

Igniting discovery: Built-for-the-cloud data integration kicks Redshift into high gear

SnapLogic empowers cloud warehouse users with up to a 10x improvement in the speed and ease of data integration



Table of Contents

In search of discovery: Massive demand for massive data	3
Why traditional data integration doesn't work in cloud data warehouses	6
SnapLogic delivers a “quanta” in improvement	8
A 10x improvement, quantified	10
Summary	11

In search of discovery: Massive demand for massive data

Every day, business users extract data-driven insights that, even a few years ago, would have required a legion of data scientists, terabytes of physical storage, and a weeks-long wait in an IT production queue. Consider, for example, data's role in the evolution of how pharmaceuticals are marketed. In the recent past, an army of sales reps would market directly to doctors, their efforts loosely coupled with consumer advertising placed across demographically compatible digital and traditional media.

This “spray and pray” approach with promotional spending, while extremely common, made it difficult to pinpoint specific tactics that drove individual product revenues. Projections and sales data factored heavily into the campaign planning stage and in reports that summarized

weekly, monthly, and quarterly results, but the insights gleaned were nearly always backward-looking and without a predictive element.

A pharmaceutical company pinpoints opportunity

Today, sophisticated pharmaceutical marketers have a much firmer grasp on how to use data to drive sales in a predictive manner, by deploying resources with pinpoint precision. A case in point: To maximize the market share of a prescription asthma medication, a leading pharmaceutical company analyzes and correlates a wide range of data on a daily basis, capitalizing on even the smallest market and environmental fluctuations.

- Each night, the marketing team takes in pharmacy data from around the US to monitor sales in each region, to learn how many units of the asthma medication sold the previous day. These numbers are processed, analyzed, and reported back to the sales team the following morning, allowing reps to closely monitor progress against their sales objectives.
- With this data, the pharmaceutical marketing team can monitor, at aggregate and territory levels, the gross impact of many variables including:
 - > Consumer advertising campaigns
 - > Rep incentive programs
 - > News coverage of air quality and asthma

“With SnapLogic’s integration platform, Amazon Redshift customers can pull massive volumes of data from multiple sources into Amazon Redshift at a fraction of the cost of a traditional data integration solution.”

Terry Wise
Director of Worldwide Partner Ecosystems
Amazon Web Services

- However, the pharmaceutical marketing team takes its exploration much deeper. Layered on top of the core sales data, the marketing team correlates weather data from the National Weather Service (NWS) and multiple data sets from the US Environmental Protection Agency (EPA), such as current air quality, historic air quality, and air quality over time. Like their sales data, the weather and EPA data cover the entire US.

By correlating these multiple data sets, the marketing team can extract extraordinary insights that improve tactical decisions and inform longer-term strategy. At a very granular, local level, the team can see:

- How optimal timing and placement of advertising across digital and traditional media drives demand
- Which regional weather conditions stimulate the most sales in specific locales
- The impact of rep incentive programs on sales
- How news coverage of air quality and asthma influences demand

Ultimately, the pharmaceutical marketing team can identify, with uncanny precision, markets to concentrate spending on local and regional media, which can change on a constant basis. In this way, prospective consumers are targeted with laser-like accuracy, raising their awareness of the pharmaceutical company's asthma medication at the time they need it most.

The results of the targeted marketing strategy are clear: the pharmaceutical company has enjoyed significant market share growth with its asthma relief medication, while reducing advertising costs due to more effective targeting.

Tools to empower business users

The pharmaceutical industry example exemplifies perhaps the biggest trend in recent business history: massive demand for massive amounts of data, to provide insight and drive informed decision-making. But five years after data scientist was named “the sexiest job of the 21st century,”¹ it's not data scientists who are gathering, correlating, and analyzing all this data; at the most advanced companies, it's business users.

“Since its launch in November 2012, Redshift has grown to dominate the cloud data warehousing market by a large margin.”

Terry Wise
Director of Worldwide Partner Ecosystems
Amazon Web Services

At the pharmaceutical company and countless others like it, the analytics explosion is ignited by “citizen data scientists,” fueled by two key enabling technologies:

- Amazon Redshift is a fast, fully managed, petabyte-scale cloud data warehouse (CDW) from Amazon Web Services (AWS) that allows users to analyze massive amounts of data with existing business intelligence tools. Since its launch in November 2012, Redshift has grown to dominate the cloud data warehousing market by a large margin.

