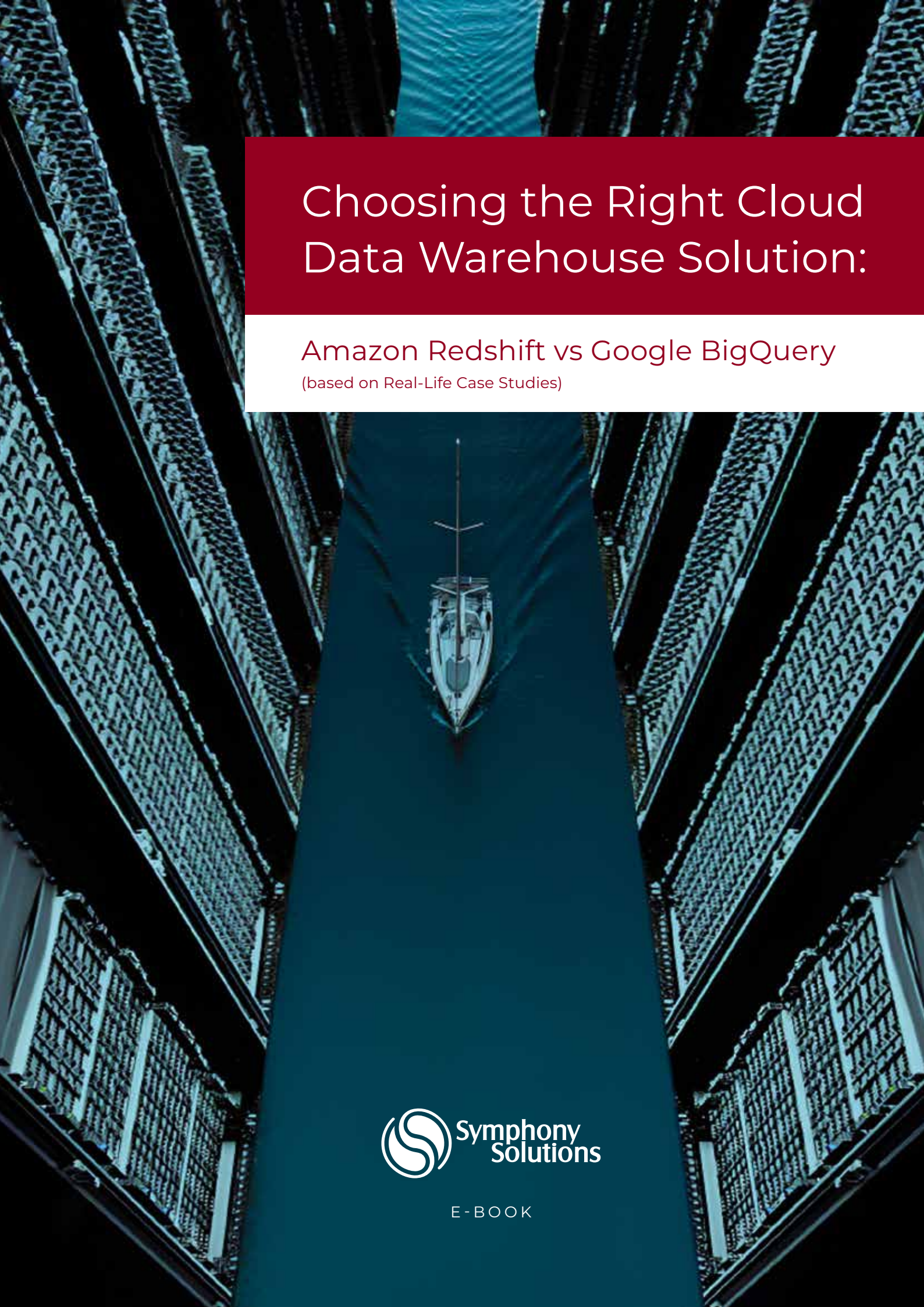


Choosing the Right Cloud Data Warehouse Solution:

Amazon Redshift vs Google BigQuery

(based on Real-Life Case Studies)

Symphony Solutions

E-BOOK

TABLE OF CONTENTS

INTRODUCTION

Today, having a cloud data warehouse solution is non-negotiable for modern businesses that want to stay ahead. From low total cost of ownership, improved performance and speed to improved access and integration, more secure data and better disaster recovery, the benefits if cloud data warehouse are undeniable.

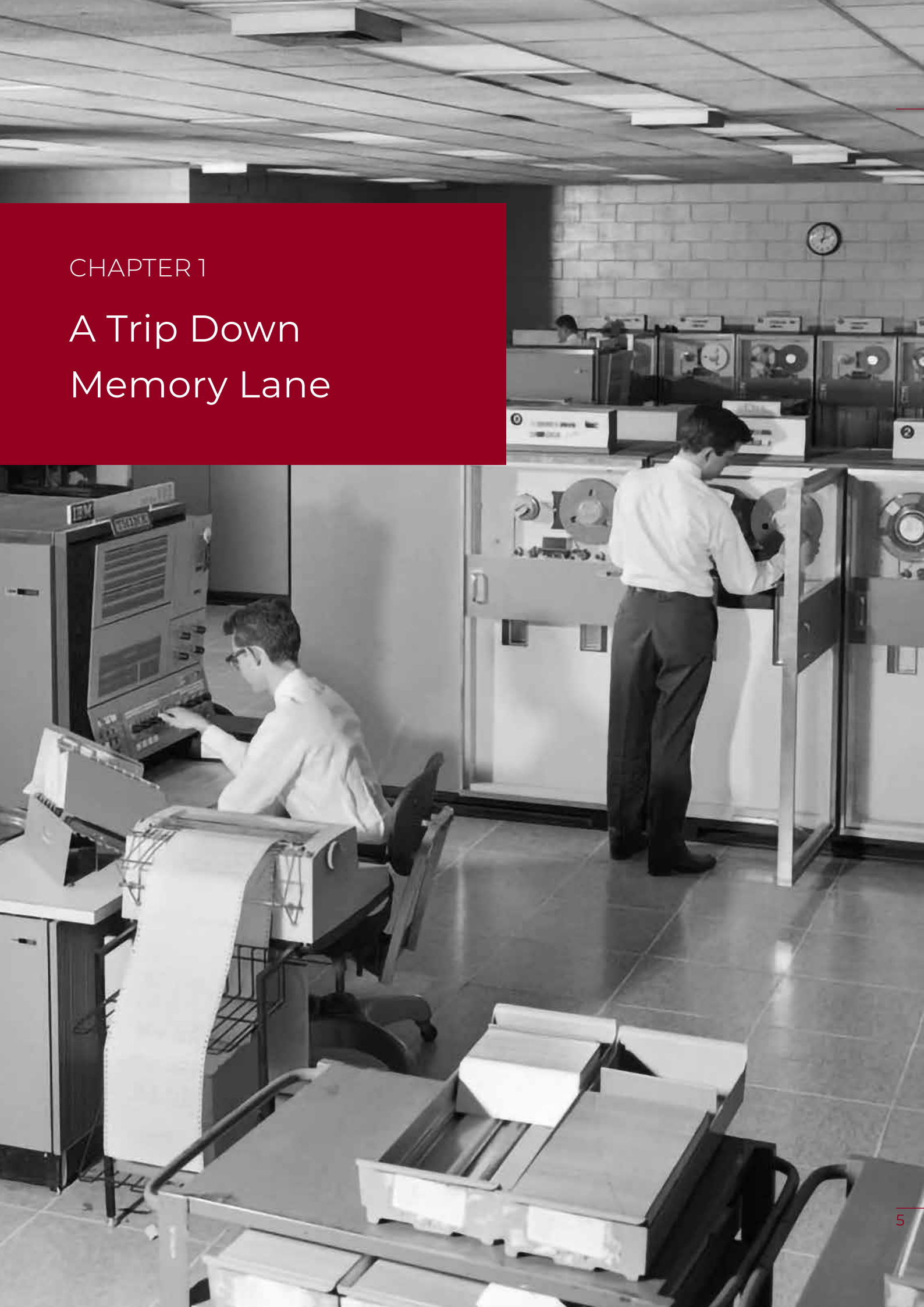
BUT even more important is choosing the right cloud data warehouse solution cloud provider. Should you go with one of the big two, Amazon RedShift or BigQuery? Or should you choose another cloud data warehouse beside these two popular ones?

