

LEARNING MADE EASY

3rd Snowflake Special Edition

Data Sharing

for
dummies[®]
A Wiley Brand



What is modern
enterprise data sharing?

How will data sharing
transform your business?

Why data sharing is your key
to the data economy

Brought to
you by:



David Baum

About Snowflake

Snowflake delivers the Data Cloud — a global network where thousands of organizations mobilize data with near-unlimited scale, concurrency, and performance. Inside the Data Cloud, organizations unite their siloed data, easily discover and securely share governed data, and execute diverse analytic workloads. Wherever data or users live, Snowflake delivers a single and seamless experience across multiple public clouds. Snowflake's platform is the engine that powers and provides access to the Data Cloud, creating a solution for data warehousing, data lakes, data engineering, data science, data application development, and data sharing. Join Snowflake customers, partners, and data providers already taking their businesses to new frontiers in the Data Cloud. [snowflake.com](https://www.snowflake.com).



Data Sharing

3rd Snowflake Special Edition

by David Baum

for
dummies[®]
A Wiley Brand

Data Sharing For Dummies®, 3rd Snowflake Special Edition

Published by: **John Wiley & Sons, Inc.**, 111 River St., Hoboken, NJ 07030-5774, www.wiley.com

Copyright © 2022 by John Wiley & Sons, Inc., Hoboken, New Jersey

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without the prior written permission of the Publisher. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at <http://www.wiley.com/go/permissions>.

Trademarks: Wiley, For Dummies, the Dummies Man logo, The Dummies Way, Dummies.com, Making Everything Easier, and related trade dress are trademarks or registered trademarks of John Wiley & Sons, Inc. and/or its affiliates in the United States and other countries, and may not be used without written permission. Snowflake and the Snowflake logo are trademarks or registered trademarks of Snowflake Inc. All other trademarks are the property of their respective owners. John Wiley & Sons, Inc., is not associated with any product or vendor mentioned in this book.

LIMIT OF LIABILITY/DISCLAIMER OF WARRANTY: WHILE THE PUBLISHER AND AUTHORS HAVE USED THEIR BEST EFFORTS IN PREPARING THIS WORK, THEY MAKE NO REPRESENTATIONS OR WARRANTIES WITH RESPECT TO THE ACCURACY OR COMPLETENESS OF THE CONTENTS OF THIS WORK AND SPECIFICALLY DISCLAIM ALL WARRANTIES, INCLUDING WITHOUT LIMITATION ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. NO WARRANTY MAY BE CREATED OR EXTENDED BY SALES REPRESENTATIVES, WRITTEN SALES MATERIALS OR PROMOTIONAL STATEMENTS FOR THIS WORK. THE FACT THAT AN ORGANIZATION, WEBSITE, OR PRODUCT IS REFERRED TO IN THIS WORK AS A CITATION AND/OR POTENTIAL SOURCE OF FURTHER INFORMATION DOES NOT MEAN THAT THE PUBLISHER AND AUTHORS ENDORSE THE INFORMATION OR SERVICES THE ORGANIZATION, WEBSITE, OR PRODUCT MAY PROVIDE OR RECOMMENDATIONS IT MAY MAKE. THIS WORK IS SOLD WITH THE UNDERSTANDING THAT THE PUBLISHER IS NOT ENGAGED IN RENDERING PROFESSIONAL SERVICES. THE ADVICE AND STRATEGIES CONTAINED HEREIN MAY NOT BE SUITABLE FOR YOUR SITUATION. YOU SHOULD CONSULT WITH A SPECIALIST WHERE APPROPRIATE. FURTHER, READERS SHOULD BE AWARE THAT WEBSITES LISTED IN THIS WORK MAY HAVE CHANGED OR DISAPPEARED BETWEEN WHEN THIS WORK WAS WRITTEN AND WHEN IT IS READ. NEITHER THE PUBLISHER NOR AUTHORS SHALL BE LIABLE FOR ANY LOSS OF PROFIT OR ANY OTHER COMMERCIAL DAMAGES, INCLUDING BUT NOT LIMITED TO SPECIAL, INCIDENTAL, CONSEQUENTIAL, OR OTHER DAMAGES.

For general information on our other products and services, or how to create a custom *For Dummies* book for your business or organization, please contact our Business Development Department in the U.S. at 877-409-4177, contact info@dummies.biz, or visit www.wiley.com/go/custompub. For information about licensing the *For Dummies* brand for products or services, contact BrandedRights&Licenses@Wiley.com.

ISBN 978-1-119- 85620-7 (pbk); ISBN 978-1-119- 85621-4 (ebk)

Publisher's Acknowledgments

Some of the people who helped bring this book to market include the following:

Development Editor: Brian Walls

Project Manager: Jen Bingham

Acquisitions Editor: Ashley Coffey

Editorial Manager: Rev Mengle

Business Development

Representative: Molly Daugherty

Content Refinement specialist:
Mohammed Zafar

Snowflake Contributors Team:

Emily Dillon, Vincent Morello,
Jennifer Belissent, Eric Dorf,
Shreyas Desai, Clarke Patterson,
Leslie Steere

Table of Contents

INTRODUCTION	1
About This Book	2
Icons Used in This Book.....	2
Beyond the Book.....	2
CHAPTER 1: Getting Up to Speed on Data Sharing Basics.....	3
Sizing Up the Data Landscape	3
Understanding Basic Data Sharing Concepts	4
Reviewing popular data sharing methods.....	5
Examining modern data sharing alternatives	5
Sharing data via a cloud data platform.....	7
Data Sharing Use Cases.....	8
Exploring Data Sharing Possibilities.....	9
Developing a 360-degree view of the customer	9
Unlocking unlimited data for data science	10
CHAPTER 2: Understanding Traditional Data Sharing Challenges.....	11
Addressing a Multifaceted Problem.....	11
Contending with Time-to-Value Delays	12
Examining the Impacts of Traditional Data Sharing	14
Eyeing the Future	16
CHAPTER 3: Recognizing the Business Value of Sharing Data.....	17
Looking Back at the Early Days of Data Sharing.....	17
Breaking Down Data Silos with Modern Data Sharing	18
Understanding Data Sharing Relationships.....	19
Sharing data across lines of business	19
Sharing data with external organizations	22
Monetizing Data	23
Controlling Access with Data Governance	24
Building a Foundation for Modern Data Sharing	26
CHAPTER 4: Sharing Data across Industries.....	27
Yielding Greater Value in Financial Services	27
Delivering Better Healthcare Outcomes.....	31
Powering the Retail Supply Chain	33

	Delivering Superior Media and Entertainment Services	36
	Exploring data sharing techniques	36
	Responding to cookie regulations	37
	Delivering Fresh Insights to Technology Companies	39
CHAPTER 5:	Tapping into Data Marketplaces	41
	Examining Traditional Data Marketplaces	41
	Embracing Modern Data Marketplaces	42
	Applying Modern Data Sharing Technology	43
	Opening new markets for data providers	43
	Unlocking new insights for data consumers	46
CHAPTER 6:	Monetizing Your Data	49
	Examining New Revenue Opportunities across Industries	50
	Evaluating Distribution Channels	51
	Pricing Your Data Sharing Service	53
	Optimizing Your Data Listings	54
	Evaluating Data Marketplaces	55
CHAPTER 7:	Enabling Live Data Sharing with a Modern Architecture	57
	Capitalizing on a Modern Architecture	57
	Decoupling storage, compute, and services	58
	Boosting concurrency	60
	Enabling sharing cross-cloud and cross-region	61
	Increasing Data Security	63
	Sharing business logic	63
	Enabling fine-grained access	64
CHAPTER 8:	Governing Your Data	65
	Knowing Your Data	65
	Controlling Your Data	66
	Governing how consumers interact with your data	67
	Protecting your data	68
	Classifying and identifying your data	68
	Applying Data Governance to Common Business Scenarios	69
	Sharing the same data with one or multiple consumers	69
	Sharing different subsets of data with multiple consumers	70
	Sharing predefined slices of data with secure views	71
	Unifying Your Data	71
CHAPTER 9:	Six Steps to Advance Your Business with Modern Data Sharing	73

Introduction

Your organization's most valuable asset, data, is siloed across dozens or even hundreds of software-as-a-service (SaaS) and on-premises applications. To make matters worse, the data you do manage to extract is copied over and over and placed in data lakes, data warehouses, data marts, and even spreadsheets — and strewn across departments, business units, and subsidiaries, creating a data governance nightmare. The result? Few insights are derived from the data sets you can corral, and decisions are made based on gut, not fact.

Then it gets worse. Some of the most useful data actually exists outside your organization. It originates from your business partners and from commercial data providers, and it's no more accessible or useful than the data scattered across your business. However, if it's combined with your internal data, you can understand and predict your customers, your company, and the markets in which you conduct business in ways you never imagined previously.

Until recently, only the largest companies with big data analytic budgets could somewhat identify, access, integrate, analyze, and share data. But the need for any organization to share data continues to intensify. In 2020, more than 70 percent of global data and analytics decision-makers were expanding their capability to use external data, and another 17 percent planned to do so in 2021, according to the Forrester report “Business Technographics Data and Analytics Survey.”

However, many organizations still use outdated methods to discover, access, and share data. The data that is shared is stale the moment it is copied and transmitted via multiple steps to its intended destination. But modern data sharing now makes it possible for any organization to easily access and share live data.

This book describes the rapidly evolving world of data sharing across your business and beyond, and the many opportunities it presents. You can unify internal data siloed across your organization. As a *data consumer*, you can take advantage of a near-limitless set of opportunities to leverage external data and combine it with your existing data sources. As a *data provider*, you can easily and securely monetize your data and create self-service relationships between your organization and an endless number of data consumers.

About This Book

Welcome to *Data Sharing For Dummies, 3rd Snowflake Special Edition*, where you can explore how modern data sharing enables any organization to share and access live, ready-to-query data in a governed and secure way — with almost none of the cost, headache, and delay that have plagued traditional data sharing methods. Modern data sharing allows an organization to easily and quickly forge one-to-one, one-to-many, and many-to-many relationships to share data in new and imaginative ways and massively reduce time to insight.

Icons Used in This Book

In this book, you'll occasionally see special icons calling attention to important information:



CASE STUDY

The case studies provide best practices from organizations that have successfully used modern data sharing methods.



REMEMBER

This icon points out information you should commit to your non-volatile memory — your gray matter.



TECHNICAL
STUFF

This icon explains the jargon beneath the jargon.



TIP

This icon points out useful nuggets of information and helpful advice.

Beyond the Book

If you like what you read in this book, visit www.snowflake.com for a free trial of Snowflake's Data Cloud or to obtain details about plans and pricing, view webinars, access detailed documentation, or get in touch with a member of the Snowflake team.

IN THIS CHAPTER

- » Defining data sharing
- » Recognizing the importance of data sharing
- » Reviewing popular data sharing methods
- » Understanding how organizations share data
- » Exploring data sharing opportunities

Chapter 1

Getting Up to Speed on Data Sharing Basics

This chapter introduces data sharing — what it is, why it matters, why organizations share data, and the tremendous advantages awaiting organizations that seize today's data sharing opportunities.

Sizing Up the Data Landscape

For many organizations, yesterday's data landscape still exists today and is populated primarily by data locked in silos. But in the age of modern data sharing, data silos are being broken down into a centralized single source of truth across and beyond the enterprise. Data comes from many sources: enterprise applications; many types of semi-structured sources, including the clicks and browsing activities of millions of website visitors; Internet of Things (IoT) data from beacons and sensors; and unstructured data, such as audio, videos, documents, and images.

Understanding Basic Data Sharing Concepts

Modern data sharing is the act of providing live access to governed data between business units inside the same organization or among organizations external to each other. The business unit or external organization that shares its data is a *data provider*. The organizations that want to use shared data are *data consumers*. Any organization can be a data provider, a data consumer, or both.

Some data providers share data. Others share *data services* that put that data to work. For example, an organization might supplement its internal customer data with third-party data to better understand the age and income of groups that have purchased from its website. The same organization might subscribe to a data service that cross-references online purchase behavior with additional third-party demographic data, enabling a more personalized understanding of each customer group or segment.

Endless data-creating scenarios exist in the modern world, generating an immense quantity of data, including an increasingly rich set of internal and external data sources. Many organizations such as yours acquire data sets from suppliers, other business partners, and *data marketplaces* — online transactional locations that facilitate the purchase and sale of data and data services. And it's now easy to access data from these second- and third-party sources and properly and securely share your own data across your organization and with external constituents.

SUPPLEMENTING INTERNAL DATA

Broadly speaking, three types of data form the foundation for business operations and analytics within most enterprises:

- First-party data is data you collect directly — for instance, from interactions with customers and prospects.
- Second-party data is produced by or in collaboration with trusted partners, such as data from a software-as-a-service (SaaS) vendor.
- Third-party data is data acquired from external sources that don't have a relationship with your organization. Common examples include demographic, weather, and financial market data.

Reviewing popular data sharing methods

Traditional data sharing methods are riddled with problems, making discovering, capturing, and using all this data a challenge. In many cases, these traditional approaches are costly, create manual overhead, and limit how much data an organization can share. Popular methods of sharing and transferring data include:

- » **Email:** A data file is emailed from a provider to a consumer.
- » **File transfer:** Data files are shared and downloaded between two computers or via the Internet through File Transfer Protocol (FTP).
- » **Application programming interfaces (APIs):** A proprietary API is used to initiate and manage the data transfer.
- » **Extract, transform, load (ETL) software:** ETL software extracts data from the provider's database, transforms it into a format suitable for consumption, and then loads it into the consumer's database.
- » **Cloud storage:** The provider stores a copy of the data and provides the consumer with credentials for accessing it.

Figure 1-1 shows how organizations have traditionally shared data — by copying and sending the shared data to their data consumers. The data consumers then download the data to analyze or combine that data with their existing data for deeper insights into who their customers are, how efficiently their business operates, and what new market opportunities they can reveal. Unfortunately, these traditional data sharing methods are slow, cumbersome, costly, and usually allow for moving only limited amounts of stale data.

Examining modern data sharing alternatives

Figure 1-2 shows a modern data sharing scenario in which a data provider makes available live, ready-to-query data to its data consumers via modern cloud data sharing. The data can be

shared across data cloud providers and regions without using ETL or other traditional procedures and is updated automatically — decreasing management overhead for both the data provider and data consumer. When sharing live read-only data, a data consumer can easily access and integrate the shared data set without changing the data provider’s original version. When the provider updates the data set, the data consumer’s read-only version is updated almost simultaneously.

