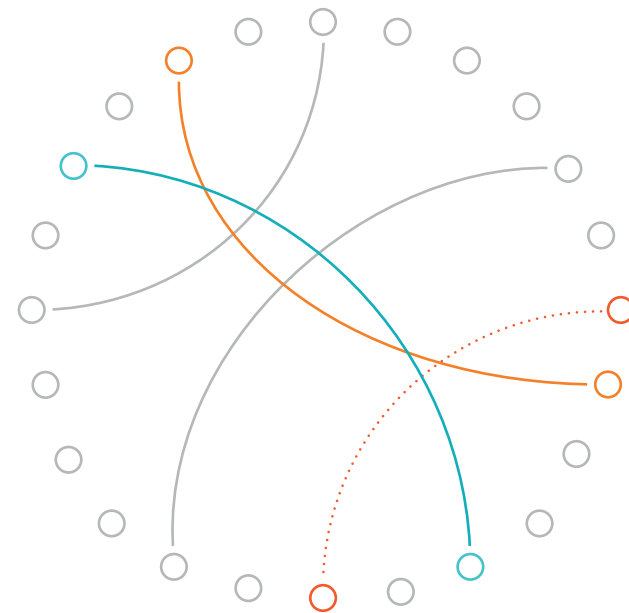
All this drives the need for a modern, agile approach to data warehousing that’s based on cloud computing and supported by a cloud data management strategy that’s optimized for a cloud-first or combined on-premises and cloud environment.

Part 3

# Data Management Best Practices for Azure Synapse

# Accelerate Data Migration with Automated Data Management

To fully seize the benefits of an Azure data architecture you need the backing of a proven data management and integration solution. You'll require a high performance, scalable data integration and management solution that provides native automated connectivity to key Azure data services such as Azure Synapse, Azure Blob Storage, Azure HDInsight, Azure Data Lake Store, Azure SQL Database and Azure CosmosDB, to accelerate your analytics and other data-intensive initiatives. These solutions simplify access to the many data sources accessible via Azure and on-premises data centers, and then automate their access.



**Successful cloud data warehouse projects share common data management characteristics. Here's a list of imperatives and best practices to guide your success:**

– **Robust connectivity architecture:**

A data integration architecture that separates data integration logic from the underlying data source APIs delivers a higher degree of code reuse and investment protection. You can simplify and accelerate your data integration development (separation of concerns), and can swap sources and targets in and out. Best of all, you don't need a lot of domain knowledge to understand the underlying APIs. This is especially important in the cloud, where SaaS applications and data services rapidly change and evolve their APIs.

- **Catalog for Data visibility:** The creation of metadata to support change management is critical. A data integration solution that automatically captures and catalogs technical, business, operations and usage

metadata speeds up search, understanding and onboarding of new data. Helping you reduce both the risk of change and the time lost by managing the change across upstream and downstream systems.

- **Operational confidence:** You can catch production problems in your data early by monitoring your data integration processes to ensure they are delivering timely data as expected. Having complete visibility into your entire enterprise data production environment is a powerful asset to ensure the right data is delivered in a timely fashion to your data consumers.
- **Security:** It's important to build the right level of security into the architecture at every level, both within technology and at the process level. Sensitive data should be protected both at rest and in motion via encryption and access controls. Regular patch maintenance and penetration testing keep you ahead of your competitors. Be wary of any data management platform that

hasn't achieved certification in the following AICPA SOC 2 and 3, HIPPA, HITECH, Privacy Shield, ISO/IEC 27001:2013 among others. Security should not be bolted-on after the fact; it must be integrated throughout the environment.

- **Agility:** The definition of an agile IT environment is the ability to meet changing requirements quickly and without excessive retooling or cost. By leveraging platforms and architectures that enhance the productivity of all your analytics stakeholders—both technical and citizen integrators—you can ensure that connected, trusted data can be rapidly delivered to support business initiatives.



