

LEARNING MADE EASY

Alteryx Special Edition

Cloud Analytics Automation

for
dummies[®]
A Wiley Brand



Democratize data
across the enterprise

—
Improve governance
and auditability

—
Enable real-time analytics
and machine learning

Compliments
of
alteryx

Lawrence Miller

About Alteryx

Alteryx powers analytics for all by providing the leading analytics automation platform. Alteryx delivers easy end-to-end automation of data engineering, analytics, reporting, machine learning, and data science processes, enabling enterprises everywhere to democratize data analytics across their organizations for a broad range of use cases, both on-premises and in the cloud.

Cloud Analytics Automation

**for
dummies[®]**
A Wiley Brand



Cloud Analytics Automation

Alteryx Special Edition

by Lawrence Miller

for
dummies[®]
A Wiley Brand

Cloud Analytics Automation For Dummies®, Alteryx Special Edition

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

Copyright © 2023 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. 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.

ISBN 978-1-394-17129-3 (pbk); ISBN 978-1-394-17130-9 (ebk)

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.

Publisher's Acknowledgments

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

Project Editor: Elizabeth Kuball

Acquisitions Editor: Ashley Coffey

Editorial Manager: Rev Mengle

Senior Client Account Manager:

Matt Cox

Production Editor:

Tamilmani Varadharaj

Table of Contents

INTRODUCTION	1
About This Book	1
Foolish Assumptions.....	2
Icons Used in This Book.....	2
Beyond the Book.....	3
CHAPTER 1: Discovering the Exciting New World of Cloud Analytics Automation	5
Why Automate Your Analytics?.....	5
Unlocking Analytics Automation Like Never Before in the Cloud	7
CHAPTER 2: Transforming Analytics Automation in the Cloud	11
Enabling Greater Data Volumes	11
Democratizing Data across the Enterprise	12
Improving Governance and Auditability.....	12
Increasing Manageability and Scalability.....	14
CHAPTER 3: Looking at Different Methods for Automating Cloud Analytics.....	17
Comparing ETL and ELT.....	17
Working with Code-Centric Approaches to Transforming Data (the “T” in ELT)	21
Step 1: Discover the data	22
Step 2: Write the data transformation rules.....	22
Step 3: Apply the data transformation rules	23
Step 4: Validate the data	23
Empowering Users with Low-Code and No-Code Solutions.....	24
CHAPTER 4: Scaling Your Cloud Analytics.....	25
Empowering People with Democratization and Self-Service.....	25
Solving Big Data Processing Challenges.....	27
Enabling Real-Time Business and Decision-Making Processes.....	27

CHAPTER 5: Exploring Common Cloud Data Automation Use Cases 31

Reporting Analytics 31

Data Onboarding..... 33

Data Modernization and Democratization..... 35

CHAPTER 6: Ten Questions to Ask Before Purchasing a Cloud Analytics Automation Platform..... 37

Introduction

In the not-too-distant past, most data storage and compute operations took place exclusively in on-premises data centers, which limited scale and was often prohibitively expensive. As the amount of data being generated continued to grow at an exponential rate, it was becoming increasingly challenging for traditional data warehouses to keep up with business requirements and scale in a timely and cost-effective manner.

With the cloud-computing era, most of these constraints became a thing of the past. The cloud enables large-scale processing of data at lower costs, increased scale, and reduced time. Today, the cloud is the de facto solution for data processing and performing advanced analytics to gain business insights.

Traditional data warehouses have increasingly moved from on-premises to the cloud, leading to the proliferation of cloud data warehouses such as Google BigQuery, Amazon Redshift, Snowflake, and Databricks. These repositories can consolidate data from multiple sources; store the data securely; maintain a high level of consistency, quality, and accuracy; and transform the way data is delivered and consumed.

About This Book

Cloud Analytics Automation For Dummies, Alteryx Special Edition, consists of six chapters that explore the following:

- » Why you should automate your analytics and how the cloud enables new analytics capabilities (Chapter 1)
- » How the cloud transforms the world of analytics automation (Chapter 2)
- » What different techniques are available for the automation of cloud analytics (Chapter 3)
- » How to scale your cloud analytics (Chapter 4)

- » Which common data automation use cases can be solved in the cloud (Chapter 5)
- » What important questions to ask when evaluating a cloud analytics automation partner (Chapter 6)

Each chapter is written to stand on its own, so if you see a topic that piques your interest, feel free to jump ahead to that chapter. You can read this book in any order that suits you (though I don't recommend upside down or backward).

Foolish Assumptions

It has been said that most assumptions have outlived their usefulness, but I assume a few things nonetheless!

Mainly, I assume that you're interested in learning how to automate your data analytics. Maybe you're a business executive with limited knowledge of cloud and automation technologies and you're trying to understand how cloud analytics automation can deliver business value. Or maybe you're a more technical reader — such as a chief data and analytics officer (CDAO), IT executive, or data engineer — with deep knowledge of analytics and analytics automation, but you're looking for more information about the specific advantages of the cloud and how to implement analytics automation in the cloud.

If any of these assumptions describes you, then this is the book for you! If none of these assumptions describes you, keep reading anyway — it's a great book, and after reading it, you'll know quite a bit about cloud analytics automation.

Icons Used in This Book

Throughout this book, I occasionally use special icons to call attention to important information. Here's what to expect:



REMEMBER

This icon points out important information you should commit to your nonvolatile memory, your gray matter, or your noggin — along with birthdays and anniversaries.



TECHNICAL
STUFF

If you seek to attain the seventh level of nerd-vana, then perk up! This icon explains the jargon beneath the jargon and is the stuff legends — well, legendary nerds — are made of.



TIP

Tips are appreciated, but never expected, and I sure hope you'll appreciate these useful nuggets of information.



WARNING

These alerts point out the stuff your mother warned you about. Well, probably not, but they do offer practical advice to help you avoid potentially costly mistakes.

Beyond the Book

There's only so much I can cover in this short book, so if you find yourself at the end wondering, "Where can I learn more?" go to www.alteryx.com.

- » Welcoming the data analytics revolution
- » Leveraging the power of the cloud for analytics automation

Chapter **1**

Discovering the Exciting New World of Cloud Analytics Automation

This chapter explores the explosive growth of the digital universe and explains why cloud analytics automation is the key to unlocking new opportunities in this bold new world.

Why Automate Your Analytics?

Data is everywhere. According to predictions by IDC, the digital universe will grow to 175 zettabytes (that's 175 trillion gigabytes!) by 2025. Enterprises everywhere are creating new data everywhere, and the explosive growth of the Internet of Things (IoT) has further contributed to the digital universe, with sensors collecting volumes of data about every facet of our lives and our world every day.