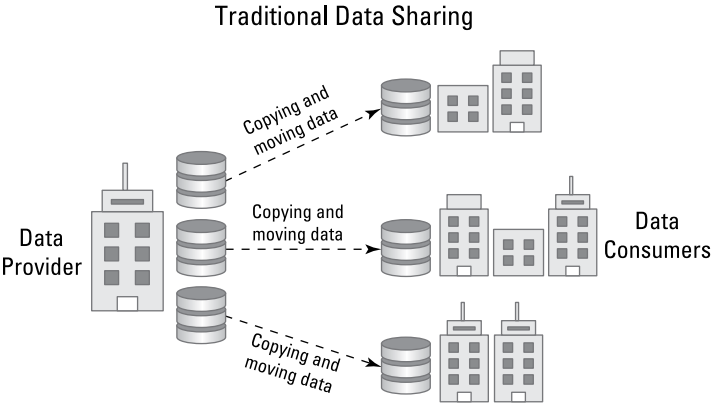


FIGURE 1-1: Traditional data sharing requires copying and moving stale data across environments from a data provider to data consumers.

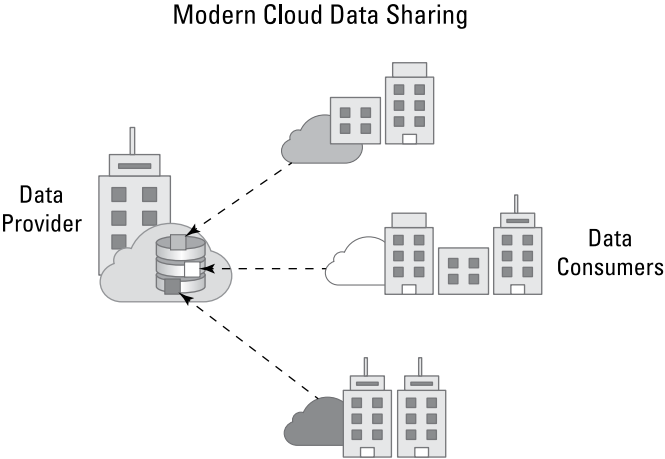


FIGURE 1-2: Modern data sharing enables live, governed data to be shared across clouds and regions without needing to move files across environments or create unnecessary copies.

With modern data sharing technology, a data provider can easily grant governed access to the data it wants to share with its intended consumers without managing cumbersome data pipelines. End-to-end security, multiparty governance, and metadata management services are systematically applied, even when the data consumers span multiple clouds. With updates made automatically, you don't have to link applications, set up file-sharing procedures, or frequently upload new data to keep data current. Chapters 5 and 7 provide specific details about these capabilities.

Sharing data via a cloud data platform

Organizations, such as Kraft Heinz, streamline data sharing activities with a *cloud data platform*, a specialized cloud service optimized for storing, analyzing, and sharing (creating access) to large and diverse volumes of data for many types of analytic workloads. A cloud data platform that offers modern data sharing simplifies setting up and revoking data sharing relationships and enabling read-only access to the data via SQL.



CASE STUDY

A GLOBAL DATA SHARING PLATFORM

When you think about the Kraft Heinz Company, you probably think of iconic brands, such as Heinz Ketchup and Kraft Macaroni & Cheese. However, Kraft Heinz is also a leader in digital technology. As part of an extensive digital transformation initiative, the consumer packaged goods company decommissioned its on-premises Hadoop data warehouse and moved half a trillion records into a modern cloud data platform, running on Microsoft Azure. A unified data hub now drives its global operations, allowing the organization's worldwide user base to share data and consume data services easily from a rapidly growing data marketplace.

According to Mani Gopalakrishnan, vice president of Digital Transformation at Kraft Heinz, sharing data with the old Hadoop platform was tedious. Data engineers had to create custom data

(continued)

(continued)

pipelines, establish integration points within the source and destination databases, conduct a security review, write programmatic processes to load and refresh data at designated times, profile the data, and monitor the data-ingestion processes to ensure adequate performance. Now, first-, second-, and third-party data is easily brought into the cloud-based data hub, including weather data, benchmark data, and economic data for Kraft Heinz's growing data science team.

Gopalakrishnan believes these modern, cloud-based data sharing practices will become progressively more important as a growing number of Kraft Heinz businesses leverage the data hub to share information securely and with respect for data privacy. "Having information is powerful, but sharing information can be even more powerful," he adds. "With a data marketplace, shared data is just a click and a credit card away."

Data Sharing Use Cases

Here are just a few of the business opportunities made possible by modern cloud data sharing:

- » **Data sharing to eliminate data silos:** Develop a single source of truth for all your internal data and share it across all your business units.
- » **Data sharing for business efficiencies:** Share live data with your business partners to optimize costs, streamline operations, and provide superior customer service.
- » **Data sharing as a product:** Monetize your data and data services by offering governed access to these assets. Use modern data sharing technology to allow consumers to purchase governed slices of your data to enrich their data.
- » **Data sharing as a product differentiator:** Software-as-a-service (SaaS) providers can offer live and comprehensive access to their subscribers' data. Those subscribers can then perform deeper analysis on more of their data — analysis previously unavailable to them.

Capitalizing on these opportunities requires data sharing capabilities with uncommon speed, power, security, governance, and simplicity. These capabilities are not usually available with traditional data platforms or data sharing methods.



REMEMBER

Astute businesses access data from business partners, suppliers, and data marketplaces. They combine internal and external data to establish new revenue streams, differentiate their brands, reveal new market opportunities, and offer compelling experiences to customers.

Exploring Data Sharing Possibilities

Across industries and organizations, leaders are increasingly examining data sharing possibilities. More than 70 percent of the respondents to Forrester’s 2020 “Business Technographics Data and Analytics Survey” said that line-of-business managers in sales, marketing, and other business domains, rather than IT, were the primary owners of external data. The following are just two compelling examples of how businesses are taking charge.

Developing a 360-degree view of the customer

Rolling out personalized marketing campaigns requires a deep understanding of customers, competitors, and industry trends. The primary path for gaining this understanding involves acquiring data you don’t already have in order to reveal what you don’t already know.

However, with the recent passage of stricter web cookie restrictions, companies must obtain consent from visitors before gathering personal data via their websites. In response, businesses must learn to establish trusted relationships, deepen those engagements, and entice people to volunteer information about themselves.

Marketers can bolster their first-party data by acquiring second-party and third-party data from external sources, including advertising occurrence and exposure data; audience targeting data; and profiles of hundreds of millions of consumers that include age, languages, hobbies, lifestyles, and more. They can cleanse

and refine this consumer data by subscribing to data enrichment services that merge third-party data with an existing customer and/or prospect database, systematically enhancing their data to enable more advanced and targeted outreach campaigns.

Whether it's demographic data, financial data, weather data, or industry benchmark data, a wide variety of data providers are stepping in to fill the gaps. These comprehensive, enriched data sets and data services allow marketers to create more complete views of customers and prospects, personalize their offers, and tailor their campaigns.

Unlocking unlimited data for data science

Data scientists require massive amounts of data to build and train machine learning models. Easily accessing and effortlessly sharing many types of data is paramount to the success of these advanced analytic endeavors. For example, a sales team may want to score accounts, which involves ingesting data from internal customer relationship management (CRM), enterprise resource planning (ERP), and sales force automation (SFA) systems. This enterprise data may be enriched with third-party account firmographic and technographic data and then fed into a scoring model to gauge each account's propensity to buy.

A modern cloud data platform can serve as the control center for sharing data among key business applications, such as connecting customer data from Salesforce with vendor data from Workday. It can also simplify incorporating third-party data and data services via a secure cloud data marketplace, such as open source machine learning models and algorithms. This superior approach fosters collaboration and ensures the organization has a scalable data environment for data science and related analytic endeavors.

IN THIS CHAPTER

- » **Confronting traditional data sharing limitations**
- » **Addressing the business impact of today's data sharing methods**
- » **Exploring opportunities for internal and external data sharing**
- » **Introducing cloud-based data sharing alternatives**

Chapter 2

Understanding Traditional Data Sharing Challenges

This chapter conveys the limitations of traditional data sharing technologies within the context of three primary use cases: data sharing within your organization, data sharing across an external business ecosystem, and data sharing as a means of monetizing your data.

Addressing a Multifaceted Problem

If you anticipate sharing data with tens, hundreds, or even thousands of internal or external data consumers — each with unique data sharing requirements — how can you easily support this challenge? Traditional data platforms were not designed to share live data (see Table 2-1). Furthermore, they lack the security, governance, and administrative advantages of today's modern cloud data platforms, and their rigid architectures drive up costs (see Figure 2-1).

TABLE 2-1 The Challenges with Traditional Data Sharing

Traditional Data Sharing Methods	Challenges
Email	Data must be copied and moved
File Transfer Protocol (FTP)	Data transmission is slow
Extract, transform, load (ETL) procedures	Data becomes stale immediately
Online file-sharing services	Complex software is required
Cloud storage	Methods are costly to maintain
Application programming interfaces (APIs)	Processes are error-prone
	Data is not secure when moved
	Lack of data governance
	Can't update data automatically
	Can't scale for large data sets

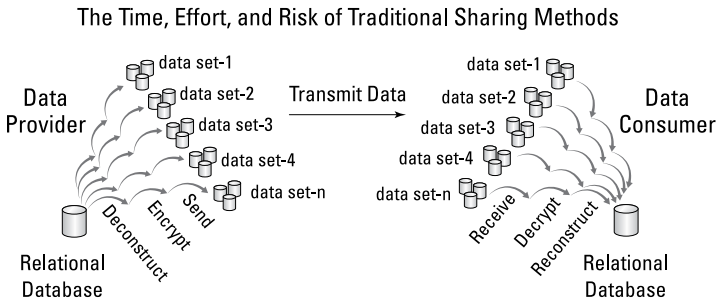


FIGURE 2-1: Multiple steps of a typical legacy FTP-based data sharing workflow that can result in increased costs and data control issues.

Contending with Time-to-Value Delays

Traditional options for sharing data at scale require provisioning and scaling complex computing platforms, even for small slices of data. Managing the associated technology infrastructure places an administrative burden on IT teams. Ongoing maintenance consumes valuable time and expertise, sapping scarce IT resources and diverting attention from other projects.

Conventional data sharing methods can create other challenges that cause more delays and require additional assistance from your IT teams, including:

- » **Handling increased data size:** The shared data set is often much larger than originally scoped, which creates problems with the data extraction process. You'll likely need a scripting language to automate breaking down the data into small parts before extracting it, which may require additional IT assistance. The reverse process must also occur for data consumers.
- » **Maintaining data pipelines:** Even after a data pipeline is in place, data engineers may need to scale the infrastructure as workloads grow, monitor data ingestion processes to ensure jobs complete without errors, and monitor the source database to ensure the data is present as expected. To protect against failures during the file transfer process, on the extraction and/or import side, both the data provider and data consumer must incorporate special software code or scripts to monitor the transfer and automatically restart the process in the event of failure.

The accumulation of these steps results in slow and painful processes for data providers and consumers. All of this must happen *before* attempting to analyze and develop insights from the data, which delays time to value. Unfortunately, the delays and difficulties don't subside with the data transfer effort. For example:

- » **Updating shared data at shorter intervals:** IT expertise may be needed if the data set is shared more frequently than anticipated. Data extraction and transfer processes must be repeated every time data changes, because shared data is always static and becomes stale immediately with traditional data sharing technologies.
- » **Cleaning data:** The import process may introduce problems, such as duplicate or extraneous data that should have been disregarded. Thus, the data provider must build increasingly sophisticated data extraction processes, resulting in rising costs and additional delays.

Examining the Impacts of Traditional Data Sharing

Traditional data sharing methods can have negative business impacts on cost, risk, and business growth (see Figure 2-2). These methods create myriad challenges, including the following:

- » Growth is limited because critical business decisions are made based on outdated, incomplete, or inaccurate data.
- » Limited discovery of third-party data and services and data locked in silos limits your capability to create a 360-degree view of customers and unlock growth opportunities.
- » Building and maintaining multiple and disparate data sharing tools results in increased costs.
- » Working with data across different silos decreases the efficiency of both technical and business users.
- » With data moving across environments, securely governing access and following regulatory and compliance requirements is nearly impossible.
- » The risk of a data breach or accidental data loss/disclosure multiplies, along with their associated costs, such as breach notifications, credit monitoring services, damage to an organization's brand, customer churn, litigation, forensic analysis, and recovery.

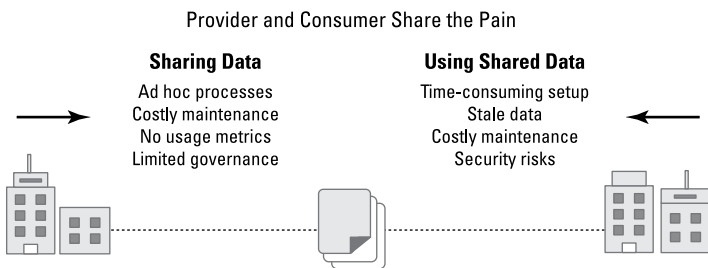


FIGURE 2-2: The difficulties of traditional data sharing methods.

The inability to extract insight from data quickly is an inhibitor to maximizing the commercial value from data. Data consumers encounter delays in developing insights, which can prevent those organizations from quickly and effectively knowing and serving customers, staying ahead of competitors, and revealing new market opportunities.



TIP

If you're using a cloud data platform that enables you to share and access live, ready-to-query data with multiparty governance controls, sharing either an entire data set or a selected subset will be easy. Universally applied security and governance controls simplify compliance mandates and reduce security risks. Additionally, diverse teams can collaborate without maintaining multiple copies of data or moving data from place to place.

LEVERAGING THIRD-PARTY DATA SETS



CASE STUDY

Petco offers an extensive catalog of pet products and services through its ecommerce website, smartphone app, and more than 1,500 physical locations. To enable data-driven decision-making across the company, Petco ingests and analyzes large amounts of transactional data. Sharing live data sets with Petco's media partners streamlines decision-making for digital advertising campaigns. Unfortunately, Petco's on-premises data architecture could not elastically scale to share large data sets while accommodating data engineering, data science, and analytic workloads.

Realizing the need for a modern data environment, Petco migrated to a cloud data platform that scales instantly to eliminate resource contention. The new data platform also allows Petco to access live, ready-to-query data from an associated data marketplace, infusing new insights and enabling the retailer to adapt to rapidly changing market conditions. For example, Petco added a third-party data set to understand and better predict COVID-19's impact on stores, motivating the company to launch curbside delivery, same-day delivery, and other customer conveniences.

Account-to-account data sharing, automated by the cloud data platform, further simplifies Petco's data sharing process by easily sharing

(continued)

(continued)

data nearly instantaneously with other cloud data platform ecosystem members. According to the senior director of data and analytics at Petco, the company can now share live data sets ten times faster than before. To achieve these performance gains with Petco's old data sharing system would have required buying at least three on-premises servers at the cost of more than \$3 million.