¹ “Data Scientist: The Sexiest Job of the 21st Century,” Thomas H. Davenport and D.J. Patil, Harvard Business Review, October 2012.

- The SnapLogic Enterprise Integration Cloud is a self-service integration platform, built for the cloud, that makes it fast and easy to connect data, applications, and devices. The building blocks of the Enterprise Integration Cloud, SnapLogic Snaps, are pre-built connectors that abstract application programming interfaces (APIs) graphically, allowing business users to quickly integrate data and processes using pre-built patterns.
- The Snaps for Amazon Redshift integration are pre-built connectors that enable IT and business users to transfer up to petabytes of data into and out of Amazon Redshift nearly instantaneously, for a fraction of the cost of traditional data integration solutions. More than 400 additional Snaps allow application data to be easily integrated.

This white paper explains how the SnapLogic Enterprise Integration Cloud delivers a CDW data integration solution that offers performance improvement of up to 10x over traditional data integration solutions, which are not native to the cloud. In doing so, SnapLogic empowers business users and citizen data scientists, igniting discovery while addressing enterprise data governance concerns.

“Sophisticated pharmaceutical marketers have a much firmer grasp on how to use data to drive sales in a predictive manner.”

Terry Wise
Director of Worldwide Partner Ecosystems
Amazon Web Services

Why traditional data integration doesn't work in cloud data warehouses

While the pharmaceutical company's expansive daily analytics operation works like clockwork now, that wasn't always the case. Several months prior to launching its current environment, with data flows powered by SnapLogic, the pharmaceutical company tried, unsuccessfully, to accomplish the integration with Informatica PowerCenter and Informatica Cloud.

An IT organization abandons Informatica

In 2016, the pharmaceutical company's IT organization recognized that with rapid data growth worldwide, the time was right to move the company's data warehouses - across US, EMEA, and Asia - to the cloud. The IT team's plan was to move data from Salesforce, Veeva (cloud-based applications for the pharmaceutical industry), and third-party sources into Amazon Simple Storage Service (S3), and then integrate the data into Redshift for sales and marketing analytics.

However, the project stalled due to difficulty with Informatica PowerCenter, the IT team's initial choice for data integration. PowerCenter, which Informatica describes as a "metadata-driven integration platform," is a data extract, transfer, and load (ETL) product rooted in mid-1990s enterprise architecture. The team found PowerCenter complicated to use and slow to deliver the urgently needed integrations.

Looking for faster results, the pharmaceutical company then attempted to use Informatica Cloud, Informatica's cloud-based integration solution. The data integration initiative was again derailed, this time by the solution's

lack of maturity and functionality. The pharmaceutical company's data was forced back on-premises, jeopardizing the entire cloud data warehouse initiative.

“We attempted to use an on-premises solution to integrate data into Redshift, but found it too difficult. We were quickly able to build powerful integrations with SnapLogic.”

Senior Business Capability Manager at a large global pharmaceutical company

Data integration aligned with the cloud

But the IT team kept searching for the right data integration solution. “Cloud was instrumental to our plans, and we needed data integration that aligned with where we were headed,” said the Senior Business Capability Manager in charge of the integration project. The pharmaceutical company chose the SnapLogic Enterprise Integration Cloud.

After a self-evaluation, the IT team was able to quickly build data integrations with SnapLogic; no specialized resources or consultants were required. To accomplish the integrations in Redshift, the pharmaceutical company used:

- Salesforce Snap
- Redshift Snap
- Various RDBMS Snaps
- ReST/SOAP Snaps
- Transformation Snaps

With the data integration accomplished in a matter of days, the IT organization was assured that current skill sets could support the company's future global BI architecture. In addition, the IT team found the SnapLogic Enterprise Integration Cloud easy enough for business users, like the marketing team, to integrate new data into Redshift - which is exactly what happened with the EPA and NWS data. Leveraging Redshift's nearly infinite availability of low-cost data storage and compute resources, the analytic possibilities are equally limitless - igniting the marketing team's discovery of new strategies to drive new insights, revenues, and operational efficiencies.

“The IT team found the SnapLogic Enterprise Integration Cloud easy enough for business users and ignited the marketing team's discovery of new strategies to drive new insights, revenues, and operational efficiencies.”