Unfortunately, we only use a small fraction of the data that is available to us. Too much of the data that comprises our digital universe is effectively lost in a black hole. Much of it is siloed

in on-premises databases and enterprise data warehouses, or cloud-based software as a service (SaaS) applications. And although nearly half of all data is processed and stored in the public cloud, we've only just begun to extract the full value of all this data.

Cloud analytics automation is revolutionizing data usage and business outcomes. This is no trivial statement. The powerful combination of cloud computing and analytics automation may be the most important advancement of the decade and could lead to new innovations and digital transformation initiatives that are, well, transformational.

Some important benefits of cloud analytics automation include

- » **Increasing data usability:** Increased insights from existing data assets are a huge, largely untapped treasure trove of value for businesses. According to a recent State of Dark Data report by TRUE Global Research and sponsored by Splunk, approximately 55 percent of an organization's data is considered "dark" — unknown, undiscovered, unquantified, underutilized, or completely untapped.
- » **Reducing manual work:** Cloud analytics automation reduces the burden on limited IT and technical resources. IT and data science resources spend as much as 80 percent of their time cleaning and preparing data for use (more on this topic in Chapter 3). Automating data preparation tasks, or even tasks like machine learning and report building, frees your limited technical resources to work on other strategic value-adding projects, reduces errors, and significantly lowers the cost of utilizing your data.
- » **Accelerating response time:** Cloud analytics automation enables faster response times, which promotes greater organizational agility to react to new information.



REMEMBER

Much of the data that is generated every day, particularly IoT sensor data, is highly perishable. If the data isn't analyzed and consumed quickly, it becomes useless. For example, although meteorological data and traffic-camera feeds may provide climatological or historical trend value, to maximize the value of this data it needs to be analyzed and consumed in near real time.

Unlocking Analytics Automation Like Never Before in the Cloud

As organizations everywhere embrace the cloud — 96 percent of respondents to the Flexera 2022 *State of the Cloud Report* are using at least one public cloud (see Figure 1-1) — they're able to realize real benefits including the following:

- » Democratized access to data automation
- » The ability to easily connect disparate data in centralized sources for new types of insights
- » Dramatically reduced costs
- » Always-on data products and pipelines delivering speed (faster time to value), lower cost of utilization, and ongoing and continuous insights

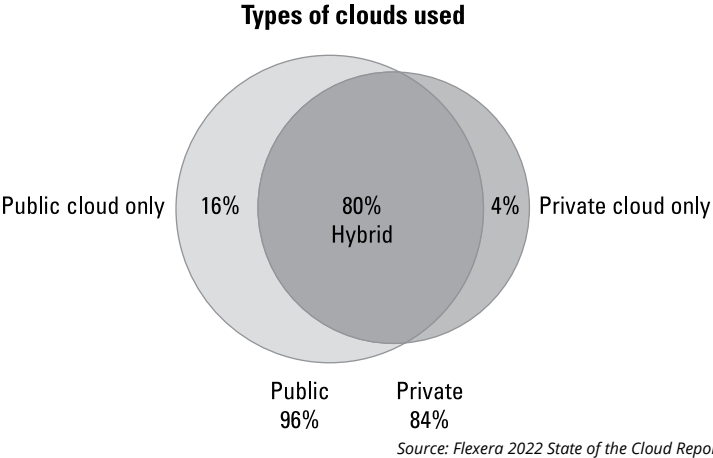


FIGURE 1-1: Public and private cloud adoption has proliferated.

Cloud computing has been with us for more than a decade. Major public cloud providers today include Amazon Web Services (AWS), Google Cloud Platform (GCP), and Microsoft Azure. These providers offer an ever-growing array of services — including SaaS, platform as a service (PaaS), and infrastructure as a service (IaaS) — that enable practically unlimited, on-demand access to cloud-based resources when and where they're needed.

Leveraging the cloud for analytics automation enables enterprise customers to connect their disparate data sources across multiple clouds (including SaaS applications) and their on-premises data centers. With access to massive compute and storage resources in the cloud, enterprises can maximize their use of artificial intelligence (AI) and machine learning (ML) technologies to glean meaningful insights from all their data. Cloud analytics automation truly has the potential to revolutionize our world!



TIP

Today, the cloud is the de facto solution for many when it comes to processing data and performing advanced analytics to gain business insights. Traditional data warehouses have moved from on premises to the cloud, leading to the proliferation of cloud data warehouses such as Google BigQuery, Amazon Redshift, Snowflake, and Databricks. These repositories can consolidate data from multiple sources; store the data securely; maintain a high level of consistency, quality, and accuracy; and transform the way data is delivered and consumed.

AUTODESK REDUCES BUSINESS INTELLIGENCE REPORTING TIME AND HEADACHES

Autodesk makes software for people who make things. If you've ever driven a high-performance car, admired a towering skyscraper, used a smartphone, or watched a great film, chances are, you've experienced what millions of customers are doing with Autodesk.

Challenges

The Market Intelligence team was tasked with creating a 360-degree view of all of Autodesk's construction customers who don't own their construction processing tool, so the company could proactively pursue upsell opportunities. Combining separate data sources with various business intelligence (BI) access points and large data sets using SQL and Microsoft Excel resulted in lengthy, unstable processes. Because the prep work was all being done manually with legacy platforms, it was impossible to report live updates, and each iteration for an individual report took at least three hours.

Solution

Autodesk built a modernized data ecosystem that included AWS for staging and storage, Snowflake as a centralized data warehouse, Alteryx's Designer Cloud for data preparation and automation, and Power BI for visual analytics. This data stack created a simple, streamlined, and dynamic workflow.

With the data in a centralized location, analysts at Autodesk could use Designer Cloud to easily join and transform data from various sources and automate the pipeline to BI dashboards. This provided their internal customers with real-time insights. If modifications were necessary, analysts could easily update Designer Cloud recipes within minutes to adjust for any changes.

"With Designer Cloud and Snowflake, we've created a new centralized data ecosystem that has cleaned up the way that we do analysis. Our reporting process has been reduced from hours to minutes, and we are able to get deeper and more proactive insights from our data. For myself as an analyst, Alteryx has just decreased a lot of headaches," says John Gardner, a senior business intelligence analyst at Autodesk.

Results

- **Faster reporting:** With Designer Cloud, Autodesk was able to automate the data workflows, which would previously take a week or two every quarter to update the business rules with legacy systems.
- **Project scalability:** The Market Intelligence team is able to maintain the same project completion rate with one analyst as they were with four analysts previously.
- **Deeper and proactive insights:** Analyst bandwidth has been freed up to focus on deeper insights, allowing the field team to drive more revenue.

IN THIS CHAPTER

- » Unlocking the potential of data
- » Making data accessible to everyone
- » Ensuring data governance and auditability
- » Maintaining the quality of your data at scale

Chapter 2

Transforming Analytics Automation in the Cloud

This chapter explains how cloud analytics automation enables greater data volumes, data democratization, better governance and auditability, easier manageability, and massive scale.

