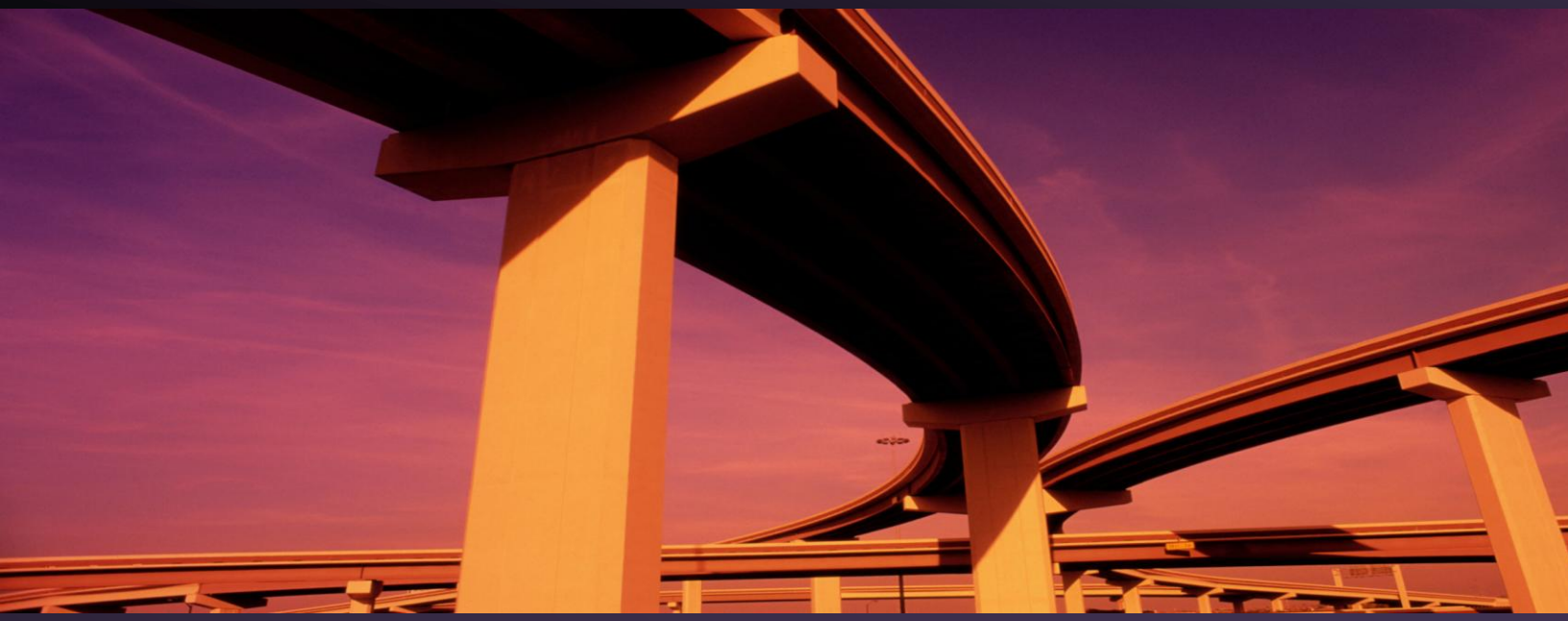
apatar beta
connecting data



Five Steps to Integrate SalesForce.com with 3rd-Party Systems and Avoid Most Common Mistakes

This white paper will help you learn how to integrate your SalesForce.com data with 3rd-party on-demand, on-premises, and legacy applications (e.g., CRM/ERP systems) and avoid the most common mistakes, which may occur during integration. In this paper, you will take a look at how to get the job done without having to write custom code or even studying SalesForce.com APIs.

Apatar Open Source Data Integration Tutorial Series

Table of Contents

1. Executive Summary

> "Corporate developers spend approximately **65 percent** of their effort building bridges between applications."
>
> *-- The Gartner Group*

Companies of all sizes are challenged to deliver their products and services to market faster and to manage more complex sales and marketing programs with limited budgets and decreasing time frames in order to accelerate revenue generation.

Salesforce.com represents a tremendous opportunity for companies to solve these challenges by leveraging its proven, non-intrusive and scalable on-demand platform.

To fully leverage the benefits of using Salesforce.com, companies must realize the need to integrate customer-facing business processes with the rest of the enterprise.

Today, when hosted CRM software is becoming mainstream, a typical company using SalesForce.com has to figure out how to connect the information residing in Salesforce.com with 3rd-party systems, such as ERP, accounting and CRM packages, custom applications, and databases.

Having the right data integration and data quality model is critical.

This white paper will guide you through the process of integrating Salesforce.com data with 3rd-party applications and databases using freely available, open source application Apatar Data Integration. It will also provide a set of best practices and tips on how to solve typical challenges and avoid the most common mistakes.

2. Integration Process

At some point Salesforce.com customers realize the need to integrate their customer information between SalesForce.com and 3rd-party systems and applications. What steps should be taken to get your customer and enterprise information (currently residing in SalesForce.com) integrated, replicated, or migrated to your new Software-as-a-Service package?

2.1. Preparation and Planning

Prior to any data migration, ask yourself some questions to clarify the goals of the oncoming integration process. For data integration specialists, it is critical to know:

- What data (tables/fields/rows) should be extracted?
- What data (tables/fields/rows) should be considered as targets?
- Do I need to integrate SalesForce.com with one single database or multiple data sources?
- Is it enough to perform a one-time migration, or do I need an ongoing synchronization?
- Do I need to have SalesForce.com data backed up?
- Do I have enough experience to do manual coding, or would the use of visual data integration tools be the best decision?

In case you are still considering manual coding, take into account the time and effort required to learn APIs, provide connectivity to both the source and the target, write transformation logic, and, most importantly, the tasks related to debugging, reporting, and future maintenance of the integration and the related metadata.

The more clearly you set the goals, the more accurate your SalesForce.com integration will be. Sometimes you may need to join data; sometimes it's all about eliminating duplications; and sometimes the data should be validated or filtered first.

Possible Mistakes:

- No strategic vision.
- Lack of evaluation criteria.