Petco's cloud-based data sharing infrastructure simplifies administration and frees up resources for more critical initiatives, such as democratizing data analytics for Petco's front-line staff. Data visualizations, powered by the same cloud data platform, provide 360-degree customer views that guide decision-making about personalized service offers, promotional campaigns, and other revenue-generating programs.



TIP

If you think cloud storage is the answer, think again. These basic cloud storage services lack essential capabilities for managing metadata, upholding data governance policies, or enforcing transactional consistency to ensure all users see a consistent, up-to-date view of the data at all times.

Eyeing the Future

As you will see throughout this book, today's modern cloud data platforms offer all the advantages of public cloud storage and much more:

- »» A single platform that supports many different workloads, including data sharing
- »» Secure, governed access to all data and all types of data
- »» Multiparty governance controls and revocable access
- »» Instant and near-infinite performance and scalability
- »» A zero-maintenance, cost-effective cloud service that's instantly available and extremely easy to use

With the right cloud data platform, data providers can even share data with data consumers across clouds and regional boundaries.

IN THIS CHAPTER

- » Tracing the history of data sharing in business
- » Sharing data with one or more accounts
- » Sharing data with a designated group or consortium
- » Commercializing and monetizing data
- » Controlling access with data governance

Chapter 3

Recognizing the Business Value of Sharing Data

In this chapter, you learn how data sharing methods have evolved in business, why data sharing is critical to any business, how organizations share data internally and externally, and how cloud delivery models have increased value for data providers and consumers.

Looking Back at the Early Days of Data Sharing

Not long ago, organizations hosting and supporting multiple business applications within their data centers was considered the norm. For example, there would be an application for finance, another for marketing, and others for sales, human resources, operations, and so on. Large companies hosted and ran hundreds of business applications from their own data centers, and many still do.

In the cloud era, many of these business applications have moved out of the data center and into public or private cloud infrastructures. However, the result of these software-as-a-service (SaaS)

apps creating their own data silos remains. In many ways, SaaS apps have exponentially complicated the problem due to the low barriers of entry for creating SaaS apps, thanks to the cloud.

Nevertheless, the desire to share data seamlessly has never been greater. According to a 2019 Forrester Research series titled “Think You’re Doing Data Sourcing Right? Think Again,” more than 75 percent of decision-makers want to find new external data sources, such as weather, news, social media data, demographic data, census data, and socioeconomic indicators. This need only increased in 2020 as the COVID-19 pandemic disrupted consumer and business behavior around the world, upending organizations’ traditional predictive analytics practices based on past behaviors. And it’s a trend many industry experts expect will be a lasting change. In times of increasingly disruptive events — be they pandemics, extreme weather, or economic crises — organizations need more knowledge.

Breaking Down Data Silos with Modern Data Sharing

Harnessing the value of data — either for consumption, mass collaboration, or new business opportunities — requires an easy method for enabling data access without copying and moving the data across environments. A modern cloud data platform eliminates data silos by allowing organizations to store structured, semi-structured, and unstructured data, tightly integrating previously siloed databases without manual integration or data pipelines.

From there, the platform’s modern data sharing technology enables a data provider to easily provide live, read-only data to its data consumers across clouds and regions. Data consumers can query live, up-to-date data from a single source, with no complex setup or integration. The data provider can deliver business results faster, with less overhead, and at a fraction of the cost of traditional data sharing methods.



REMEMBER

A fundamental component of modern data sharing is the ability to share live, ready-to-query data with data consumers no matter their cloud provider or region. Modern data sharing technology allows data providers to share data cross-cloud and cross-region without any extract, transform, and load (ETL) or application programming interfaces (APIs), all while maintaining strong data governance.

Understanding Data Sharing Relationships

Data sharing across and beyond an organization consists of three basic workflows:

- » **Across lines of business (LOBs):** Sharing data among departments, business units, and subsidiaries within the same organization
- » **Between external organizations:** Sharing data with and receiving data from a separate organization, such as a supply chain, business ecosystem partner, or data vendor
- » **Monetized data and data services:** Sharing and accessing live data as a service — often via a data marketplace

With modern data sharing, a customer's account on a cloud data platform is the enabler of multiple data sharing scenarios, maximizing options for both data providers and data consumers. Any such organization can exchange data across business units and ecosystem partners and with other third parties.

Sharing data across lines of business

Sharing data and data services across an organization drives timely and informed business decisions. Data sharing scenarios within an organization include:

- » Sales groups sharing data with finance groups to track sales and revenue to forecast an organization's performance
- » Marketing teams accessing and analyzing customer data to predict behavior and align demand-generation programs
- » Different subsidiaries of an organization sharing data to align their go-to-market plans better and gain more understanding of the separate areas of the business

When functional groups within an organization can't share data effectively, each group creates its own *data mart* — a copy of some of the data from the corporate data warehouse. Organizations use data warehouses and data marts to store and integrate transactional data collected from internal business applications, including marketing, sales, production, and finance. The data is

modeled and analyzed to reveal trends, patterns, and correlations that provide information and insight.

The proliferation of these data stores — sometimes called data *sprawl* — complicates data management and curtails data accuracy. For example, if different teams use different versions of the same data, they may produce conflicting analytic outcomes, such as discrepancies in revenue forecasts.

When data remains in different data stores, each LOB may depend on email, spreadsheets, shared network drives, APIs, and other outdated methods for communicating and sharing data. Moreover, without visibility into a centralized source of all data, business units may be unaware of existing resources and make redundant purchases of external data.

A modern cloud data platform with its modern data sharing breaks down these barriers by centralizing data and simplifying sharing data across LOBs (see Figure 3-1). The platform permits one-to-one, one-to-many, and many-to-many data sharing relationships, even when consumers use different cloud providers and operate in different cloud regions. These relationships can be structured in three primary ways:

- » **One-to-one:** A data provider shares data with an individual data consumer. For instance, a marketing manager shares live data about event registration with a sales manager.
- » **One-to-many:** A data provider shares data with multiple data consumers within a designated or circumscribed group of cloud data platform account holders. For instance, the product team shares key performance metrics with the leaders of every business unit.
- » **Many-to-many:** Multiple parties within the designated group can contribute, discover, and access data via a publicly available data marketplace or internal data exchange. For example, LOBs share data related to the customer base to form a 360-degree view of the customer.

Data is automatically up to date, discoverable to designated accounts, and easy to govern and revoke access to. In addition to sharing data and metadata, modern data sharing technology allows organizations to share business logic, such as multi-language user-defined functions (UDFs) and external functions, ensuring your organization has the tools it needs.

Sharing Data Across Lines of Business

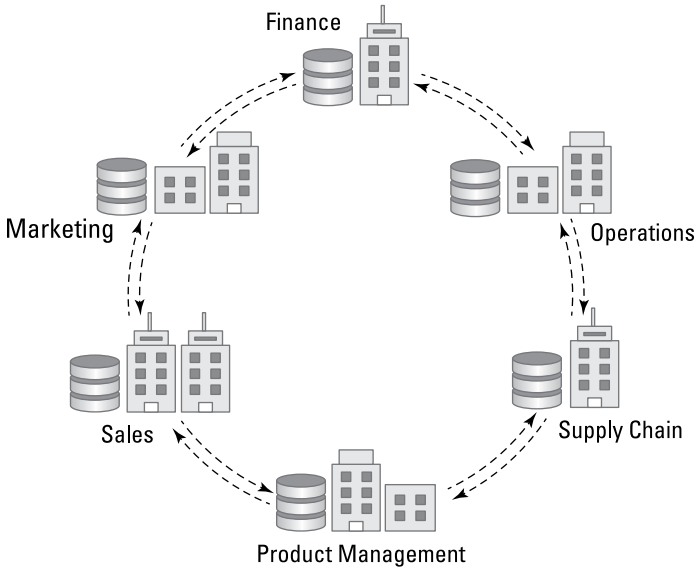


FIGURE 3-1: Any combination of departments and subsidiaries can share live data for nearly any business use case.

STREAMLINING OPERATIONS INSIDE AND OUT WITH MODERN DATA SHARING



CASE STUDY

Rakuten Rewards provides cash back and shopping rewards, operating the largest loyalty program in the United States. Its 12 million members have earned more than \$1 billion in cash back at their favorite stores. Each company division or subsidiary has specific legal requirements and permissions related to sharing data, creating a complex data sharing infrastructure.

Previously, Rakuten Rewards had to export millions of rows of data into a comma-separated values (CSV) file and then transfer it to stakeholders. Additionally, ongoing data maintenance and updates necessitated pulling the data, formatting it, recombining it, and performing quality assurance.

(continued)

(continued)

Rakuten Rewards chose to migrate to a modern cloud data platform that enables its business units and subsidiaries to access and combine governed data easily with their data sources, which soon served as a foundation for offering secure bidirectional data access to vendors and merchant partners. For example, Rakuten Rewards works with specialist companies for media planning and buying and can easily exchange large data sets with these companies to optimize media spending.

With this new data sharing landscape, partners have secure access to the data, eliminating a complex, multistep data sharing process. As a result, ongoing costs of providing data to strategic partners have effectively dropped to zero.

Sharing data with external organizations

Sharing data outside of an organization is increasingly common as business leaders look to leverage external data. A vendor-supplier relationship, a partner relationship, a developer-producer relationship, or any number of other business relationships require two or more organizations to collaborate with data to drive business (see Figure 3-2).

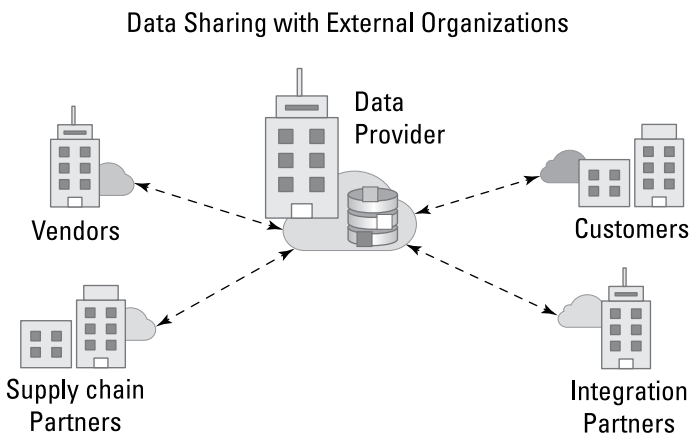


FIGURE 3-2: Modern data sharing enables organizations to share live data to streamline operations and create new business opportunities.

If organizations can't effectively exchange data with business partners, they miss opportunities to collaborate seamlessly at an operational level. Sharing limited or stale data is just another roadblock in situations where partners rely on each other for timely information.

Monetizing Data

For decades, monetizing data was the exclusive domain of sophisticated data originators, data aggregators, and data service providers, such as Nielsen, Experian, and Dun & Bradstreet — partly because the cost of entry was so high. Today, with advances in data sharing technology, the barriers to monetizing data have dropped significantly. Many companies see the value in their data and have joined *data marketplaces*, enabled by cloud data platforms, to bring their data and data services to market. For example, a data provider that gathers mobile phone location information and usage data might share governed slices of that information with advertising agencies and marketing groups so they can execute highly targeted campaigns to specific consumers.



REMEMBER

Data is a business asset that can yield different types of value depending on the organization that wants to consume that data. With modern data sharing, any organization can monetize this asset by charging for access to governed slices of its data. Consumers can then use the data to advance their own business objectives.

Within these data sharing relationships, data consumers can use shared data without having to capture and collect it themselves. They can benefit directly from analyzing that data or combining it with other data to enhance its value.

The rise of data marketplaces, powered by cloud data platforms and modern data sharing technologies, allows businesses to list their data easily on a marketplace, meet interested data consumers, and share data seamlessly and responsibly (see Figure 3-3). Today's data marketplaces not only make sharing data easy but often include such aids as data listing services and search engines that help consumers discover valuable data and try out sample data sets before they purchase them. Additionally, data providers can share data across clouds and regions by leveraging global, transactionally consistent replication technology, which keeps the data in sync in all consumer accounts.

Data Sharing via a Modern Cloud Data Marketplace

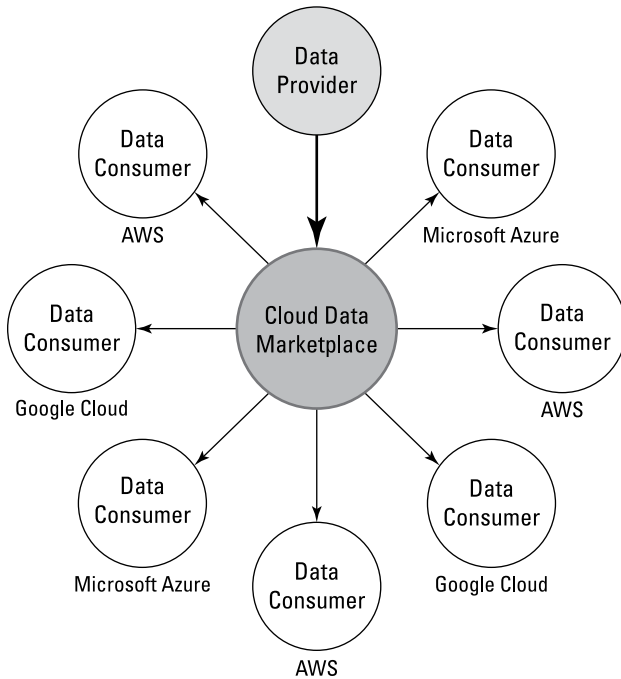


FIGURE 3-3: Create new revenue opportunities by monetizing your governed data via a modern data marketplace that spans the major cloud providers.

These cloud-based marketplaces accelerate time to value for both providers and consumers by lowering data management costs and providing consumers with real-time data updates — without continually copying data or setting up resource-intensive ETL procedures. Whether predictive maintenance data for farm equipment or energy consumption data for mining operations, data marketplaces present an opportunity to discover insights across industries and categories.

Controlling Access with Data Governance

Data governance ensures data is properly classified, accessed, protected, and used. It also involves establishing strategies and policies to ensure all data sharing activities comply with data privacy and regulatory requirements. Privacy regulations, such as the European Union's General Data Protection Regulation (GDPR),

the United States' Health Insurance Portability and Accountability Act of 1996 (HIPAA), and the California Consumer Privacy Act (CCPA), are increasingly rigorous, and large organizations have teams of people devoted to ensuring their firms adhere to these regulatory mandates.

Governance policies establish rules and procedures to control the ownership and accessibility of your data, with special attention to personally identifiable information (PII), such as credit card information, Social Security numbers, names, dates of birth, and other personal data.

