Implementing a Cloud Analytics Solution on Amazon Web Services with Amazon Redshift and Informatica Cloud

Reference Architecture Guide
Introduction to Cloud Analytics

The abundance and variety of data available to companies today holds the promise of maximizing every aspect of the business, from customer acquisition and retention to better operational performance. But harnessing data for business success relies on your ability to rapidly combine and analyze massive volumes of disparate data residing in cloud and on-premises in meaningful ways.

Traditional approaches for building and evolving on-premises data warehouses are robust, but can take months and cost millions of dollars just to get started. Then, add in maintenance and growth costs of the on-premises system, and the cost and effort increase even further during the life cycle of the system.

With Amazon Redshift and Informatica Cloud, you can rapidly and cost-effectively build a cloud first data warehouse and extend your data warehouse to any on-premises SaaS or to other data sources. With the connectivity and power of this solution, you can deliver the data-driven agility required for business success today and tomorrow.
Amazon Redshift provides a petabyte-scale cloud data warehousing service that’s fully managed, fast, and cost-effective. It also serves as a powerful launch pad for analytics applications such as Tableau, Qlik, Birst, and Amazon QuickSight. The elasticity and speed of this environment changes the dynamics of data warehousing by making it easy to provision nodes, scale on demand, and query datasets securely. Clusters can be resized dynamically without downtime, and distributing workloads across compute nodes can optimize I/O time.

Amazon Redshift’s architecture leverages massive parallel processing (MPP) capabilities with columnar storage and data compression to enable timely execution of even the most complex queries and the resulting business insights.

However, a data warehouse is only as good as the data inside. To fully realize the benefits of Amazon Redshift, it’s imperative to rapidly load it with trustworthy, connected, meaningful, and timely data.

Informatica Cloud is a multi-tenant fully managed integration platform as a service (iPaaS), which can be rapidly and cost-effectively deployed, and connects all your disparate data sources with Amazon Redshift. It enables you to access, aggregate, synthesize and load data into Amazon Redshift from numerous systems, including cloud sources such as SaaS, social and IoT, as well as on-premises systems, such as SAP and relational databases. It gives you the agility to rapidly spin up a small cloud analytics Redshift project and seamlessly scale it up or down as data volume and your business needs demand. Combining a cloud integration solution from Informatica with Amazon Redshift enhances and expedites your analytics initiative and unlocks the power of Amazon Redshift.

This guide provides design and implementation recommendations, an architectural reference, quick start guidance, and tutorials to help you start operating in a hybrid cloud environment using Informatica Cloud with AWS services.

Informatica PowerCenter and Informatica Cloud provide a single solution for integration data between AWS data stores and a wide variety of traditional on-premises resources and popular cloud-based services like Salesforce, Workday, and Marketo. In this document, we focus on Informatica Cloud capabilities for data integration with Amazon Redshift.

This guide is written for IT administrators, data engineers, and business analysts. It discusses planning topics, architectural considerations and options, and configuration steps for dealing with a hybrid cloud environment that uses Informatica Cloud Services (ICS) and AWS services, such as Amazon Redshift, Amazon Elastic Compute Cloud (Amazon EC2), Amazon Virtual Private Cloud (Amazon VPC), Amazon Simple Storage Service (Amazon S3) and Amazon Relational Database Services (RDS), to run a scalable and reliable managed data warehousing solution. This guide discusses the planning and architectural considerations in the context of several deployment Amazon Aurora scenarios. It considers several approaches that have tradeoffs in terms of latency, cost, complexity, and security.

This document offers step-by-step instructions for the following tasks:
- Creating an Amazon Virtual Private Cloud (VPC) instance.
- Launching a Windows virtual instance with Amazon Elastic Cloud Compute (EC2) service.
- Launching an Amazon Redshift data warehouse cluster.
- Launching an Amazon MySQL Relational Data Service (RDS) instance.
- Setting up Informatica Cloud.
- Loading data from various sources into Amazon Redshift with Informatica Cloud.
- Integrating cloud analytics data sources with Informatica Cloud.
Before You Get Started

Implementing a cloud analytics platform in the AWS cloud is an advanced topic. If you are new to AWS, see the Getting Started section of the AWS documentation at http://docs.aws.amazon.com/gettingstarted/latest/awsgsg-intro/intro.html.