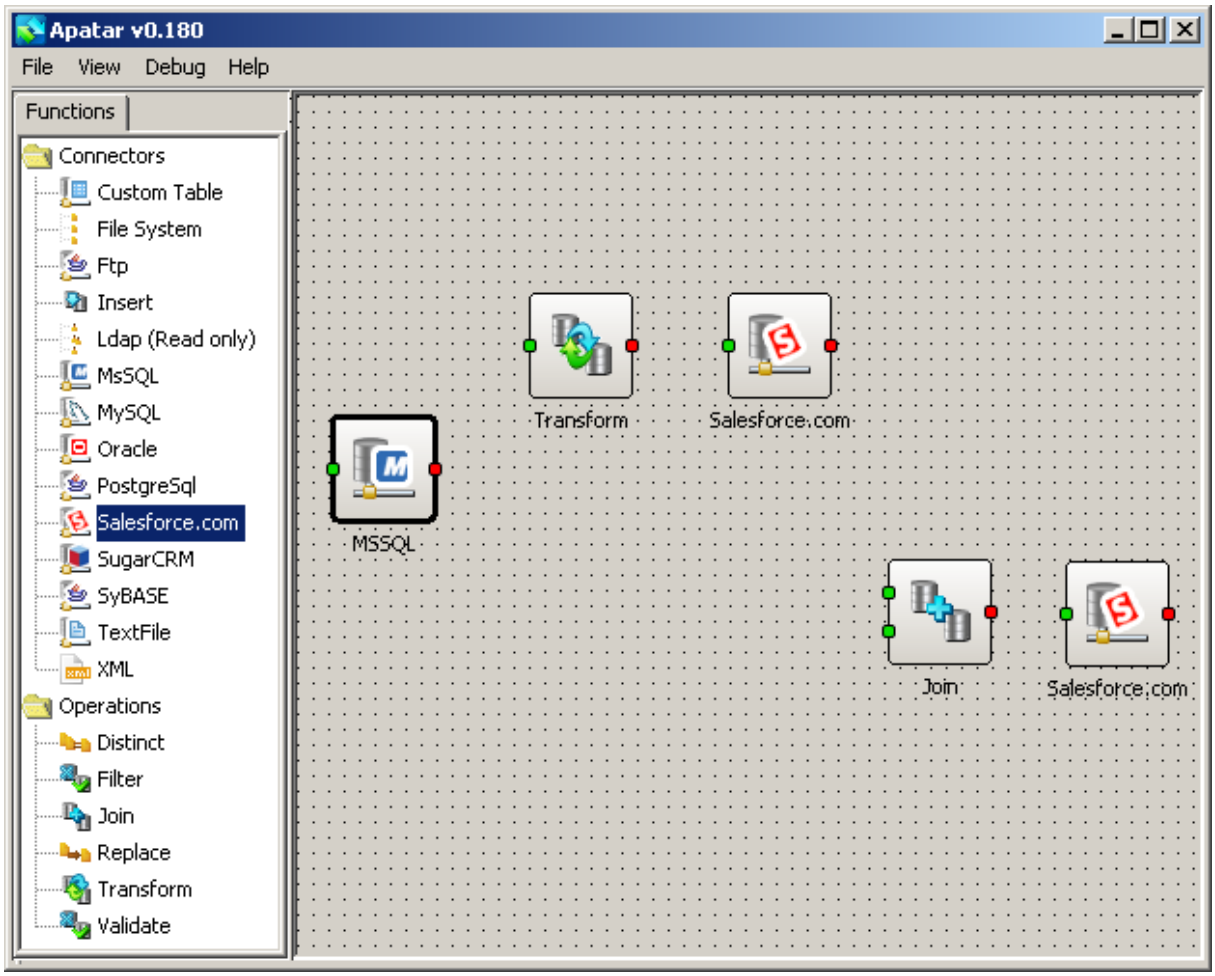
Possible Solution:

- Try to set your objectives and goals properly.

2.2. Data Source Connection Maintenance

To start reading source data, you need to establish connections to the source databases. In other words, you need to gain access to data tables, data structures, and data entries. This is where data integration actually starts. With visual tools like Apatar, for instance, you can do it without having to write a single line of a code. Just open the "drag-and-drop" job designer, choose the necessary data connectors, enter SalesForce.com authorization details, and provide the paths to the database servers or storage files. The application is ready to operate with data.

Figure 2.2.1 Data Source Connections



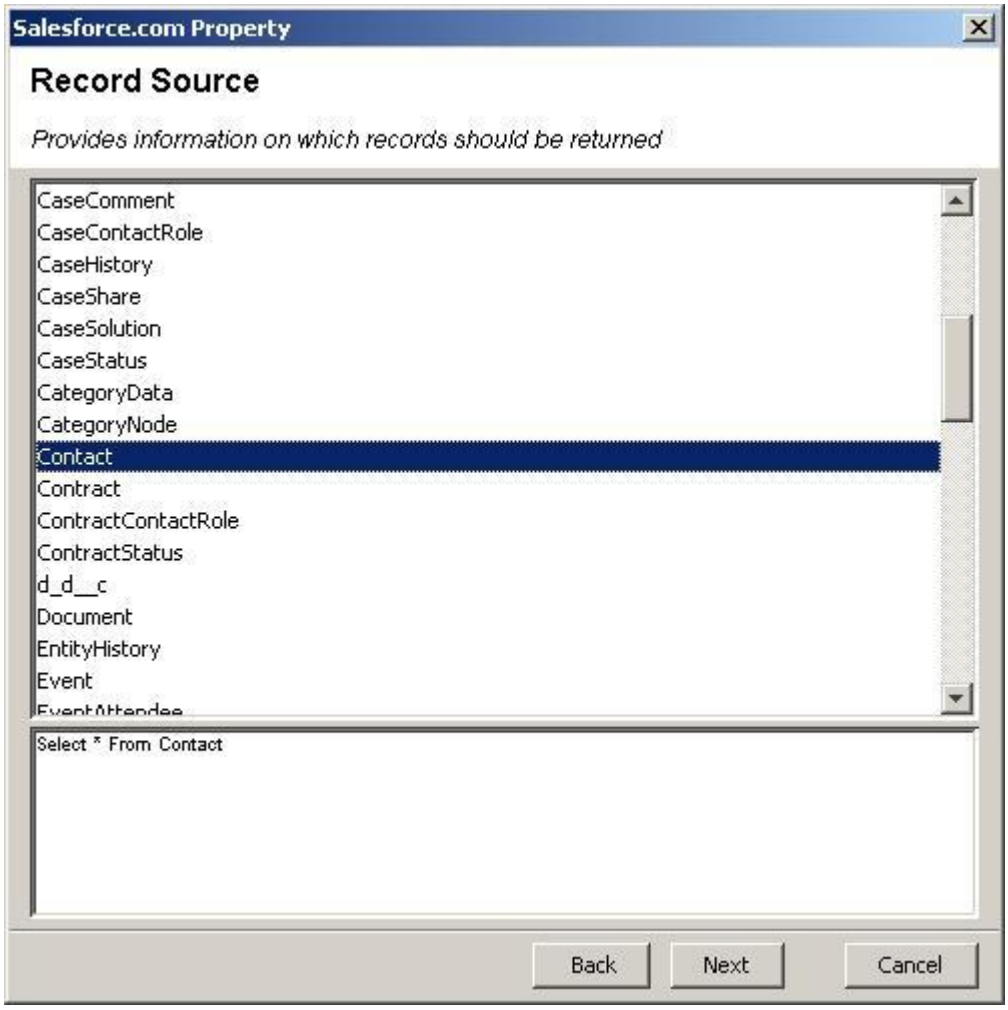
Possible Mistakes:

- Company's data integration map is not comprehensive and lacks a number of data sources being used during decision-making.

Possible Solution:

- Explore the needs and communications between the company's departments to identify the most critical data flows.
- Structure your in-house data and join it with the integration map.

