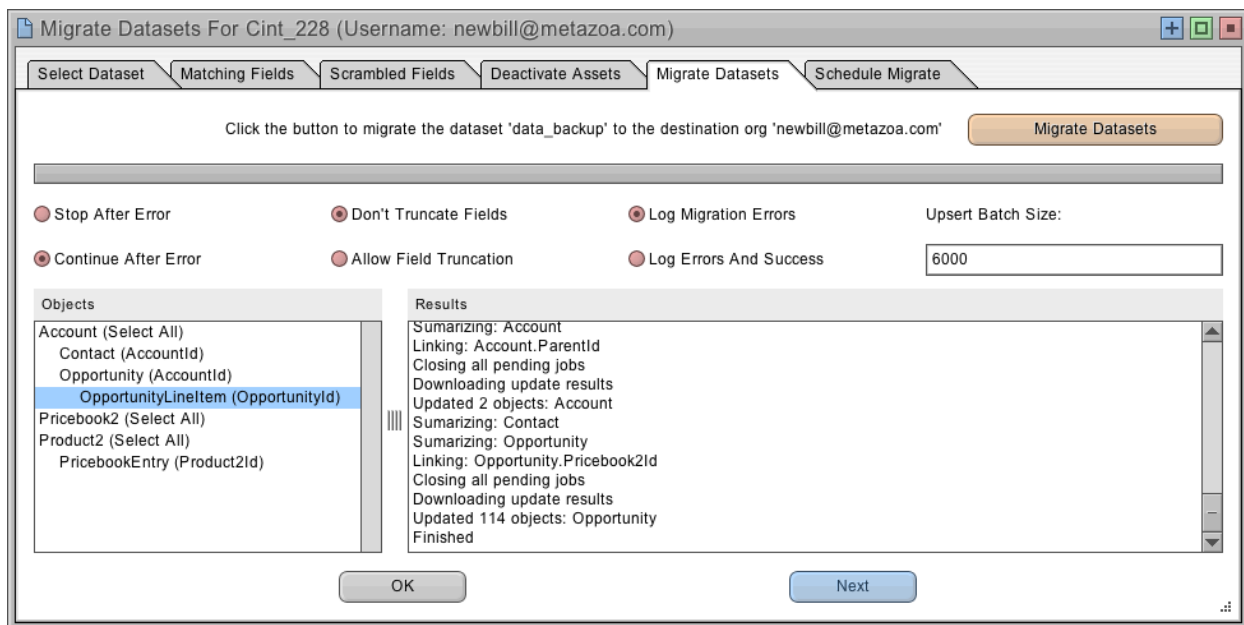


Snapshot Best Practices: Dataset Migration

Snapshot provides powerful tools for migrating connected sets of data records between Salesforce orgs. This capability is useful for backing up data, refreshing Sandboxes, merging orgs, and populating orgs with test data for application development. The Bulk API is used for all transactions to ensure that very large Datasets can be moved efficiently. We have moved hundreds of thousands of records while testing the data migration tools in Snapshot.



The screenshot shows a web application window titled "Migrate Datasets For Cint_228 (Username: newbill@metazoa.com)". The interface includes several tabs: "Select Dataset", "Matching Fields", "Scrambled Fields", "Deactivate Assets", "Migrate Datasets", and "Schedule Migrate". The "Migrate Datasets" tab is active, displaying a message: "Click the button to migrate the dataset 'data_backup' to the destination org 'newbill@metazoa.com'". A "Migrate Datasets" button is located to the right of this message.

Below the message, there are several configuration options with radio buttons:

- ☐ Stop After Error
- ☐ Don't Truncate Fields
- ☐ Log Migration Errors
- ☐ Upsert Batch Size: 6000
- ☐ Continue After Error
- ☐ Allow Field Truncation
- ☐ Log Errors And Success

There are two main panels: "Objects" and "Results".