Enabling Greater Data Volumes

The cloud creates new opportunities for enterprise customers by providing practically unlimited access to compute and storage resources.

Unlike on-premises data-center environments that are too often constrained by server and disk capacity and ineffective performance and capacity planning, the cloud offers on-demand access to powerful capabilities when and where they're needed.

But compute and storage capacity are just part of the story when it comes to data in the cloud. The cloud also enables massive volumes of data to be collected and generated from potentially

billions of data sources from the Internet of Things (IoT). Unlocking new insights in these massive treasure troves of data is the next frontier in data analytics, creating new and exciting opportunities for businesses in every industry.

Democratizing Data across the Enterprise

Too often, data analytics is relegated to the realm of IT and data scientists that have a relatively limited understanding of the business and its requirements. Unfortunately, the business users typically have a relatively limited understanding of data management and lack the skills required for technical tasks like preparing data or building machine learning models (see Chapter 3 for more on this topic).

Data democratization is about empowering business users with easy access to data and intuitive yet powerful data analytics and reporting tools. It enables business users to help themselves with self-service capabilities that accelerate time to value in business data insights.

Data democratization drastically reduces, and often completely eliminates, the traditional back and forth between IT team members and business users trying to align business and technical requirements. By empowering the users who understand their data needs best, data democratization enables better decision-making and greater productivity across the organization.

Improving Governance and Auditability

Everything in the world of data is growing at an alarming speed, meaning questions of governance and auditability must be considered up front — this isn't something you can easily fix later on.

Effective governance ensures you can trust your data. The ability to trust your data hinges on trusting the process you use to clean it. Among other things, this means having a full audit trail to understand lineage and chain of custody.

It's not enough to just communicate your results — you need to communicate the steps that got you there. Show your work. This is critical for meeting external compliance requirements (take the regulatory reporting needs of financial services firms as an example — banks are required to fully document their data systems and data transformation efforts), as well as for your own internal credibility. To ensure your results can be reproduced, understood, and trusted, you have to be able to audit how and when the data was used, as well as who used it. Be transparent about data usage in order to build trust, ensure consistency, and remove potential bias.

The cost and complexity of this kind of data governance necessitates a metadata-driven approach that is self-documenting and provides built-in lineage, auditing, and controls.

For example, take a look at data governance for data transformations.

Legacy approaches that involve writing tons of custom code or ad hoc manipulation in spreadsheets make it impractical and, in some cases, impossible to provide the right level of transparency. For example, when scripting is used, understanding the data transformation process requires a full code review, often walking through thousands of lines of code to ensure integrity.

At the other end of the spectrum, in situations where spreadsheets are used, the changes are not rules driven, so they're neither consistently repeatable nor collectively verifiable. Worse yet, without any metadata, many of the changes made in spreadsheets are destructive, leaving no clues as to what changed, how it changed, and who changed it.

This is why it's essential to use a data preparation platform that will systematically track any trace of data transformation, building in the governance and controls that will establish and maintain data provenance in an automated way. If it's a separate process, divorced from the work itself, it will add unnecessary overhead and slow you down, or it won't get done. If it's locked up in complex code, it will be largely inaccessible — too hard to review, share, and validate. Solving these problems involves balancing efficiency in doing the work with the need to make changes to

the data self-evident and unambiguous. Certain data-preparation platforms can ensure that auditability is a natural by-product of the act of cleaning the data.



REMEMBER

Because of the speed at which data architectures are changing, today's best-of-breed tools are built with the cloud in mind — but that doesn't mean you can sacrifice governance and security. Technology must support best-in-class security and governance standards, so you can track and manage your data across projects with built-in authentication, authorization, and encryption capabilities.

Increasing Manageability and Scalability

Cloud analytics automation also increases data manageability and enables massive scalability. Data manageability is crucial to maintaining data quality and ensuring your organization is able to maximize the value of its data. Deploying a SaaS solution across an organization greatly simplifies manageability compared to desktop software that must be installed and regularly updated. Likewise, scalability ensures that your data analytics capabilities grow as your business needs grow. The cloud offers massive on-demand scalability to meet these requirements, allowing users to run data jobs at any scale. A common problem for organizations is that users often must spend hours processing a large data workflow. With the cloud, the same processing can be done in minutes, allowing your users to work seamlessly with your data, regardless of how large the data set is that you're working on.

Data stored in software as a service (SaaS) applications is a prime example of how the cloud enables better manageability and scale. Unlike enterprise software that is locally installed on desktop and laptop computers, SaaS applications run in the cloud and allow users to access the applications and their associated data from anywhere. SaaS applications run in the cloud instead of relying on the local computer's processing capabilities. Additionally, the data in SaaS applications remains in the cloud, enabling better security and governance (see the previous section).

CALLAHAN INCREASES MEDIA IMPACT WITH GOOGLE CLOUD DATAPREP

Callahan is a brand strategy and digital marketing agency. Founded in 1982, the company is headquartered in historic downtown Lawrence, Kansas, with satellite offices in Plano, Texas, and Rancho Dominguez, California.

Challenges

Callahan sets itself apart as a marketing agency with its use of data. The agency collects data from a huge variety of sources that extend well beyond a client's marketing function in order to best assess baseline practices and identify opportunities. This approach, while successful, was not without its challenges — preparing and integrating data pulled from SAP, LinkedIn, Shopify, Facebook, Salesforce, and so on was time-consuming and tedious. Working with demanding clients only exacerbated these challenges; it became a race against the clock for Callahan to prepare a given client's data sources fast enough to communicate campaign results. As an agile team without dedicated data engineering resources, Callahan needed a solution that would allow everyday analysts to accelerate data preparation.

Solution

Callahan assembled a modern cloud data warehouse stack built on Google Cloud that includes Google Dataprep (the Google-native version of Designer Cloud), Google BigQuery, and Fivetran. Because the stack is dominated by SaaS technologies, Callahan got up and running in minutes — there was nothing to install or configure. Google Dataprep, in particular, has greatly accelerated all the cleansing and transformation work required to integrate multiple data sources. Google Dataprep automatically suggests the best transformations for the data set, such as removing null values or standardizing state names, and it allows Callahan to easily join columns together or remove them from an analysis. The result is faster, more robust data preparation, which leads to better analysis and more impactful client results.

(continued)

(continued)

“This technology structure has fundamentally changed the way we do business, and it has given us a competitive advantage,” says Zack Pike, vice president of data strategy and marketing analytics at Callahan.

Results

- **Seventy percent of Callahan analyst time now spent on analysis.** Callahan analysts get to do what they do best — analyzing data, not managing or maintaining it.
- **Ninety percent increase in media impact for clients.** Callahan was able to more quickly identify successful ads versus those performing poorly.
- **2X increase in new customers for clients.** Pinpointing the leads that mattered allowed Callahan to convert those leads more readily.

