

Integrating AWS Redshift and Salesforce

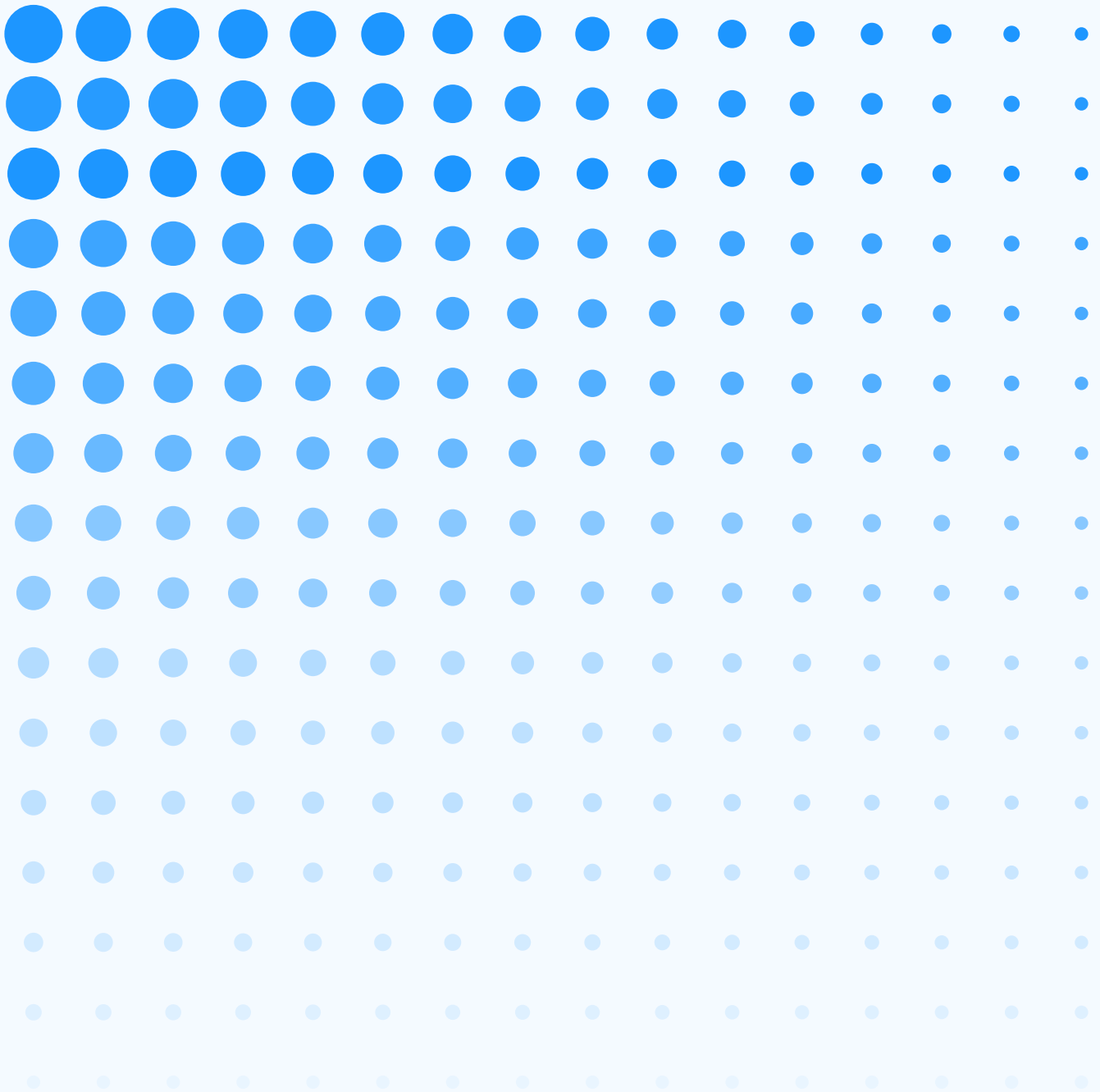
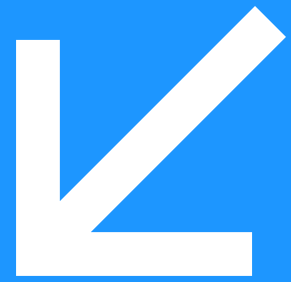