Objects:

- Account (Select All)
- Contact (AccountId)
- Opportunity (AccountId)
- OpportunityLineItem (OpportunityId)
- Pricebook2 (Select All)
- Product2 (Select All)
- PricebookEntry (Product2Id)

Results:

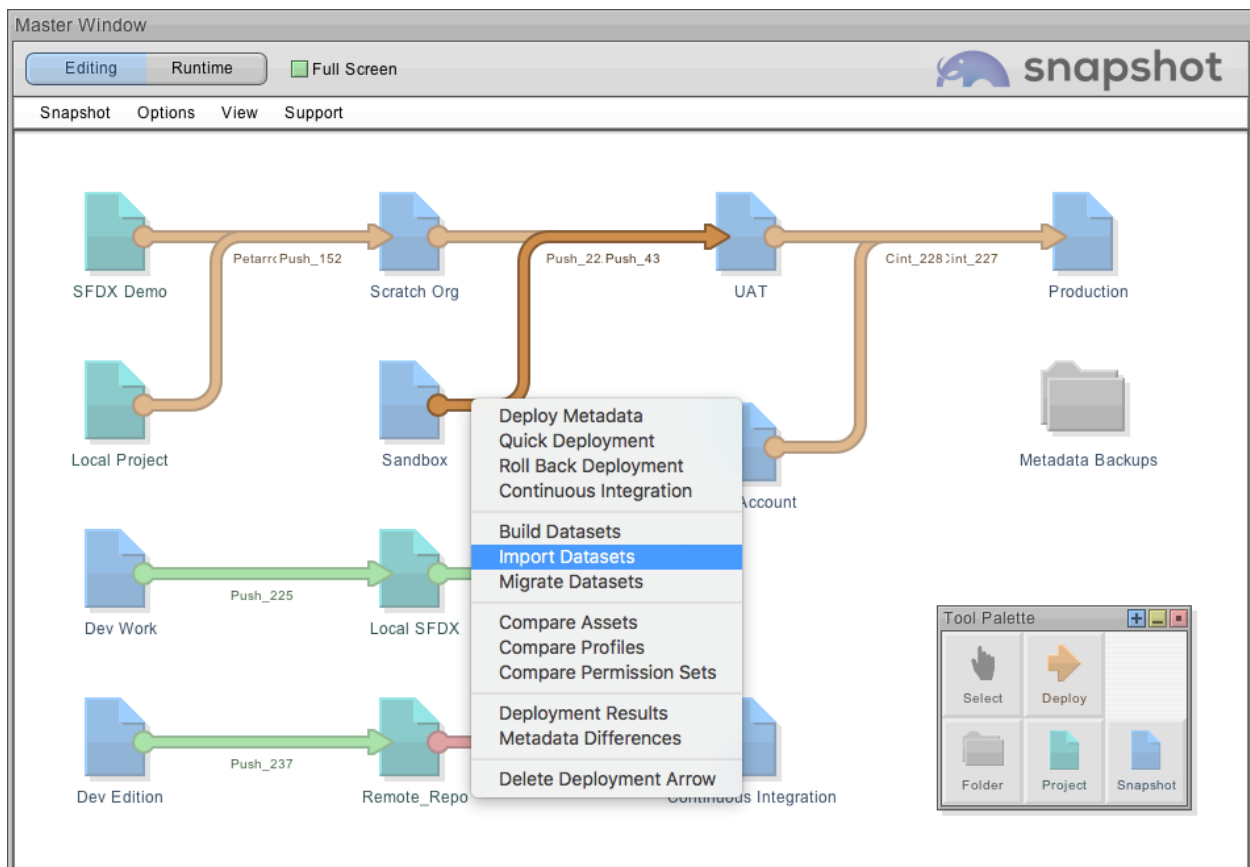
```
Sumarizing: Account
Linking: Account.ParentId
Closing all pending jobs
Downloading update results
Updated 2 objects: Account
Sumarizing: Contact
Sumarizing: Opportunity
Linking: Opportunity.Pricebook2Id
Closing all pending jobs
Downloading update results
Updated 114 objects: Opportunity
Finished
```

At the bottom of the window, there are "OK" and "Next" buttons.

When records are migrated between orgs, all of the internal relationships are preserved. External references in the Dataset are also connected to matching objects on the destination. This whitepaper presents detailed information on how to build and migrate Datasets. We also document the CSV format that Snapshot uses for importing Datasets from other systems or creating them with a spreadsheet editor.