Senior Business Capability Manager at a large global pharmaceutical company

SnapLogic delivers a “quanta” in improvement

“Fundamentally, I believe that SnapLogic is 10 times better than Informatica. That’s a design goal, and it’s also a necessary and sufficient condition for success. If a startup is going to survive, it’s got to have some 10x factor, some quanta of a value proposition.

The quanta over the state of the art - the best-of-the-best of the incumbents - is vital. SnapLogic can fluently solve enterprise data problems almost as they are happening. That has a ‘wow’ factor people experience when they harness the power of our data integration technology.”

Gaurav Dhillon
Founder and CEO, SnapLogic

With cloud data warehousing growing at a rate 4.5 times faster than on-premises data warehouses through 2020,² the need for “data integration aligned with the cloud” is urgent. The SnapLogic Enterprise Integration Cloud is a mature, full-featured Integration Platform-as-a-Service (iPaaS) built in the cloud, for the cloud. Through its visual, automated approach to integration, the SnapLogic Enterprise Integration Cloud uniquely empowers both business and IT users, accelerating cloud data warehouse and analytics initiatives on Redshift and other cloud data warehouses

Unlike on-premises ETL or immature cloud tools, SnapLogic combines ease of use, streaming scalability, on-premises and cloud integration, and managed connectors. Together, these capabilities present a 10x improvement over legacy ETL solutions like Informatica or other “cloud-washed” solutions originally designed for on-premises use, accelerating CDW integrations from months to days.

With cloud data warehousing growing at a rate 4.5 times faster than on-premises data warehouses through 2020,² the need for “data integration aligned with the cloud” is urgent.

² “IDC Futurescape: Worldwide Big Data and Analytics 2016 Predictions,” Dan Vesset, et. al., November 2015.

The vast majority of products marketed as cloud data integration solutions were not, in fact, built for the cloud.

Why “cloud-washed” isn’t good enough

The vast majority of products marketed as cloud data integration solutions were not, in fact, built for the cloud. As noted above, at its core, Informatica PowerCenter is built on 25-year-old legacy technology, modified for the cloud. Other “cloud-washed” solutions are built on relatively newer technology foundations such as XML and SOAP, which are not optimized for today’s cloud environments. These solutions are characterized by:

- Complex interfaces
- Poor or non-existent handling of unstructured data
- Time-consuming integrations that often require specialized internal resources or consultants
- Open source connectors that present enterprise IT risk
- “Big data” capabilities sold separately, as part of a restrictive legacy licensing approach that often includes per-user licensing. This limits usage by citizen integrators and citizen data scientists.

Of the data integration solutions purporting to be built for the cloud, most are immature and offer limited numbers of application connectors; Informatica Cloud, for example, has fewer than 150 connectors, as compared to more than 400 SnapLogic Snaps.

Strong benefits beyond fast integration

While SnapLogic customers routinely cite speed of integration as a primary reason for choosing the Enterprise Integration Cloud, they automatically receive additional benefits including:

- Ease of use for business and IT users: SnapLogic’s attractive, graphical user experience is browser-based, offering a familiar approach to self-service data integration in Redshift.
- Built for scale: Bulk data movement and streaming data integration are easily accomplished with the SnapLogic Enterprise Integration Cloud, which offers big data and native Hadoop support without additional subscription fees.
- Ideal for hybrid environments: With over 400 pre-built connectors for cloud and on-premises data sources, SnapLogic can readily handle relational, document and legacy data, as well as today’s unstructured data.
- Cloud data warehouse-ready: SnapLogic accelerates self-service analytics data governance initiatives with native support for popular cloud data warehouses including Amazon Redshift, Microsoft Azure, Snowflake, and BigQuery. Snap Patterns provide pre-built processes, accelerating data integration within the warehouse.

Designed for data governance

Data governance is a significant topic, and a major concern of IT organizations charged with maintaining the consistency of data routinely accessed by citizen data scientist and citizen integrator populations. Gartner estimates that through 2016, less than 10% of self-service BI initiatives will be governed sufficiently to prevent inconsistencies that adversely affect the business.³

Data discovery initiatives using desktop tools risk creating inconsistent silos of data. Cloud data warehouses afford increased governance and data centralization. SnapLogic helps to ensure strong data governance by replicating source tables into Redshift clusters, where they can be periodically synchronized at any time interval desired, from real-time to overnight batch. In this way, data drift is eliminated, allowing all users who access data, whether in Redshift or other enterprise systems, to be confident in its accuracy.