The reference architecture will provide high-level references and links to specific areas of documentation. Your understanding of this reference architecture guide will be easier if you are already familiar with the following topics:

- Amazon EC2 and Amazon VPC
- Amazon Redshift, Amazon S3, Amazon Aurora, and Amazon RDS
- Informatica Cloud
- Windows Server 2008 or higher or Red Hat and SUSE Linux
- Windows Server Active Directory and DNS

This document briefly discusses Informatica Cloud setup tasks. It focuses on configuration topics that require careful consideration when you are planning and deploying a cloud analytics warehouse in the AWS cloud that is part of a hybrid cloud environment, and you use Informatica Cloud.

For Informatica Cloud configuration guidance, functionality, and best practices see the Informatica Cloud product documentation: https://community.informatica.com/docs/.

Informatica Cloud Overview

Informatica Cloud is a multi-tenant, on-demand Integration Platform as a Service (iPaaS) delivered as a software subscription service. Informatica Cloud provides native, high volume, high performance data integration with Amazon Redshift and supports out-of-the-box connectors to any cloud and on-premise data system and application, including Amazon S3, Amazon Aurora, Amazon RDS and Amazon DynamoDB.

Informatica Cloud supports secure data movement between on-premises environments and AWS using the Secure Agent. The Secure Agent is a lightweight application that runs in both the EC2 environment and an on-premises data center system, to securely access the Informatica cloud services located in the Informatica Cloud managed environment.

Informatica Cloud enables you to develop and run data integration tasks (also known as mappings), data synchronization tasks, task flows, and schedules.

Informatica Cloud integration is a visual, easy-to-use metadata-driven solution, enabling self-documenting code, improved data visibility via metadata, extensive reuse in development, and automation in deployment.

With Informatica Cloud and Amazon Redshift, you will be up and running with your new cloud data warehouse on AWS and loading it with meaningful data within hours, not days. The result is that you can set up a fully functional cloud analytics warehouse that will help you understand your company’s data, quickly and effectively.
Informatica Cloud includes the following components:

- **Informatica Cloud Service** – a hosted application service that stores all task and organization information (metadata). The application metadata is configured through a wizard-based web browser interface.

- **Informatica Cloud applications** – purpose-built data integration applications, such as Data Synchronization, Data Replication, Contact Validation, Data Assessment, Data Masking, and Mapping Configuration tasks.

- **Informatica Cloud Secure Agent** – a small footprint application that enables secure communication across the firewall between the client organization and Informatica Cloud. It is installed on a local machine or on an Amazon EC2 instance. It is a functionally equivalent, run-time version of the enterprise-class Informatica PowerCenter execution component. All Informatica Cloud data integration services use the Secure Agent to traverse the firewall and access sources and targets such as application, relational database, and file stores over the network.

- **Informatica Cloud provides connectivity to various data sources through their native or generic API.** The system can connect natively to AWS services such as:
  - Amazon S3
  - Amazon Redshift
  - Amazon Aurora
  - Amazon RDS
  - Amazon EMR
  - Amazon DynamoDB
  - Amazon QuickSight

**Amazon Redshift Overview**

Amazon Redshift is a petabyte-scale clustered data warehouse managed as a service. Redshift delivers fast query performance by using columnar storage technology to improve I/O efficiency.
and parallelizing queries across multiple nodes. You access Redshift by using standard SQL technology, which allows you to use a wide range of familiar SQL clients and analysis solutions.

Most of the common administrative tasks associated with provisioning, configuring, and monitoring a data warehouse are automated with Redshift. Backups to Amazon S3 are continuous, incremental, and automatic. Restores are fast—you can start querying in minutes while your data is spooled down in the background. Enabling disaster recovery across regions takes just a few clicks. For more information about Redshift, see [http://aws.amazon.com/redshift/](http://aws.amazon.com/redshift/).

You load data into Redshift in parallel in order to achieve high speed loading of large data sets. The most common way to achieve parallel loading is to place files into the Amazon Simple Storage Service (S3), and then issue a copy command. For best performance, split your load data into multiple files based upon the size of the cluster. Redshift then assigns each file to nodes in the cluster to operate upon in parallel.