Source and Destination Orgs

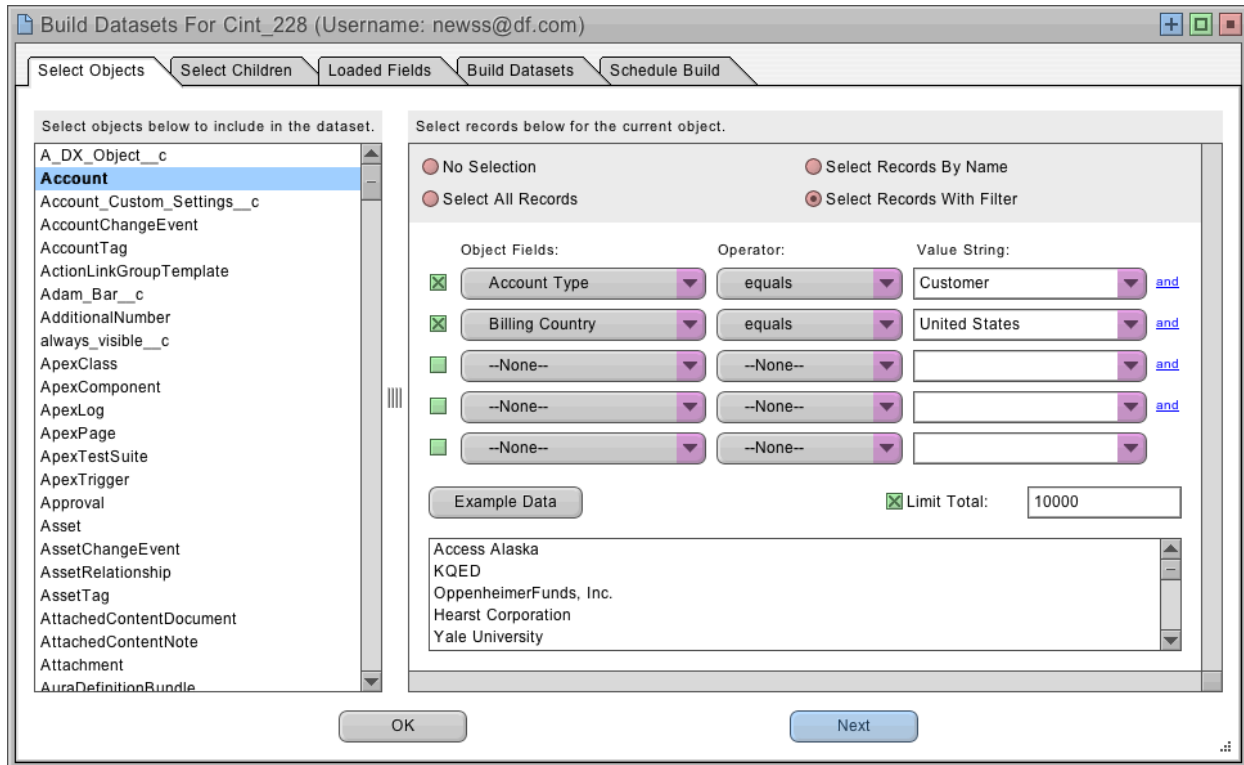
When you select the Deployment Arrow between any two connected Snapshots, the Options Menu will display various commands. You can also right-click the Deployment Arrow to see the options in a popup menu. The second set of options will be to Build, Import, and Migrate Datasets. If you do not see these options, then perhaps the Deployment Arrow is connected to a Developer Project. They do not have any actual data and cannot be used as a source to build a Dataset or as a destination for migration.



The option to Build a Dataset will use the source Salesforce org to download multiple records in the form of XML files to your local machine. The option to Import a Dataset will use CSV or XML files from any source to create a new Dataset as well. Lastly, the option to Migrate a Dataset will insert and update records from the selected Dataset into the destination Salesforce org.

Build Datasets Dialog

The first tab of the Build Dataset dialog allows you to select parent objects that you want to include in the Dataset. These records are available on the source Salesforce org. You can select all records, a subset of records by name, or a subset of records using a complex filter. The total number of downloaded records can be limited. This is useful for grabbing a random subset of records for acceptance testing or application development. The selected parent objects are shown on the list at left in bold text.