- **Elasticity:** One of the great benefits of cloud environments is the ability to scale up and down to match the workload needs of the business. You can leverage this elastic capability offered by both cloud data warehousing and cloud data integration solutions to support variable data volumes and almost infinitely scale your solution as your data grows. Pipeline optimization, parallelism, clustering technology, Push Down Optimization (PDO) and partitioning are components of the entire stack, including the data warehouse and the data integration solution, all of which allows you to scale to meet workload requirements.
- **Performance:** Easy-to-scale performance is an important benefit of cloud data warehouses. To get the best performance, leverage database software and schema architecture optimized for data warehousing and analytic processing. Ensure underlying cloud infrastructure (compute nodes, cluster capabilities, storage, and networking) are modern and appropriate to the level of performance you require. Employ parallel processing, clustering, and PDO on modern multi-node columnar architectures and ensure that your data management solution automates and fully supports these types of data warehousing use cases.
- **Flexible Integration:** The data integration solution should connect all the data sources and applications you need regardless of their location on-premises or in the cloud. High quality, mature data integration solutions with robust capabilities in cloud as well as on-premises promote this flexibility, leading to increased business agility.
- **Standardization:** Within cloud data warehousing, the practice of standardization applies to multiple areas: data integration, governance procedures, application processes, code and skills reuse, and data preparation. Standardization is a prerequisite step to repeatability and automation, which enhances agility and time to market of new solutions.
- **Repeatability:** Processes and procedures should be well-defined, documented, and repeatable so they can be automated. Ad-hoc and one-off processes requiring manual intervention or monitoring return less value to the business. Moving and keeping your data updated is a multi-phase process. First, you have your initial data load (migration), but after that, ongoing data synchronizations need to occur to keep the target system(s) current. Implement solutions that support the automation of repeatable processes so your resources are freed up to work on other, higher impact projects, such as business analytics.

## Intelligent Data Management

Leveraging a metadata-driven, microservices-based architecture enables the continuous delivery and deployment of data. Providing visibility, improved productivity, reduced risk, and superior security and governance. You will achieve increased operational confidence by having a single point of control for end-to-end data flows. To maximize the full capabilities of Azure Synapse, ensure your data management platform is certified to run on Azure.

- IT and business collaboration: Teamwork, communication, and sound project management principles are often the difference between success and failure. If the collaboration between IT and business is insufficient, communication is weak, or expectations and goals are not managed, then even a great technical solution will struggle to be successful. Close collaboration between stakeholders is the key to success for any IT project, regardless of technology. Look for solutions that promote business and IT collaboration. For example, role-based tools allow different stakeholders to leverage their unique expertise (e.g. data knowledge vs. technology knowledge) and transparently “swap” work products back and forth as part of the data management work stream.
- Data governance: Implementing strong data governance procedures becomes even more important when you move data off-premises into a cloud environment. While cloud computing brings a new set of concerns for data governance, they are entirely manageable with careful planning, forethought, and with the right technology stack. Azure provides robust security and access controls in the cloud. Understanding your data, the impact of changes to source and target systems and having a holistic view of data governance is just as important as building out your data integration architecture.

### What is iPaaS?

Integration Platform as a Service (iPaaS) is a cloud service that gives IT a single platform from which to manage application, data, and process integration. It allows enterprises to rapidly execute any integration pattern, logically manage any data, and easily serve any user in need of integration.

In the end, having a robust data migration and data integration strategy is critical, especially if you are moving a lot of data, so select your data migration and integration solutions carefully. Diligently plan the initial data migration, utilizing data management tools to profile and understand all your on-premises data, but also consider on-going data integration, data quality and security requirements, maintenance, and accessibility.

Despite moving your data warehouse, in many situations you will still have an environment with on-premises and cloud-based data sources. This is where intelligent data integration solutions yield the greatest benefits; consider both on-premises data integration and cloud-based iPaaS (Integration Platform as a Service) solutions for your environment.