A cloud data platform can simplify every aspect of data governance. Rather than moving files across environments and potentially losing track of its usage, the platform should enable read-only access to a governed portion of a live data set, accessible via SQL. When sharing data, multiparty governance controls, such as row-level access policies and data masking, allow you to control your data in a centralized and flexible way. Additionally, data access is revocable, which isn't always the case in traditional data sharing scenarios. With modern data sharing, you can easily remove access to the data altogether, enabling you to comply better with industry regulations. See Chapters 7 and 8 for more details.

DATA ANALYTICS AS A SERVICE THANKS TO MODERN DATA SHARING



CASE STUDY

Heap provides web and mobile analytics for companies across several industries. Its Heap Connect digital insights platform, used by more than 7,600 companies worldwide, helps users enrich the 360-degree view of customers by automatically capturing and syncing a complete set of digital engagement data.

Previously, to share raw data with customers, companies had to drop files into cloud storage, or host files with a cloud-based vendor for customers who weren't sophisticated enough to handle all the ETL

(continued)

(continued)

activities required to put the data in a form that was suitable for analysis. These costly, resource-intensive efforts included the following:

- **Engineering work:** Five to ten hours of work per week debugging pipelines, performing cluster maintenance, handling unique customer setups, and resolving resource contention
- **Customer service work:** Five or more hours per week spent obtaining access; scheduling cluster operations; and creating, maintaining, and resizing hosted clusters with a traditional cloud data warehouse

By leveraging a modern cloud data platform with modern, secure data sharing capabilities, Heap eliminated the engineering work and reduced customer service work significantly. Secure data sharing consolidates onboarding to a matter of minutes — not weeks or months, as before — by eliminating time-consuming cluster connections, data replication, and ETL pipelines. System administrators spend less time on permissions and security approvals.

Heap now shares live, ready-to-query data directly with customers. Customers can query live event data within moments of authorizing Heap as a data provider, and they pay only for the computing resources needed to query the shared data, with no storage costs.

Building a Foundation for Modern Data Sharing

As you'll discover throughout this book, a modern cloud data platform serves as the foundation for modern data sharing and eliminates traditional and time-consuming methods for managing a legacy data warehouse or platform and the associated data sharing procedures, thus decreasing costs, mitigating risks, and accelerating growth. Near-unlimited performance at any scale should be built into a modern data platform, so there's no infrastructure to tweak, no knobs to turn, and no tuning required. With near-zero management, you can pursue far-reaching data sharing strategies to target a larger base of data consumers — across business units, with business partners inside your ecosystem, and with other organizations as part of the burgeoning data economy.

IN THIS CHAPTER

- » Yielding greater value in financial services
- » Generating positive outcomes in healthcare and life sciences
- » Powering the retail supply chain with fresh insights
- » Delivering superior media and entertainment experiences
- » Offering comprehensive data management capabilities to technology firms

Chapter 4

Sharing Data across Industries

Modern data sharing allows organizations across all industries to deliver insights not possible from just their own data to boost revenues, reduce costs, and deliver previously unimagined experiences to customers, patients, and citizens. This chapter describes how data leaders in such industries as financial services; healthcare and life sciences; advertising, media, and entertainment; retail and consumer packaged goods; and technology are utilizing first-, second-, and third-party data to create future-proofing business outcomes.

Yielding Greater Value in Financial Services

The financial services industry is rapidly transforming. Firms that comprise the sub-industries of financial services, from insurance to asset management, are looking to enhance investment

decisions, highly personalize their offers, and automate more and more internal processes. Data, both internal and external, plays a critical role in each of these trends. The rise of modern data sharing technologies via a cloud data platform offers financial services firms new opportunities to leverage real-time data from across their business ecosystems and beyond, allowing them to make more data-driven business decisions, increase efficiencies, and deliver on regulatory and compliance requirements.

A PWC report titled “Financial Services Technology 2020 and Beyond: Embracing Disruption” showed that more than 75 percent of data leaders invest in putting their customers at the center of every interaction. To deliver hyper-personalized experiences, these organizations are centralizing first-party data that may have previously been siloed in different business units, second-party data from partners and vendors, and third-party data from external data providers to develop 360-degree views of their customers.

SHARING DATA TO IMPROVE CUSTOMER EXPERIENCES AND ENHANCE INVESTMENT DECISIONS

Financial services companies can deliver timely and personalized customer, investor, and policyholder experiences and enhance investment decisions with modern data sharing. Common data sharing use cases within the financial services industry include:

- **Retail banks and insurers:** Unify and securely share customer data, removing silos across all departments, to create an enterprise-wide 360-degree view of each customer.
- **Investment and wealth management providers:** Leverage market data and alternative data sources to generate insights to enhance investment decisions.
- **Financial data providers:** Increase your offerings by sharing relevant customer and investment data so institutions can create deeper customer relationships.

Second-party data from vendors and partners, and third-party data and services, help any financial services firm better understand the economic and societal trends that affect their business. For example, shared data and data services assist with critical regulatory and governance tasks, such as fraud detection, anti-money laundering, risk management, and credit assessments. These firms integrate second-party data and data services from vendors and partners, including modern analytics services and machine learning models. Analyzing data from diverse sources is much simpler when these firms can store all their data in a centralized, globally available, and consistently governed data platform without the need to maintain extract, transform, and load (ETL) pipelines or application programming interfaces (APIs) to share data.

For example, insurance companies share data with partner companies and access publicly available data sets from industry marketplaces. A property and casualty (P&C) insurer that wants to minimize claims leakage, for example, can use data services from Quantifind to investigate potentially fraudulent auto repair claims. Available via modern data marketplaces, Quantifind uses external data sources, such as sanctions and blacklists, in conjunction with predictive risk-typology models that indicate risk levels. This data helps claims adjusters establish fraud risk scores for each individual that submits a claim.

Maintaining data in a cloud data platform that offers modern data sharing technologies and a data marketplace gives insurance adjusters immediate and direct access to third-party data and services, facilitating collaboration across the insurer's business ecosystem. For example, insurance executives might want to know how their financial metrics, such as *total annual claims paid*, compare with similar metrics from other P&C insurance companies in their region. S&P Global, a marketplace vendor, addresses this with comprehensive statutory financial data from the National Association of Insurance Commissioners.

Secure data sharing technology makes third-party data instantly available, without the need to transfer data or set up custom APIs, making it easier to mobilize data for business intelligence or data science endeavors. For example, the P&C insurer might want to combine the Quantifind data mentioned above and the S&P Global data into an executive dashboard, along with information from its internal customer relationship management (CRM) system about each auto body repair shop.

BUSTING BOTTLENECKS FOR DATA SCIENTISTS



CASE STUDY

Pacific Life helps millions of individuals and families with their financial needs through a wide range of life insurance products, annuities, and mutual funds. Previously, Pacific Life had an on-premises data warehouse. When data science teams attempted to perform key business processes, such as analyzing equity, interest rate returns, or customer activity, their large siloed data sets presented challenges. In some instances, data scientists and business analysts had to wait a month for data to be refreshed.

Pacific Life sought a more versatile data platform that would allow its data science, sales, marketing, and product design teams to fully leverage their data. By consolidating key data sets onto a centralized cloud data platform, Pacific Life dramatically reduced data latency and minimized the dependency on its database administrators. Now, predictive analytics are available almost immediately, allowing for actionable intelligence to help drive sales, allocate resources, and influence other key business decisions.

Secure data sharing technology within the cloud data platform enables Pacific Life to access third-party data from Experian and other data providers, eliminating file transfers and ETL. The cloud data platform serves as a single source of truth, enabling governed data to be shared freely but securely across the organization. Now, each team can easily create workspaces to run its analytics. If the data science team curates a valuable data set, for instance, that data can be published for the larger organization to use.



REMEMBER

Financial services companies use a cloud data platform to securely access second- and third-party data for broader and more precise customer and investor views, to reduce fraud and risk exposure, and to enhance investment decisions. Having a centralized, secure, and shareable source of governed data simplifies identifying best practices for developing new products and services, and making new investments.

Delivering Better Healthcare Outcomes

The rapid rise of patient health data generated from electronic health record (EHR) systems and connected health devices paves the way for new approaches to care delivery, clinical diagnostics, medical innovation, and regulatory decision-making. Immense data volumes flow between healthcare providers, payers, life sciences companies, and medical research institutions. Poised at the center of this data-rich landscape are two distinct industries with similar attributes: healthcare and life sciences.

Healthcare focuses on delivering health services for members and patients, whereas life sciences focus on developing and commercializing medicines and therapies. Due to the critical nature of this work, these are highly regulated industries. In the U.S., the most notable health information security requirement affecting these industries is the Health Insurance Portability and Accountability Act (HIPAA), which governs the use of protected health information (PHI), and personally identifiable information (PII). In the European Union, health data must be secured per General Data Protection Regulation (GDPR) requirements.

By housing their data on the same cloud data platform, multiple organizations can easily, securely, and responsibly share sensitive health data. Each entity can be granted seamless, near real-time, governed access to data. Each of these organizations can securely access live data while safeguarding protected health information. Centralized governance policies facilitate compliance with data security and data privacy regulations. Additionally, if the cloud data platform also hosts a multi-industry data marketplace, each organization can use healthcare-specific data sets and data services, from physician credentialing and physician-based market share data to public health data.



REMEMBER

Complying with industry regulations governing PHI and PII security presents ongoing challenges for many healthcare and life sciences companies. Astute organizations use modern data platforms to securely share live health data internally and with their ecosystem of partners to provide quality patient outcomes, drive growth, shorten time to market, and reduce costs while deploying controls that enable compliance with HIPAA and other data-governance regulations.

SHARING DATA TO IMPROVE PATIENT OUTCOMES

Tremendous opportunities exist for healthcare and life sciences companies that can reimagine how to store, manage, and share clinical and administrative data. Popular data sharing use cases include the following:

- **Personalized medicine to improve patient outcomes:** Healthcare providers can create a single source of patient data, including electronic health records, along with clinical, administrative, operational, and Internet of Medical Things (IoMT) data.
- **Fraud detection and prevention:** Health insurance companies can share claims, billing, and other data with contracted healthcare providers, as well as share formulary and diagnostics data with pharmacy and lab partners to identify anomalies and flag suspicious claims.
- **Product development:** Life sciences companies can share clinical, sales, marketing, and other data across business units and externally with medical device manufacturers.
- **Collaborative research:** Research institutions and university hospitals can aggregate and share health-outcomes data collected from providers, life sciences companies, and other partners.



CASE STUDY

SHARED DATA FOR PATIENT CARE

Healthcare data and analytics provider Komodo Health believes that smarter, more innovative use of data and analytics is essential for reducing disease burden — the impact of a disease on a population's quality of life. To that end, the company has applied artificial intelligence and other advanced data science techniques to its Healthcare Map, which it claims is the industry's most precise view of the U.S. healthcare system. By doing so, Komodo tracks the unique journeys of more than 320 million patients, augmented by analytic algorithms and clinical expertise.

Previously, Komodo had difficulty linking data across thousands of siloed sources. Komodo's IT team struggled to build access controls, maintain quality control processes, and scale the ingestion of terabytes of data each day. Now, by using a cloud data platform with modern data sharing technology, Komodo Health can leverage more than 150 payer data sets in conjunction with more than 65 billion clinical and pharmacy encounters.

The data platform enables Komodo customers — primarily health-care organizations — to use frameworks, such as R, and programming languages, such as Python, to directly query these data assets. Komodo also offers data services that allow these healthcare organizations to run custom analytics on top of their data and develop data assets they can leverage to glean their own insights and identify unmet need throughout the healthcare value chain.

Powering the Retail Supply Chain

Meeting customer needs during the COVID-19 lockdowns required retailers and consumer packaged goods (CPG) companies to improve online delivery channels to ensure unwavering speed, convenience, and quality. Companies across the industry are now discovering how to share data to forecast customer demand, manage supply chains, and conduct business efficiently.

For example, retailers and their CPG partners each collect siloed information that is valuable to the other. Retailers have granular point-of-sale (POS) data that CPG firms can't readily access. CPG companies want to analyze that data and combine it with data from marketing applications, supply chain management applications, inventory control systems, and CRM systems to make category and brand management recommendations. They need to know precisely what customers are buying to produce the right goods in the right quantities. Meanwhile, CPG firms can share data about their production plans to help retailers make better decisions about warehousing, stock keeping, distribution, and merchandising strategies.

Whether trying to learn more about customers, manage inventory levels more efficiently, or forecast demand for popular items, a

complete source of live, governed data is critical to glean these consumer-level and product-level insights. Modern data sharing technology enables retailers and CPG companies to easily share all this data while building mutually beneficial relationships and forging extended partner networks.



CASE STUDY

DATA SHARING SERVICE OPENS NEW REVENUE STREAMS

Maintaining product quality in the CPG industry requires rapid decisions related to inventory, warehouse management, shipping, and delivery. By helping grocery suppliers understand trends in their retail-customer data and share insights with those customers, Atheon Analytics keeps goods moving to the right place at the right time. Its SKUtrak service presents machine learning analytics via intuitive, interactive visualizations.

Fresh groceries are a byproduct of fresh data, and Atheon's new data sharing service, based on its cloud-built data platform, allows Atheon to easily share data to SKUtrak customers, so they always see the latest insights but without complex data-copying or data-moving procedures. "SKUtrack DataShare provides direct access to the data that powers SKUtrak dashboards, enabling our customers to perform ad hoc analysis, build their own custom dashboards, or feed their IT systems with fresh, clean, curated data whenever they need it," says Atheon CEO Guy Cuthbert.

Having a cloud-built analytics platform to securely share real-time data is now an essential part of Atheon's business. When CPG staff develop a detailed understanding of product performance across the supply chain, they are confident in the decisions they make, and actions they take, to improve availability, reduce wastage, and better meet true customer demand. "Data sharing offers a great new revenue stream," Cuthbert adds. "It is an essential part of our product portfolio. Developers and consultants aren't waiting around for data, and productivity has gone through the roof."

Tech-savvy retailers and CPG firms also depend on cloud data platforms to conduct their analyses, often accessing third-party data available from the platform's data marketplace. For example, some data providers offer consumer purchase information from millions of U.S. households, informing go-to-market strategies. Other data providers offer advertising data from Google, Facebook, Snapchat, and other large websites to help companies better understand customer preferences and purchasing trends.

Data marketplaces present opportunities for retailers and CPG companies to create new revenue streams by monetizing their data and selling it to partners that might wish to target those same consumers with tangential offerings, including insights about popular items, optimal price points, seasonal demand fluctuations, and geographic trends.