Build Datasets For Cint_228 (Username: newss@df.com)

Select Objects | Select Children | Loaded Fields | **Build Datasets** | Schedule Build

Select objects below to include in the dataset.

- A_DX_Object__c
- Account**
- Account_Custom_Settings__c
- AccountChangeEvent
- AccountTag
- ActionLinkGroupTemplate
- Adam_Bar__c
- AdditionalNumber
- always_visible__c
- ApexClass
- ApexComponent
- ApexLog
- ApexPage
- ApexTestSuite
- ApexTrigger
- Approval
- Asset
- AssetChangeEvent
- AssetRelationship
- AssetTag
- AttachedContentDocument
- AttachedContentNote
- Attachment
- AuraDefinitionBundle

Select records below for the current object.

☐ No Selection ☐ Select Records By Name

☐ Select All Records ☒ Select Records With Filter

| Object Fields: | Operator: | Value String: |
|---|-----------|---------------|
| <input checked="" type="checkbox"/> Account Type | equals | Customer |
| <input checked="" type="checkbox"/> Billing Country | equals | United States |
| <input type="checkbox"/> --None-- | --None-- | |
| <input type="checkbox"/> --None-- | --None-- | |
| <input type="checkbox"/> --None-- | --None-- | |

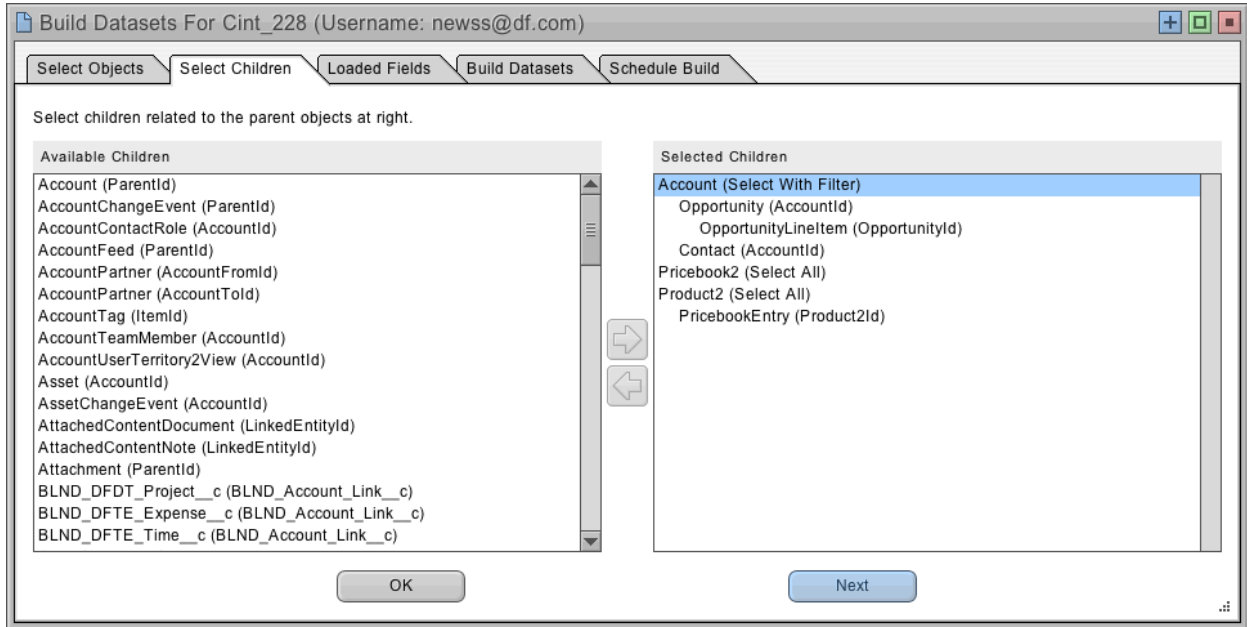
Example Data ☒ Limit Total: 10000

Access Alaska
KQED
OppenheimerFunds, Inc.
Hearst Corporation
Yale University

OK Next

Selecting Parents and Children

The next tab allows the selection of connected child objects for each parent object. When a Dataset is created, the selected parent records are loaded first, followed by all of the children connected to that parent. You can specify multiple child objects in a hierarchy. The relationship field used to associate each parent and child is shown in parenthesis. The internal relationships between parent and child are always preserved when the Dataset is migrated.