- » Reviewing legacy ETL and ELT processes
- » Recognizing challenges in code-centric approaches
- » Preparing data with low-code/no-code solutions

Chapter 3

Looking at Different Methods for Automating Cloud Analytics

This chapter explores traditional approaches to building data pipelines and their limitations, including extract-transform-load (ETL), extract-load-transform (ELT), and code-centric approaches. It also explains how low-code/no-code solutions can accelerate time to value for your business data.

Comparing ETL and ELT

ETL tools were created roughly 25 years ago to automate much of the tedious coding that was required to retrieve and cleanse data for business use. ETL was originally designed to handle data that was generally well structured, often originating from a variety of operational systems or databases from which the organization wanted to create reports (see Figure 3-1). ETL pipelines were built for a specific set of users and use cases.

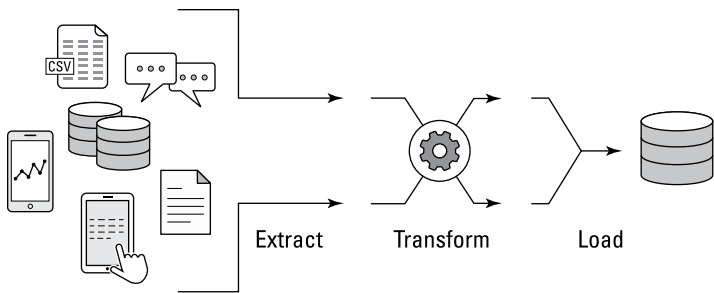


FIGURE 3-1: ETL refers to the process of extracting data from disparate sources, transforming data into standard schemas, and loading data into repositories.



TECHNICAL
STUFF

ETL is a data integration process used to copy, combine, and convert data from different sources and formats and load it into a new destination such as a data warehouse or data lake. It consists of the following individual processes:

- » **Extract:** During the extraction process, structured and unstructured data is pulled from multiple sources and likely in multiple formats, such as JavaScript Object Notation (JSON), Extensible Markup Language (XML), non-relational databases, scraped websites, and so on.
- » **Transform:** Data transformation brings together data of different formats and changes it to a more useful format so that it can be used across an organization. For it to be successful, the technical requirements of the target destination and the needs of users must be considered. This could mean checking which character sets are supported by the system, checking what type of coding the warehouse uses, or creating a new value relevant to a specific analysis. Data cleansing is another vital step to transformation and includes removing duplicates, unwanted nulls, and whitespaces, as well as modifying data type and size. Data transformation logic can be developed/built with either no-code or code-centric approaches.
- » **Load:** Loading involves writing transformed data to its storage location, whether a data warehouse or a data lake, on premises or in the cloud. After data is loaded, you must ensure all data was migrated and check for errors to verify the data quality.

Today, much of the architecture and data surrounding ETL has changed. In addition to data warehouses, data lakes and data lakehouses are the common end target. The data itself has become much bigger and messier. And even the use cases, which were once clearly defined, have become more experimental in nature.

Perhaps the biggest difference today is that instead of providing data for a few business groups, ETL pipelines are expected to serve a wide variety of users and use cases across an organization. Each of these users requires different data that has been cleansed and transformed differently. But there's one commonality: They all want the data fast.

Traditional ETL pipelines have struggled to extend support for the self-service agility required by these emerging analytics use cases. ETL tools were built for IT users rather than business users, which often leaves business users waiting in line to get data cleansed and then passing requirements back and forth with IT until they receive their desired output.

Meanwhile, IT teams — once considered the target end user for all data operations — are struggling to offload some of the ETL cleansing and standardization tasks that business users are eager to take on.



REMEMBER

ETL refers to three steps (extract, transform, load) used to integrate data from multiple sources into a centralized repository. Data is extracted from its source, transformed into a usable format, and loaded into a system for analysis.

Although traditional ETL may be considered a bottleneck today, that doesn't mean it isn't valuable. The same basic challenges that ETL tools and processes were designed to solve still exist, even if many of the surrounding factors have changed.

For example, at a fundamental level, organizations still need to extract (E) data from legacy systems and load (L) it into their data lakes or warehouses. And they still need to transform (T) that data for use in analytics projects. ETL work needs to get done, but what *can* change is the order in which it's achieved and the new technologies that can be used to support this work.

Instead of an ETL pipeline, many organizations are now taking an ELT approach. This ELT approach follows a larger IT trend. Whereas IT architecture was once built in monolithic silos,

many organizations today are decoupling individual system components so that they can function independently in the ELT process. Decoupled technologies require less up-front work — technology stacks can be deployed without necessarily understanding all possible uses and outcomes — and enable more efficient maintenance.

ELT is popular in the cloud era because it uses the processing of the destination system to handle the transformation. When done as ETL instead, the transformations must be done beforehand, thus preventing this benefit from being realized. ELT is also popular in data lakes because it lets organizations perform the “T” after loading (you can load raw data into a data lake before you know what you need out of it, whereas data warehouses generally required you to transform it and get it in a specified form before loading).



TIP

The public cloud enables organizations to further decouple IT architecture by delivering on-demand consumption-based resources to automate analytics.

A clean separation between data movement and data preparation (often through the ELT process) takes advantage of the processing power of the source system and delivers specific benefits, including the following:

- » **Less friction:** The person or process loading the data isn't responsible for transforming it to a specification at load time. Postponing transformation until after data is loaded creates incentive for sourcing and sharing data. It also preserves the raw fidelity of the data.
- » **More control:** Loading data into a shared repository enables IT to manage all of an organization's data under a single application programming interface (API) and authorization framework. Thus, there is a single point of control at a granular file level.
- » **More flexibility and transparency:** Information can be lost as raw data is “boiled down” for specific use cases. In contrast, untransformed data can be reused for different purposes, leaving a record for auditing and compliance.

Working with Code-Centric Approaches to Transforming Data (the “T” in ELT)

Code-centric approaches to working with data are time-consuming, error-prone, and require expensive technical resources that are often difficult to recruit, retain, and replace. There simply aren’t enough developers or coders with knowledge of common data manipulation languages, such as SQL and Python, to meet the modern demand for data.

It’s also challenging for new IT resources to build and maintain custom code written by other developers. Documentation is often woefully inadequate, and companies are left struggling with custom-built applications when developers leave the organization — taking their tribal knowledge with them. This challenge is somewhat analogous to your grandmother’s recipe for chocolate chip cookies that has been handed down for generations but isn’t written anywhere — except we’re talking about a SQL script that may require several hundred lines to accomplish your data transformation tasks. So, instead of an easy-to-follow 20-step recipe, everything is made from scratch — you also need to know how to mill the flour, extract the sugar, and ferment the cocoa beans!