Of course, strict consumer privacy laws, such as GDPR and the California Consumer Privacy Act (CCPA), and cybersecurity threats drive the need for relentless data governance and advanced security practices. All retailers desire to know their customers and anticipate their needs, but failing to secure PII data can result in significant fines for companies noncompliant with prevailing data privacy laws.

Having a secure data platform that enforces strict governance is essential. Chapter 8 has more on this.

EXCHANGING DATA WITH RETAIL PARTNERS

Throughout the retail supply chain, organizations are learning to centralize their data and securely share live data with supply chain partners, while optimizing pricing and inventory strategies, increasing margins, and ensuring consumer privacy.

- **Retailers:** Centralize siloed data across supply chain, inventory, point of sale, CRM, customer loyalty, and marketing analytics systems.
- **CPG:** Access transaction data from retailers and combine it with other data to influence supply chain, inventory, and marketing decisions.

Delivering Superior Media and Entertainment Services

The wholesale shift to online, digital, and streaming services has forever changed the media industry. Yesterday's business models have evolved to accommodate digital content and online services, such as the video-on-demand (VOD) services that have transformed today's television markets. Consider YouTube, which began as a website for sharing consumer videos. Acquired by Google, it has morphed into YouTube TV — a content-delivery service that now competes with major broadcasters and cable TV companies. Google, Apple, Amazon, and other tech titans continue to grab market share from cable providers and satellite services, crowding the market with hundreds of viewing options.

All digital business models yield a steady stream of data that can help target customers, deliver personalized content, and measure the impact of marketing and outreach campaigns. Unfortunately, data silos and integration challenges prevent many advertisers and brands from accessing and enriching the data they need. This fragmented data landscape complicates achieving crucial business goals and adhering to regulatory compliance — a costly and risky prospect as data privacy laws toughen.

Exploring data sharing techniques

Market leaders in media and entertainment understand the imperative of deploying modern data sharing to make up for these deficiencies. Data sharing is the starting point for unifying consumers' identities, gaining awareness of their online activities, and ultimately revealing each subscriber's customer lifetime value (CLV).

After creating these complete customer views, media and entertainment companies rely on centralized data repositories to democratize access to pertinent insights, enabling teams to make more-informed decisions on content production and acquisition, distribution, advertising, and product experience. Furthermore, media firms can share deeper insights about their audiences with brand advertisers, enabling advertisers to better predict the return on ad spend and increase the value of ad inventory.

For example, advertisers collect granular customer data across many sources to create better ads and measure returns from each ad campaign and attribution across channels. Media and entertainment companies use modern data sharing technology to exchange first-party, second-party, and third-party data across business units, with advertisers, and with data enrichment partners.



REMEMBER

Today's media companies need to acquire, retain, and engage audiences to earn their business via subscriptions, incremental purchases, and advertising revenue streams. Data holds the key to making the right content decisions, acquiring customers, preventing churn, and optimizing viewer experiences.

Responding to cookie regulations

Digital advertisers have long used cookies to gather consumer data, measure ad penetration, and personalize their marketing campaigns. Websites use *first-party* cookies to monitor what users do on a site so that people aren't asked to perform the same tasks again and again. *Third-party cookies* are created and placed by companies other than the website a user is visiting. These cookies are commonly used for retargeting, in which users' search activity is tracked to determine which ads to deliver based on products and services in which they've shown interest. For example, you might search for LED light bulbs on Amazon and then suddenly see ads for light bulbs in your Facebook newsfeed.

In response to consumer privacy demands, the major web browsers are phasing out support for third-party cookies. To make up for the loss of cookie data, media companies are forging partnerships to share user data in a privacy-compliant way. Many create data *cleanrooms* to share data among media companies, advertisers, and other partners, without moving data or exposing PII. Some cloud data platforms offer data marketplaces that allow media and entertainment companies to solidify and streamline these data sharing relationships as well as to access relevant third-party data sets, such as identity graphs, demographics data, and audience data, in conjunction with data enrichment services from commercial data providers.

SUPERIOR EXPERIENCES FOR MEDIA AND ENTERTAINMENT

Media firms rely on cloud data platforms to securely share governed data across brands, advertisers, ad platforms, and data enrichment providers. As a result, these organizations make data-driven decisions to increase subscriber lifetime value and increase return on advertising spend. For example:

- **Media firms** unify subscriber behavior data across platforms and channels with purchase information from brands to create 360-degree customer views.
- **Brands** unify customer touchpoints across media channels and advertising platforms to optimize campaigns that improve conversions and sales.
- **Advertisers** connect advertising spend with media consumption and purchase behavior to demonstrate marketing programs' return on investment.
- **Ad-Tech companies** help media firms increase the effectiveness of digital and online offerings (see the sidebar "Creating New Revenue Opportunities").

CREATING NEW REVENUE OPPORTUNITIES



CASE STUDY

SpringServe offers a full-stack ad-serving platform to enable automation and optimization of video ads across devices. Trusted by leading publishers and advanced TV distributors, its platform delivers control, transparency, and analytics to increase ad performance and revenue from media sales. SpringServe's platform serves 200,000 ad requests per second.

Previously, when SpringServe collaborated with another company or partner, it had to set up a custom pipeline to synchronize data. Each partnership required a unique but traditional data sharing solution, with accompanying engineering, development, and management.

Now, SpringServe uses a cloud data platform to share data with clients and partners with minimal setup, management, and overhead. Each partner can access specific tables in the platform's database, allowing them to obtain a unique slice of data without requiring data exports, custom engineering work, or management overhead.

SpringServe now can offer new services with virtually zero setup, according to the organization, giving clients access to their raw data without jumping through hoops or setting up systems.

Delivering Fresh Insights to Technology Companies

Today's modern web and mobile apps have disrupted legacy technology vendors in marketing automation, CRM, enterprise resource planning (ERP), and across vertical markets, such as healthcare, logistics, manufacturing, and finance. In the cloud era, many organizations are jettisoning expensive, on-premises software applications and saying goodbye to tedious upgrade cycles. Software-as-a-service (SaaS) providers now offer subscription-based apps and services much easier to acquire and maintain.

One of the biggest growth areas involves embedding analytics into apps to allow customers to sort, summarize, and visualize data for better decision-making. Many SaaS app providers discover new revenue streams by sharing this data and selling these data analytic services. For example, a CRM application that tracks and categorizes sales leads might combine first-, second-, and third-party data to enrich a sales team's understanding of customers and prospects. The SaaS provider might enrich basic CRM data with technographic and firmographic information from industry data sources, helping to classify and add propensity scores to each account lead. By looking at this data-driven score, salespeople know immediately which accounts to contact and in which order so they can focus on the strongest opportunities.

Of course, most CRM applications aren't designed to accommodate these data enrichment processes. They're designed for booking deals and serving customers. When the SaaS provider uses the right data platform, it can increase value for customers by

integrating, analyzing, and correlating this information to deliver the insights salespeople need to discover and close high-value opportunities.

When the platform is built around a centralized data repository, SaaS companies can securely share governed data without copying that data or developing infrastructure to move the data. The SaaS provider simply authorizes customers to access read-only versions of the data set, eliminating the need to create data pipelines and code custom interfaces. The customer can then leverage comprehensive data that is richer than what would be available via the user interface of the SaaS app — for data science and data analytics, for example.

This modern data sharing approach also simplifies governance and compliance. The SaaS provider can determine which parts of a database to share, while also protecting sensitive information. Customers enjoy better experiences powered by more complete data, while the SaaS provider upholds important governance, security, and data privacy requirements.



REMEMBER

A modern data platform allows software companies to share unique slices of raw data with each customer, list new products and data services on a data marketplace, and offer differentiated apps and services.

HALLMARKS OF A COMPLETE DATA SHARING ARCHITECTURE

SaaS providers should standardize on a cloud data platform that fulfills the following data sharing requirements:

- Easily share data with customers across cloud providers and regions, without any ETL or API.
- Enable an almost limitless number of a SaaS customer's users to access a single copy of its data.
- Provide customers with dedicated and independent compute resources, so they don't experience slowdowns or disruption to queries.
- Allow SaaS customers to pay only for the computing resources needed to query the shared data, with no additional storage costs.

- » Capitalizing on modern data marketplaces
- » Opening new markets for data providers
- » Simplifying access for consumers

Chapter 5

Tapping into Data Marketplaces

Data marketplaces aren't new, but those built on modern data sharing technology deliver new and inspiring insights and opportunities for data providers and consumers. This chapter illustrates how your organization can participate in a modern data marketplace as a data consumer to leverage essential third-party data that is easy to access, easy to integrate with your existing data, and always up to date. You also learn how, as a data provider, you can easily and securely offer governed slices of your data to potentially thousands of data consumers to create new revenue streams.

Examining Traditional Data Marketplaces

Data marketplaces are accessible to participants via a portal or app store-like environment. In traditional data marketplaces, the data sets offered are often static copies of the data that require the data consumer to move the data via extract, transform, and load (ETL) pipelines or application programming interfaces (APIs). Some traditional data marketplaces are cloud-based. However, they typically require both the data provider and data consumer to

use the same cloud provider. As a result, data providers must publish their data across multiple data marketplaces to reach as many consumers as possible. Likewise, data consumers are required to set up and manage new data sharing workflows for each data set in each cloud.

In recent years, a new type of data marketplace, powered by modern data sharing, has sprung up, creating new opportunities for both data providers and data consumers.

Embracing Modern Data Marketplaces

With modern data marketplaces, data providers can improve the customer experience by giving consumers a fast, secure, and cost-effective way to access live data sets and data services.

Data consumers can access external data and data services, including data enrichment services that enhance customer records by appending demographic data or detecting potential security threats based on the analysis of data created by website traffic.

Modern data sharing makes all of this possible (see Figure 5-1).

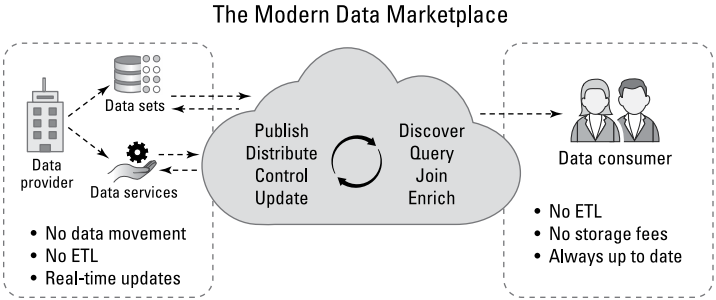


FIGURE 5-1: An architecture for an efficient, real-time data marketplace.

As organizations begin to explore using external data and data services, two common questions often arise:

- » How do I find relevant data I can easily source from third parties without managing ad hoc data pipelines?
- » How do I scale my external data use to include a larger ecosystem of data providers and partners?

Likewise, as organizations begin to explore providing data to external consumers, they often find themselves asking:

- »» How do I make data available to partners and customers?
- »» How do I scale data sharing to broader ecosystems without overwhelming my data infrastructure?

It all hinges on having modern technology in place for sharing and discovering third-party data. Modern data marketplaces don't require moving data files across environments, ETL technology, or constant updates to share data with consumers. There is no need to transfer data via deploying File Transfer Protocol (FTP) or to develop and maintain APIs.

Applying Modern Data Sharing Technology

In contrast to the heavy toll traditional marketplaces place on both data providers and data consumers, the ideal data marketplace uses modern data sharing technology to provide the following benefits:

- »» Access to live data sets across cloud providers and regions, without APIs or ETL processes
- »» Automatic updates to shared data sets every time a data provider refreshes the data
- »» The ability to create personalized, governed views so each data consumer sees only the data they're authorized and required to see based on roles and access rights
- »» Compliance and regulatory controls, including the capability to revoke access to data



REMEMBER

Older data marketplaces usually depend on FTP downloads, ETL transformations, and API connections. Modern marketplaces should use modern data sharing technology to allow you to easily tap into live, ready-to-query data no matter your cloud provider or region.

Opening new markets for data providers

Modern marketplaces aren't just data repositories: They're vehicles for growth. Today's cloud-built technologies have lowered the barriers to entry, allowing nearly any organization to securely

and responsibly share and monetize its data assets. From small startups to global conglomerates, many commercial entities can leverage these technologies to use a data marketplace for launching new revenue streams. Examples include:

- » **Marketing analytics businesses** gather, analyze, and share user, session, and event-level data from websites and applications.
- » **Financial market vendors** provide information to help businesses identify the creditworthiness of potential business partners.
- » **Manufacturers** provide insights into equipment operations to predict maintenance needs and reduce downtime.
- » **Aviation businesses** offer global air travel intelligence access, spanning flights from thousands of airports and hundreds of countries.

These data products have opened up new revenue streams for companies that wish to appropriately monetize their data.

One big advantage of modern data marketplaces is the capability to market and seamlessly provide data and services to consumers across regions and cloud providers — opening a wider network of potential customers.

Modern data marketplaces also accelerate time to market for data providers. For example, modern data marketplaces allow data providers to create product listings with rich information about the data product, such as example use cases and sample queries. These marketplaces also offer data providers insight into performance via such metrics as clickthrough rates and query volumes, which reveal how popular the data set is among consumers. This all adds up to higher customer satisfaction and better customer retention.

Finally, with multiparty governance controls and revocable access, providers can easily share data in safe and governed ways. Granular access controls can determine which parts of a database to share, while other database security policies can protect sensitive data in individual fields. In addition, when a data provider shares data with a data consumer and the consumer chooses to terminate the data sharing contract, the provider can easily revoke access to the data. Chapter 8 elaborates on how these technologies enable data providers to share data while maintaining governance, security, and privacy.

BENEFITS FOR DATA PROVIDERS

Here are some of the reasons data providers are gravitating to data marketplaces built on modern cloud data platforms that enable data sharing:

- **Reach:** A data marketplace provides a single place to easily publish, market, and distribute data to new data consumers across industries, regions, and clouds.
- **Value:** Providers can deliver data products and services to a broader audience without needing to individually manage data sharing infrastructure.
- **Control:** Embedded security and governance technologies allow providers to control who can see their data listings and which parts of each data set to reveal to each consumer.
- **Simplicity:** Real-time data distribution means there is no need to move data through custom pipelines via ETL, FTP, APIs, or other programmatic methods.
- **Insight:** Instant access to telemetry and usage data on data products help to inform future offerings.

DISTRIBUTING REAL-TIME DATA VIA A CLOUD DATA PLATFORM



CASE STUDY

FactSet provides financial data, market data, and analytics to tens of thousands of investment professionals. These clients integrate the data with their applications, web portals, and statistical packages to make crucial decisions.

Previously, FactSet used API calls, FTP, and Secure FTP methods to transfer data, which was time-consuming and required significant compute resources. In some instances, ingesting very large data sets took weeks. To both expedite and alleviate this process, FactSet sought a more efficient way to distribute its content to clients.