Table Of Contents



INTRODUCTION	1
SALESFORCE OVERVIEW: THE #1 CRM	2
AWS OVERVIEW: THE #1 PUBLIC CLOUD	3
SALESFORCE AND AWS REDSHIFT: BETTER TOGETHER	4
HOW TO INTEGRATE AWS REDSHIFT WITH SALESFORCE	5
FINDING THE RIGHT SALESFORCE/REDSHIFT INTEGRATION TOOL	6
HOW INTEGRATE.IO CAN HELP INTEGRATE SALESFORCE WITH REDSHIFT	7

01

INTRODUCTION

It's hard to overstate just how valuable of an asset your customer data is.

Data-driven businesses are engaged in a digital “arms race,” collecting and analyzing everything from customers’ basic demographic information to their entire website visits and purchases history. To wrangle all this data, you’ll need an excellent [customer relationship management \(CRM\)](#) tool—and here, it’s hard to go wrong with the Salesforce CRM. Salesforce is the world’s most popular CRM platform, helping companies of all sizes and industries keep track of their customers and prospective customers.

Salesforce has its own native capabilities for customer data analytics. In many cases, however, it’s a good idea to export this data to a powerful, centralized, dedicated solution for data storage and analysis.

That’s precisely where Amazon Redshift comes in. Redshift is Amazon Web Services’

data warehousing solution hosted in the cloud. With AWS Redshift, users can store, process, and crunch massive quantities of information, helping deliver cutting-edge insights to data-driven businesses.

All of this sounds great—so what’s the catch? The biggest barrier is that Salesforce and Amazon Redshift are two separate platforms from two different IT providers. So how can you migrate your Salesforce data into Amazon Redshift (without getting a computer science Ph.D. or pulling your hair out in frustration)?

In this white paper, we’ll discuss how to do exactly that. First, we’ll offer an overview of Salesforce and Amazon Redshift, as well as the benefits of combining them. We’ll then discuss several methods of integrating your Salesforce data with Amazon Redshift so that you can make the choice that’s right for your situation.

02

SALESFORCE OVERVIEW: THE #1 CRM

[Salesforce](#) is by far the world's leading customer relationship management (CRM) solution. As of the last count, Salesforce had a [20% share of the CRM market](#). That figure is more than three times its nearest CRM rivals from companies such as Oracle, Microsoft, and SAP.

What makes Salesforce so popular? The Salesforce CRM platform comes with many different features for tracking your sales workflows, accounts, and customers. Some of the key benefits of Salesforce include:

1 EASY TO USE

User-friendliness is an essential factor in software adoption, so you'll be pleased to hear that Salesforce makes ease of use a priority. Not only is the Salesforce software easier to install than competitors, but it also has a gentler learning curve. With Salesforce, even non-technical users have full visibility into your company's sales pipeline.

2 CLOUD-BASED

Salesforce is a cloud-based CRM solution, which comes with multiple advantages. For one, users can access Salesforce from anywhere and at any time, using a desktop, laptop, smartphone, or tablet. Second, cloud-based CRMs like Salesforce offer greater scalability and reliability than on-premises CRM deployments, growing alongside you as your business expands.



3 FEATURE-RICH

Salesforce offers a wide variety of features that are enough to suit the needs of any organization. With Salesforce, you can construct a panoramic view of all your accounts, leads, opportunities, and customers. The Salesforce platform also

includes reporting, dashboards, and forecasting capabilities, so that key decision-makers have a better sense of what's happening and what's coming down the pipeline.