In this insight-packed eBook we take a practical approach from our hands-on experience with our customers to show you the stark differences between Amazon RedShift and BigQuery as well as how they stack up to other cloud data warehouse solutions out there.

To help you make an informed decision, our experts share their experiences building data warehouses using either Amazon Redshift and BigQuery. Read on to find out why they chose either of the two options and make a regret-free decision.

ABOUT THIS BOOK

This ebook thoroughly explores cloud data warehouse solutions, using practical, real life case studies of our clients. Gain a better insight from how other companies in your industry decided (with our guidance) to choose either Amazon RedShift or Google BigQuery



CHAPTER 1

A Trip Down Memory Lane

The CDC 6400 Computer System, a 1960s mainframe computer used at Kitt Peak National Observatory, is shown here. Credit: NOIRLab/AURA/NSF

LOOKING BACK AT THE EARLY DAYS OF DATA MANAGEMENT

Data management entails the steps and process for gathering, organizing, and analyzing information. The concept dates back to the 1960s, after the rise of the Association of Data Processing Service Organizations (ADAPSO).

Businesses operating during the era leveraged outdated computers that required entire warehouse floors to sort data stored in punch cards. Tools used to manage data then included tabulators, sorters, and card punches. Data professionals use Absolute Machine Language or First Generation Programming Languages in a binary form to manage business information.

UNDERSTANDING THE ORIGIN OF DATA WAREHOUSING

The concept of data warehousing started gaining attention in the late 1980s after IBM researchers Barry Devlin and Paul Murphy built the first data warehouse. However, the industry credits American computer scientist Bill Innmon through his various forums and publications.

Innmon wrote and published the first book that sensitized users on the building process, usage, and maintenance of data warehouses. In his first conferences and in-person classes, Innmon describes a data warehouse as "a subject-oriented, non-volatile, integrated, time-variant collection of data in support of management decisions."

THE RISE OF CLOUD DATA WAREHOUSE SOLUTIONS

Cloud data warehouse solutions became a reality in 1999 (although the concept has been around since the 1960s) after Salesforce came up with an idea to offer application delivery via its website. Amazon then popularised the idea in 2002 with cloud-based internet services that extended to storage.

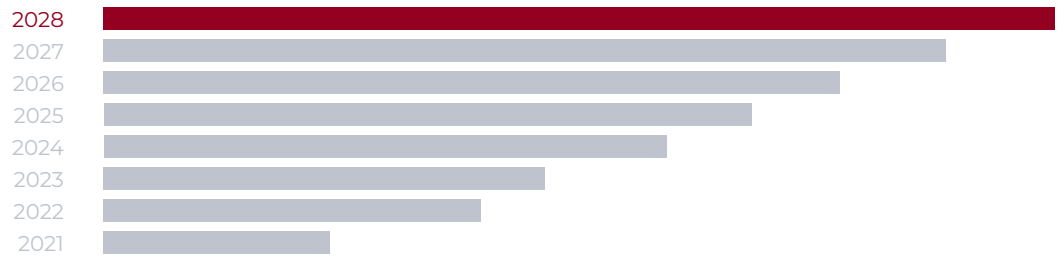
Since then, cloud warehouse solutions have emerged as a go-to cost-efficient way of storing and managing ever-growing enterprise data. As a result, Allied Market Research forecasts the global data warehousing market size to grow at a compound annual growth rate of 10.7% and peak at $51.18 billion by the end of 2028.

Global Data Warehousing is expected to reach

$ 51.18 billion

Growing at a CAGR of 10.7%*

GLOBAL DATA WAREHOUSING MARKET *OPPORTUNITIES AND FORECAST, 2021-2028*



* Global Opportunity Analysis and Industry Forecast, 2021–2028 by Allied Market Research®

WHO NEEDS A CLOUD DATA WARE-HOUSE?

A cloud data warehouse is for any business that wants to harness the benefits of powerful computing capabilities for real-time data analytics. Unlike on-premise data warehouse solutions, cloud management is faster and more accurate when gathering data and generating insights from multiple sources.

THE IMPORTANCE OF A CLOUD DATA WAREHOUSE

A cloud data warehouse (DWH) has various business benefits, including:

- Greater flexibility and secure accessibility, thanks to access control features
- High scalability and elasticity to complement the dynamic needs of the modern enterprise business model
- Enhanced performance, given that multiple servers in the cloud share the computation load in hand
- Increased data storage options for both small and large businesses—you can choose from a variety of pricing models

> UNLIKE ON-PREMISE DATA WAREHOUSE SOLUTIONS, CLOUD MANAGEMENT IS FASTER AND MORE ACCURATE WHEN GATHERING DATA AND GENERATING INSIGHTS FROM MULTIPLE SOURCES.