(continued)

(continued)

Now, FactSet deploys its data sets once via a cloud data platform, making them instantly available to a nearly unlimited number of clients. Investment professionals obtain access to FactSet data in near real time. This includes access to more than 20 proprietary data feeds and dozens of third-party data feeds, without incurring storage costs. Clients can query structured and semi-structured data using SQL, join it to their existing data, and analyze the merged data sets with their own tools — without using cumbersome ETL procedures.

File-sharing processes that used to take hours — or even days if the team had to provision cloud resources — are now nearly instantaneous. Some FactSet clients use the marketplace to join disparate data sets, such as joining FactSet queries with data from other leading content providers.

Unlocking new insights for data consumers

Modern cloud data marketplaces simplify discovering and accessing third-party data products and services from providers across industries, regions, and clouds. Data sets from hundreds of providers can be accessed from within the consumer's account with just a few clicks. Personalized listings in the marketplace enable consumers to request access to customized slices of data based on their specific needs.

A comprehensive cloud data platform enables a modern data marketplace built on the platform's modern data sharing technology. Organizations can meet their data sharing needs via direct sharing or a marketplace while also enabling their other critical data workload needs, such as data engineering, data lakes, data warehousing, data science, and building data applications. With internal and external data unified in a single platform, it's easy for a virtually unlimited number of teams to tap into and collaborate using these shared resources. Everything is available via a highly scalable and flexible low-maintenance platform (see Figure 5-2).

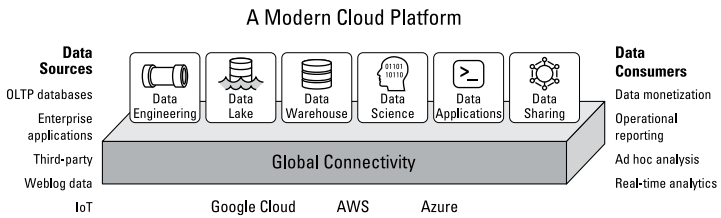


FIGURE 5-2: A modern cloud data platform supports data sharing and other critical workloads, even when the data spans multiple public clouds.

BENEFITS FOR DATA CONSUMERS

Here are some compelling reasons that consumers gravitate to modern data marketplaces:

- **Ready to query:** Consumers can immediately query the data, join it with their own data, feed it into machine learning (ML) models, or visualize it via dashboards and other business intelligence tools.
- **Always up to date:** Updates made by the data provider reflect almost immediately in the consumer's account, without having to establish update procedures and schedules.
- **Cost-effective:** No need for resource-intensive data loading and transformation procedures; no need to incur data movement, storage costs, or the risks associated with traditional data marketplaces.
- **Versatile:** The best data marketplaces seamlessly span multiple major public cloud providers to maximize deployment options and minimize vendor lock-in.
- **Unified:** All data — including first-, second-, and third-party data; Internet of Things (IoT) data; and other operational data — can be unified into one platform, making it easy to utilize and derive insights.
- **Centralized:** All other data workloads are centralized and supported by a modern data cloud platform and its modern data sharing, offering increased efficiency and performance.



TIP

As a data provider, if you want to ensure long-term flexibility and market expansion for your data listings, work with a cloud data platform provider that can offer seamless access to data among multiple geographic regions in multiple clouds — including AWS, Microsoft Azure, and Google Cloud — all as part of one seamless marketplace. That way, you won't be locked into one particular cloud vendor as your data needs evolve.

Turn to Chapter 6 to learn about monetization opportunities unlocked through modern data sharing, how to evaluate distribution and pricing models, and how to optimize your presence in modern data marketplaces.

IN THIS CHAPTER

- » Discovering data monetization opportunities
- » Pricing your data as a service
- » Optimizing data listings and accelerating time to value

Chapter 6

Monetizing Your Data

Data monetization is a process by which an established commercial data provider, or any organization that wants to monetize slices of its data, charges data consumers a fee to gain access to its data or data services, which can include data models, data enrichment, and data analytics.

Nielsen is a pioneer in this space, and its service is so pervasive, it has become a household name. For more than 90 years, the company has collected, analyzed, and sold consumer data to media companies, advertising firms, retail organizations, and many other industries.

Nielsen's proven business model has inspired numerous data sharing paradigms. For example, financial services providers, such as S&P Global, collect stock market data, package it, and sell it to brokers, hedge funds, and other investors. Gaming industry data service providers can offer data analytics services that help gaming companies determine which game features are most popular. Business intelligence (BI) dashboards and custom reports about players' activity can help studios optimize their games, improve the gaming experience, and drive more business.

Examining New Revenue Opportunities across Industries

Today's data marketplaces are growing fast, fueled by the demand for external data (see Figure 6-1). In addition to acquiring external data for internal use, organizations take their data to market to generate incremental revenue. For example:

- » **Telecommunications companies** can offer location data to help retailers target and personalize advertisements.
- » **Logistics companies** can offer data about transportation patterns and shipping activity to help manufacturers optimize supply chains.
- » **Retailers** can mine point-of-sale data to generate valuable insights about consumer purchasing trends.
- » **Real estate companies** can offer market data to alternative asset management firms to identify new investment opportunities.
- » **Health insurance companies** can combine clinical and claims data to identify population health trends and then provide that data to health plans so that they can avoid unnecessary procedures and deliver better patient care.

Access to Data Unlocks New Value across Industries



FIGURE 6-1: Organizations in every industry are discovering new business opportunities and new ways to serve customers best, thanks to reliable, responsible, and efficient ways to monetize data.

Sometimes the data that a product generates is more valuable than the product itself. For example, consumer electronics manufacturer Fitbit rose to prominence on the strength of its activity trackers, smartwatches, and other wireless gear. But the long-term value wasn't merely in the devices themselves, but in the data they collected from the people who wore them. According to a November 1, 2019 article in *Los Angeles Times*, Google paid \$2.1 billion to acquire the data on Fitbit's 28 million active users.

The demand for data outside an organization continues to grow. For example, customer service departments may want external data about callers to segment prospects and knowledgeable route opportunities. Marketing professionals look to external data sources to target their messages and tailor their campaigns. Sales teams utilize third-party data about the organizations they sell to, so they can properly qualify leads. Risk analysts access data from cybersecurity experts to help identify network intruders and circumvent fraud. Modern data sharing is growing every year as the volume of data that organizations collect expands exponentially.



TIP

From Internet traffic to weather data, social media trends to purchasing patterns, your data may be valuable to third parties. To determine where your data monetization opportunities lie, start by identifying the core use cases for your potential customers. Doing so helps you determine which type of data and access to give them and how you might extend your current business practices to maximize future data sharing opportunities.

Evaluating Distribution Channels

Several potential distribution channels exist for selling data and data services:

- » **Direct to customers:** The advantage of this approach is that you're selling to an audience you know well, allowing you the opportunity to deepen those relationships. The disadvantages are that it's a limited market, and you have to manage data sharing with each customer individually.
- » **Through a data broker or aggregator:** The advantage here is the ease of adoption because the broker supplies a built-in clientele. The disadvantages include a lack of control over the data and an inability to forge direct relationships with users.

» **Via a data marketplace:** Chapter 5 defines a *data marketplace* as an online store where commercial data providers and any organization that wants to monetize slices of its data can easily and securely publish, market, and distribute data products and services so data consumers can discover and access them. A data marketplace is a great way to build your brand and establish new direct relationships with customers.



TIP

Be sure to choose a marketplace that can help generate brand awareness with potential data consumers. The marketplace should have industry-specific listings that describe the value of provider offerings within retail, healthcare, financial services, and other sectors. For example, a section of the marketplace devoted to media and advertising can showcase offerings that help subscribers reduce customer churn and increase advertising impact by using third-party identity data. A section dedicated to healthcare and life sciences can help subscribers find data on COVID-19 incidence and vaccine distribution rates.



REMEMBER

No matter the channel you choose for distributing your data, ensure that channel leverages modern data sharing. In doing so, you will decrease the time and cost required to share data with your customers.



CASE STUDY

FASTER MARKETING ANALYTICS

Simon Data empowers brands to deliver data-driven, personalized customer experiences. Its customer data platform (CDP) includes a reporting and insights dashboard that its clients use to understand the effectiveness of their marketing campaigns. They can define start conditions, milestones, and goals, and the dashboard enables them to understand conversion rates through the sales funnel.

Simon Data uses a cloud data platform that enables modern data sharing technology to manage inbound and outbound data flows to the CDP, such as data from its clients' CRM systems and marketing analytics apps. This platform allows clients to develop and deliver personalization capabilities without building and maintaining complex data sharing infrastructure. Clients can consume marketing data

from Simon's data warehouse simply by subscribing to pertinent data sets in a data marketplace.

In addition to modern data sharing capabilities, a data marketplace is part of the cloud data platform, which Simon Data also uses to store and analyze internal data. One-third of Simon Data's clients also use this platform to manage their data flows, making it very easy to turn on secure data sharing within these clients' accounts.

Thanks to the easy exchange of data enabled by the data marketplace, Simon's clients obtain a quantifiable return on their investment in Simon Data's platform weeks and sometimes months sooner than they did before, when Simon Data was manually managing the data sharing infrastructure.

Pricing Your Data Sharing Service

Can you monetize your data? If so, how should you charge for it? To determine the value of your data, ask yourself three questions:

- »» What value can customers derive from it?
- »» What are customers willing to spend for access?
- »» How have other organizations priced similar data?

Answers to these questions are not always obvious. Data valuation often includes iterations with prospective customers to test the value of a data service. By applying the data to specific use cases, typically centered around how the data helps consumers generate revenue or reduce costs, chief data officers can develop guidelines for establishing the economic value of data in specific circumstances. Use-case questions may include the following:

- »» Does it improve customer prospecting efforts?
- »» Can it minimize the risk of regulatory violations?
- »» Can it help a company avoid expensive equipment failures?

You can bill customers and partners in many ways for using your data. Common pricing models include the following:

- » **Free:** Access to data at no charge.
- » **Fixed or “flat” fee:** Consumers pay a one-time fee. Fees can be per record, per data set, or per attribute (if enriching data).
- » **Subscription-based:** Consumers pay on a recurring basis, typically monthly or annually.
- » **Consumption/usage-based:** Consumers pay based on the volume of queries they run.

For data that is streamed or continuously updated, monthly subscription fees are common. However, if your data set is relatively static, then most of the value is derived at the initial time of consumption, with less value over time. In these instances, asking customers to pay a one-time fee rather than charging them a subscription may be more cost-effective.

You also can consider offering tiered pricing based on the type of usage. Consider a “freemium” model in which a limited data set is available for free exploration and development, along with premium access to additional data and data services. Across all paid pricing models, it is helpful to include a usage-based component. This ensures that your customers pay for the value they derive.



TIP

When vetting data marketplaces, look for one that provides in-platform purchasing to decrease the burden on you to manage billing infrastructure.

Optimizing Your Data Listings

If you've chosen to monetize your data products and services via a data marketplace, be sure to optimize your data listings for both discoverability and time to value. Your goal in doing so is twofold: Get as many consumers as possible to view your listing, and make it as easy (and fast) as possible for consumers to see the value of your product.

The features available for optimizing your listing will vary based on your chosen data marketplace. However, you'll want to follow these tips:

- » **Customize your listing to your target audience.** Increase the visibility of your listing by improving search and discovery, and demonstrate product relevancy by including industry-specific use cases and relevant example queries.
- » **Make data broadly available.** To reach as many consumers as possible, make your data products available in all major clouds and their regions and industries relevant to your target audience.
- » **Offer free samples.** Vetting the quality and relevancy of a data product can be a lengthy process. To accelerate time to purchase, consider offering a free sample of the data product. Some marketplaces even support a “try before you buy” option through which you can allow consumers to preview a subset of the data on a self-service basis.
- » **Document your product.** To help consumers quickly determine your product's relevance, be sure to include clear documentation, such as a data dictionary and data format descriptions, in addition to such information as possible use cases and industry applicability.
- » **Iterate.** Optimizing your data listings isn't a one-time job. Instead, you should continually experiment and evolve your strategy. Although traditional data marketplaces don't provide visibility into listing performance, modern data marketplaces provide view, click-through, and usage data to help you refine your marketing tactics.

Following these tips will help you increase awareness of and demand for your listing, ultimately accelerating time to value for both you and your consumers.

Evaluating Data Marketplaces

When evaluating data marketplaces, look for one that offers the following features to help you optimize your listings:

- » Modern data sharing technology for seamlessly sharing data products and services across cloud providers and regions

- » User-friendly search and discovery tools
- » Rich listing information, such as sample use cases, example SQL dictionaries, visualizations, and data dictionaries
- » Governance and security controls, such as fully revocable access
- » Visibility and usage metrics to provide transparency into demand and performance
- » Try-before-you-buy functionality to enable consumers to vet a sample of your product prior to purchase
- » In-platform purchasing to decrease the burden of managing billing infrastructure

IN THIS CHAPTER

- » Exploring a modern data sharing architecture
- » Allowing access to database tables and views for data consumers
- » Looking at a modern data sharing example
- » Protecting sensitive data with secure views

Chapter 7

Enabling Live Data Sharing with a Modern Architecture

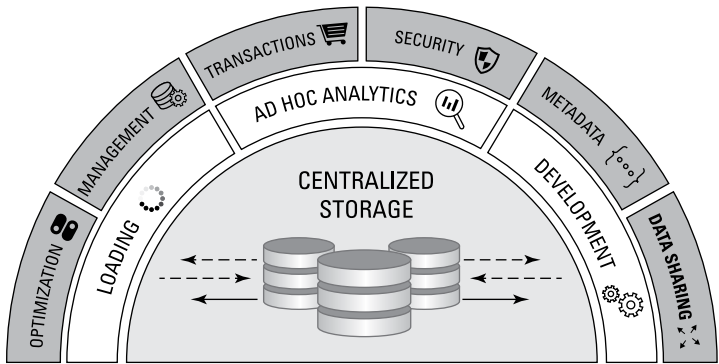
This chapter discusses the architecture that enables a modern cloud data platform and its data sharing and data marketplace capabilities. You'll also learn how real-time data sharing actually works in order to easily and securely share and access data.

Capitalizing on a Modern Architecture

With a traditional data platform, fixed compute and storage resources limit *concurrency* — the capability for a near-infinite number of users and workloads to simultaneously access the same data and resources. As shown in Figure 7-1, with a *multi-cluster, shared data architecture*, compute and storage resources are separate and can be scaled independently to leverage the near-infinite resources of the cloud. Ideally, this architecture should be cloud

agnostic, providing a consistent layer of services to each cloud provider and region for a seamless experience.