For more technical details on Redshift, see [http://docs.aws.amazon.com/redshift/latest/dg/welcome.html](http://docs.aws.amazon.com/redshift/latest/dg/welcome.html).

### Integration Use Cases for AWS Redshift Analytics

Database administrators, who are well-versed with relational database constructs, will find it easy to create and provision new Amazon Redshift nodes and clusters. When it comes to loading data into Redshift, however, getting bogged down with writing manual SQL scripts makes it hard to be agile and adapt to rapidly changing requirements from business stakeholders.

In a world that’s exploding with usage of SaaS applications, each one has its own specific API which gets upgraded multiple times a year. Keeping up with those upgraded APIs may require modifying hand-coded SQL scripts when there’s a change.

Many other complex on-premises applications have their own proprietary protocols that you need to leverage to push data into Redshift. SAP, for instance, can be accessed through regular web services, through tables, as well as proprietary protocols known as IDoCs and BAPIs. All of these connectivity issues can be managed through Informatica Cloud’s out-of-the-box native connectors.

Moreover, front-office SaaS applications such as Salesforce, Netsuite, and Marketo have customer data that is constantly changing. To identify trends quickly, these changes need to be
captured and analyzed in Redshift. Informatica Cloud can use its change detection capabilities to only replicate Salesforce data that has changed since the last job, and move it into a staging area or directly to Redshift.

Even regular relational databases have complex integration scenarios with Redshift. Many relational databases, whether on-premises or in the cloud (such as RDS) are used as a staging area to join data from disparate data warehouses, data marts, and other applications. They are then transformed and aggregated before being pushed into Redshift for analysis. Analysis of this data is imperative to help understand your customers, improve marketing, sales and service, to bridge the business and IT gap by automating business processes.

Implementing a Cloud Analytics solution on AWS and moving data to and from Redshift, RDS, S3, or other AWS solutions can be quickly designed in an easy-to-use canvas, so a business user can perform analytics on their data in hours, not days.

Informatica Cloud’s Mapping Designer canvas has numerous transformation routines for normalizing, joining, aggregating, merging, and sorting data. These transformations can vastly reduce the time it takes to get data into Redshift. Those who have used PowerCenter in the past will be inherently familiar with this approach.

Informatica Cloud also offers the less technical user a step-by-step path to simply choose a source, a target, set up some high-level data filters, create a new Redshift table, and schedule that integration job to run on a repeatable schedule.

Informatica Cloud automatically manages the complexity of these operations for you when you choose Redshift as a target. Behind the scenes, Informatica Cloud optimizes the load for the cluster size, stages data files to S3 on your behalf, and then issues SQL commands to complete the loading process.

The following diagram shows the load process steps:
Cloud Analytics Scenarios and Solutions Overview

Many customers are migrating applications to the cloud and implementing new functionality in a “cloud first” model to achieve cost savings and improved agility. These applications are either operated by the customer (typically by using approaches commonly called IaaS or PaaS), or operated as a service by third parties (commonly called SaaS). Often both on-premises and SaaS applications share common data such as customer, product, and order information.

With the adoption of SaaS applications, the need to analyze more data more frequently increases, and it needs to be in conjunction with data from existing on-premises systems. Traditional data warehouses tend to have long development cycles, primarily because of the effort necessary to set up and provision on-premises databases.

Agile data warehousing in the cloud provides an easy, cost-effective means of augmenting existing environments through the full cycle of deploying development, test, and production environments.

This reference architecture guide provides insight on creating and building a data warehouse with Amazon Redshift and using Informatica Cloud for data integration. It focuses on the specific considerations of quickly starting in the cloud analytics environment using AWS, but this is not a general guide on implementing Informatica Cloud solutions.

Many of the techniques and topologies described in this guide closely align with other types of problems, like initial and incremental load into a data warehouse, although they are not specific targets for this guide, and therefore the guidance might not be complete.

This guide is designed for the IT administrators and business users who:

- Have used Informatica products like PowerCenter in the past
- Have Informatica products in their organization, but don’t have access to these services and want a product that allows them quick integration and analytics with Redshift
- Are new to data analytics and want similar results to those other users.