03

AWS OVERVIEW: THE #1 PUBLIC CLOUD

Amazon Web Services (AWS) is the world's leading public cloud provider, with an estimated [32% market share](#). Today, AWS is one of the most profitable components of Amazon's vast business portfolio.

The IT research and advisory firm Gartner rates AWS as a "Leader" in the field of cloud infrastructure and platform services. What's more, AWS ranks first among all cloud providers on both of Gartner's metrics: "completeness of vision" and "ability to execute"—even ahead of public cloud giants such as Microsoft Azure and Google Cloud Platform.

AWS offers an extensive suite of cloud services, from compute and storage to analytics and artificial intelligence. One of the most popular products is Amazon Redshift, AWS' purpose-built cloud data warehousing solution. Companies such as McDonald's, Pfizer, GE, and Deloitte all use Redshift for their big data workloads, from storage to processing.

Redshift is a fully managed, tremendously scalable, petabyte-scale data warehouse service that can store millions of gigabytes of information. The Redshift platform also has built-in reporting and real-time business intelligence and analytics functionality, making it easier to crunch these massive datasets.



The benefits of Amazon Redshift include:

1 PERFORMANCE

Amazon claims that Redshift is the world's fastest cloud data warehouse. It achieves this blistering-fast performance with architectural decisions such as columnar data storage and massively parallel processing (MPP).

2 COST MODEL

Depending on your needs and use cases, Redshift can be a more cost-effective storage solution than maintaining an on-premises data warehouse. Redshift uses a “pay as you go” on-demand cost model. On-premises infrastructure, on the other hand, typically requires a considerable upfront capital expense and can't be easily scaled without making another large purchase.

3 SECURITY

With the average cost of data breaches now in the millions of dollars, businesses can't afford to leave their sensitive information unprotected. The good news is that Redshift has an impressive array of data security credentials, including encryption, access management, and VPC (virtual private cloud).

Want to learn more about Amazon Redshift? For the full details, check out our article [“Amazon Redshift: Comprehensive Guide.”](#)

04

SALESFORCE AND AWS REDSHIFT: BETTER TOGETHER

Thus far, we've discussed the benefits of Salesforce and Amazon Redshift in isolation.

So why should businesses use these two solutions in the same data management workflow?

When used to their fullest capacity, Salesforce and Redshift are complementary tools:

- Salesforce CRM serves as the standard repository for your customer data and has enough powerful data-crunching features to suit your day-to-day needs.
- However, you can also see benefits from moving this information into a centralized data warehouse, such as Amazon Redshift. By combining your Salesforce data with other records from across the organization (such as finance,

IT, manufacturing, and logistics), you can uncover hidden trends and insights that you wouldn't have otherwise discovered.

The predominant method of performing data integration is known as [ETL \(extract, transform, load\)](#). Data is first extracted from one or more sources, then transformed as necessary to fit the Redshift data schema, and finally loaded into the Redshift data warehouse.

In some cases, you may even want to perform data replication in reverse, migrating your data out of Redshift and back into its original Salesforce location. This process is called, appropriately, [reverse ETL](#).



Why do companies perform reverse ETL when Redshift is already a robust data warehouse? Pushing data from Redshift to Salesforce can help make it more visible and accessible to non-technical employees who don't typically interact with the data warehouse itself. Thanks to the Salesforce CRM platform's user-friendliness, people of all backgrounds and skill sets can take advantage of its data storage and analytics features.

The potential business use cases of integrating Salesforce and Redshift include:

- Combining Salesforce data with information from related platforms. Software such as HubSpot, Marketo, or Mixpanel can hold valuable bits of intelligence that help clarify or illuminate your Salesforce data.
- Running faster queries on Salesforce data. For example, queries such as "Which Salesforce accounts have purchased which products?" are slow in Salesforce due to the structure of the Salesforce database.
- Pulling data out of Salesforce for deep analytics and machine learning workflows and then pushing the results back to Salesforce. [This video](#), for example, discusses how to enrich your Salesforce data using the Amazon SageMaker cloud machine learning platform.
- Streamlining and centralizing your business intelligence and analytics infrastructure. Too many businesses rely on scattered Excel spreadsheets and ad-hoc analyses that don't get to the right people.
- Creating a "[customer 360](#)" view of your customers and prospects. For example, your Redshift data warehouse can aggregate product analytics and usage data from tools and platforms such as Blaze, Amplitude, and Snowplow. You can then push user analytics data back into Salesforce for each customer.
- Performing reverse ETL on Redshift data by pushing it to Salesforce. This breaks down silos and lets a wider variety of users access these insights.