Also, consider performance—how long will it take to migrate the data, and once there, how long will it take your ETL tools to load data into the tables? Does the target software stack support parallel processing and deliver high performance? Make sure your data integration solution can capitalize on the strengths and features of Microsoft Azure Synapse.



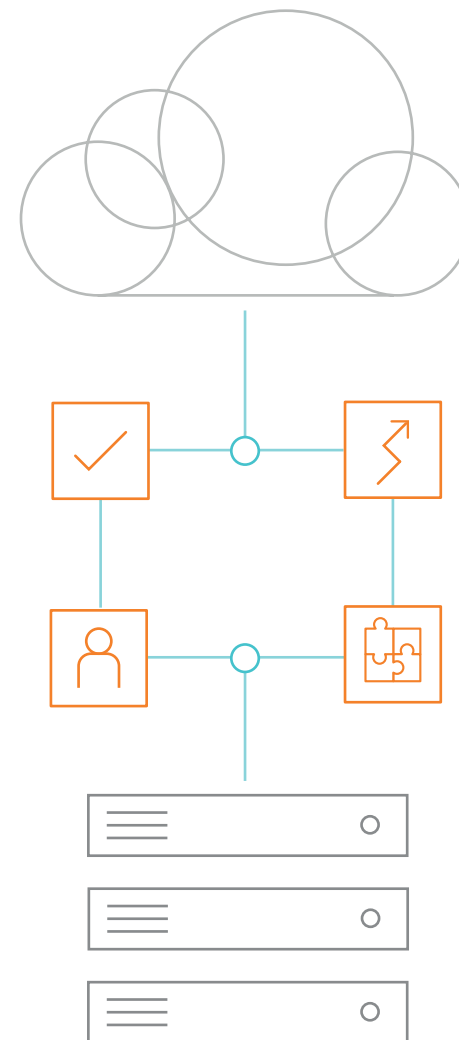
Part 4

# Cloud Data Warehouse

# Checklist for Cloud Data Warehousing Projects

In this section, we've designed a checklist to help you identify what you need to consider as you develop cloud data warehouses with Azure Synapse. Most of the questions are applicable to any use case (new cloud data warehouse, data warehouse extension to cloud, and data warehouse migration) although some topics (such as bursting) are more applicable to specific use cases.

The checklist isn't intended as a 1-2-3 stepwise install guide, but rather a way to identify and consider key points as you move forward. Check the boxes that apply to your specific case. There is no one size fits all approach. These questions are designed to help you identify gaps and areas where you can seek more information.



# Identifying Your Stakeholders

## Questions to ask yourself

- The team:** Have you identified the people and groups that will most directly benefit from this work? Who are your key stakeholders? Who will be indirectly impacted (positively or negatively)?
- The requirements:** Do you understand what stakeholders are looking for? Can you distinguish between what they absolutely need (their requirements) and what they want (their wish list)?
- The communication plan:** How will you communicate changes from your stakeholders, how often will you update them, and do they understand the balance of scope vs cost vs time?
- Self-service considerations:** What can you do to provide self-service for your different stakeholders based on their skillsets? For example, can you add self-service analytics and self-service data integration for your citizen integrators?
- Customer impact:** Have you considered customers as stakeholders? How could this benefit your company's "stakeholders"? And ultimately your company?

# Migrating and Integrating Your Data

It's helpful to understand the difference between data migration and data integration. Here are some definitions in the context of cloud data warehousing.

- **Data migration** is a one-time event to move data from a data source to a target (Azure Synapse in this case). Note that data is moved, not merely copied. Once complete, you decommission the original data source.
- **Data integration** is an ongoing process of combining data, often from multiple data sources, into a target data warehouse, commonly also transforming the data in novel ways, to support analytical needs. Comprehensive data integration capabilities can move data at any latency, ranging from scheduled batch and change data capture, to real-time, and streaming. Typical data integration architectures provide services spanning all of these latency requirements so a variety of service level agreements (SLAs) can be effectively met.