Even though we focus on Informatica Cloud as data integration tool, the same implementation concepts apply to Informatica PowerCenter.
Cloud Analytics with Amazon Redshift and Informatica Integration

Scenario

Getting Started with Redshift and Analytics

Scenario 1: Loading Flat File Data from S3 into Redshift

The first scenario is a simple scenario to help get you going. Loading data from flat files into a data warehouse or other systems is a common operation. Very often when you run data transformations on-premises, you will read files from a drop location in a file share to perform a load.

A similar pattern when using the cloud is to read files out of a location in Amazon Simple Storage Service (S3). In this scenario, a user will read data from flat files in S3 and load the data into various targets including Redshift, Relational Database Services, and SaaS applications like Salesforce.
Scenario 2: Loading Data from Relational Databases

In a common pattern, customers use Informatica Cloud to move data from a transactional system that resides on-premises or data already residing in AWS into an Amazon Redshift data warehouse.

This scenario shows how to pull data from multiple tables and apply filters. Considerations for sizing and topology for running Informatica Cloud appliance on-premises or in AWS are also covered.

Scenario 3: Integrating Cloud and Transactional Data Sources

This scenario extends Scenario 2 by integrating data from a SaaS based source. In the example, data originating in Salesforce is merged with data from transactional systems and then pushed into a Redshift data warehouse. Data transformations are also applied.
Security Considerations

Security is of paramount importance to virtually every enterprise. If you run workloads in AWS, then you have already considered security implications. If this is your first venture into the AWS, then take the time to understand the AWS security model.

You can run workloads as securely as or even more securely in the cloud than you would on-premises. AWS addresses many of the common concerns and considerations in the AWS Security Center at http://aws.amazon.com/security and lists resources at http://aws.amazon.com/security/security-resources. Whitepapers are available on security processes, best practices, compliance, Microsoft platforms, and HIPAA. The AWS compliance center http://aws.amazon.com/compliance/ lists the commercial and federal certifications that AWS obtained.

There are many different possibilities when configuring your network security. You can choose to either run the Informatica Cloud Agent on AWS or on-premises. When extracting data from disparate databases and data applications that reside on-premises, these underlying systems need to be analyzed and the applications running on these systems need comply with your company’s security guidelines. You have several options for how you configure your environment. The following sections outline several common approaches.

Informatica Cloud uses 128-bit Secure Sockets Layer (SSL) technology to protect data. It uses authentication and encryption to ensure that data is secure and available only to users within the organization. When you log in to Informatica Cloud, https precedes the URL in the address field of the web browser to indicate that the connection is secure. See http://trust.informaticacloud.com/security for details.
Informatica Cloud Secure Agent On-Premises Connecting to Amazon Redshift Using HTTPS

The simplest configuration is to deploy the agent on-premises alongside your transactional systems and securely push the data across the public internet. You run your Amazon Redshift cluster in a Virtual Private Cloud (VPC) that can access the internet. The data from the Secure Agent to the Amazon Redshift data warehouse is encrypted in transit by using HTTPS for communications. In order to use this configuration, the Amazon Redshift cluster must also be accessible over the public internet.

For higher security, you can configure the security group (firewall) for the Amazon Redshift cluster to only allow connections from the public IP address range of your company. You must allow public IP address ranges originating from the Informatica Cloud to connect to the Secure Agent through your firewall on-premises. The public IP address ranges required for access are documented in the Informatica Cloud User Guide, available at https://community.informatica.com/docs/.
Informatica Cloud Secure Agent On-Premises Connecting to Amazon Redshift Using VPN

This is similar to the first case, only you use a VPN connection from on-premises to an AWS Virtual Private Cloud for additional security. You deploy the agent on-premises alongside your transactional systems, and the agent pushes data across a virtual private connection established between your on-premises data center and your VPC in AWS. The virtual private connection encrypts data in transit.