Build Datasets For Cint_228 (Username: newss@df.com)

Select Objects | **Select Children** | Loaded Fields | Build Datasets | Schedule Build

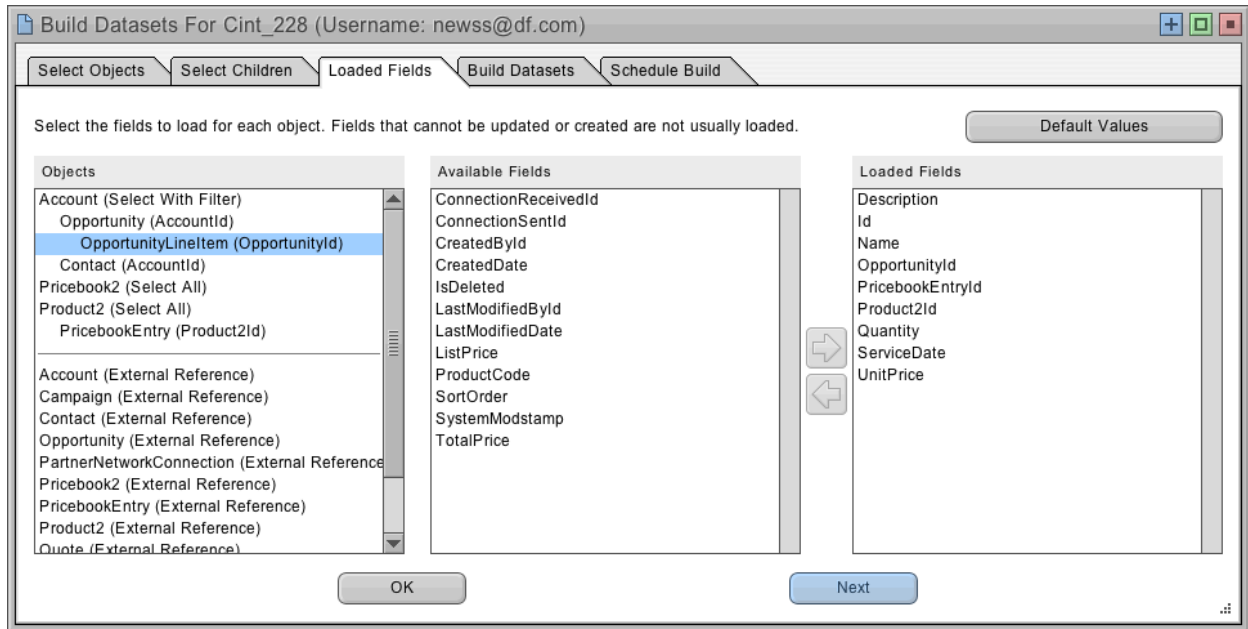
Select children related to the parent objects at right.

| Available Children | Selected Children |
|---|-------------------------------------|
| Account (ParentId) | Account (Select With Filter) |
| AccountChangeEvent (ParentId) | Opportunity (AccountId) |
| AccountContactRole (AccountId) | OpportunityLineItem (OpportunityId) |
| AccountFeed (ParentId) | Contact (AccountId) |
| AccountPartner (AccountFromId) | Pricebook2 (Select All) |
| AccountPartner (AccountToId) | Product2 (Select All) |
| AccountTag (ItemId) | PricebookEntry (Product2Id) |
| AccountTeamMember (AccountId) | |
| AccountUserTerritory2View (AccountId) | |
| Asset (AccountId) | |
| AssetChangeEvent (AccountId) | |
| AttachedContentDocument (LinkedEntityId) | |
| AttachedContentNote (LinkedEntityId) | |
| Attachment (ParentId) | |
| BLND_DFDI_Project__c (BLND_Account_Link__c) | |
| BLND_DFTE_Expense__c (BLND_Account_Link__c) | |
| BLND_DFTE_Time__c (BLND_Account_Link__c) | |

OK Next

Loaded Fields

After that, you can select fields to load for each parent and child object. The parent and child objects that were selected will be above the dividing line at left. You can choose fields that need to be loaded by moving them to the list at right. Snapshot will automatically figure out the best fields to load, so you may not need to do anything on this tab. Fields that cannot be created or updated on the destination do not usually need to be loaded. Removing unwanted fields makes your Dataset smaller in size and easier to migrate.