- **ETL (Extract, Transform, Load)** is used for high-performance transfer of huge volumes of data outside of Azure Synapse (Teradata, Netezza, Oracle) into that environment. Once you land the data into Azure Synapse staging tables, further data processing needs to be done to transform the data into facts and dimensions for analysis.
- **ELT (Extract, Load, Transform)** The data transformation between staging, intermediate, and analytics Azure Synapse tables can be accomplished in a highly performant way using SQL-based pushdown optimization (PDO - also known as Extract, Load, and Transform - ELT) wherein the transformation is converted to SQL predicates and pushed down to SQL queries.

Now, consider how your data will be loaded and how long that process will take. Moving away from manual data loads to automated methods makes your life easier so leverage data management solutions to simplify and automate these tasks.

**Tip:** Data migrations and data integration can be complex; especially if you rely on in-house, manual solutions. Leverage powerful, prebuilt solutions and methods, which enhance and automate tasks to reduce complexity while accelerating your project.

For more information on solutions to help you migrate and integrate data, visit <https://www.informatica.com/products.html>



# Migrating and Integrating Your Data

## Questions to ask yourself

### **Have you considered your data needs for production?**

- a) The data sources required to integrate into your data warehouse
- b) Your access to all these data sources
- c) Your understanding of the data from each data source
- d) The expertise you have in-house to connect to each source
- e) How long it takes to onboard each data source
- f) How you will maintain independent and reusable data integration code

as the underlying data sources and analytics requirements change

g) How you will evolve your data integration as business needs change

h) What data warehouse changes will be managed by developers and what can you enable your citizen integrators to do themselves

### **Do you know what data you are migrating and what are you leaving behind and/or archiving?**

Consider both your data sources and your data warehouse data. Can you migrate data from multiple systems so you decommission multiple data sources (databases)?

# Migrating and Integrating Your Data

## Questions to ask yourself

- Do you know what data profiling and cleansing procedures and tools you have in-house to prepare data prior to migration to ensure only trusted, high quality data is moved?**
  - a) What data management solutions you will use for this data load
  - b) Whether your migration will be manual or automated, and how long you think it will take
  - c) Whether this time will impact your overall project and if so, what you can do to address that
  - d) What your data management Service Level Agreements (SLAs) are, and whether you have a need for real time data updates to support business requirements
  - e) Whether you plan to use an Operational Data Store (ODS) technology to gather the data prior to loading into the data warehouse
  - f) How complex the process of joining and merging the data would be, and what tools you plan to use
- Do you have data cataloging tools and expertise to help you to understand your data and the associated metadata to determine what to migrate?**
- Have you considered each of these data migration factors?**

# Migrating and Integrating Your Data

## Questions to ask yourself

### **Have you planned your data integration?**

a) How you are addressing data integration

b) Whether your integration will be manual or automated, and what the time and complexity involved is. Consider the resources you have today as well in the future

c) How this impacts your agility and what you can do to optimize your processes

d) How you will scale out your data integration projects as the complexity of your projects scales

e) Whether some of your data sources will be external to your organization

f) Whether your data management will need to work across both cloud and on-premises data sources and targets

g) What your expected quality of data is, and whether you need tools to manage and monitor data quality in the new system

# Getting the Most Out of Azure Synapse

Azure Synapse is a powerful, public cloud data warehouse that is well suited to support modern cloud analytics in the worlds of big data and IoT. Several features ensure Azure Synapse provides the agility, speed, performance, and security that businesses need to deliver intelligent great data for business decision makers, while maintaining cost efficiencies. These include massive parallel processing (processing spread across multiple nodes), elastic scalable processing nodes for high performance, columnar storage for fast return of large warehouse-size data sets, built-in fault tolerance, and data encryption in transit for security, among others.