It's not just users who notice the benefits of integrating Salesforce and Redshift—the companies themselves realize the benefits of collaboration. In June 2021, Salesforce and AWS [announced](#) that they would strengthen and expand their strategic partnership, making it easier for users to build, deploy, and manage their enterprise applications. This relationship will bind Salesforce and AWS more tightly together: Salesforce data can be connected to AWS technologies, while AWS data and services can easily be linked to Salesforce apps.

05

HOW TO INTEGRATE AWS REDSHIFT WITH SALESFORCE

By now, you should know how Salesforce and Amazon Redshift work, as well as the benefits of integrating them. The next question is: how can you perform this integration in practice for your own CRM data?

In this section, we'll delve into the technical, nitty-gritty side of things, going over the two major ways to load data from Salesforce to AWS Redshift.

1 MANUAL DATA INTEGRATION SCRIPTS

Writing your manual Salesforce/Redshift data integration scripts is a plausible option if you have the technical chops to pull it off. Here's a quick look at what this might entail:

1. First, you'll need to write code that connects to the right Salesforce APIs. Salesforce has a wealth of different APIs

depending on the use case, from REST to SOAP and from bulk to streaming data. Your developers will need to select the correct API for your use case and data.

2. Next, you'll need to set up a way to extract this information according to your desired schedule. You might want to connect to Salesforce regularly (e.g., weekly or hourly) or only when you detect changes to the database. For regularly scheduled jobs, for example, you might use a Unix command-line utility such as cron.



3. After handling the data extraction phase, the next step is to transform your data appropriately. Using the JSON files you get from the Salesforce API, you'll need to map each JSON field to the right table and column in Redshift. Here, you can define a variety of custom data transformations. At a minimum, you'll need to verify that your Salesforce data is valid and that you can map all of it to a Redshift data type.
4. The final step in the manual ETL process is to load the transformed data into Redshift. In most cases, it's easier to use an intermediate repository such as Amazon S3 cloud object storage. Once your Salesforce data is inside S3, you can run a Redshift COPY command to get it inside the data warehouse.

If this all sounds like a lot of work, that's because it is. However, the most significant advantage of this approach is that it gives you maximum flexibility. Developing these scripts in-house gives you ultimate control over their behavior and performance.

Nevertheless, manual data integration scripts also come with several downsides:

Time and budget: Like any software development project, writing custom data integration scripts can be a lengthy, involved endeavor. Spending time on this manual effort is time that you're not enjoying the benefits of Salesforce/Redshift integration. You'll also need to consider the cost—either of hiring your own in-house team of data integration experts or working with a third-party development firm.

Complexity: Writing your own custom data integration code isn't for the faint of heart. Your developers will need to know both Salesforce and Redshift front to back.

Maintenance: Although manual data integration gives you ultimate control and flexibility on the developer side, it's also, in a sense, inflexible and fragile. If Salesforce APIs ever change or Redshift receives an update, you may have to rewrite or refactor your code to account for these modifications. This is especially concerning if you outsource development: your partner may not be able or willing to update the codebase, leaving you with a broken pipeline.

2 SALESFORCE ETL TOOLS

Rather than writing custom integration scripts, most businesses use a dedicated ETL tool that can automatically connect Salesforce to Redshift. The advantages of this approach include:

User-friendliness: The best ETL tools handle all of the technical complexity for you, connecting different software and platforms behind the scenes. Low-code and no-code ETL tools allow users to construct production-ready data pipelines without touching the internal code.