The Modern Data Sharing Infrastructure



■ Services □ Compute □ Storage

FIGURE 7-1: The architecture of a modern data platform — built for the cloud with compute, storage, and services completely separated but logically integrated.

Decoupling storage, compute, and services

The separation of storage and compute resources is a fundamental differentiator of a modern cloud data platform architecture and modern data sharing. All data is stored in the cloud, in optimized form, and without any loss of data fidelity.

A single copy of the data stored in a modern cloud data platform — a single source of truth — can be accessed concurrently by virtually any number of independent compute clusters, enabling an organization to perform practically any number of concurrent data workloads, including data sharing.

Decoupling storage and compute is critical for sharing data. Decoupling enables nearly any number of data consumers to directly access shared data, with each user accessing their own dedicated compute power available from their modern cloud data platform. But data consumers don't pay for storage costs because they are accessing read-only versions of data that doesn't move. And the data provider doesn't pay for any of the compute resources that a data consumer uses to access and analyze shared data.

This modern architecture offers many advantages. Query speeds on shared data can be exponentially faster and fortified with near-limitless storage and compute resources. Data providers can grant read-only access to their live, ready-to-use data in a secure and governed environment. Data consumers can then combine (JOIN) shared data to augment and deepen their data analytics.

Only the scalability, elasticity, and flexibility of a multi-tenant cloud data platform, supported by a multi-cluster shared data architecture, makes it possible to store data from diverse sources and share that data among a large number of data consumers without contention for resources.

Making shared data usable requires access to data and coordination across all data consumers to ensure consistency, security, and performance. That's why the services layer is so important. It ensures everything is secure, properly maintained, tuned, and optimized for self-service access. This type of architecture makes it possible to efficiently exchange data from one centralized system, with dynamic elasticity. Data providers can share a virtually unlimited amount of data, yet data consumers pay only for the data and resources they use. Global metadata, transactions, and security are all managed from here, making the services layer the control tower that tracks, logs, and directs access to data for every database element and object within the data platform, as shown in Figure 7-2.

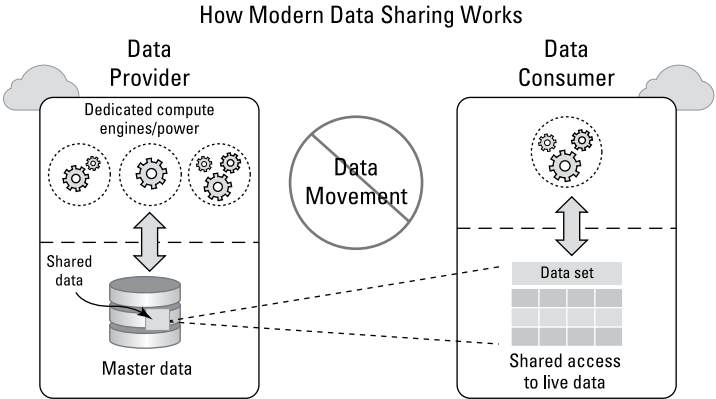


FIGURE 7-2: The architecture that underpins modern data sharing, enabling live data access between a data provider and data consumer, without resource contention.

Additionally, the services layer provides transactional consistency across all data providers and consumers, ensuring all data users see a consistent view of live and up-to-date data. A data provider can update shared data in real time. Likewise, after transactions are committed, data consumers can simultaneously view the data provider's updates and immediately query the shared data — all with transactional, ACID-based consistency.



TECHNICAL
STUFF

ACID is a consistency model that defines a set of properties to ensure transactions in a relational database are valid, even in the event of multi-statement transactions, processing errors, power failures, and crashes. The properties of ACID are:

- » **Atomicity (“all or nothing”):** Every operation in a transaction must succeed for the transaction to complete. If a single operation fails, the entire transaction rolls back, and the database state is left unchanged.
- » **Consistency:** The completion of any transaction brings the database from one valid state to another valid state.
- » **Isolation:** Concurrent transactions don't contend for access to the data and run as if each transaction executed sequentially.
- » **Durability:** After a transaction is committed, it remains committed.

Data providers depend on ACID properties to ensure data integrity for all transactions.

Boosting concurrency

With modern data sharing, large numbers of concurrent data consumers can access the same shared data, as shown in Figure 7-3. Automatic scaling of concurrency takes simultaneous query processing even further in modern data sharing by automating the scaling of additional compute engines dedicated to each share transaction without manual intervention. In contrast, the architecture of traditional data platforms forces all providers and consumers to compete for shared resources, creating a struggle to deliver optimum performance and consistency.

How Modern Data Sharing Works — Concurrency

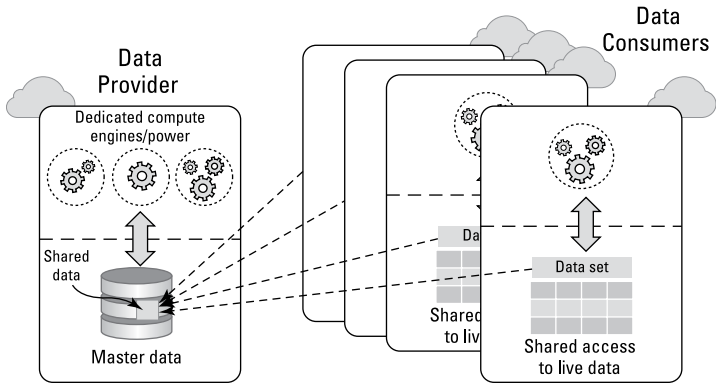


FIGURE 7-3: Near-unlimited data sharing concurrency with a modern data sharing architecture.

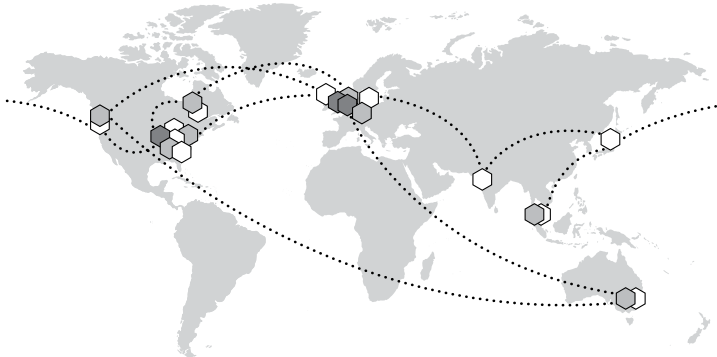
Enabling sharing cross-cloud and cross-region

Flexera’s “2021 State of the Cloud Report” found that 92 percent of enterprises have a multi-cloud environment. Likewise, a business rarely operates within the confines of a single geography. Sharing data with consumers agnostic to their cloud provider or region is necessary for effective collaboration.

When modern data sharing occurs within the same geographic region, regardless of the cloud provider, data is shared or accessed without moving or copying it. If the data provider and data consumer are in different regions of the same cloud provider or different clouds altogether, modern data sharing enables you to automatically provision data to the consumer’s region. In these two cases, data is replicated on demand to minimize *egress costs* (fees charged when data is exported from a cloud provider) and *latency* (the time it takes to retrieve the data).

Modern data sharing technology utilizes live, transactionally consistent shares and allows the data provider to configure refresh cadences down to the minute, saving data providers money and ensuring the most current data is available at all times. Only a modern cloud data platform with seamless, cross-cloud data sharing technology allows providers to share data simply by creating roles that grant consumers read-only access to the data, as shown in Figure 7-4.

Securely Share Data Within and Across Clouds and Regions



- ① **Cross-Cloud and Cross-Region**
 - Live, ready-to-query, no ETL
 - On-demand provisioning
- ② **Data, Logic, and Services**
 - Personalized data and metadata
 - Multilanguage UDF/external functions
- ③ **Secure Sharing**
 - Multiparty governance
 - Revocable access

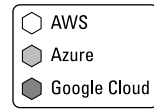


FIGURE 7-4: A modern cloud data platform allows you to share live, ready-to-query data across clouds and regions without extract, transform, and load (ETL) and other legacy data sharing procedures.

Modern data sharing uses extremely economic data replication methods to perform these operations. Incremental updates follow an initial bulk replication procedure as changes to the shared data occur. The initial upload and all secondary transactions should occur automatically without impacting query performance.

If your data platform lacks these cross-cloud and cross-region capabilities, then sharing data is much more expensive and involved because the data provider must work at the data storage layer and the data warehouse layer. For example, within a cloud provider's environment, the data provider would have to replicate data from one storage bucket to another, which creates a great deal of avoidable labor. The data provider has to launch a compute cluster to ingest data, and the consumer has to pay for the compute resources required for these processes.



With modern sharing, you can share data wherever your consumers are. Within each geographic region, all accounts can access a single live version of the data. When data sharing spans multiple regions or multiple public clouds, data should be automatically replicated and managed by a data platform that spans public clouds, reducing latency and improving performance.

Increasing Data Security

Chances are, you have sensitive data in your database. With modern data sharing, you're not limited to sharing entire databases or entire database tables. If portions of a table are subject to strict security and confidentiality policies, sharing the entire table exposes the sensitive data. With *secure views*, you can control access to shared data and avoid security breaches, as shown in Figure 7-5. Chapter 8 describes secure views in greater detail.

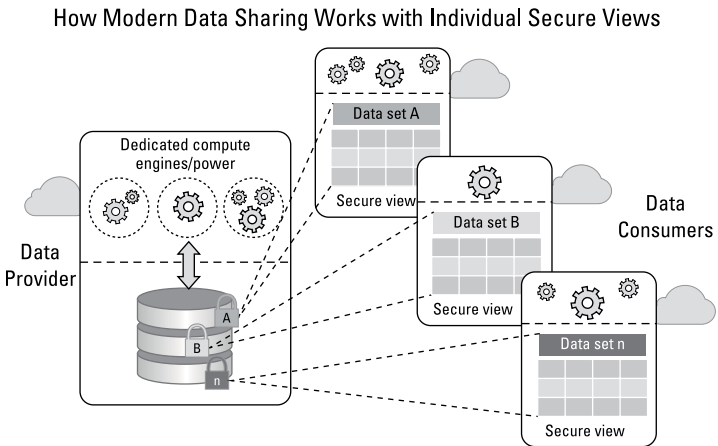


FIGURE 7-5: Modern data sharing that enables secure views allows data providers to protect against access to sensitive data.

Sharing business logic

Modern data sharing allows you to share personalized data and metadata, but also business logic to ensure your ecosystem has the data and tools it needs to collaborate. This may include user-defined functions (UDFs) written in Java, Python, or another

procedural language, with robust governance and security controls, such as row-level access policies and data masking.

Secure shared UDFs allow the data owner to limit the specific types of questions and analyses that can be performed against their data. They also let a user ask specific questions of detailed data without giving that user the ability to directly view or export the raw underlying data. Chapter 8 provides examples of how to use secure shared UDFs.



REMEMBER

Data providers can easily share data and logic while controlling access with a secure view. Sensitive data is protected, and data consumers gain access to non-sensitive data for their own analytics, without having to replicate data in the same cloud and region.

Enabling fine-grained access

In traditional data sharing environments, the data provider controls what data consumers can and can't see in the database. Modern data sharing introduces *database roles* and *database shares* that give data providers and the data consumers more control over who can see what data. A database administrator on the consumer side can set up fine-grained access to certain parts of the data set via roles.

For example, the administrator might set up a finance role to allow users in the accounting department to see the database tables pertaining to payables, receivables, and revenue, and a marketing role that lets marketing professionals view customer data to monitor advertising campaigns. Setting up this type of fine-grained access with modern data sharing technology is a simple two-step process:

1. The data provider's administrator specifies what data the data consumer organization can access within a database.
2. The data consumer's administrator, familiar with the users and departments at his or her organization, then matches the individuals with the roles to grant them access to relevant database objects.



TIP

Within these secure data sharing scenarios, data providers should have controls to fully revoke access, enabling them to easily comply with industry regulations, such as the "right to erasure." Additional control is possible with row-level access, which restricts users to query only certain rows within each table. Chapter 8 further discusses these concepts.

IN THIS CHAPTER

- » Knowing your data
- » Complying with government and industry regulations
- » Keeping sensitive data secure
- » Selectively sharing and controlling access to a data set

Chapter 8

Governing Your Data

Breaking down data silos across your ecosystem and centralizing data in a common repository makes governance easier. This chapter focuses on the three major aspects of good governance: knowing your data, controlling your data, and unlocking your data to share it securely across teams and with external consumers (see Figure 8-1).

Knowing Your Data

Data governance entails knowing what data you have, where it resides, who is authorized to access it, and how each person is permitted to use it. This has become increasingly difficult as enterprises attempt to rationalize first-, second-, and third-party data spread across disparate systems and often locked in silos. Organizations need to securely integrate their data with crucial data generated by their partners, suppliers, customers, and industry peers, while complying with data security and privacy regulations.

Organizations that rely on older technologies, such as legacy data warehouses, must copy and share the same data many times, creating a data governance nightmare. Additionally, File Transfer Protocol (FTP) methods and application programming interfaces (APIs) are hard to track when distributing data sets via cloud storage. To make data governance worse, employees often manually share files and spreadsheets via thumb drives or email. Without

the right technology, fulfilling compliance requirements is difficult. When part or all of a data set is copied, you have to apply your efforts to multiple copies of data sets, possibly in multiple locations. Protecting, classifying, and tracking the usage of data becomes progressively more difficult.

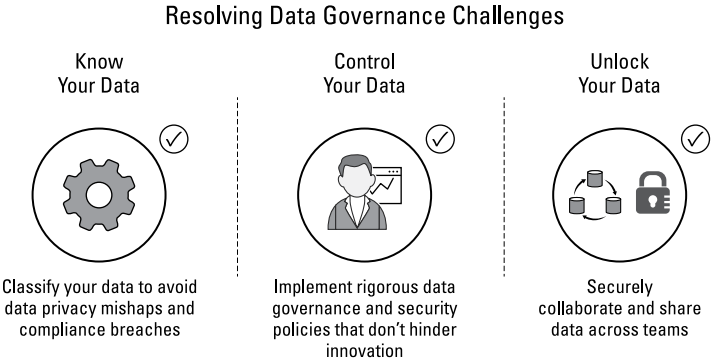


FIGURE 8-1: Comprehensive data governance is the bedrock of successful data sharing practices.

Many organizations have concerns about the proper use of personally identifiable information (PII), protected health information (PHI), competitive data, and other types of sensitive information. In some cases, they must adhere to strict regulations governing the security and privacy of consumer data, such as the European Union’s General Data Protection Regulation (GDPR), the United States’ Health Insurance Portability and Accountability Act of 1996 (HIPAA), and the California Consumer Privacy Act (CCPA). These regulations must be observed throughout the entire life-cycle of your data — from creation and storage to usage and sharing to archiving and deletion.