REMEMBER

Code is hard to maintain and optimize at scale. The complex process of cleaning and transforming data into a suitable format can consume up to 80 percent of your time.

Preparing data for analysis or any other initiative also requires constant collaboration between IT and the people who know the data best. Coders love to code — and some do it well and quickly — but they usually lack business domain expertise. On the other hand, business domain experts understand the business requirements for the data but typically don’t know how to write code.

There are four sequential steps to cleaning and transforming data with code (see Figure 3-2):

1. Discover the data.
2. Write the data transformation rules.
3. Apply the data transformation rules.
4. Validate the data.

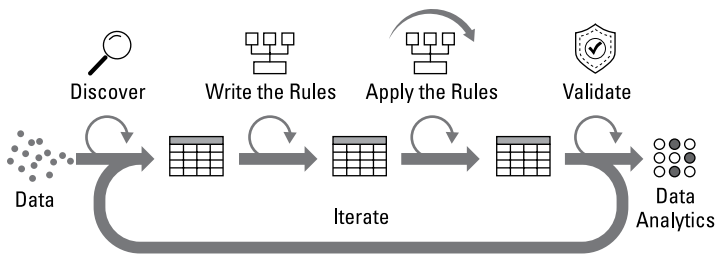


FIGURE 3-2: The four steps to preparing data for analytics initiatives.

These four steps are also iterative — there’s no “one and done.” The result of one step informs the next, and you frequently have to loop back to repeat earlier steps based on validation results and the need to further refine your data. When you’re relying on code to transform data, you have to wait for each step in the routine to be completed before proceeding to the next.

Let’s look more closely at each of these four steps to understand how and where code falls short.

Step 1: Discover the data

The first step is to assess the state of the data before it can be transformed. This is also called data profiling. Relying on code means creating algorithms to provide such data statistics as:

- » How many rows are in the data set?
- » What are the missing values?
- » Are the date formats correct?
- » Are there outliers in the data set? Are there duplicates?
- » What is the cardinality between two data sets?



TIP

Designer Cloud provides visual, automatic data profiling that dynamically adjusts as you work with the data.

Step 2: Write the data transformation rules

The next step is to develop rules to transform the data. But when you rely on code, you’re essentially “flying blind” — you don’t see the results of your data transformation rules until data is processed.

It's only at runtime that the data can be visualized. And this can take anywhere from hours to days, depending on data volumes.

Another drawback to hand-coding data transformation rules is that you often need a staging area to store intermediary results for reuse. Buying and maintaining these extra data storage structures consumes more resources, both human and financial.



TIP

Interactive data prep solutions allow you to see the results of data transformations in real time on samples of the data you're working with.

Step 3: Apply the data transformation rules

After writing the data transformation rules, you run the queries . . . and wait.

Applying the query means you run the code and then wait until the query is executed to look at the end of your data pipeline and validate that the result matches your expectations.

If you're dealing with large volumes of data, your wait time may be quite long. Changing your code or SQL queries to process a subset of the data or staging the data for a part of the algorithm to validate a subset of the data logic requires extra work and risks further delays.

Step 4: Validate the data

After data is processed, you need to validate its accuracy based on the specifications you defined. Here are some examples of validation questions:

- » Does the total sales figure account for individual transaction details?
- » Is a month of data missing?
- » Were any rows lost when combining two data sets?

The validated data has a different structure and content than the original source data. It's important to make sure no data issue is left undiscovered. If data issues remain, you have to iterate back to Step 1 or Step 2 and refine the queries.



WARNING

Hand-coding new pipelines for each new data set at the point of validation is tedious and time-consuming. It introduces potential errors and discrepancies in your data that can compromise the quality of your output and put your entire analytics initiative at risk.

Empowering Users with Low-Code and No-Code Solutions

Low-code and no-code solutions empower a new set of business users to access data, explore it to assess its content and quality, prepare it for use, get data insights, and build machine learning algorithms — while even handling some of the transformation functions of traditional ETL.

Low-code and no-code platforms are built to meet the needs of business users. For example, modern data preparation platforms like Designer Cloud incorporate visualization techniques and machine learning in order to make the data transformation process as intuitive as possible. Modern data insights tools like Alteryx Auto Insights can use machine learning to automate the tedious process of building insight-rich reports. And modern machine learning tools like Alteryx Machine Learning provide approachable machine learning for business users to build their own machine learning models.

The business benefits of a low-code/no-code data platform include the following:

- » Reduces time to value — even for those who know how to code
- » Less error-prone and easier to set up and reuse
- » Expands usability to nontechnical audiences, allowing organizations to scale their analytics automation
- » Reduces the iterative “back and forth” steps between business users and IT trying to clarify requirements until they get a workable data product
- » Allows business users to do most of the traditionally time-consuming work themselves, freeing IT to leverage their unique skills on other value-added initiatives

IN THIS CHAPTER

- » Maximizing your people resources
- » Leveraging pushdown optimization to address technology scaling challenges
- » Automating your business processes and workflows

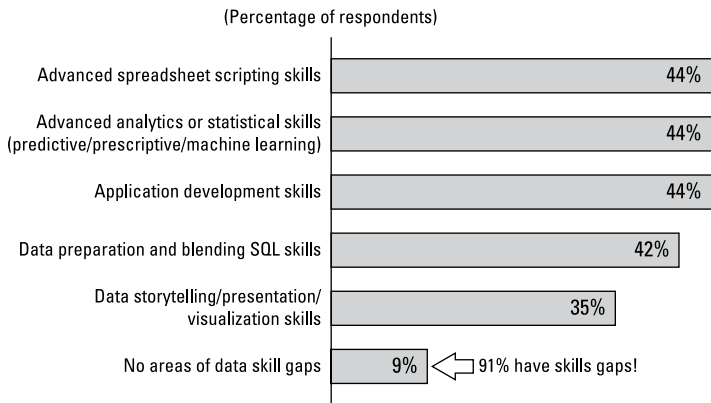
Chapter 4

Scaling Your Cloud Analytics

The cloud enables organizations to leverage high-volume data to gain powerful insights that were never before possible. However, with great power comes great scaling challenges. This chapter explores how democratization and self-service help you address people scaling, how big data processing helps address technology scaling challenges, and how real-time business and decision-making capabilities help address process scaling challenges.

Empowering People with Democratization and Self-Service

What good is high-volume data if you don't have people with the right technical skills to take advantage of it? According to IDC, advanced skills are among the biggest gaps reported by data workers (see Figure 4-1), indicative of a need for tooling that can simplify complex data and analytics activities — in other words, enable data democratization and self-service with a low-code, no-code cloud analytics automation solution.



Source: IDC Worldwide Data Science and Analytics, November 2021

FIGURE 4-1: Most organizations lack advanced skills in data and analytics.