- Optimized data recovery with timely and regular backups to ensure business continuity in case of a disaster
- Seamless integration with third-party tools and applications

CLOUD DATA WAREHOUSE VS. DATA LAKES

A data warehouse is a repository for unified and highly structured data that drive business intelligence and analytics. On the other hand, a data lake is a centralized repository that stores all business data, whether structured or unstructured.

Data Warehouse	Data Lake
Data Storage	
Stores, clean, structured, and unified data	Stores raw, unstructured data for current or future use
Users	
Often accessed by business-end users, such as managers	The typical users include data engineers and scientists
Analysis	
Data visualization, business intelligence, and analytics	Machine learning, predictive analytics, and big data analytics
Processing	
Extract, Load, Transform (ETL) format	Extract. Load, Transform (ETL format)
Schema	
Schema definition happens before data storage	Schema definition happens after the data storage
Cost	
Expensive due to increased time and management needs	Relatively inexpensive when compared to DHW

CHAPTER 2

Key Considerations and Criteria for Choosing a Cloud Data Warehouse Solution

KEY FACTORS TO CONSIDER WHEN CHOOSING A DATA WAREHOUSE SOLUTION

Innovative data warehouse solutions can help businesses keep up with the dynamic market needs of their respective industries. However, taking the decision of choosing the right solution lightly can culminate in a myopic business strategy, given that the insights will most likely be inaccurate or outdated. That said, here are the factors to consider:

DATA TYPES

Compare the technology and capabilities of your go-to data warehouse solution against the data types and sources that you want to connect it to. Prevalent data types that you'll often encounter include structured, unstructured, and JSON formats.

EASE OF USE

A go-to data warehouse solution should be easy to use, with a lower learning gap, enabling you to get started and secure the future of your business as soon as possible. Moreover, the deployment process and costs shouldn't be overwhelming.

SCALABILITY

Consider a data warehouse solution that matches your business analytics need in the future, so that you can scale without reinvesting in new storage/ analytics altogether. In other words, you should seamlessly scale upward or downward, depending on the forgoing data computation needs.

SPEED, PERFORMANCE, LATENCY, AND CONCURRENCY

The level of speed, performance, latency, and accuracy of your cloud DWH solution determines how fast you can respond to market changes and gain a competitive edge. In that case, watch out for a solution that offers parallel ETL processing capabilities for maximum availability and uptime.

SUPPORT

24/7 support will ensure that the address of your in-house business teams are addressed as soon as they arise. Depending on the level of support in your subscription plan, this function can also extend to regular updates and ongoing maintenance.

INTEGRATIONS

An ideal solution should support third-party integrations for combining data from multiple sources into a unified warehouse. Integrations also allow other users to access structured data seamlessly for analysis purposes.

COST AND PRICING STRUCTURE

Check out the pricing structure and consider other ongoing costs to determine whether your projected return on investment (ROI) is feasible. This includes data management, opportunity, procurement, as well as ongoing maintenance expenses.

CHAPTER 3

How to Set Up a Cloud Data Warehouse Solution?

Migrating to a cloud data warehouse is a priority if you want boundless capacity, efficient performance, or flexible scaling. The goal of this architect is to optimize the availability of your data on-demand, everywhere—making it accessible to data engineers, business analysts, and your entire data team.

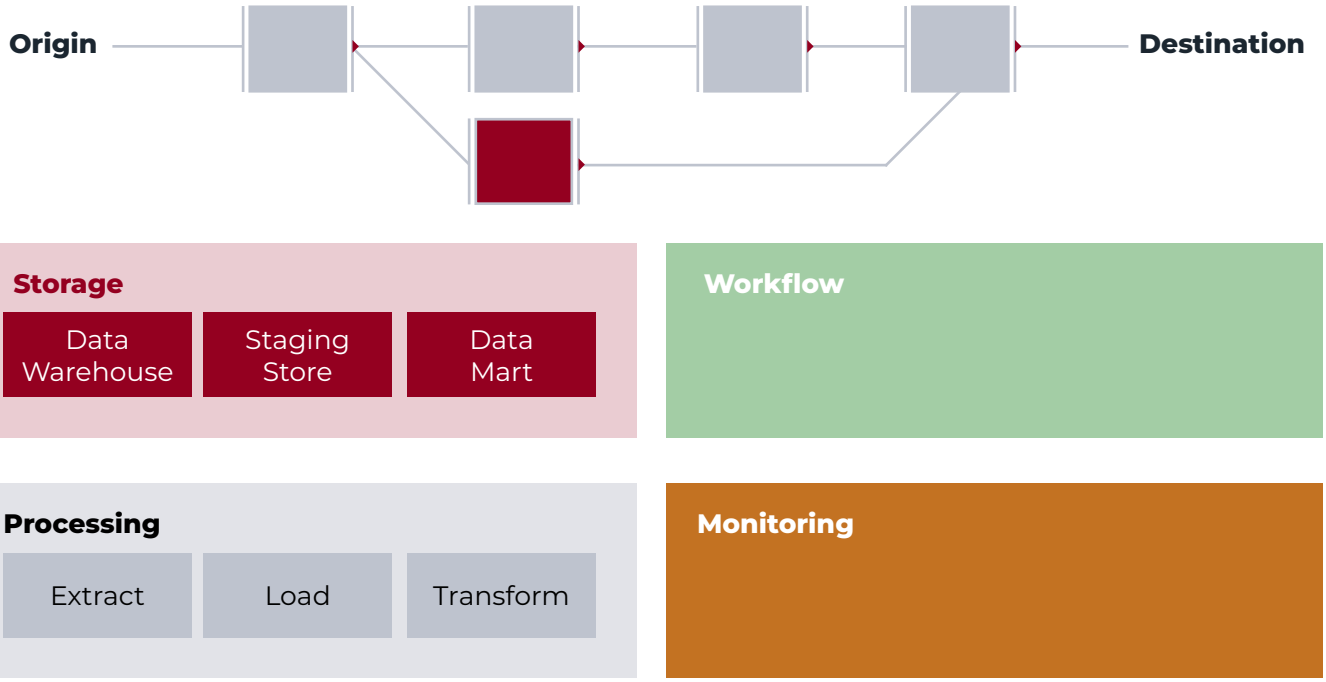
However, this is not a “plug-and-play” migration, where you and replicate your data legacy infrastructure in the cloud. Your legacy system will just be that in the cloud—a complex, ineffective architecture with no power to meet the needs of modern IT-driven businesses. In other words, legacy systems are not optimized for cloud-native environments.