Figure 2.2.2 Choosing SalesForce.com Source Data Tables

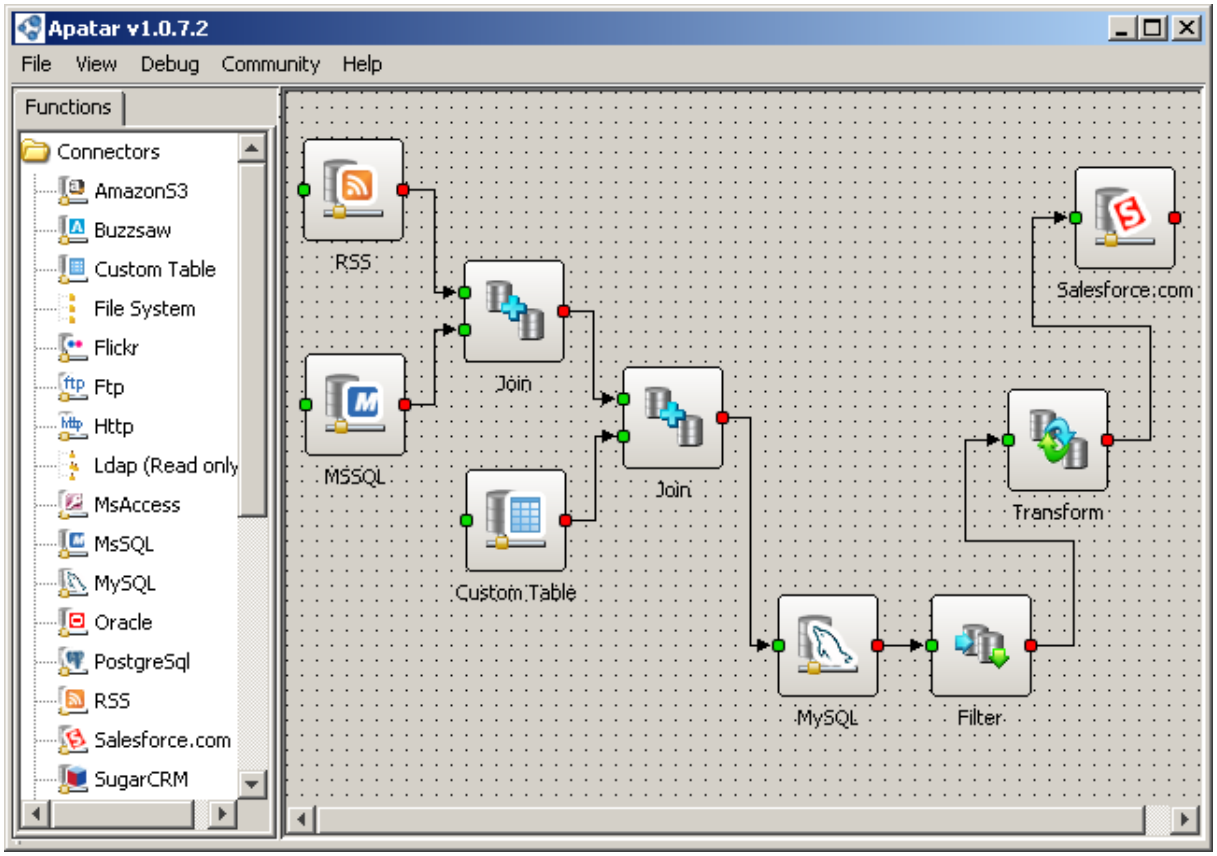


2.3. Staging Database Usage

You may also want to mashup data from multiple sources. For instance, take news from an RSS feed, extract customer information from GoldMine CRM, standardize and cleanse data, verify e-mails, names, and addresses, and then mix it all up and throw it across your SalesForce.com accounts. Consider this step if you have multiple data targets from which to aggregate information, or if you have to apply complex data cleansing or enrichment rules to the data on its way between the source and the target. This step is optional, but sometimes it's worth considering.

In this example we will use MySQL to host staging data. To connect to MySQL database, you may use Apatar's embedded MySQL connector. The process of establishing the connection is identical to the one described in the previous step. Just drag-and-drop the connector to Apatar's work panel, enter database authentication details, and provide the paths to MySQL

Figure 2.3.1 Staging MySQL Database Usage



Possible Mistakes:

- Wrong validation or enrichment rules.
- Data duplications.

Possible Solution:

- Set the filtration criteria, avoid loading odd information, and define data enrichment rules.
- If necessary, consult with a data warehousing specialist.

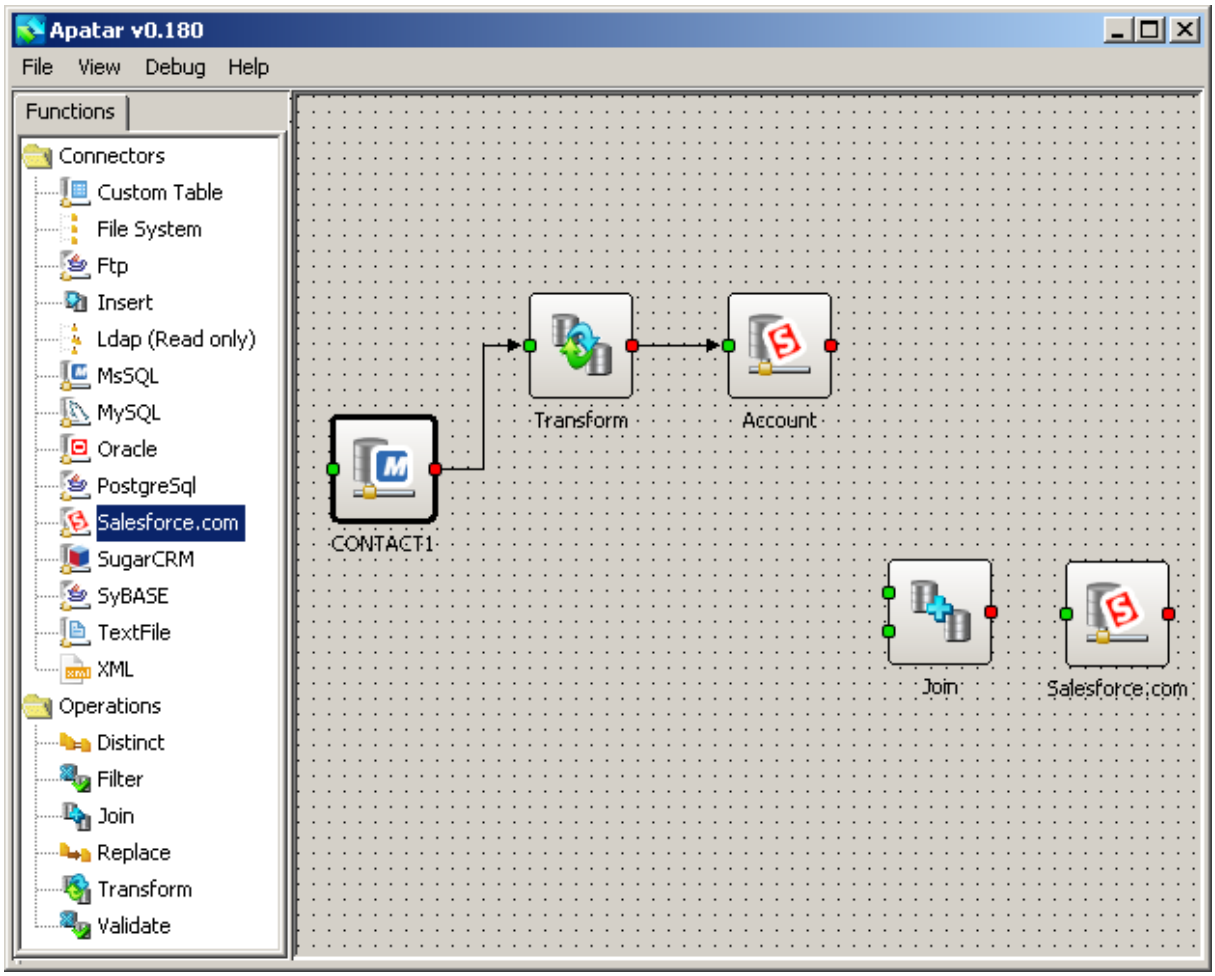
2.4. Map the Source with the Destination