TIP

To efficiently scale people and avoid missed opportunities associated with unused or underutilized data, organizations can undertake the following activities:

»» **Democratize data and analytics with tool reconciliation, consolidation, and worker upskilling:**

- Prioritize solutions that offer unified and automated data analytics experiences.
- Seek solutions that enable upskilling to improve data literacy and analytic outcomes.

»» **Understand the skill sets of data workers, and implement solutions that enable upskilling:**

- Data workers want to improve analytics and data application development skills.
- Data literacy is a challenge, but it's mandatory because decisions are increasingly data-driven.

»» **Review the data and analytics environment to understand the challenges that data workers face:**

- Multiplicity of tools impacting effective data operations and analytic processes.
- Complexity of data sources, formats, quality, scale, compliance, processing, and consumption.

- Manual and repetitive activities being performed in the absence of analytic process automation.
- Lack of self-service data preparation, report building, and machine learning due to technology and skills constraints.

Solving Big Data Processing Challenges

In the modern data stack, the hard work of data transformation — the (T) in extract-load-transform (ELT) — can be pushed into powerful cloud data warehouses.

Pushdown processing enables workloads to be pushed down to cloud data warehouses to leverage their large-scale computing power. With pushdown processing, your data workflows can run faster inside your cloud data warehouse. And business users, who are always hungry for more data, have a natural low-code/no-code ELT companion to get data ready for business intelligence (BI) initiatives faster than ever.

There are two main benefits of pushdown processing:

- » **Decreased processing times** (generally up to 90 percent faster)
- » **Decreased processing costs**



TIP

Alteryx offers a specialized product feature — pushdown optimization — designed to harness the power of your cloud data warehouse to do all your self-service data cleansing and transformation through SQL-based ELT.

Enabling Real-Time Business and Decision-Making Processes

The data we work with today has matured in terms of diversity, scale, and complexity — and this applies to data scientists and academic researchers, as well as the rest of us. Business analysts

across a spectrum of industries are asked to include larger volumes of data in their work, now pervasive due to the diminishing costs of collection and storage. Answering real analytic questions that drive business value means adapting methodologies to the reality of the data at hand.

To address this need, new self-service data analytics tools are gaining adoption, helping business users bring their domain expertise to bear on bigger, thornier data challenges. To effectively scale business processes and maximize the value of their data, organizations need to automate their existing processes as much as possible — build once and then move on (rinse and repeat, if you will). With the right foundation, data-dependent business processes can utilize live data that is fed automatically to inform key decision-making processes at scale.

MALWAREBYTES ACCELERATES MARKETING INTELLIGENCE

Malwarebytes protects you against malware, ransomware, malicious websites, and other advanced online threats that have made traditional antivirus software obsolete. In order to provide high-quality malware remediation, Malwarebytes has to sift through extremely large volumes of data to detect anomalies and quickly address those issues, processing more than 80+ petabytes of data over the lifetime of the company. Malwarebytes' success as an organization is in large part credited to its ability to work with large volumes of data.

Malwarebytes' marketing intelligence strategy centered around the following objectives:

- Lead sourcing through marketing activities
- Targeted marketing based on product usage
- Data standardization and scoring
- Campaign analytics and recommendations
- Multi-channel attribution blending Google Analytics with Malwarebytes' other sources

Pulling all this information together for effective analytics and quick turnaround requires flexible tooling and efficient processes.

To ensure optimal sales engagement, it's critical for Malwarebytes to reach out to leads within a short time frame after they come into their database. However, Malwarebytes had data-quality issues with the leads coming in and required complex lead scrubbing and lead scoring to ensure efficient sales outreach. Prior to Alteryx, Malwarebytes relied on a team of five marketing analysts doing manual lead scrubbing and scoring in Microsoft Excel. This process was error prone, time consuming, and not adequate to ensuring quick sales outreach.

Business challenges

- Malwarebytes marketing analysts spent the majority of their time manually cleaning data, taking away time from any meaningful analysis and process improvement.
- Manual work led to longer turnaround times, which impacted sales efficiency.
- A lack of automation and validation led to errors in the lead-scrubbing process, driving inconsistent sales decisions.

Technical challenges

- The Excel-based data transformation process was manual, inefficient, and time consuming, leaving data issues hidden.
- The process couldn't scale to the growing volume of data to deal with.
- There was a lack of lineage and governance in desktop-based Excel work.

Solution

Malwarebytes chose Snowflake as its data lake and data warehouse solution for its unlimited scale, on-demand execution clusters and support of a rich set of data formats. This capability has allowed Malwarebytes to centralize all data from transactions to billions of telemetry logs, to marketing leads and customer relationship management (CRM) data. For its less technical users, Malwarebytes needed a platform that combined ease of use, automation of redundant tasks, collaboration, and governance.

(continued)

(continued)

Results

- **Improved efficiency:** Malwarebytes was able to increase worker productivity by 80 percent, taking the work of five analysts and condensing it into one, freeing up time to work on other strategic initiatives.
- **Greater sales effectiveness:** 3x improvement in turnaround time, from weeks down to 48 hours. High-priority leads cut down to under 24-hour turnaround time. Eighty-five thousand leads have gone through vetting and qualifying with the new process.
- **Better data quality:** Automated and repeatable processes enabled by Designer Cloud have reduced the errors due to manual processes, which were more prone to human error.

Malwarebytes accomplishes all this using Amazon Web Services (AWS) as the platform foundation for its cloud technologies. The flexibility and cost effectiveness of cloud technologies has allowed a smaller company like Malwarebytes to create state-of-the-art data processes that used to only be available to companies with the resources to manage all the technology in-house. Malwarebytes leverages Kafka for streaming real-time data feeds into Amazon Simple Storage Service (S3); Qubole for data processing and Amazon Athena for data integration and extract-transform-load (ETL); Designer Cloud for data preparation; Snowflake as its analytics data warehouse; both Looker and Tableau for analytics and reporting; and Amazon Sagemaker, R, and Python for machine learning (ML) and artificial intelligence (AI).

- » Creating meaningful reports
- » Onboarding data
- » Modernizing and democratizing data

Chapter 5

Exploring Common Cloud Data Automation Use Cases

This chapter explores several common use cases for data automation across different industries, including reporting analytics, data onboarding, and data modernization and democratization.

Reporting Analytics

Reporting analytics enables powerful insights that drive important business decisions in every industry. Reporting tools allow report writers and analysts to transform raw data into meaningful and relevant information that can be easily consumed by business users. Some reporting tools, like Alteryx Auto Insights, even go a step further by automating the process of building and sharing reports.

Automating reporting analytics can reduce delays that inevitably occur in manual processes that require report writers and analysts to create reports with limited context about the data that is needed and its business purpose.