You can make your Virtual Private Cloud only accessible from your corporate network, and you can also make your Amazon Redshift cluster only accessible using a private IP address. You can also choose to use https for additional encryption across the VPN tunnel.
**Informatica Cloud Secure Agent on AWS Connecting to On-Premises Sources Using VPN**

You can deploy the Secure Agent on AWS EC2 and pull the data across a private VPN connection established between your corporate network and the virtual private cloud, which encrypts all data in transit. You can choose to make your Amazon Redshift data warehouse only privately accessible. You must have public connectivity in order for the Secure Agent to communicate with the Informatica Cloud, and you can restrict the public IP address ranges to the ranges specified in the *Informatica Cloud User Guide*.

**Informatica Cloud Agent Running in AWS Connecting to Cloud Data Sources**

When all of the data sources being analyzed originate in the cloud, the best practice is to run the Secure Agent in the cloud as well. In this case you might choose to omit a VPN connection back to your corporate network, although often this is still a requirement since users analyzing the data still need a secure connection.
For more information on AWS security, see https://aws.amazon.com/documentation/vpc/.

**Instance Type Overview**

When you purchase virtual instances from AWS to run your workloads, you specify an instance type. Instance types determine the amount of compute, memory, network capacity, and ephemeral storage you will have available for your workload. Each of these aspects impacts your ability to process and transport data when using EC2 to host the Secure Agent.

There are many options for instance types available in AWS for a wide variety of workloads. The most important aspects for Secure Agent performance are computing and network capacity. The C4 instance type family on AWS is tuned for compute intensive workloads while the M4 type provides a balance of performance, memory, and network performance. For smaller workloads, the M4.large is recommended.

The following table describes the minimal and recommended instance types for each type of workload:

<table>
<thead>
<tr>
<th>Workload Type</th>
<th>Storage Type</th>
<th>Recommended Instance Type</th>
<th>Minimum Instance Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small subset of rows, simple transformations</td>
<td>any</td>
<td>T2.small</td>
<td>T2.small</td>
</tr>
<tr>
<td>Thousands of rows, some transformations</td>
<td>any</td>
<td>M4.large</td>
<td>T2.medium</td>
</tr>
<tr>
<td>Hundreds of thousands of rows with complex transformations, joining of multiple data sources</td>
<td>SSD Elastic Block Storage</td>
<td>C4.large</td>
<td>M4.medium</td>
</tr>
<tr>
<td>Replication or synchronization of existing data sources, millions of rows, few transformations, lookups, or joins.</td>
<td>SSD Elastic Block Storage</td>
<td>C4.2Xlarge</td>
<td>C4.large</td>
</tr>
</tbody>
</table>

For information about Amazon instance types, see: http://aws.amazon.com/ec2(instance-types/).
Hardware and Software Setup for Informatica Cloud Secure Agent

The Secure Agent operates either on-premises or running in AWS. The Secure Agent runs on either a Windows or Linux operating system. The Informatica Cloud User Guide describes how to set up the Secure Agent on a Linux operating system.

There are network and software considerations for using the Secure Agent. How you address the network considerations issues depends upon whether you run the agent on-premises or in the AWS cloud. The software considerations are the same in both cases.

Hardware and Network Configuration

There are many potential topologies, four of which are outlined in the security considerations section. In this tutorial, we will run through how to setup a configuration on Amazon EC2 and provide automation that will help you create this environment. In this case, automation refers to the ability within AWS to describe and script a network and server environment that can be created on your behalf by defining templates and using scripts.

The generic considerations for any hybrid approach are:

- Providing connectivity from the Secure Agent to the Informatica Cloud
- Providing connectivity from the Secure Agent to source and target systems
- Provisioning connectivity between on-premises and an AWS Virtual Private Cloud (VPC)
- Creating a server instance to host the Informatica Cloud Secure Agent. In some simple non-production or limited-production cases, you might choose to run this from your desktop. That scenario is not considered herein.

Often you will want to setup a VPN. For this more advanced setup scenario, go to: http://docs.aws.amazon.com/AmazonVPC/latest/NetworkAdminGuide/Welcome.html.

This environment topology is configured so that the Secure Agent Security Group allows access for Informatica Cloud to control the Secure Agent that you will install on the a sample AWS EC2 Windows instance. Instances running in the public DMZ are accessible from the Internet through
the Internet Gateway (IGW). The private subnet containing an RDS MySQL database and an Amazon Redshift cluster cannot be reached from the public internet directly, but instances in the public subnet are able to reach the database subnet.