According to The Gartner Group, corporate developers spend approximately 65 percent of their effort building bridges between applications. Luckily, today's data integration tools allow even non-technical users to integrate data between systems, databases, and applications. Imagine you could visually design ("drag and drop") a workflow to exchange data between files (Microsoft Excel spreadsheets, CSV/TXT files), databases (such as MySQL, Microsoft SQL, Oracle), applications (Salesforce.com, SugarCRM), and the top Web 2.0 destinations (Flickr, RSS feeds, Amazon S3), all without coding.

As an example of such visual tools, Apatar embeds a visual job designer to enable users to create integration jobs called DataMaps, link data between the source(s) and the target(s), and schedule one-time or recurring data transformations. These data "mashups" can be saved for future re-use, or sharing, or even re-distribution. In other words, now you have all the integration settings saved and will not have to waste your time again and again if you want to perform a similar task or repeat exactly the same transformations.

With installation taking 60 seconds or less, users connect to data destinations and then match appropriate source (e.g., legacy CRM) and target (SaleForce.com) fields to accomplish the data integration job.

Figure 2.4.1 Visual Mapping



To do this, simply open the Apatar desktop client application, connect to data destinations, and then match appropriate source (e.g., GoldMine CRM) and target (SaleForce.com) fields to start populating the tables.

Why is the mapping so critical?

The truth is that **most** data integration projects in today's enterprises **never** get built.

Firstly, you need to point where the source of data is and where the target is. Secondly, sometimes you need to transform 3rd-party database table formats to SalesForce.com table formats. For instance, if you have Time or Binary objects at the source and need to save them as text at the target.

Using mapping, you tell your data integration tool what transformations you want to do, and where exactly you want them to be done.

Possible Mistakes:

- Input and output data types are contradictory.
- Overly complex data maps and relevance models.

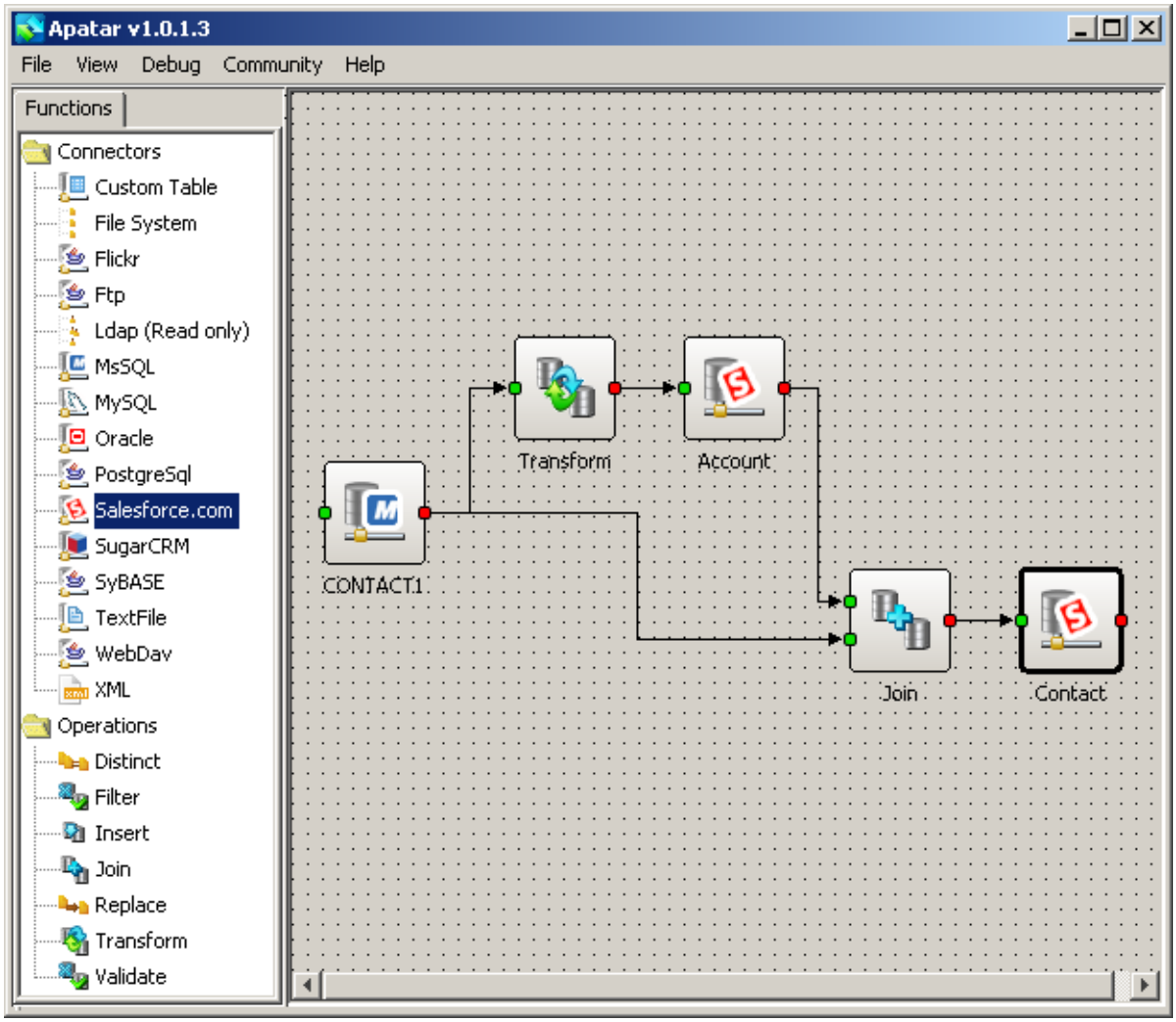
Possible Solution:

- Perform necessary data type transformations.
- The more simple your models, the more stable they are.

2.5. Solve SalesForce.com IDs Issue

Some tables in SalesForce.com database are related by means of key IDs, i.e., an Account table, which contains information about a customer's company, is related to a Contact table, which contains the customer's name and phone number. SalesForce.com's engine generates these IDs automatically, and the problem occurs when a data integration specialist wants to populate both tables together.