READING DATA FROM THE SOURCE

A typical set up process of a cloud data warehouse solution begins by reading information from on-premise, cloud, or any other source. Data experts will then use innovative tools to sort the raw information into relevant data formats, enrich their records, and finally ingest them into the cloud.

However, taking this standard route comes with its share of challenges—structural drift and semantic drift. Structural drift happens when sudden changes are made to the data schema, while semantic drift is associated with updated data meaning. If these scenarios happen while you’re setting up your cloud DWH, the data may drop or disappear before reaching its new location.

DATA PIPELINE COMPONENTS



Smart data pipelines bring a big difference when setting up a cloud data warehouse solution for the first time. Unlike the standard data pipeline, this model comes with in-built intelligence that identifies threats and opportunities in your on-demand data. This approach puts you ahead of the competition while strategizing for foreseen challenges that could potentially result in untenable losses.

At the same time, intelligent data pipelines outsmart proprietary loading tools when it comes to delivering reliable results. For instance, they decouple data types, sources, as well as their new location, enabling you to make updates without rewriting. Moreover, smart data pipelines are automated to spot drift in real-time.

When setting up a cloud data warehouse solution, smart data pipelines will:

- Enhance performance and leverage the in-built snapshot feature for real-time debugging
- Transform data in real-time, regardless of its source, new location, processing mechanism, or even format

> **THE INNOVATION OF SMART DATA PIPELINES CAME JUST IN TIME WHEN BUSINESSES WANT INSTANT INSIGHTS TO ADDRESS REAL-LIFE CHALLENGES AND MEET THE HEIGHTENED EXPECTATIONS OF THEIR CUSTOMERS.**

- Automate the process of data formats conversion into relevant data types while simultaneously creating new fields
- Create multiple tables automatically, regardless of the knowledge of the schema source
- Enable reboot of the entire pipeline in the event of a failover

UNDERSTANDING THE CAPABILITIES OF SMART DATA PIPELINES

The innovation of smart data pipelines came just in time when businesses want instant insights

to address real-life challenges and meet the heightened expectations of their customers. Here is what this innovation is capable of:

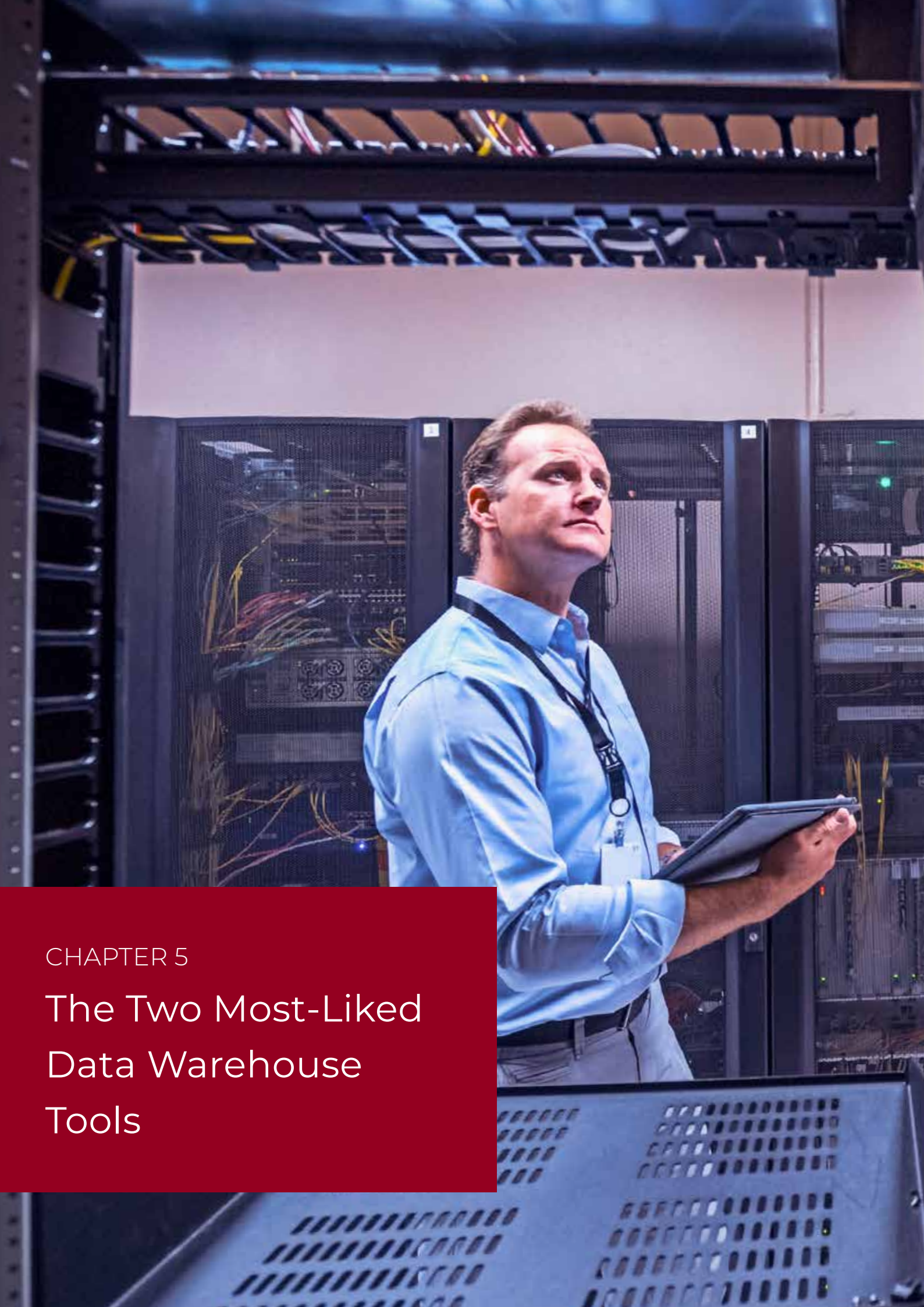
- **Pipeline monitoring:** in-built dashboards in smart data pipelines enable you to monitor data flow in the infrastructure and spot any challenges in real time. Moreover, you can validate whether your data is delivered, thanks to enhanced visibility into the end-to-end lag.
- **Real-time data integration:** You can seamlessly connect disparate sources with targets in smart data pipelines. This includes IoT devices, the data warehouse, utility applications, and even messaging systems.
- **Build applications on streaming data:** Leverage an SQL-based engine to filter, transform, and enrich streaming data for various applications. Prevalent use cases are data denormalization and PII masking.
- **Machine learning integration:** Take data forecast to the next level while detecting anomalies in time using machine learning integrations. ML integrations also facilitate live monitoring dashboards for critical events.
- **Swift scalability:** Expand your business horizons exponentially by computing resources in independent, distributed clusters. This means you can easily integrate multiple smart data pipelines into the architect.
- **Reliability:** Experiencing zero downtime with smart data pipelines means your application has failed over to independent noses, which is nigh impossible. In other words, you get at least one guaranteed processing.
- **Decentralization:** Unlike outdated monolithic data infrastructures, smart data pipelines run on decentralized and decoupled architectures. This gives you analytical insights into data products that are relevant to your business domain.