Typically you will also configure a VPN. Once the VPN is configured, you can reach the instances in the private subnet from your corporate network. If you are using the approach described in the section “Informatica Cloud Secure Agent On-Premises Connecting to Amazon Redshift Using HTTPS,” then the Amazon Redshift cluster and RDS Database should be created in the public subnet.

**Informatica Cloud Service Setup**

**Informatica Cloud Secure Agent Setup**

Use the following instructions to set up Informatica Cloud on Windows. The setup process for Linux is quite similar. See details on Linux in the Informatica Cloud User Guide.

Log in to Informatica Cloud by accessing: [https://app.informaticaondemand.com/ma/login](https://app.informaticaondemand.com/ma/login)

If you don’t have an Informatica Cloud account, visit [https://aws.amazon.com/marketplace/seller-profile/ref=dtl_pcp_sold_by?ie=UTF8&id=d13da660-f20f-4a0b-9f73-77a73706ce96](https://aws.amazon.com/marketplace/seller-profile/ref=dtl_pcp_sold_by?ie=UTF8&id=d13da660-f20f-4a0b-9f73-77a73706ce96)

Once logged in select “Download the Secure Agent” and select your appropriate operating system:
Once you have downloaded the agent either locally on your laptop, server, or onto an EC2 instance, walk through the step-by-step wizard to install the agent and configure it to your Informatica Cloud user name and password.

**Note:** See the *Informatica Cloud User Guide* for details on installing the Secure Agent on Linux.

When the Secure Agent installation is complete, you will be prompted to register it.

1. Enter your ICS account name and password.
2. Verify that the secure agent has Status set to “Active”.

   The Secure Agent name will display your computer or EC2 instance name:

   ![Secure Agent Interface](image)

   **Note:** To access Informatica Cloud online help, click on the question mark button on the right-side of the window:
To view the navigation panel, click on the folder icon in the upper-left corner of the online help page:

Informatica Cloud AWS Redshift Connector Setup

1. Once logged into Informatica Cloud, select Configure → Add-on connectors.
2. Select the connectors you want to use.

<table>
<thead>
<tr>
<th>Add-On Connectors</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Amazon DynamoDB</strong></td>
</tr>
<tr>
<td>Publisher: Informatica Cloud</td>
</tr>
<tr>
<td>Updated On: Jun 18, 2016</td>
</tr>
<tr>
<td><strong>Amazon S3</strong></td>
</tr>
<tr>
<td>Publisher: Amazon Web Services</td>
</tr>
<tr>
<td>Updated On: Jun 18, 2016</td>
</tr>
<tr>
<td><strong>Amazon Redshift</strong></td>
</tr>
<tr>
<td>Publisher: Informatica Cloud</td>
</tr>
<tr>
<td>Updated On: Jun 18, 2016</td>
</tr>
</tbody>
</table>

3. Click “Free Trial” for any connector you want to add, whether it be Amazon S3, Amazon Redshift, or Marketo, Tableau, etc.
4. Once you have added Amazon Redshift into your org, enter the proper credentials to access your Redshift cluster.

Use this table for more information about connections settings:

<table>
<thead>
<tr>
<th>Connection Property</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Runtime Environment</td>
<td>The name of the runtime environment where you want to run the tasks.</td>
</tr>
<tr>
<td>Username</td>
<td>User name of the Amazon Redshift account.</td>
</tr>
<tr>
<td>Password</td>
<td>Password for the Amazon Redshift account.</td>
</tr>
<tr>
<td>Schema</td>
<td>Amazon Redshift schema name. Default is public.</td>
</tr>
<tr>
<td>AWS Access Key ID</td>
<td>Amazon S3 bucket access key ID.</td>
</tr>
<tr>
<td>AWS Secret Access Key</td>
<td>Amazon S3 bucket secret access key ID.</td>
</tr>
<tr>
<td>Master Symmetric Key</td>
<td>Optional. Amazon S3 encryption key. Provide a 256-bit AES encryption key in the Base64 format.</td>
</tr>
</tbody>
</table>
**Connection Property** | **Description**
--- | ---
Cluster Node Type | Node type of the Amazon Redshift cluster. You can select the following options:  
- ds1.xlarge  
- ds1.8xlarge  
- dc1.large  
- dc1.8xlarge  
- ds2.xlarge  
- ds2.8xlarge  
For more information about nodes in the cluster, see the Amazon Redshift documentation.