Time and cost savings: Rather than spending time on development, businesses can hit the ground running with an ETL tool, leveraging the insights they gain to make crucial business decisions. ETL tools are also likely more cost-efficient than custom development and come with multiple pricing models (such as per connector or data volume).

Compatibility: Of course, ETL tools don't just work exclusively with Salesforce and Redshift data. Consumer ETL platforms offer dozens of pre-built connectors for other CRM software and data warehouses, as well

as other types of data sources and targets. This means that if the structure of your data pipeline ever changes—say, switching to Google BigQuery from Redshift—you can adjust it in your ETL tool with a few clicks, rather than having to overhaul your entire codebase.

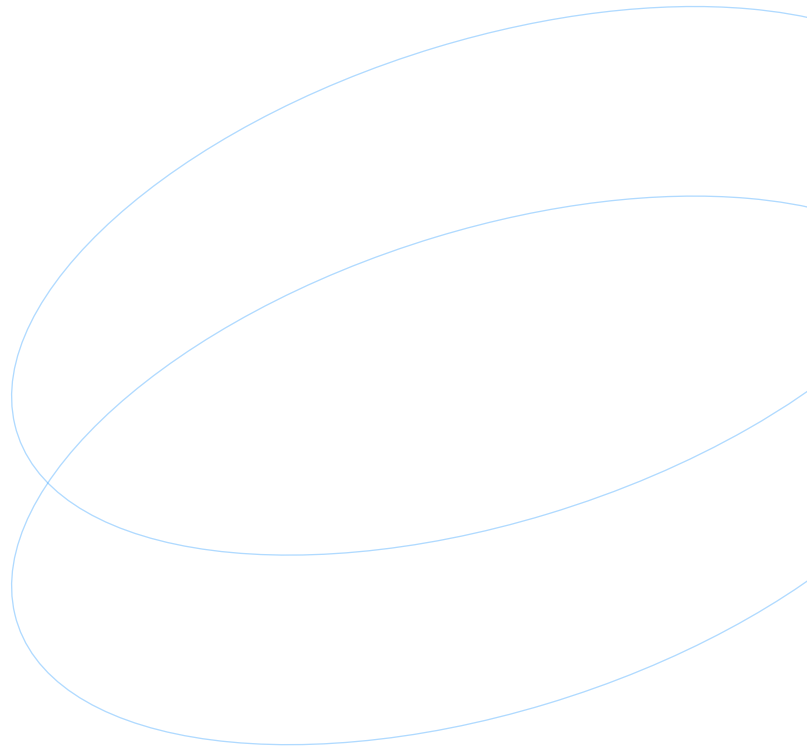
To demonstrate just how easy Salesforce/Redshift integration is with the right ETL tool, let's use Integrate.io as an example. Below is an overview of how to migrate Salesforce data to Redshift with the Integrate.io platform:

1. Integrate.io provides the “Salesforce Source” connector for extracting data from Salesforce objects. Once you select this component, you'll need to define a few key settings: the name of the table to import from, the table fields to import, an optional SQL query to filter these records, etc.
2. After capturing this Salesforce data, you can define a number of transformations to cleanse and validate this information. Integrate.io provides a rich array of transformations, from cloning and

splitting data flows to selecting a random sample of data records. To see a more complete list of the possibilities available to you, read our article [“What is Data Transformation?”](#).

3. Finally, you'll use Integrate.io's Redshift connector to move this transformed data to its ultimate destination. First, you'll need to configure Redshift to be publicly accessible and accept incoming connections from Integrate.io. Then, within the Integrate.io interface, create a new Redshift connection and enter your credentials.

Check out our article on [Integrate.io's Salesforce integration](#) for the full details. Our article [“Allowing Integrate.io access to my Redshift cluster”](#) covers the process in-depth on the other end of the pipeline.