SOLVING THE PROBLEM OF STRUCTURAL DRIFT

A custom-built smart data pipeline can help you manage structural drift when setting up your cloud data warehouse solution. The architect will automatically generate tables and columns to edit upstream data and detect deletion simultaneously.

“

A CUSTOM-BUILT SMART DATA PIPELINE CAN HELP YOU MANAGE STRUCTURAL DRIFT WHEN SETTING UP YOUR CLOUD DATA WAREHOUSE SOLUTION. THE ARCHITECT WILL AUTOMATICALLY GENERATE TABLES AND COLUMNS TO EDIT UPSTREAM DATA AND DETECT DELETION SIMULTANEOUSLY.



CHAPTER 5

The Two Most-Liked Data Warehouse Tools

Amazon Redshift

Amazon Redshift is Amazon's cloud data warehouse solution that leverages SQL to analyze structured and semi-structured data. The solution works with data from data lakes, operational databases, and data warehouses. It also uses machine learning for best-price performance delivery at any scale.

BEST FOR

Amazon Redshift is ideal for businesses that want to analyze humongous data sets of at least a petabyte scale. The cloud solution DWH tool is also ideal for real-time analytics, management of multiple data sources, business intelligence, and log analysis.

TOP FEATURES

- Federated query capabilities for operational relational databases
- AWS data exchange in a single cluster or multi-cluster deployments
- Integration for Apache Spark applications
- Native integration with machine learning, databases, and other AWS services
- Native support for advanced analytics
- Auto-copy from Amazon S3

PROS

- Greater performance and scalability
- Ease of use and administration
- Improved security
- Columnar storage
- Reputable AWS ecosystem
- Flexible pricing models
- Widely adopted cloud data warehouse tool

CONS

- The platform is not a serverless architecture
- Doesn't support multi-cloud capabilities
- Concurrent execution

PRICING

The lowest paid plan for Amazon RedShift is $0.25 per hour but can be scaled to accommodate thousands of concurrent users and petabytes of data. In addition, the pricing model is based on On-Demand Instances, where users pay for services by the hour instead of upfront fees.

Google BigQuery

Google BigQuery is a popular, fully-managed enterprise DHW solution that helps businesses analyze their data. As a serverless architecture solution, BigQuery allows users to leverage SQL queries for business solutions without the need for infrastructure management.

BEST FOR:

Google BigQuery is ideal for businesses that handle big data sets and require “heavy” queries. The more expansive your data set, the better the chances of gaining a positive ROI with this cloud data warehouse solution.

TOP FEATURES

- Supports real-time analytics and data querying
- Hadoop and Sparks integrations
- Data visualization and workflow dashboard
- Notebooks for task creation and collaboration
- In-built database machine learning
- Data lakes analytics and integration

PROS

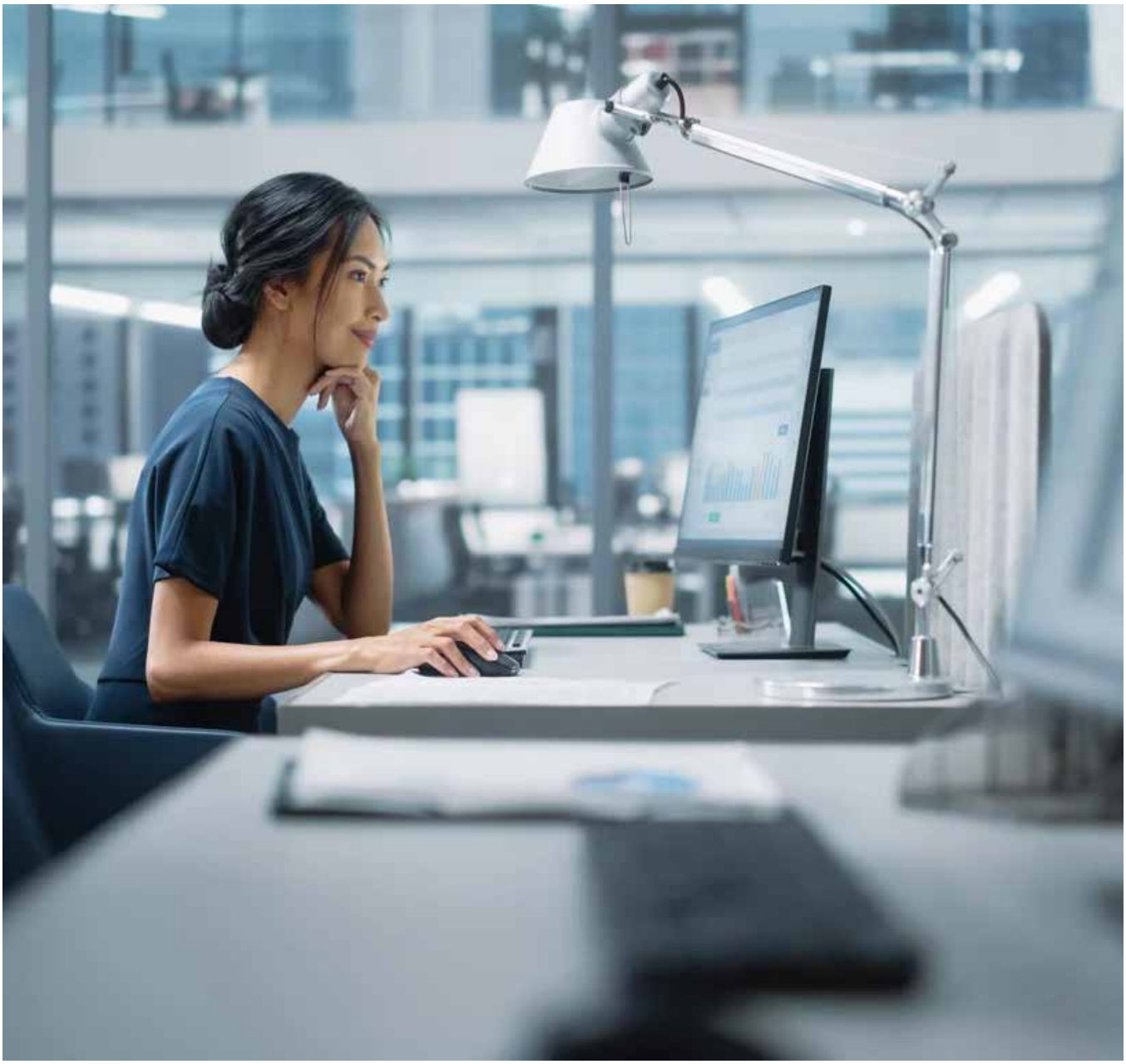
- Saves storage capacity with improved query performance
- Ensures high data integrity by eliminating duplications and inconsistencies
- Business Intelligence (BI) tools integration for actionable insights
- Swift scalability
- Stringent data governance and security measures