BIZDOG SAVES ITS CUSTOMERS \$300,000 A YEAR WITH DESIGNER CLOUD

BizDog's business journey started more than ten years ago when its founder sold his accounting practice, wrote a book about using data analytics to grow revenue, and started speaking to and consulting for companies around the world.

Today BizDog Group provides revenue and profit-driving insights for clients worldwide by analyzing customer lifetime value from their very first interaction with the company.

Challenges

BizDog focuses on midsize companies. A lot of work comes from private equity firm referrals. Almost every company has data sources from customer relationship management (CRM) systems, enterprise resource planning (ERP) systems, billing and customer support systems, and so on. Their clients are using manual processes to ingest data, including Google Analytics. Due to the data size with millions of rows and VLOOKUPs, the process was not scalable, making it challenging for their clients to meet their business goals. BizDog also used other methodologies for assembling large data sets together to build scripts and clean and process data. In the process, they faced challenging bottlenecks and were unable to easily provide insights on data lineage to clients when they were looking for it. With a customer base growing at 20 percent annually, BizDog needed to develop a platform-based service offering.

Solution

Designer Cloud delivers a collaborative self-service data engineering platform that has empowered BizDog and given them the ability to automate and enable business intelligence (BI) capabilities to allow for resultant data for client insights. This expanded marketing analytics has provided granular insights down to the keyword and acquisition level, giving a lifetime value to customer spend. BizDog Group's platform-based solution is now being marketed to companies globally.

"At BizDog, we like to give opportunities to people to enter the analytics profession who may not have traditional 'data' backgrounds.

Remote work allows us to provide opportunities to various groups, such as stay-at-home parents and single parents who need flexible schedules. Alteryx helps us with this as well, since we don't have to insist team members have strong technical skills. Instead, we can hire based on the person's values, attitude, and life experience and grow them quickly into high-performing team members," says Steve Wilkinghoff, founder and president of BizDog.

Results

- **Enriched data:** The marketing analytics platform helps optimize keyword selection, resulting in savings of up to \$300,000 per year.
- **Pull-forward revenue:** Timely onboarding of customers has allowed for growth; the base is growing at 20 percent.
- **Faster insights:** Ability to provide daily ad-spending insights instead of ad hoc analysis.

Data Onboarding

As discussed in Chapter 3, manual data onboarding processes such as extract-transform-load (ETL) and extract-load-transform (ELT) delay time to value.

Automating the data onboarding process accelerates the data pipeline and reduces opportunities for costly errors.

PEPSICO DELIVERS SMARTER SALES FORECASTING WITH ALTERYX

PepsiCo products are enjoyed by consumers more than a billion times a day in more than 200 countries and territories around the world. PepsiCo's product portfolio includes a wide range of enjoyable foods and beverages, including 23 brands that generate more than \$1 billion each in estimated annual retail sales.

Challenges

In order to supply the right product quantities to its retailers, PepsiCo must constantly calibrate sales forecasting. Any given sales forecast

(continued)

(continued)

may incorporate warehouse data, store stock data, or promotional forecast data, all of which is provided by retailers in a variety of different file formats and delivered to PepsiCo using a variety of different methods. The name of the game for PepsiCo is speed — prepare sales forecasts fast enough, and the team could anticipate spikes in demand and increase product shipments. But under Microsoft Access- and Microsoft Excel-based processes, the huge amount of time required to prepare this data meant that analysts only leveraged it once a month (or not at all), which risked under- or oversupplying retailers.

PepsiCo built a Microsoft Azure data lake that allows for centralized data storage and access. Analysts trained in Excel and SQL easily adapted to Alteryx and can now prepare and standardize diverse retailer data significantly faster for visualization in Tableau. Due to Excel's size limitations, PepsiCo analysts often had to stitch together multiple spreadsheets; with Designer Cloud, PepsiCo can see the full picture of its data and quickly spot any errors or inconsistencies. Because they're spending less time spent on data preparation, PepsiCo analysts are now able to dedicate the majority of their time to analyzing data and predicting trends, driving more value for the company.

"Designer Cloud brought an entirely new level of productivity to the way our analyst and IT teams explore diverse data and define analytic requirements. Our users can intuitively and collaboratively prepare the growing variety of data that makes up PepsiCo's analytic initiatives," says Ben Sokol, a data integration analyst at PepsiCo.

Results

- **Accelerated analytics:** Analysts were able to reduce the total reporting time by 70 percent and build dashboards for new customers 90 percent faster than with Excel and Access.
- **Expanded reporting:** Less time spent on collecting, cleaning, and preparing data means that analysts now have time to create dashboards for more customers, including online retailers.
- **Reduced errors:** Designer Cloud's visual profiling capabilities allow the team to easily identify gaps and outliers in the data to get ahead of larger issues.

Data Modernization and Democratization

Modernizing and democratizing data makes it more accessible to business users across the organization. Self-service creates new opportunities to expand automation across the organization.

AMWAY ADAPTS TO PRODUCT HIERARCHY CHANGES WITH ALTERYX

Amway is the world's largest direct selling company (with reported sales of \$8.8 billion in 2018). It manufactures and distributes nutrition, beauty, personal care, and home products, which are sold exclusively in 100 countries through Amway independent business owners (IBOs).

Challenges

Amway manufactures more than 450 different products, each of which needs to be carefully categorized and organized in its product hierarchy. Updates to the hierarchy need to be made daily as categories expand, business lines evolve, and new products develop. However, Amway's original desktop-based model was complex and required numerous manual updates, ultimately proving itself slow and unsustainable. Amway smoothed out some of these kinks after switching to a virtualization solution, but the solution's SQL-based transformations required too much hand-holding from the engineering team. Analysts and engineers had to communicate back and forth about data requirements until, days later, the outcome produced was as expected. Neither solution allowed for flexibility or agile changes to the product hierarchy.

Solution

Amway migrated its data to Google BigQuery and centered its product hierarchy around Google Dataprep. The transition to a cloud-based, self-service data engineering technology has dramatically reduced time to insight. Before, Amway would have to wait two to

(continued)

(continued)

three days before a change could be realized; now it's a matter of minutes. Much of this acceleration is credited to Google Dataprep's ease of use. Analysts can easily obtain product hierarchy data and make direct changes without communicating back and forth with engineers. Plus, Google Dataprep has improved visibility into the product hierarchy with clear audit trails and data lineage. The result is a more streamlined and agile product hierarchy that can quickly respond to new changes each day.

"We can update daily — it's allowed us to reduce that time to insight as far as making changes. Because as you know, the product hierarchy needs to be updated daily. We have new items coming in and new changes to items, and, through this process, we can allow this to happen in a rapid way. . . . We don't have to rely heavily on any technical staff to make this happen," says Kevin Schaefer, a senior data engineer at Amway.