06

FINDING THE RIGHT SALESFORCE/REDSHIFT INTEGRATION METHOD

When integrating your Salesforce data with AWS Redshift, you have essentially two choices: use an off-the-shelf ETL platform for automatic data integration or write your own code.

For most businesses, it's better to leave this work in the hands of the professionals by going with an ETL tool. The right decision here will depend on factors such as:

- The complexity of the data transformations that you need to perform.
- The frequency of the data integration (e.g., a one-off project or a regular occurrence).
- Integrations with other third-party data sources and targets.
- Developers' knowledge and technical background (i.e., familiarity with Salesforce and Redshift schemas, APIs, data formats, etc.).

If you decide to use a tool for ETL automation, another question immediately arises: which tool will you use? The good news is that there's no shortage of options. Below are some considerations for you to narrow down the search:

Integrations: ETL tools offer pre-built integrations and connectors to simplify the process of building data pipelines. However, no two tools provide precisely the same suite of connectors. Make sure the tool you choose is compatible with your current and future ETL workflow.

Cost model: As mentioned above, ETL tools differ in terms of pricing and their pricing model. Integrate.io, for example, charges a flat fee based on the number of connectors in your data pipeline, not on the amount of data you consume. This makes Integrate.io a smart choice if you expect your data consumption to grow over time while keeping your pipeline the same.

If you decide to use a tool for ETL automation, another question immediately arises: which tool will you use? The good news is that there's no shortage of options. Below are some considerations for you to narrow down the search:

Integrations: ETL tools offer pre-built integrations and connectors to simplify the process of building data pipelines. However, no two tools provide precisely the same suite of connectors. Make sure the tool you choose is compatible with your current and future ETL workflow.

Cost model: As mentioned above, ETL tools differ in terms of pricing and their pricing model. Integrate.io, for example, charges a flat fee based on the number of connectors in your data pipeline, not on the amount of data you consume. This makes Integrate.io a smart choice if you expect your data consumption to grow over time while keeping your pipeline the same.

Security and compliance: Redshift has multiple security techniques for protecting your Salesforce data at rest inside the data warehouse. However, you'll also need to secure this data in transit while it's being transferred between Salesforce and Redshift. In addition, if you're in an industry or region covered by specific data regulations (such as GDPR or HIPAA), look for an ETL tool that's compliant with these regulations.



07

HOW INTEGRATE.IO CAN HELP INTEGRATE SALESFORCE WITH REDSHIFT

The advantages of integrating Salesforce with Redshift are clear, but the specific data integration method will depend on your unique situation.

Using a dedicated ETL and data integration tool is the right choice for many businesses. But with so many ETL tools for integrating Salesforce and Redshift, how can you select the right one?

[Integrate.io](https://integrate.io) is a powerful, user-friendly, feature-rich ETL tool that makes it easy to unite Salesforce and Redshift in a single data integration workflow. The advantages of using Integrate.io to integrate Redshift and Salesforce include:

Low-code interface: Integrate.io has a simple, user-friendly, drag-and-drop visual interface. Users of any technical level can

build robust, production-ready data pipelines with Integrate.io, requiring little to no coding.

Pre-built integrations: Integrate.io comes with more than 140 pre-built connectors and integrations. These include everything from data warehouses and SaaS platforms to CSV files and analytics tools. No matter what your data integration workflow looks like, Integrate.io can help you build it.



Top-notch security: The Integrate.io platform is compliant with popular data security and privacy regulations such as GDPR, HIPAA, CCPA, and SOC 2. Integrate.io also enforces network data security with firewalls and permissions-based access and uses encryption for data at rest inside the platform.

Excellent support: If you ever have a question, issue, or concern, Integrate.io is here to listen. Our friendly, knowledgeable support agents are available via email, live chat, and Zoom calls.

Want to learn more about how Integrate.io can help with your Salesforce/Redshift integration? [Get in touch with our team of data experts today](#) to chat about your business needs and objectives. If you'd rather try it out for yourself, you can also schedule a 7-day pilot of the Integrate.io platform, no strings attached.