CONS

- Complex user interface for newbies
- No tooling support outside the GCP ecosystem
- Flat table-centric, which can make data management challenging

PRICING

Users can enjoy up to two pricing models under Google BigQuery, including on-demand and flat-rate. After a free 1TB of query data processing for the first month, users incur $5 per TB in the on-demand pricing model, while a flat rate is available per second, monthly, or annual commitments.



WHICH CLOUD DATA WAREHOUSE SOLUTION IS BEST FOR YOU?

A cloud data warehouse solution that works for one business may not be ideal for yours. The right choice boils down to the current state of your business, budget, and future data management needs. That said, there is no absolute right or wrong choice—do your research, and you'll find an ideal solution tailored to your business.

WHAT ABOUT OTHER CLOUD DATA WAREHOUSE SOLUTIONS?

Top cloud data warehouses providers

Snowflake	Amazon RedShift	Google BigQuery	Azure Synapses
Deployment			
Continuous delivery deployment.	Automatic modular deployment	Continuous data integration	Continuous integration and delivery deployment
Data types			
String & binary data, date & time data, geospatial data, semi-structured data	Raw, uncompressed data	String, bytes, array, struct, numeric	Structured and unstructured data
Scalability			
Automatic concurrency scaling	Can boost throughput by more than 35 times	Elastic for rapid scaling	Supports auto-scaling
Maintenance			
The service provider offers ongoing maintenance, upgrades, and tuning	Weekly schedule for latest patches and upgrades	Optional periodical maintenance	Users schedule maintenance
Integration partners			
Matillion, Talend, Tableau, Informatica, Fivetran	Apache Spark integrations	Hadoop and Sparks	Ab Initio, Aecorsoft, Alooma, Alteryx, and Xpert BI, among others
Pricing model			
Based on actual computing and storage usage	On-demand instances model	On-demand and flat-rate models	Hourly or monthly billing
Best for			
Ideal for data management and analytics	Humongous data sets of at least a petabyte-scale	Big data sets that require heavy query	Complex queries and aggregations

CHAPTER 5

How to choose the right Cloud Data Warehouse Solution in 2023 (With Case Studies)

Find out why our client chose Amazon Redshift

Our client, a leading developer of casino gaming products, has a portfolio of pool-based games with a unique twist, powered by a custom engine. These games, including table and sportsbook options, are sophisticated, engaging, and offer dynamic odds, social elements, and immersive entertainment. The company works with top-notch operators to offer these games on its own platform as well as third-party platforms.

BACKGROUND

To support the growth and success of their business, our client recognized the need for a robust data warehouse solution that could handle the vast amounts of data generated by its gaming operations. A data warehouse is a specialized database system that is designed to support the efficient querying and analysis of substantial amounts of data and is typically used to support business intelligence and data analytics activities.

To design and build a data warehouse solution that met their needs the company turned to Symphony Solutions, a leading provider of data engineering and data analytics services.

CHALLENGE

One of the main challenges faced by the Symphony Solutions team was the fact that the data warehouse solution was being developed for a product that was not yet in production and actively developed. This meant that the team had to work closely with the client to understand the future needs of the product and design a solution that would be able to support it. In addition, Symphony team had to consult with the client on the system needs and use, as well as better ways of leveraging the system, as the client was lacking knowledge in this area.

Another challenge faced by the Symphony Solutions team was the

“

TO DESIGN AND BUILD A DATA WAREHOUSE SOLUTION THAT MET THEIR NEEDS THE COMPANY TURNED TO SYMPHONY SOLUTIONS, A LEADING PROVIDER OF DATA ENGINEERING AND DATA ANALYTICS SERVICES.

need to cooperate with various vendors who were contributing data to the data warehouse. This required careful coordination and communication to ensure that the data was being ingested and transformed in a consistent and reliable manner. The team also had to make architecture adjustments according to the changing environment and business needs, as well as support the evolving nature of data sources, which were being developed at the same time as the data warehouse.

In addition to the above challenges, the Symphony Solutions team had to design a scalable batch data ingestion process with the possibility of introducing near real-time data in the future. This required the team to carefully consider the performance and scalability requirements of the data warehouse and design a solution that would be able to handle the expected volume and velocity of data. The team also had to create a flexible data warehouse baseline that would allow the business to build insightful analytics on full-system data and deliver it to multiple third-party systems using a single integration point.

SOLUTION

To address the challenges the team faced with data warehouse development, they implemented several technologies and approaches.

“

THIS APPROACH REQUIRED MORE EFFORTS TO DEVELOP THE DATA INGESTION PART, BUT IT ALLOWED THE TEAM TO DELIVER DATA TO THE DATA WAREHOUSE AS QUICKLY AS POSSIBLE WITHOUT ADDITIONAL OVERHEAD FOR DEVELOPMENT TEAMS FROM OTHER SYSTEM COMPONENTS.

First, the data warehouse had a direct connection to all data sources. This approach required more efforts to develop the data ingestion part, but it allowed the team to deliver data to the data warehouse as quickly as possible without additional overhead for development teams from other system components.

To further improve the data ingestion process, all data sources were forced to push data to EventBridge, a fully managed event bus service that makes it easy to process and transfer data between different applications and services. Some data sources were able to do this automatically using MongoDB change stream or DynamoDB stream, while others had

to develop dedicated EventBridge push mechanisms. This allowed the team to unify the data ingestion process and use Kinesis Firehose, a fully managed service for real-time streaming data ingestion and load data into data stores and analytics tools, for scalable data stream consumption.