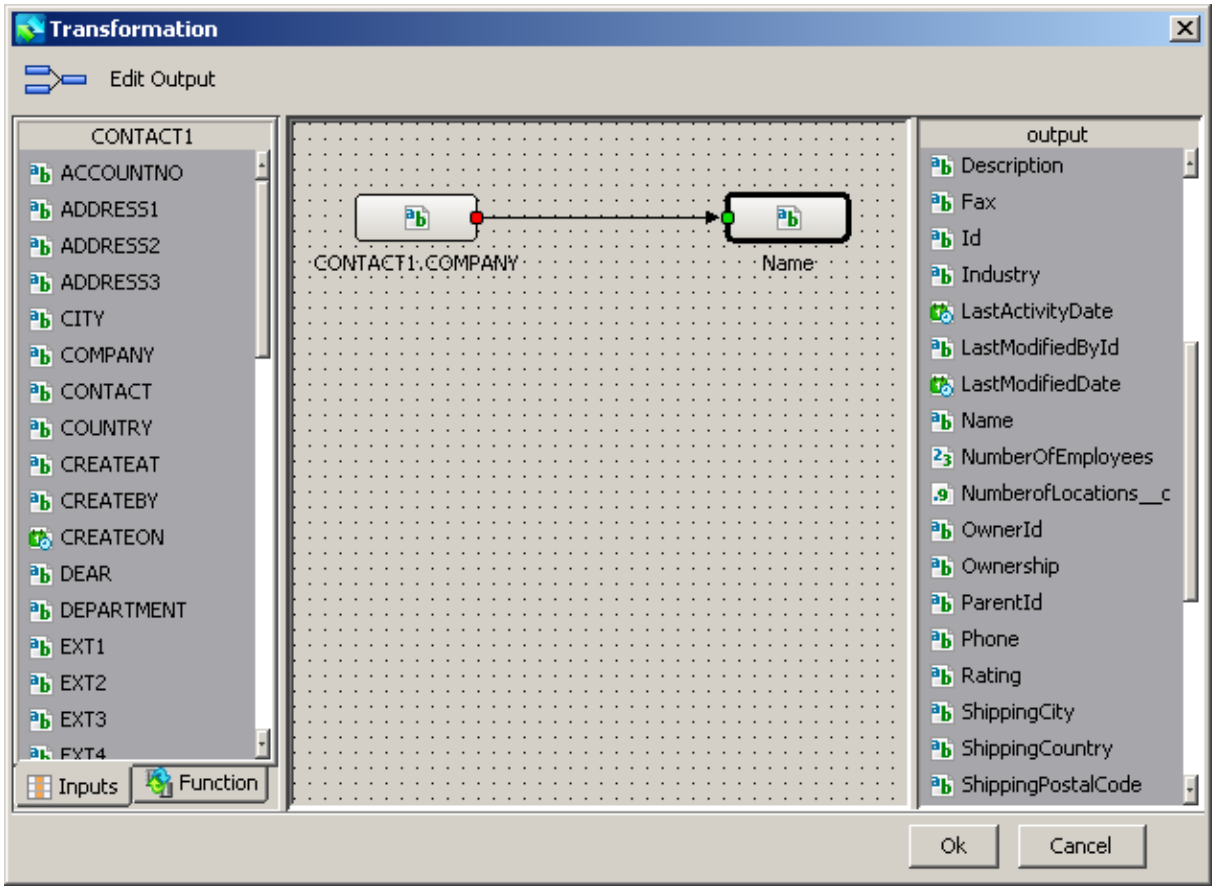
Figure 2.5.1 Final Connector Mapping



Let's now explore how Apatar can solve this problem. To gain control over SalesForce.com IDs, you need to use two SalesForce.com connectors. Configure one connector for the SalesForce.com Account table, and another one for the SalesForce.com Contact table. Map all the nodes as shown in the picture above.

SalesForce.com IDs will be generated in the connector configured for the Account table, where the customer company names are stored. To set the rules for this part of the integration process, you need to open Transform module, drag-and-drop the input and output fields containing the customer company names, and map these fields together.

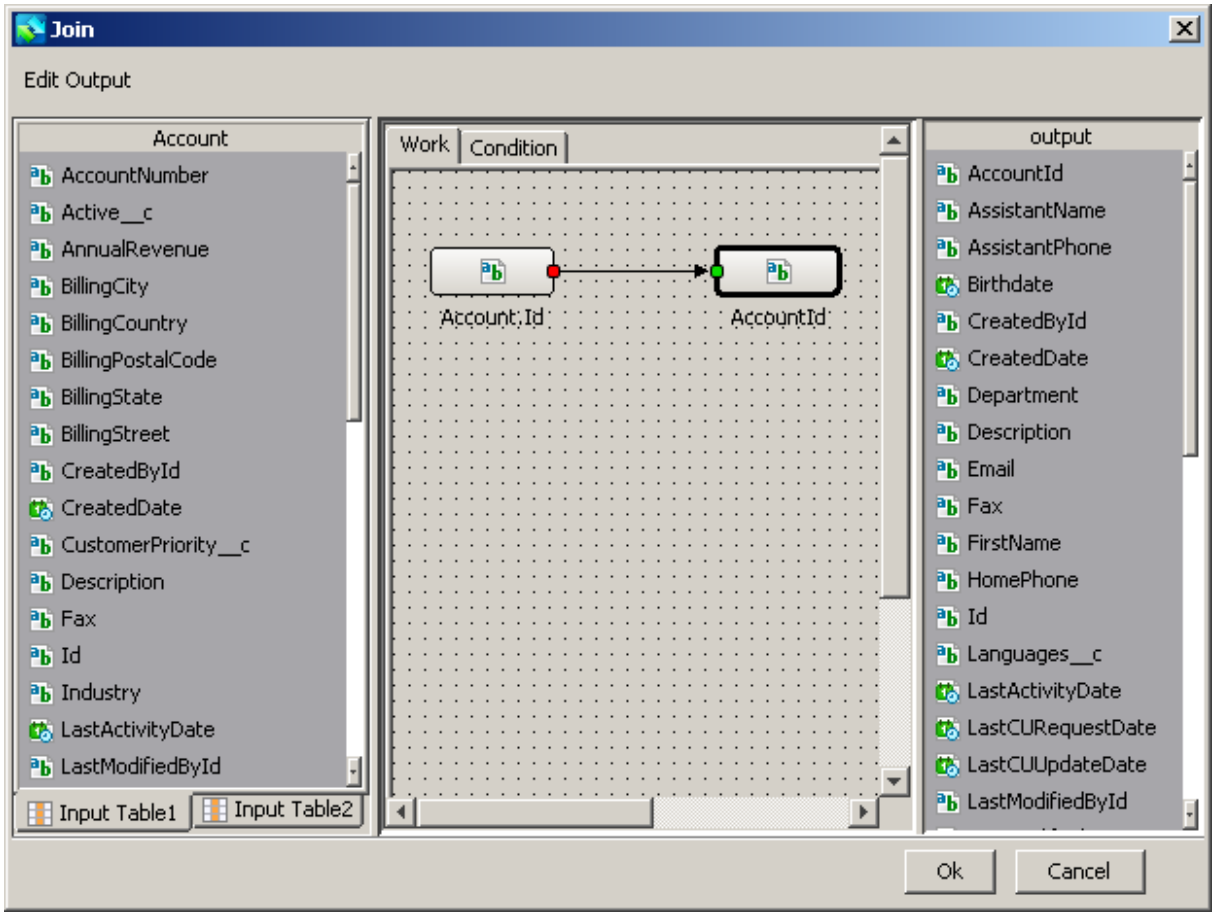
Figure 2.5.2 Mapping the Name Fields



This would allow you to have SalesForce.com-generated IDs for the Account table, all based on the name of the customer's company. Now these IDs should be transferred into another connector, configured for the Contact table. For this purpose, Apatar uses the Join module with the join condition inside.