A 10X IMPROVEMENT, QUANTIFIED

The SnapLogic Enterprise Integration Cloud presents an improvement of up to 10x over solutions based on legacy ETL or “cloud-washed” on-premises technology. With SnapLogic, Snap Patterns users can:

ACCELERATE

Accelerate cloud data warehouse adoption with prebuilt patterns that can be configured by an automatically generated series of steps

CONNECT

Connect Redshift rapidly to a variety of relational database services including Amazon RDS for MySQL, PostgreSQL, Oracle, and SQL Server

LOAD

Load data into an Amazon S3 bucket quickly and kick off the Amazon Redshift import process in a single step

REPLICATE

Replicate source tables into their Amazon Redshift clusters easily and detect daily changes to keep data synchronized

PERFORM

Perform sophisticated ETL operations such as slowly changing dimensions Type 2 (SCD2) and database lookups without any coding

DESIGN

Design a variety of data operations visually using a set of core Snaps such as Binary, Flow, Script, Transform, and XML

CAPITALIZE

Take advantage of core REST and SOAP Snaps for broader connectivity

³ “Predicts 2017: Analytics Strategy and Technology,” Kurt Schlegel, et. al., Gartner, November 30, 2016

Summary

Big data, once a novelty to be dealt with and analyzed by IT wizards and experienced data scientists, is now a mainstream business commodity, accessed by everyday business users. But while some of the excitement over the sheer size of big data may have worn off, the discoveries it yields are stoking enthusiasm in every department of the organization. At the pharmaceutical company profiled in this white paper, the notion of marketers performing sophisticated, multilayer data analysis to pinpoint opportunities was unthinkable even five years ago. With mainstream use of cloud data warehouses like Amazon Redshift, it's now routine.

SnapLogic operationalizes innovation

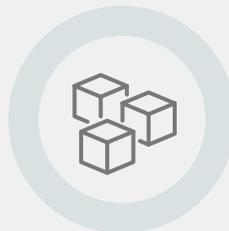
SnapLogic plays an important role in operationalizing innovation by providing critical enabling technology for cloud data warehouse integration. By making CDW data integration fast and easy, SnapLogic is helping to ignite discovery at companies in every industry. Customers look to the SnapLogic Enterprise Integration Cloud as a solution ten times better than “cloud-washed” legacy data integration solutions, and products billed as “cloud-ready” that, in fact, are not.

By enabling users to more quickly build, deploy, and efficiently manage multiple high-volume, data-intensive integration projects, SnapLogic uniquely delivers:



EASE OF USE

Ease of use for business and IT users through a graphical approach to integration



SCALABILITY

A solution built for scale, offering bulk data movement and streaming data integration



HYBRID CAPABILITIES

Ideal capabilities for hybrid environments, over 400 Snaps handle relational, document, unstructured, and legacy data sources



CLOUD DATA

Cloud data warehouse-readiness with native support for Redshift and other popular cloud data warehouses



GOVERNANCE

Built-in data governance by synchronizing data in Redshift at any time interval desired, from real-time to overnight batch

For more information on how to use SnapLogic with Amazon Redshift to ignite discovery within your organization, visit snaplogic.com/solutions/amazon-redshift-integration

Other customers using SnapLogic for enterprise integration



SnapLogic is the global leader in self-service integration. The company's Enterprise Integration Cloud makes it fast and easy to connect applications, data, and things. Hundreds of customers across the Global 2000 - including Adobe, AstraZeneca, Box, Capital One, GameStop, Verizon, and Wendy's - rely on SnapLogic to automate business processes, accelerate analytics and drive digital transformation. SnapLogic was founded by data industry veteran Gaurav Dhillon and is backed by blue-chip investors including Andreessen Horowitz, Capital One, Ignition Partners, Microsoft, Triangle Peak Partners, and Vitruvian Partners. Learn more at snaplogic.com.



2 West Fifth Avenue Fourth Floor San Mateo, CA 94402 +1 (888) 494-1570 info@snaplogic.com snaplogic.com