For more information about Azure Synapse, visit <https://azure.microsoft.com/en-us/services/synapse-analytics/>

**Tip:** Understanding key components of Azure and how Azure Synapse works will only make you more effective as you pursue your data warehousing projects. You don't need to be a technical expert, but a reasonable time investment in learning core fundamentals will yield dividends as you move forward.



# Getting the Most Out of Azure Synapse

## Questions to ask yourself

- Do you know how Massively Parallel Processing (MPP) of Azure Synapse aids your initial data load and reduces your load time?**
- Have you considered how you can leverage elastic cloud computing resources and modern infrastructure to fully enable and compliment your analytic processing activities?**
- Does your data management solution easily connect to all your data sources within Azure, the cloud, and on-premises?** Are these connections out-of-the-box, easy to maintain, and high performance or are they manual and cumbersome?
- Is your data management platform well-integrated into Azure?** Is your solution metadata-driven using role-based tools? Does it provide data visibility and a single point of control for end-to-end production data-flows?

# Bursting to Meet Temporary Workload Surges

Azure Synapse is built on a robust infrastructure that allows for rapid addition of storage, memory, and processor compute power in times of need. When no longer needed, these resources are released back into the cloud to be used by another customer. This scaling up and down of resources is the definition of elasticity and is a major benefit of cloud architecture. A unified data integration platform simplifies your bursting of data into the cloud. Additional compute and storage resources are added to the cloud environment as needed to meet the processing demand.

## Reuse Your Integrations

When you use an automated tool that's supported by a robust connectivity architecture to load your on-premises data warehouse, you can easily reuse the integrations to integrate the same data to Azure Synapse. Because the integration logic is separated from the underlying sources or targets, there is minimal change needed to adapt the code (or data mappings) to a new data warehouse.

**Tip:** Don't simply assume your workload will automatically take advantage of cloud bursting. Do your homework and test to ensure the bursting workload is processed as expected.

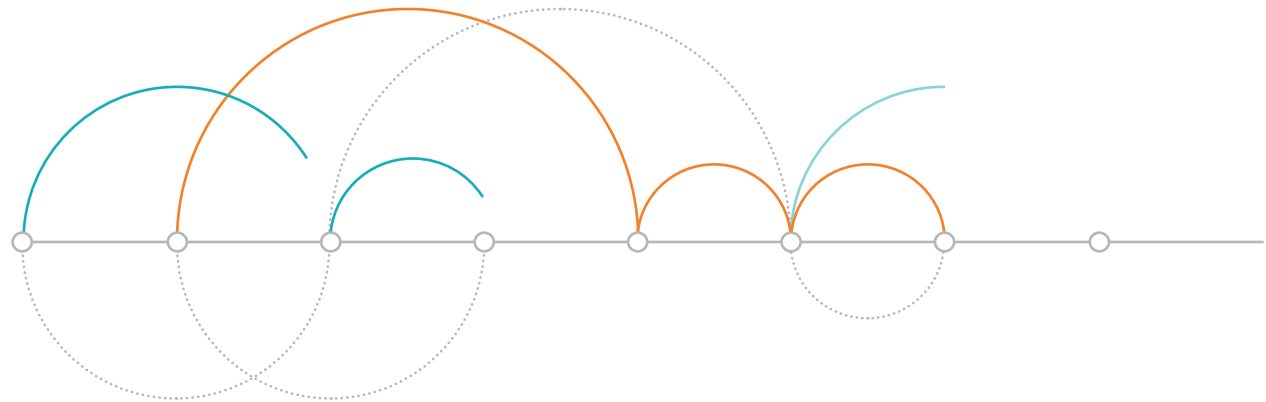
# Bursting to Meet Temporary Workload Surges