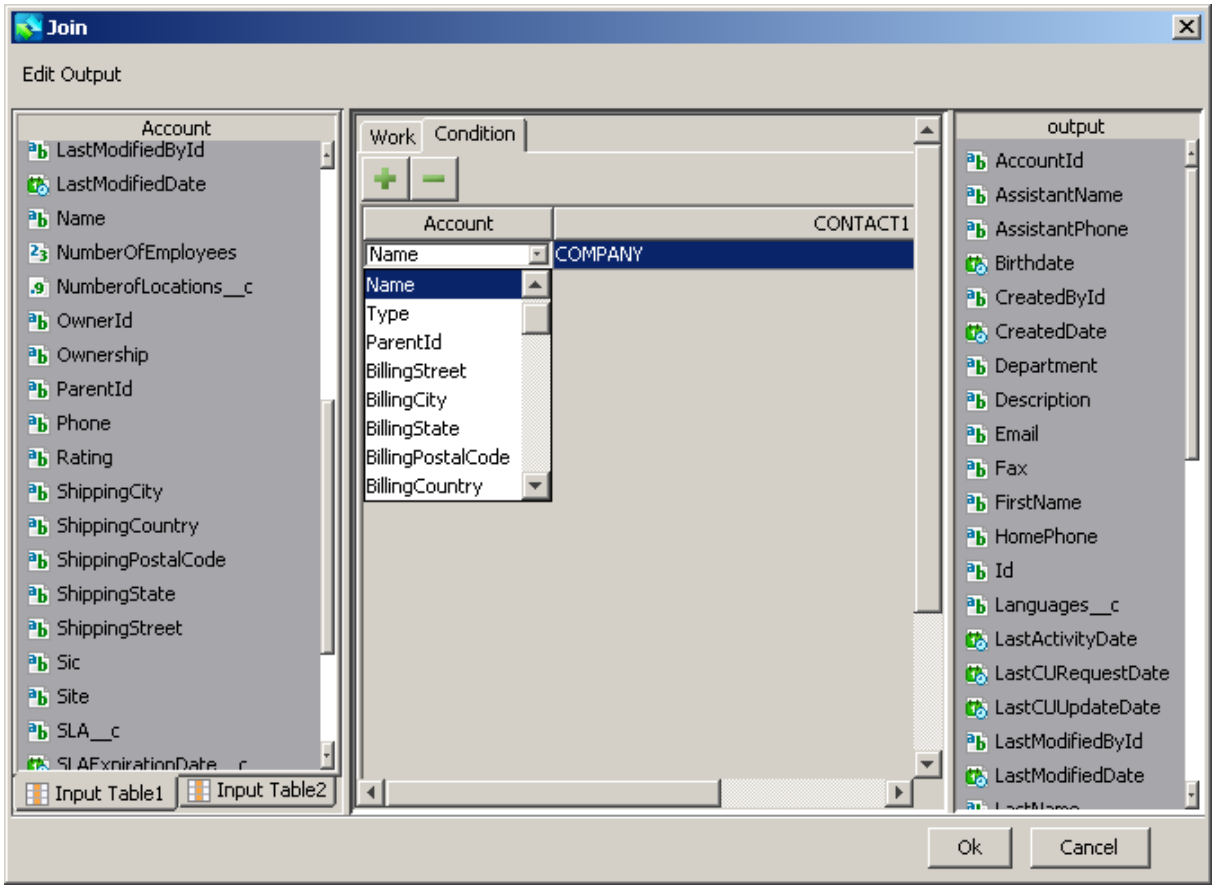
To leverage the Join module, the user opens the Join window, drag-and-drops "ID" fields from SalesForce.com Account and Contact tables, maps these fields together, and sets the condition to link the tables.

Figure 2.5.3 Mapping Key IDs



During this step, the user needs to specify the fields, which should be equal in both the SalesForce.com Account and GoldMine CONTACT1 tables. To provide a relation based on a customer's company name, choose the Name field from the Account table and the COMPANY field from the CONTACT1 table.

Figure 2.5.4 Setting the Join Condition



After that, drag-and-drop into Join's work panel all the input and output fields you want to transfer from GoldMine to SalesForce.com. Map the fields together.

Now the integration job is almost complete.

Finally, run the transformation and let your SalesForce.com accounts be populated. If all initial settings and mappings were accomplished accurately, you may now just sit back and relax. The data integration tool will do the integration.

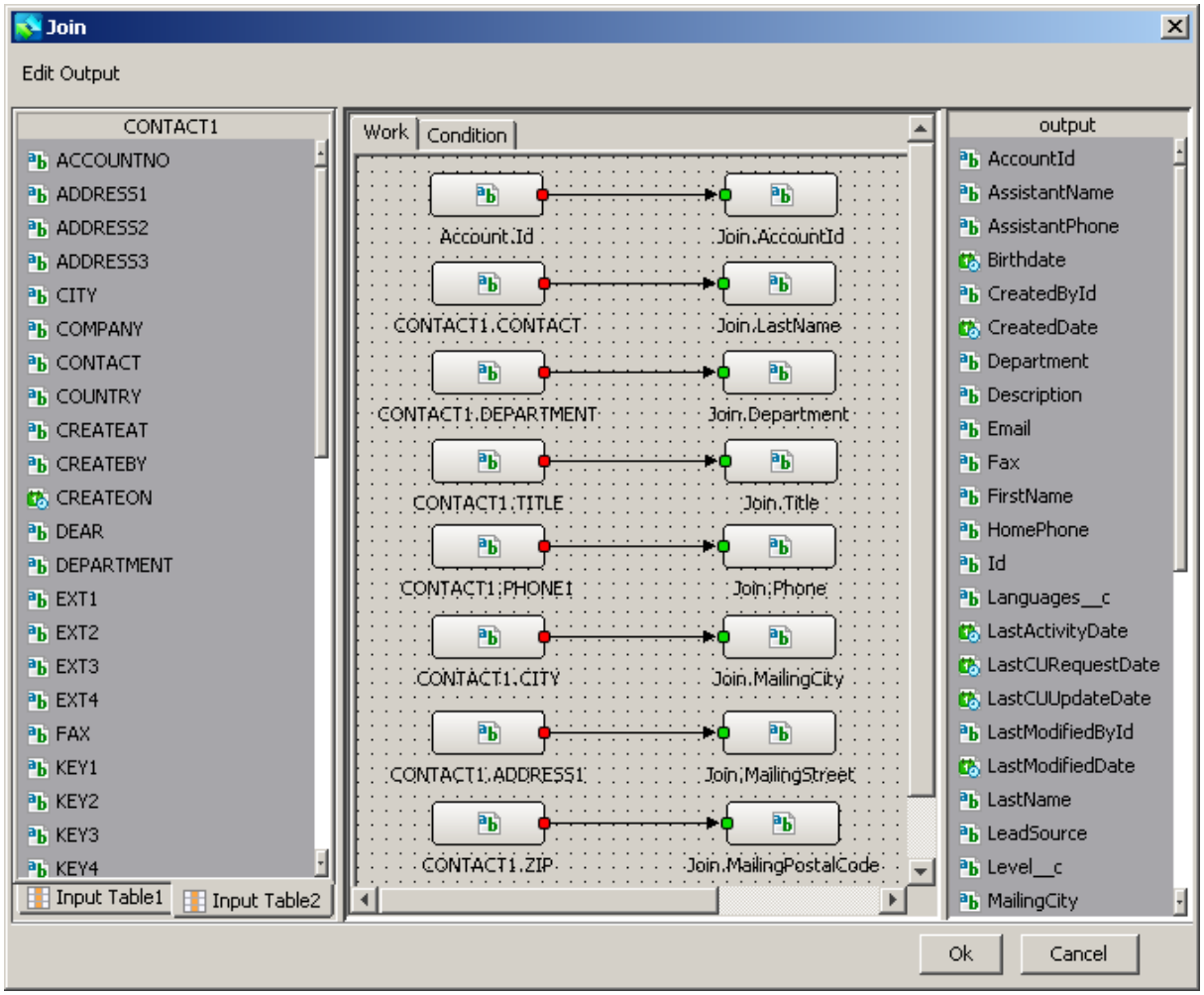
Possible Mistakes:

- The wrong primary or foreign key vision.

Possible Solution:

- If necessary, request support from SalesForce.com team to clarify data table structures and relations.

Figure 2.5.5 Fields Mapping



3. Recurring Integration

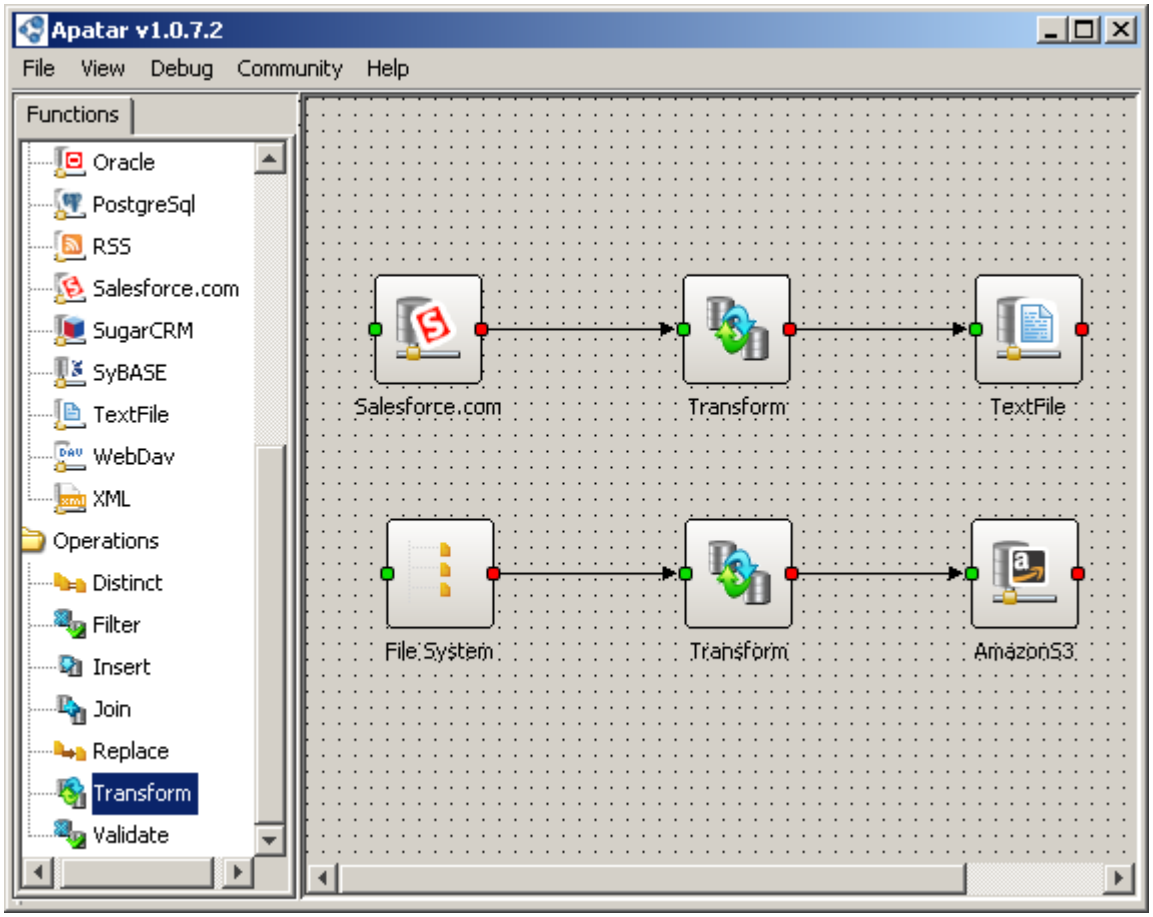
Business data is never consistent; real-time updates, new data entries, and other changes require maintenance. If you need recurring integration jobs, you may also use the scheduling function to automate the data integration process. Tools, such as Apatar, enable you to configure automated scheduling by entering the frequency of transformations and specifying the lifetime of a scheduled job. For example, you can set data migrations to launch at midnight daily and last until the end of this year. Each morning, you will find the entire set of customer data synchronized.

4. Back Up SalesForce.com Data

To have your customer data backed up to a secure remote location, you may want to explore Amazon S3 (Simple Storage Service), which lets you easily store and retrieve virtually any amount of files anytime, anywhere. Amazon S3 deploys the same highly scalable, reliable, fast, and inexpensive data storage infrastructure that Amazon.com uses to run its own global network of websites. Apatar's Amazon S3 connector brings the power of Amazon S3 to Salesforce.com users who may want to store or back up Salesforce.com customer data and documents.

For instance, if a company's executive wants to have his or her company's most significant customer information backed up every day (e.g., extracted into flat files, and saved to Amazon S3), the Apatar tool allows for this data to be backed up and then uploaded to Web storage at a specified time. All you need to do is configure the Amazon S3 connector and enter the frequency and the moments of SalesForce.com data backups in Apatar's Scheduling module. The ETL engine will do the rest automatically.