Number of Nodes in the Cluster | Number of nodes in the Amazon Redshift cluster. For more information about nodes in the cluster, see the Amazon Redshift documentation.

JDBC URL | Amazon Redshift connection URL.

Number of bytes needed to support multibytes for varchar | Applicable to Create Target. Reads the Varchar precision of the source table and creates the target table with 1x/2x/3x/4x times of the source precision to successfully write multibyte characters in the target table.  
**Note:** You cannot create a target table if the Varchar precision exceeds 65535, as that is the maximum allowed.

**Informatica Cloud Integration Tasks**
Once you have setup your connection to Informatica Cloud and the Secure Agent on your laptop, server, or onto an EC2 instance, walk through the step-by-step Data Synchronization Task Wizard.

1. Click “Task Wizards” → Data Synchronization.
2. Set the task operation to “Insert” and click Next.

3. Select your “Source” from whatever connection type you choose and click Next.
4. Select your AWS Redshift connection as the Target.
5. Optionally, click “Create Target” to build the source schema as a new Redshift table. You can specify which fields you want to keep and remove. Click Next.

6. Set up any necessary data filters and click Next.
   For example, you can filter records by any field where the date that an account was created is greater than 2016 or by any records where the field name is equal to a specific string.

7. Optionally, select the fields that you want to map and click Next.
If you are creating a new Redshift table, select all fields to map. You can also set up row-level field expressions by selecting the “fx” symbol next to each target field. Once you have done so, you can take the source field and create an expression on that data as pictured below:
8. Set the task you are using to run on a repeatable basis, as often as every minute. You will also need to set up some advanced configuration properties to integrate with AWS Redshift. To do so see the next section.

**Informatica Cloud AWS Redshift Advanced Properties**

Once you have entered in the necessary properties in the Advanced Properties section, you can select “Save and Run” which will start your integration task and create or update a table based on the naming convention you chose in Step 3 above.

After that you can do some quick analytics on that data to make sure it’s the correct data and then modify as necessary afterwards.

**Informatica Advanced Users**

**Informatica Cloud and PowerCenter Interoperability**

Informatica Cloud leverages PowerCenter assets. It is powered by the same “Intelligent Data Platform” as PowerCenter. This empowers PowerCenter users to run their PowerCenter mappings on Informatica Cloud, and vice versa, via a simple export process.
To run PowerCenter mappings on Informatica Cloud:

1. Log in to your Informatica Cloud account.
3. Create a PowerCenter Task.
5. Click Save and Run the task.
Informatica Cloud for Redshift -- Free Trial on AWS Marketplace

Informatica Cloud for Amazon Redshift is available on the AWS Marketplace for both Linux and Windows. You can start your 15-day free trial or buy it for a low hourly rate.

Find us on the AWS Marketplace at one of the following links:


About Informatica.

Digital transformation is changing our world. As the leader in enterprise cloud data management, we’re prepared to help you intelligently lead the way. To provide you with the foresight to become more agile, realize new growth opportunities or even invent new things. We invite you to explore all that Informatica has to offer—and unleash the power of data to drive your next intelligent disruption. Not just once, but again and again.

About AWS.

For 10 years, Amazon Web Services has been the world’s most comprehensive and broadly adopted cloud platform. AWS offers over 70 fully featured services for compute, storage, databases, analytics, mobile, Internet of Things (IoT) and enterprise applications from 33 Availability Zones (AZs) across 12 geographic regions in the U.S., Australia, Brazil, China, Germany, Ireland, Japan, Korea, and Singapore. AWS services are trusted by more than a million active customers around the world – including the fastest growing startups, largest enterprises, and leading government agencies – to power their infrastructure, make them more agile, and lower costs. To learn more about AWS, visit http://aws.amazon.com.