## Questions to ask yourself

- How are you planning for workload surge events within your organization?** Are they predictable and well planned or are they unpredictable and disruptive?
- Have you considered the cost of not reacting to a surge workload in terms of lost opportunity or failure to meet a regulatory requirement?** If your on-premises resources are sized for peak workload, do you know how much capacity is unused during normal operations?
- Have you considered how you will manage the data integration aspect of cloud bursting?** Can you seamlessly move the same data integration processes from your on-premises data warehouse to the cloud data warehouse with minimal effort?
- Since Azure does not require you to define memory and processing, the question to ask yourself is how much storage is required?**

# Reducing Your Risks and Promoting Success

Leveraging Azure Synapse for cloud data warehousing to support analytics yields many benefits and, with the right data integration solutions, is fast and relatively easy. That said, there are some tips that improve your odds of success. At the core of cloud data warehousing is the data, consider the following within that context.



# Reducing Your Risks and Promoting Success

## Questions to ask yourself

- How will you validate that all your data has been moved?** What prepared validation checks and tools do you have in place?
- Do your key stakeholders believe that all the data has been moved?** How do the stakeholders define success and have you managed their expectations?
- Have you given thought to the impact to the organization of moving its production data store?** Think in terms of technical, but also non-technical governance issues. Cloud computing is secure, but are there regulatory or contract restrictions on the placement of some data elements?
- Are you using the right tools and automation?** Are the data management and integration solutions working together as a complete solution or are they fractured and stove-piped?  
  
Consider what features in a data management platform benefit you the most (e.g. connectivity, ease-of-use, rapid integration, complete solution, scalability, etc).

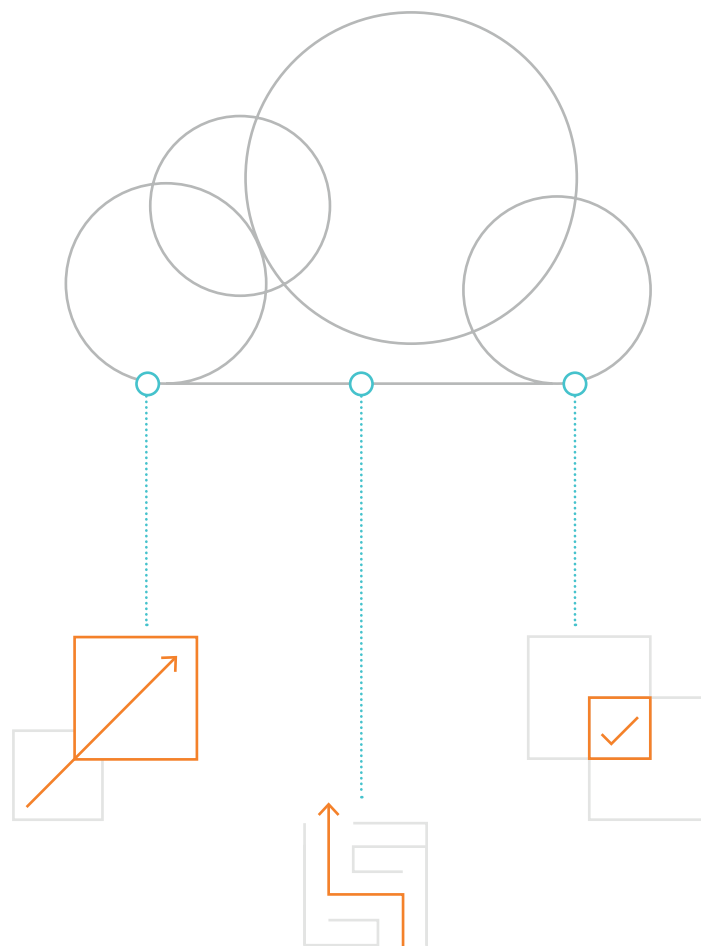
# Accelerate Data Warehouse Modernization to Azure

# Accelerate Data Warehouse Modernization to Azure with Informatica's AI-Driven Platform

**When you combine the scalability of public cloud data warehousing with cloud data management strategies you can increase your organization's IT agility, and accelerate time to insights for your business decision makers.**

Informatica, the leader in Enterprise Cloud Data Management, helps you to modernize your data architecture with intelligent data management, purpose-built for Microsoft Azure, that easily adapts and scales as your data types, volume, applications and architecture changes.