The use of EventBridge and Kinesis Firehose, combined with Redshift, a data loading utility that enables fast data ingestion from various data sources into Amazon Redshift, ensured scalability at all stages of the data ingestion process. In the future, the team plans to extend the data ingestion process to serve near real-time data by leveraging more features of Kinesis.

In addition to the data ingestion process, the Symphony Solutions team also focused on designing a smart data model that would add value to the data warehouse. The data model included multiple layers, including historical and keep-latest tables as a core layer and a star schema as a top layer. This design allowed the team to serve different business cases and provide the client with the flexibility to build insightful analytics on full-system data.

Finally, the Symphony Solutions team chose Rudderstack, an open-source event stream processing platform, to ease the integration of the data warehouse with multiple third-party tools such as Braze and Google Analytics. This helped the client to deliver data to these systems using a single integration point, streamlining the data flow, and reducing the complexity of the system.

To ensure the smooth flow of data into the data warehouse, the Symphony Solutions team implemented a few technical solutions, including EventBridge, Kinesis Firehose, and Airflow DAGs. Multiple data sources, some of which have a CDC (change data capture) nature, are now sending data to EventBridge, where it is buffered into batches by Kinesis Firehose and stored in AWS S3. From there, Airflow DAGs run Redshift commands to ingest the data files into the data warehouse.

One of the key benefits of the data warehouse system implemented by Symphony Solutions is the ability to build an OLAP (online analytical processing) layer that is ready for consumption from QuickSight reports and other systems. This layer is created through a process of extracting, loading, and transforming data from multiple sources, using Redshift as the data warehouse and Airflow as the orchestration tool.

By using Redshift orchestrated by Airflow, the Symphony Solutions team was able to build an OLAP layer that is scalable, flexible, and easy to use, providing the client with a powerful tool for gaining insights into its business and driving growth.

RESULTS

The data warehouse services provided by Symphony Solutions will play a crucial role in the success and growth of the client's business. With access to a centralized repository of structured data, the company will be able to gain insights into the performance and popularity of its games, understand customer preferences and behavior, and optimize business operations.

The data warehouse will enable the client's teams to have ready-to-use BI dashboards to analyze product performance and make informed decisions. The marketing team will have the capability to run conversion-optimized campaigns thanks to the data integrations made by Symphony Solutions, while the CRM team will be able to create engaging communication campaigns using many different channels, target future users based on their segments, run trigger-based campaigns, and personalize messages with relevant content.

Overall, the data warehouse solution provides the client with the tools and capabilities it needs to drive growth and success in its gaming operations, and to better understand its customers and optimize its business.

Find out why GOAT Interactive chose BigQuery

GOAT Interactive is a part of the Editec Group. Its goal is to become the number one operator in the African market. It aims to fuse the world of football and betting, creating an authentic, highly engaging community of sports fans who love the thrill of a bet. The brand is closely partnered with renowned football club brands and leverages these partnerships to provide a unique player experience, enabling GOAT to provide unique betting propositions, content and events to the rapidly growing African market – an experience that brings fans closer to clubs, leagues, and the sport they love.

BACKGROUND

Driving relevant conclusions from the data on hand is key to making the right business decisions. But for an organization to make an accurate decision, it has to have full information. The data needs to be collected and managed from diverse sources. That's why many businesses today are turning to data warehousing as a way to organize data. Our client, GOAT Interactive needed a solution that would facilitate the growth of current business and potential new markets. Their request was to build a modern serverless data warehouse that would help them better understand and grow their product. The company wanted to understand business performance of their solutions both from marketing and business performance point. Prior to approaching Symphony, they were using an on-premise solution that had a low performance.

CHALLENGE

Our client, being one of iGaming industry's leaders, is receiving large amounts of data on market behavior, company business performance metrics, as well as some historical data. However, when they approached Symphony Solutions this data was not centralized. Thus, they were not able to correlate metrics to drive some insights and make the right business decisions on time. The list of tasks included the following:

- Migrate large amounts of data (about 10 TB) from on-premises databases.
- Build cloud data warehouse (DWH) and Data Lake on top of Google Cloud Platform (GCP) ecosystem.
- Create BI dashboards using Data Studio and Looker. Implement email reporting services to data stakeholders.
- Integrate DWH with CRM tools like Atlas, Optimove, and Raventrack. •

- Implement real-time data streaming and analytics within DWH

SOLUTION

Based on client's cloud preferences, Symphony started working on building data warehouse on Google Cloud. The initial stage was creating PoC, which was followed by adding detailed reports, migrating all data, enabling complex integrations with CRM providers thus empowering users' data enrichment to be able to conduct more effective behavior analysis.

Based on the client's request, Symphony's team set out to work on developing the following:

- DWH and Data Lake have been built on top of BigQuery and Google Cloud Storage / Dataproc respectively
- Cloud Composer is used for ETL pipelines orchestration
- Data Studio and Looker are used for BI dashboards
- Dataflow is used for real-time data streaming and analytics (in progress)

"THE CLIENT OBTAINED AN AUTOMATED, RELIABLE AND SCALABLE SOLUTION. THEIR BUSINESS PROCESS BECAME ROBUST AND PREDICTABLE WITH THE BUILT CLOUD DWH

As a next phase which is currently in progress, the work is evolving from batch processing to real-time data streaming as a way of improving the client's services and automating business processes.

RESULT

The client obtained an automated, reliable and scalable solution. Their business process became robust and predictable with the built cloud data warehouse, including integrations with different CRM tools. The current list of outcomes includes:

- Automated cloud data warehouse with batch processing and real-time data analytics features. 15 countries.
- Our teach accomplished 20 different important reports for We built more than 22 Data Studio dashboards.
- We built up to 10 models containing 30+ explores within Looker.
- Client Administrator can perform analytics in real-time.
- Cloud ecosystem allowed to get all the data in one place and orchestrate ETL pipelines in unified way, and prepare on-demand OLAP cubes.
- Migrated more than 10TB of external data.
- A reliable, scalable, and well maintainable solution.

Currently, we are working on developing and extending real-time data streaming and providing the support of the existing solution.

Over to you

Choosing between Google BigQuery and Amazon Redshift can be a challenging task as both solutions offer great ecosystems. It's important to note that there is no right or wrong choice between the two. Factors like security and customization might make Amazon Redshift a better fit for some cases, while BigQuery might be more suitable for startups. Other factors such as developer expertise or pricing model preference may also drive the choice for a cloud solution. In the case of big companies, the partnership with Amazon or Google often predefines the choice.