Results

- **More agile analytics:** With Google Dataprep, Amway was able to reduce time spent making changes to its product hierarchy from two to three days down to five to seven minutes.
- **Reduced resourcing costs:** Amway's engineering team has been freed up to focus on more complex deliverables instead of responding to data requirements.
- **Improved business strategy:** Amway analysts can now obtain data at different hierarchy levels and, using their unique business knowledge, remodel the hierarchy for the most profitable option.

IN THIS CHAPTER

- » Supporting multiple clouds and cloud data warehouses
- » Enabling self-service analytics, collaboration, and sharing
- » Ensuring data governance, auditability, and security
- » Interacting with data in real time
- » Triggering automation with APIs and webhooks
- » Orchestrating data efficiently
- » Maintaining data quality over time

Chapter 6

Ten Questions to Ask Before Purchasing a Cloud Analytics Automation Platform

Here are ten important questions you need to ask when considering a cloud analytics automation platform for your enterprise:

- » **What cloud platforms does your product support?**
Enterprise data is everywhere — from on-premises data centers to multiple public clouds, such as Amazon Web Services (AWS), Google Cloud Platform (GCP), and Microsoft Azure. You need a cloud analytics automation solution that

allows you to utilize your data wherever it lives, on any cloud data warehouse or cloud data lake/lakehouse, or in on-premises databases and data warehouses.



TIP

Because the Alteryx Analytics Cloud is not tied to any one storage architecture, execution environment, or security framework, you have the freedom to choose the best solution for your team and business.

- » **Can your product leverage a cloud data warehouse for pushdown processing/execution? If so, which ones?** In addition to supporting a variety of cloud data warehouses, it's helpful if your vendor's product can leverage the processing power of your cloud data warehouse for job execution (known as *pushdown processing*). This can significantly reduce costs while increasing processing speeds by up to 90 percent or more. Ask your vendor if they offer pushdown processing, and for which cloud data warehouses.



TIP

As of this writing, the Alteryx Analytics Cloud currently offers optimized pushdown support for both Snowflake and Google BigQuery.

- » **Can your platform be used by line-of-business users for self-service analytics?** Look for a low-code/no-code data preparation platform that empowers line-of-business users with little to no technical skills to work directly with the data they need.



TIP

The Alteryx Analytics Cloud is designed for self-service use by business users. Intuitive, no-code approaches to data preparation, report building, and machine learning increase the efficiency of technical users and expand high-value data capabilities to those who don't know how to code.

- » **How does collaboration or sharing between team members work in your product?** An enterprise-grade cloud analytics automation solution should provide easy, yet powerful collaboration and sharing capabilities that allow team members to be more efficient and productive.



TIP

The Alteryx Analytics Cloud offers collaborative products where you can leverage the collective intelligence of your teams and the community. For example, Designer Cloud allows users to share workflows, recipes, and pipelines with the right permissions to streamline data operations at scale.



TIP

» **How does your product handle data governance and auditability?**

Effective data governance and immutable auditability are always critical to maintaining data quality throughout its life cycle. When seeking to empower self-service, strong data governance becomes crucial.

The Alteryx Analytics Cloud, consisting of Designer Cloud, Alteryx Auto Insights, and Alteryx Machine Learning, places a high priority on data governance. For example, with full audit trails and lineage, versioning, and software development life cycle (SDLC) support, Designer Cloud tracks and manages changes automatically across projects and environments.

» **How does your product ensure data security?**

The modern cybersecurity landscape is increasingly sophisticated and includes threats to both security and privacy. Your cloud analytics automation solution must safeguard data security and privacy while ensuring compliance with industry standards and regulations.



TIP

» **Does your platform allow users to interact directly with data in real time?**

Building data transformation logic from the command line, or in tools that require you to build and then test, can bog down workflows with endless back-and-forth iteration. The ability to interact with data in real time is crucial when it comes to accelerating time to value — especially when seeking to empower nontechnical users.

In the Alteryx Analytics Cloud, Designer Cloud uses intuitive data samples to allow business users to interact directly with data in real time. Users can see the results of their changes as they go, greatly accelerating time to value.



TIP

» **Does your product allow for automation triggered by webhooks and/or application programming interfaces (APIs)?**

Automation is only as good as its ability to integrate with your existing processes. The best analytics automation platforms provide built-in connectivity, as well as extensibility through webhooks and APIs, for maximum usability.



TIP

The Alteryx Analytics Cloud can integrate with existing workflows through software development kits (SDKs) and OpenAPI standards available in a multitude of languages. Orchestrate across third-party applications — from source control, ingestion, and replication tools to catalogs and business glossaries.

- » **Does your product allow for orchestration sequentially or in parallel?** When orchestrating data pipelines, you don't want your automation to be bogged down by preliminary tasks. Many analytics automation providers only offer sequential orchestration requiring earlier tasks to be fully executed before new tasks can be started — even if they're part of different backend processes.



TIP

In the Alteryx Analytics Cloud, Designer Cloud leverages the cloud to offer parallel orchestration, allowing users to automate data workflows in unlimited directions at the same time, without any backend process getting bogged down by a different process.

- » **What mechanisms does your product have to ensure that data quality stays consistent over time?** Every data workflow is subject to things like data drift or context changes in the meaning of underlying data. The best analytics automation platforms can detect these changes automatically, and alert you when data quality issues threaten the integrity of your data pipelines.



TIP

The Alteryx Analytics Cloud makes it easy to discover and validate data quality issues. Statistical data profiles are used to identify complex patterns, automatically suggesting possible quality rules such as integrity constraints, formatting patterns, and column dependencies. In addition, monitoring point-in-time and historical data quality trends provides the context for proactive alerting (via email, Slack, PagerDuty, and other platforms) to changes of schema and data distributions that may affect data fitness for downstream use.



alteryx

Data Transformation and Analytics in the Cloud

Quickly profile, transform, and pipeline your data for analytics and machine learning.

Start your **free trial** today

Unlock your data with cloud analytics automation

Our modern world is saturated with data, but few organizations are able to maximize the value of their data investments. It's time to make your data work for you with agility, flexibility, and scale. Organizations that successfully leverage cloud analytics automation with modern low-code/no-code solutions gain many advantages. Learn how the cloud can make data useful and readily accessible for business users, and start your journey toward building a solid foundation for future data and analytics success. *Cloud Analytics Automation For Dummies* shows you how to get started.

Inside...

- Explore users with low-code/no-code solutions
- Explore cloud data automation use cases
- Understand ETL and ELT limitations
- Increase manageability and scalability
- Enable greater data volumes and insights
- Solve big data processing challenges

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

alteryx

Lawrence Miller served as a Chief Petty Officer in the U.S. Navy and has worked in information technology in various industries for more than 25 years. He is the coauthor of *CISSP For Dummies* and has written more than 250 *For Dummies* books on numerous technology and security topics.

ISBN: 978-1-394-17129-3

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.