TIP

You can ease your compliance efforts by keeping a single “source of truth” of live data in a single location and granting on-demand access to governed slices of the data.

Controlling Your Data

Fulfilling data privacy and protection requirements is much simpler when you have the right technologies in place. Instead of dealing with copies of data in many locations, with modern data

sharing, you can establish a single copy of your data, governed by advanced technologies for accessing and sharing it.

As described throughout this book, a modern cloud data platform enables data providers to control access to data and secure views. Once the data provider makes data available, the data consumer can then query the data. Without moving data across environments, instant access is made possible because all data is maintained and updated in a centralized repository, orchestrated by global metadata management services.

You will need strong data security and flexible governance controls no matter the types of data sharing use cases your organization deploys. Several common scenarios are described below.



TIP

Instituting comprehensive controls reduces the risk of compliance violations. All data governance strategies should seek to protect sensitive data while it is accessed and shared.

Governing how consumers interact with your data

To maximize data availability while minimizing risk, organizations must create flexible data-access policies — with centrally enforced protections and controls. If these procedures lack flexibility and are not universally enforced, then the responsibility for governance falls on the individual teams that wish to provision and share data, creating extra work for employees. To improve flexibility and minimize exposure of sensitive data, a modern data sharing environment should include the following protections and controls:

- » **Interaction controls**, such as secure views, secure joins, and secure user-defined functions (UDFs), are applied as people interact with the data.
- » **Traceability tools** allow users to track data where it lives to ensure protections are continually applied and allow for data deletion where appropriate (such as the “right to be forgotten”).

Secure views allow people to control access to data and avoid potential security breaches, such as allowing customers to see only specific rows of data from a table and not to see rows that pertain to other customers.

Secure joins can establish discrete linkages (to people, devices, cookies, or other identifiers) without exchanging or making visible any PII.

Secure UDFs allow data consumers to link, join, and analyze fine-grained data while preventing other parties from viewing or exporting the raw data.

Protecting your data

Organizations concerned about protecting sensitive data can control access at a granular level. Common methods of data protection include the following:

- » **Row access policies** allow users to see only the information that is relevant to them. For example, you might implement a policy that ensures sales reps can see rows of customer data only for their own accounts, while regional managers can see all the customer data from their regions.
- » **Dynamic data masking** selectively masks data at query time. You can use this technology to store PII without making it visible to unauthorized users while still enabling robust analytics on that data.
- » **External tokenization** turns the data into a random string of characters with no meaningful value if the organization's system is breached. You can then dynamically detokenize the data at query runtime. Once data is tokenized, visibility is very similar to dynamic data masking: If you're authorized, you see the clear data; if you're unauthorized, you see the tokenized data.

Classifying and identifying your data

Classification and identification policies help you avoid data privacy mishaps and compliance breaches by tracking the types of data in use, the lineage of that data, and how that data changes. For example, you can use *object tagging* to control access to confidential and sensitive information, such as salary amounts and Social Security numbers. You might set permissions so that line-of-business managers can see salary information for employees within their department, while an HR manager can see this information for all employees.

Applying Data Governance to Common Business Scenarios

Nearly every organization recognizes the strategic value of data to streamline operations, deliver personalized customer experiences, and open up new market opportunities. Acquiring second- and third-party data allows companies to better understand their customers, refine their go-to-market strategies, and make more informed decisions.

In all these scenarios, data governance ensures data is properly classified, accessed, protected, and used. To determine what type of data governance and security you need, start by identifying the scenario that most closely matches your use case.

Sharing the same data with one or multiple consumers

In this situation, the need is simple: You have a table or view you'd like to share with one or many data consumers (see Figure 8-2). For instance, if a retail chain wanted to share its entire database with 50 franchises, it could use this method. The process is relatively straightforward. First, the retailer needs to create a *share* — a database object that grants permissions and data access to the data consumers.

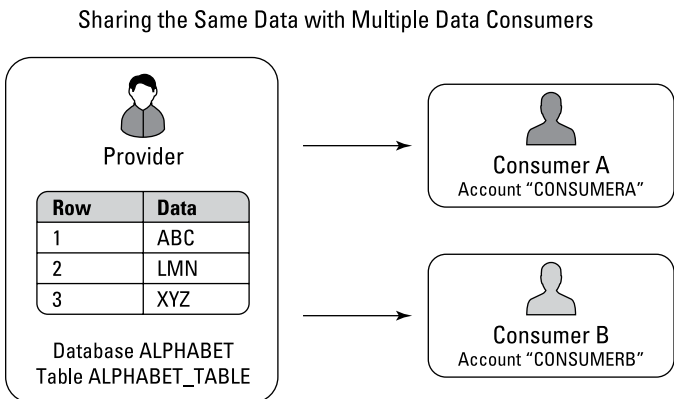


FIGURE 8-2: Using permissions to share the same data with multiple consumers.

To complete the process, the retailer would add permissions for all related database objects it wants to share and then add the 50 franchise accounts to that share. This could be automated via a scheduled script, or a workflow, or provisioning system using Python or SQL.

Sharing different subsets of data with multiple consumers

This scenario is more sophisticated because it involves sharing different subsets of data with multiple consumers. For instance, a car manufacturer might want to share data with its dealers about production plans. Some dealers may be confined to a single state, whereas others span many states. Using this method, the manufacturer could share the pertinent data with each dealer, regardless of dealer location.

For example, Figure 8-3 shows data in one table divided into two groups. Group 1 contains row 1, visible only to Consumer A. Group 2 contains rows 2 and 3, visible to all consumers.

Sharing Different Data Subsets with Multiple Data Consumers

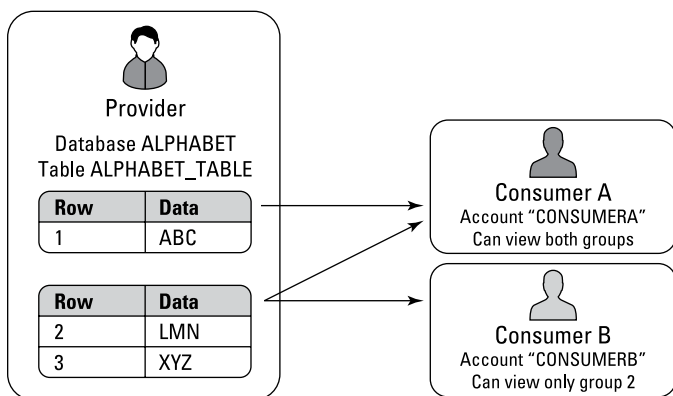


FIGURE 8-3: How to share the different subsets of data with multiple consumers.

Facilitating this kind of sharing requires a couple of additional steps. First, the data provider creates an *entitlements table* to match the dealer account names to the set of objects each dealer needs to access. Second, the manufacturer must create a *secure view* that controls the data each dealer can access when that dealer runs queries.

Sharing predefined slices of data with secure views

If any portion of a table is subject to security and confidentiality policies, sharing the entire table could expose sensitive data (Chapter 7 touches on this). *Secure views* allow you to control access to the data and avoid potential security breaches.

Secure views are an effective way to enforce table-, column-, row-, and even cell-level security when sharing data between organizations external to each other. Data providers can create secure views of their data and share access to those views with other users of the cloud data platform, even if these users are in other organizations. This is also a useful method for organizations that grant database access directly to multiple end customers. Secure views allow these customers to see only their specific rows of data from each table, but not the rows that contain data about other customers.

When should you consider using secure views? Consider the following guidelines:

- » To allow data consumers to access data without compromising security
- » When PII and PHI are involved
- » To restrict data access when multiple consumers need access to the same database

For example, for online retailers to plan inventory levels, they need to share merchandise and sales data with their distributors. However, if the data set that contains the sales data also contains sensitive customer ID information, that information must be blocked and protected.

Unifying Your Data

Your data access policies must be simple, transparent, and universally applied across all workloads so that they don't require any extra effort for data scientists, business analysts, data engineers, and countless other individuals who interact with the data. Data governance policies must make life easier for these workers, not more difficult.

For example, data scientists might use Spark to create and train machine learning models, whereas business analysts use a data visualization tool to query the same data. If governance is decentralized, each team has to set up access policies to govern how data is used in their respective environments. A modern data sharing environment gives you a centralized place to define these policies one time and apply them everywhere. All database objects are centrally maintained and updated by the cloud data platform, in conjunction with end-to-end security, governance, and metadata management services. Rather than physically transferring data to each group for its particular project, you can authorize people in all groups with read-only access to a governed portion of a live data set, centrally maintained and accessible via SQL.

Working on a common, shared data set simplifies bringing workloads together and applying centralized governance frameworks. With a cloud data platform, all data management capabilities are available in a single cohesive environment. By sharing data within this modern platform, you can enable *live access* to any subset of your data for virtually any number of data consumers, inside and outside your organization, to support analytics endeavors and other data-driven initiatives. Shared data can be accessed by large numbers of concurrent users, without competing for resources.



REMEMBER

If you have a single copy of your data in one location, controlling the data and updating it becomes easy. It's also easy to monitor who interacts with your data.

HOW TO CENTRALIZE GOVERNANCE

Unifying the policies that apply to data, workloads, and people enforces consistency in the following ways:

- **Data:** Unify data in a universal repository to simplify data management and make establishing and enforcing access policies easier.
- **Workloads:** Grant live access to any subset of your data so that various project teams can collaborate on data-driven initiatives.
- **People:** Unify governance policies so they can be applied to people of all skill levels and functions.

Chapter 9

Six Steps to Advance Your Business with Modern Data Sharing

This chapter outlines six key steps to help your organization get started with modern data sharing.

Step One: Identify Opportunities

What types of data and insights are valuable to internal and external consumers? Will you allow these consumers to access just data, or will you enrich that data by adding data analytics services? Will you look for data monetization opportunities? Start by identifying these data sharing scenarios internally. Then, reach out to your business partners to get their thoughts on how they can benefit from using your data — and what kinds of data they can share with you. Focus on identifying the data that has the

potential to produce the most value. Here's what to look for as you formulate your data sharing business plans:

- » **Examine existing data flows:** Identify the data sharing already taking place across your organization and with external stakeholders to understand who generates the data, who consumes the data, and how these processes are executed.
- » **Evaluate workflow processes:** Identify the tools and methods currently used to share data. This, combined with the mapping of data flows, will help you determine the current costs, technical hurdles, and limitations associated with data sharing.

Step Two: Define Stakeholders and Roles

Who will spearhead your data sharing initiatives? Do you have data stewards in place to manage the data? Are there incentives to encourage teams to share data? Appoint stakeholders to evangelize the data culture. Make sure to identify the organizations, internal and external, that must be brought on board. All data sharing relationships have at least two stakeholders — the data provider and the data consumer(s).

Step Three: Verify the Capabilities of Your Data Sharing Platform

Once you've identified outdated data sharing methods, such as FTP and email, consider how you can replace them with modern data sharing technology. Do you have a modern cloud data platform that can easily and cost-effectively enable modern data sharing? It must be able to support these key principles:

- » **Data does not move across environments:** Modern data sharing enables data consumers to access live, ready-to-query data across clouds and across regions from a single cloud data platform.

- » **Real-time updates:** Consumers should see changes as soon as a data provider updates its data, which should happen without extra work by the data provider or consumer. Having this level of consistency and integrity increases the value of your data.
- » **Centralized data security, privacy, and governance:** Modern data sharing requires granular control so you can meet the needs of each data consumer while protecting sensitive data and complying with industry regulations, such as the General Data Protection Regulation (GDPR). Your cloud data platform should give you security, flexibility, and control.
- » **Supports a viable data marketplace:** Tapping into an existing data marketplace via a cloud data platform allows you to access third-party data and data services with the same modern data sharing technology.

Step Four: Implement a Proof of Concept

Data consumers want to experiment with new data sets and verify the value of data products and services. Execute a proof of concept (PoC) to determine how well the data platform serves your needs — and the needs of your constituents. Your PoC should validate assumptions about all or most high-value requirements, including ease of migrating your data to the new solution, loading new data, running queries, and handling your most critical data workloads at speed and near-limitless scale. Most importantly, your PoC must demonstrate the business value derived via data sharing.

Step Five: Operationalize Your Data Sharing Strategy

Internal and external data sharing require clearly defined business terms, licensing conditions, and audit processes. Develop a flexible data sharing agreement that documents which data is shared and how the data can be used. Each agreement should identify pertinent roles and responsibilities, stipulate usage rights, and list any restrictions. Establish a data governance council to coordinate cross-company policies and best practices, oversee data

sharing rules, encourage cross-department collaboration, and track how data is used.

Do you plan to use third-party data and data services from a data marketplace or monetize your data and data services by offering them to customers via a marketplace? Follow the steps below to participate in a data marketplace.

For data providers:

- » Select a data marketplace powered by modern data sharing.
- » Create one or more marketplace listings, along with a provider profile.
- » Optimize your listings by including data dictionaries, use cases, and sample queries.
- » Monitor listing usage to inform future offerings.

For data consumers:

- » Discover third-party data and services that will add value to your business.
- » Validate the data or services' relevancy and quality by using try-before-you-buy features.
- » Enrich your data by securely accessing ready-to-query data or data services via the marketplace.

Step Six: Evangelize to Develop a Data Sharing Culture

To demonstrate the benefits of modern data sharing to your stakeholders, include the concept in your data literacy program. Demonstrate the productivity improvements your organization will gain and, if applicable, forecast the revenue potential for monetizing your data. You should be able to develop a complete picture of the ROI potential for modern data sharing. You will then be well on your way to taking data sharing to new levels of capabilities and opportunities for your organization.

Reveal previously unimagined business opportunities with modern data sharing

The Internet, smart phones, and cloud computing have profoundly transformed business. Modern data sharing has already begun to do the same. At the heart of all these groundbreaking technologies is the ability to use data in ways never imagined before. Modern data sharing enables any two or more organizations to seamlessly share live data across any major cloud and its regions, and between clouds, locally and globally. With near-unlimited access to data, you can easily reveal untold insights about your business, your customers, and the many opportunities that await you. Read on to learn about modern data sharing and what's possible for your business.

Inside...

- How modern data sharing works
- Why early methods limit what's possible
- The business value data sharing delivers
- Why a data marketplace is key to success
- How to monetize your data
- Ways to govern and protect shared data
- Real-world, industry-specific case studies



David Baum (david@dbaum comm.com) is a freelance business writer specializing in science and technology.

Go to **Dummies.com**™
for videos, step-by-step photos,
how-to articles, or to shop!

ISBN: 978-1-119-85620-7

Not For Resale



for
dummies®
A Wiley Brand

WILEY END USER LICENSE AGREEMENT

Go to www.wiley.com/go/eula to access Wiley's ebook EULA.