However, choosing the right solution between the two is still critical. If you are on the fence and having a hard time making a decision, our experts can make this very easy for you. Go ahead and schedule a free consultation with us below. Even better, whatever solution that you end up choosing, Symphony Solutions' all-in-one expert team can help you get the most out of your cloud data warehouse solution.

Contact us today for free consultations on BigQuery or Amazon Redshift and transform your business to match or beat the competition.

WHY **SYMPHONY SOLUTIONS** IS THE BEST PARTNER WHEN IT COMES TO **CLOUD DATA WAREHOUSE SOLUTIONS**

Symphony Solutions is a Dutch-based agile-driven partner for custom cloud data warehousing services. We help businesses extend their portfolio into the cloud for enhanced access and integration, seamless self-service capabilities, and heightened data security.

Here are the benefits of choosing Symphony as your cloud data warehouse solutions partner:

DYNAMIC CAPACITY

We can dynamically scale up or scale down, to some extent but this means you don't have to make an extended commitment.

DYNAMIC TEAM COMPOSITION

If at some point you need a specific profile or expertise, we should have it on board and ready to join your team in a completely transparent manner.

READY TO START TOMORROW

Literally. This is not an exaggeration but something we are ready to commit to.

DOMAIN KNOWLEDGE

We know the business and the nature behind your data. We have considerable experience in iGaming domain, as well as healthcare, retail and renewables and our domain knowledge and expertise keep growing.

AVAILABLE BLUEPRINTS WITHIN SYMPHONY DEVOPS STACK

We have some ready-to-use scripts that allow to deploy and deliver value fast

LET'S TALK FURTHER

This eBook provides the best insights about cloud data warehouses. But knowledge can only go so far. Our cloud data warehouse experts stay on top of all the latest tech trends to deliver the best data solutions to our partners.

AND, they are just one click away. Contact us today to get expert insights on which cloud data warehouse to choose from, how to set it up and how to get the best out of it. Get in touch!

GET YOUR FREE CONSULTATION

YEVHEN TANCHIK
Data and Analytics competence center lead

A highly talented and experienced cloud data warehouse expert, fully focused on Amazon Redshift and strategic partnerships. Talk to Yevhen to hear his hands-on experience on the development, launch and strategic promotion of Amazon Redshift, BigQuery and Snowflake among many others

FAQ

BESIDES COMPATIBILITY, WHAT ARE THE TOP REASONS WHY CUSTOMERS CHOOSE AMAZON REDSHIFT

Amazon Redshift is a go-to solution for businesses that want to accelerate time-to-insights with a powerful analytical tool. The solution also supports the integration of machine learning services and scaling of data warehouse capacity to match the computation needs of demanding, unpredictable workloads. Most importantly, you get to benefit from performance innovation at no extra charges.

Amazon RedShift integration for Apache Spark allows businesses to build Apache Spark applications on their cloud data warehouses for running complex analytic sets and ML models. This integration also facilitates seamless monitoring and troubleshooting to identify issues in Apache Spark applications during ongoing maintenance.

BESIDES COMPATIBILITY, WHAT ARE THE TOP REASONS WHY CUSTOMERS CHOOSE BIGQUERY?

Google BigQuery reportedly offers 27% lower three-year TCO than the competition. Businesses that collect and store different types of data based on simple SQL can leverage the solution's in-built machine learning capabilities for online predictions. The solution's BigQuery Analytics Hub offers a secure platform for sharing data assets across the organization or with third parties for commercial and public analyses.

BigQuery also supports geospatial analysis with native integrations for augmenting your data analytics workflows with location intelligence. With this approach, you can draw real-time insights from polygon, multi-polygon, arbitrary point, or line geospatial data formats to conquer new markets and expand your operations overseas.

HOW DOES THE USE OF CLOUD COMPUTING AFFECT THE SCALABILITY OF A DATA WAREHOUSE?

Modern data warehouse solutions leverage cloud computing to store current and historical information in a centralized repository. In cloud computing, clients can add or delete storage, computing, or even network services as they scale vertically or

horizontally to enhance availability, performance, and the prevailing demand.

HOW TO SET UP A CLOUD DATA WAREHOUSE?

Businesses can set up a cloud data warehouse in three simple steps:

- Extract transactional data from internal and external sources
- Transform the transactional data
- Load the transformed data into dimensional databases in the cloud

WHICH AWS SERVICE ALLOWS YOU TO BUILD A DATA WAREHOUSE IN THE CLOUD?

Amazon RedShift database is an AWS service that allows clients to build and deploy scalable data warehouses in the cloud, using native or in-house business intelligence tools.

IS BIGQUERY BETTER THAN RED-SHIFT?

When comparing Amazon Redshift vs gc bigquery, the latter stacks up better, especially for clients who run rapid queries a few times within a day. RedShift charges this service per hour. Also, BigQuery would be ideal for data mining or data science operations that require ML to run.

CAN REDSHIFT HANDLE BIG DATA?

Yes, RedShift can handle big data, allowing clients to scale from a few hundred GB of data to a petabyte, or even more, depending on your data analytics needs.

SOURCES

https://symphony-solutions.com/services/data-and-analytics
https://symphony-solutions.com/cases/building-centralized-data-management-solution-for-igaming
https://symphony-solutions.com/insights/redshift-vs-bigquery-comparing-data-warehouses *https://www.dataversity.net/brief-history-data-management/*
https://dbpedia.org/page/First-generation_programming_language
https://dbpedia.org/page/First-generation_programming_language
https://www.alliedmarketresearch.com/data-warehousing-market
https://www.alliedmarketresearch.com/data-warehousing-market
https://www.alliedmarketresearch.com/data-warehousing-market
https://cloud.google.com/bigquery/?utm_source=google&utm_medium=cpc&utm_campaign=emea-emea-all-en-dr-bkws-all-all-trial-p-gcp-1011340&utm_content=text-ad-none-any-DEV_m-CRE_574805387464-ADGP_Hybrid%20%7C%20BKWS%20-%20PHR%20%7C%20Txt%20~%20Data%20Analytics%20~%20BigQuery%23v1-KWID_43700072654569871-kwd-301529154162-userloc_1009824&utm_term=KW_bigquery-NET_g-PLAC_&gclid=Cj0KCQiAn4SeBhCwARIsANeF9DL13y-o4N_QzuTwo8LmlRfyAHKSF-JU42BqjcCnvz1lVFSJ1drlPrMa