Informatica's AI-driven Intelligent Data Platform has a modular microservices architecture that accelerates your Azure Synapse project deployment by automating your data integration development lifecycle, including connectivity, development, deployment, and management.



# Simplify with Intelligent Cloud Data Management for Azure

Optimize cloud data warehousing with Informatica's Azure certified cloud data management solutions and purpose built Azure connectivity. Benefits of the Intelligent Data Platform include:

- Data connectivity eliminates silos: Optimize cloud data warehousing with out-of-the box connectivity to hundreds of on-premises, cloud data sources, including Azure Synapse, Azure Blob Storage, Azure SQL Database, Azure Data Lake Store, Azure CosmosDB, and Azure HDInsight.
- Development agility accelerates time to market: Role-based SaaS development tools, optimized for both developers and business users are accessible anywhere, any time.
- Holistic view of your enterprise data landscape expedites migration: Machine learning-based data asset discovery, visibility and preparation solutions help you transform big data into fit-for-purpose data sets.
- Ensure deployment scalability and operational confidence: Support for both ETL and ELT patterns for end-to-end high-performance data pipelines.
- Unleash big data insight: Ingest, prepare, catalog, master, govern, and protect your big data to deliver successful data lakes.
- Deliver high-quality clean, trusted and secure data: Scalable enterprise-class data quality, data security and governance solutions that can address any size and format of data; platform; or technology.

The Informatica Intelligent Data Platform offers comprehensive solutions for Azure, including data integration, data discovery and preparation, data lakes and big data management, master data management, data quality and data security to accelerate Azure deployment and deliver trusted data in the cloud and on-premises.

Learn more at: [www.informatica.com/azure](http://www.informatica.com/azure) or Contact us [through this form](#)



# About Informatica

Digital transformation changes expectations: better service, faster delivery, with less cost. Businesses must transform to stay relevant and data holds the answers.

As the world's leader in Enterprise Cloud Data Management, we're prepared to help you intelligently lead—in any sector, category or niche. Informatica provides you with the foresight to become more agile, realize new growth opportunities or create new inventions. With 100 percent focus on everything data, we offer the versatility needed to succeed.

We invite you to explore all that Informatica has to offer—and unleash the power of data to drive your next intelligent disruption.

## Worldwide Headquarters

2100 Seaport Blvd, Redwood City, CA 94063, USA

Phone: 650.385.5000

Fax: 650.385.5500

Toll-free in the US: 1.800.653.3871

[informatica.com](https://informatica.com)

[linkedin.com/company/informatica](https://linkedin.com/company/informatica)

[twitter.com/Informatica](https://twitter.com/Informatica)

[CONTACT US](#)

IN08-1220-03373

© Copyright Informatica LLC 2020. Informatica and the Informatica logo are trademarks or registered trademarks of Informatica LLC in the United States and other countries. A current list of Informatica trademarks is available on the web at <https://www.informatica.com/trademarks.html>. Other company and product names may be trade names or trademarks of their respective owners. The information in this documentation is subject to change without notice and provided "AS IS" without warranty of any kind, express or implied.



Informatica®

# About Microsoft Azure

For more than 40 years, Microsoft has been the world's most comprehensive and broadly adopted enterprise technology platform. Microsoft Azure offers over 100 fully featured services for compute, storage, databases, analytics, mobile, Internet of Things (IoT) and enterprise applications from 42 global regions and comply with more than 50 government and industry certifications. Azure services are trusted by more than a million active customers around the world – including the fastest growing startups, largest enterprises, and leading government agencies – to power their infrastructure, make them more agile, and lower costs.

To learn more about Azure, visit [www.azure.microsoft.com](http://www.azure.microsoft.com)