Figure 4.1 Backing up SalesForce.com Data to Amazon S3

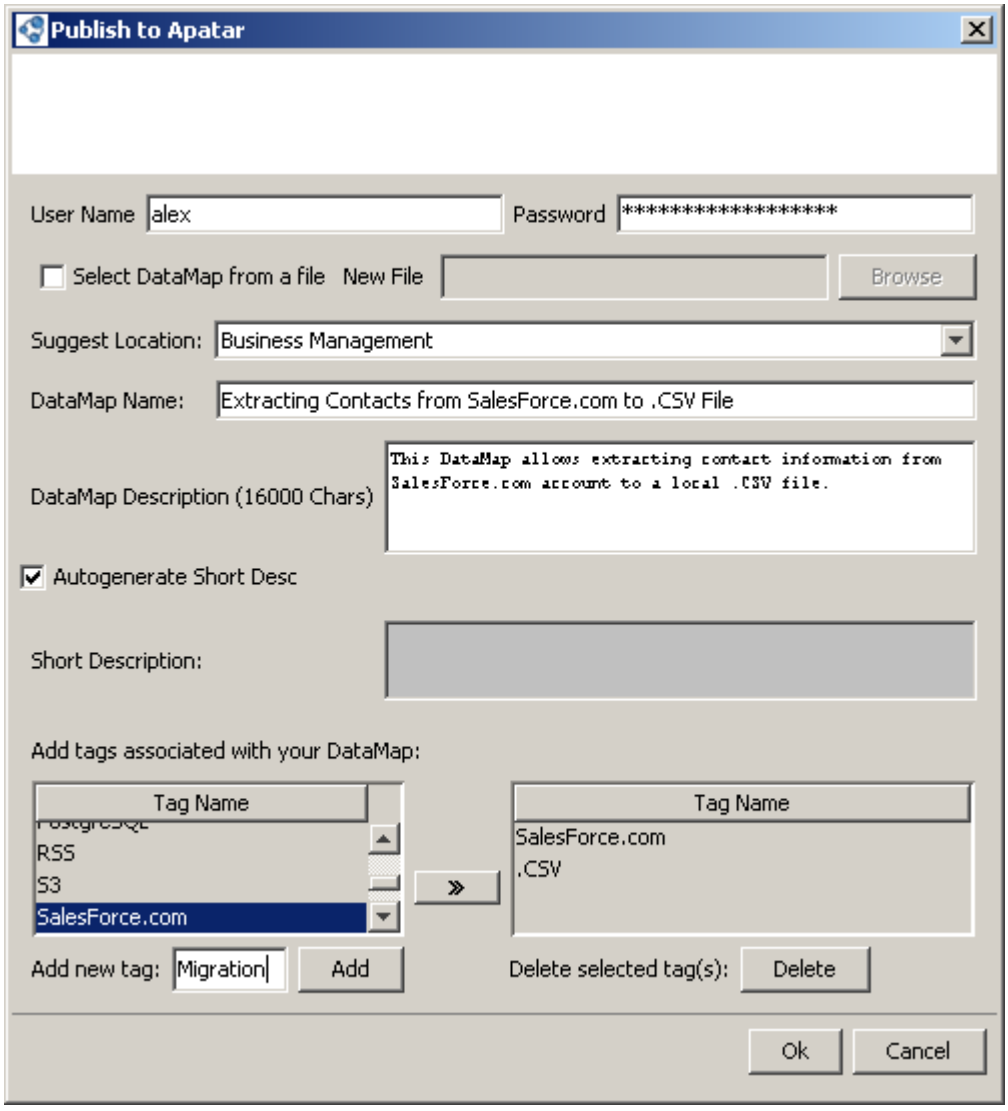


5. Mashup Multiple Sources

In many cases, you need to repeat the migration later, or re-do it from the start. With Apatar, developers can create integration jobs, called DataMaps, to store the data links between source(s) and target(s).

You may also want to mashup data from multiple sources. For instance, take news from an RSS feed, filter news related to Salesforce.com accounts, and throw it across your SalesForce.com accounts.

Figure 5.1 Publishing Your DataMap to www.apatarforge.org's Repository



6. Best Practices

1. Formalize Data Schemas	Custom objects, data fields, and tables created by individuals should be documented and align with all applications and processes within the integration environment, as well as be visible to other users. To enforce data standardization, it is recommended to start with defining initial data schemas and setting how the processes of subsequent changes to the data schemas will be made in the future.
2. Update the Information	Information should be updated on a regular basis or, if possible, in real time. Out-of-date views are useless, so keep an eye on this.
3. Maintain the Integration	Even the most defined integration process requires maintenance. New tables may be added, data structures may change, and so on. If there's no one to take responsibility for the long-term success of the integration initiative, the process may fail sooner or later. Having no plan or budget for an ongoing integration is a mistake, which may become expensive to fix.
4. Go Open Source	The truth is that most data integration projects in today's enterprises never get built. The ROI (Return On Investment) on these small projects is simply too low to justify bringing in expensive middleware. That's why you may consider using commercially supported open source tools for your integration projects. You may want to consider Apatar's application to design and orchestrate data integration processes, as well as MySQL database to host data warehouse and staging tables.
5. Verify and Clean Up the Data	You also need to perform data cleansing and verification required by your business and industry. For instance, it is critical to check the names, addresses, and e-mail details of your prospects and customers. Each industry will have its own baseline, inputs/outputs, and best practices for such data quality management. Mature data integration tools provide data cleansing capabilities. Consider the Name and Address Verification, as well as the E-mail Verification features embedded into the Apatar Data Integration application.
6. Transform Raw Data into Business Information	Business users are typically looking for useful information that can be applied across the enterprise and provide business decision-making. That's why raw data needs to be aggregated, filtered, enriched, and summarized. In order to enable business analytics, executives must be involved in designing the schemes of raw data transformations as well as creating appropriate business data models and views.
7. Consistence is a Key to Providing Valuable Data	Make sure that the output data is consistent and reliable. For example, a company may need a single view of customers, products, employees, or boards of directors. Sometimes it becomes a challenge to obtain an agreement on the criteria of these views.
8. Stick to Business Value	Finally, don't forget that the integration processes should bring value and align with your business processes. There's no use creating even a complex integration model, if it doesn't increase your revenue or, at least, save you money and time.

7. Summary

Integration with SalesForce.com may be easier than you think. You may want to try Apatar's open source tools and professional services for your next data integration project. With powerful data integration capabilities, thousands of Apatar users enjoy the benefits of using a visual interface and mapping capabilities, which provide even a non-programmer with all of the means to plan, design, and execute various data integration, migration, and replication jobs. Scheduling capabilities allow one-time or recurring transformation of information between Salesforce.com and 3rd-party systems. With today's data management solutions, the business user has a powerful toolset not only to manage data streams within the enterprise, but to join data with the Web, keep it safely in SalesForce.com, and exchange information with partners globally.

For a free download of Apatar Open Source Data Integration and more information on how to integrate SalesForce.com with other applications or data destinations, please visit www.apatar.com or www.apatarforge.org.

8. About the Authors

Renat Khasanshyn is CEO of Apatar, Inc.
Alex Khizhnyak is Apatar Evangelist, Apatar, Inc.

Apatar (www.apatar.com) is the leading provider of open source software tools for the data integration market. With powerful Extract, Transform, and Load (ETL) capabilities, Apatar enables its users to easily link information between databases (such as MySQL, Microsoft SQL, Oracle), applications (Salesforce.com, SugarCRM), and the top Web 2.0 destinations (Flickr, Amazon S3, RSS feeds). Apatar provides support, training, and consulting services for its integration solutions. Headquartered in Western Massachusetts, Apatar operates a development center in Minsk, Belarus. For more on Apatar, please visit www.apatar.com and www.apatarforge.org.