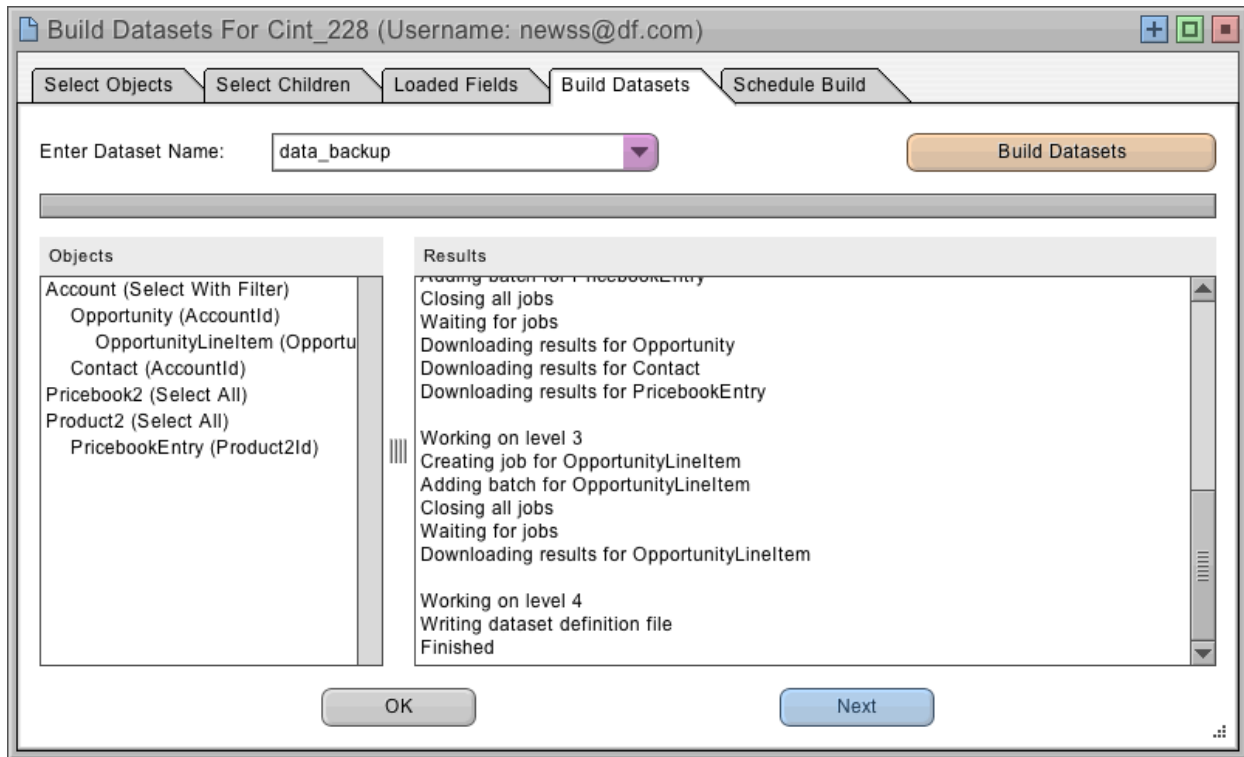
Snapshot also uses the loaded fields to identify matching records in the destination org during migration. If a source record matches a destination record, then the corresponding destination record is updated. If a source record does not match any destination record, then a new destination record is created.

Snapshot automatically selects the most common matching fields for you. However, you can also manually select matching fields. For example, if you want to match Accounts by Name and BillingCity, then be sure that both Name and BillingCity are loaded. You will be able to specify the exact list of matching fields that you want to use in the Migrate Datasets dialog.

Underneath the dividing line you will see “external references” to other objects in the destination org. For example, if you select Opportunity objects for migration then you will see an external reference to Campaigns, because the Opportunity object contains a CampaignId reference field. Snapshot will automatically connect external references to matching objects on the destination org when the Dataset is migrated. You can manually specify additional fields needed for matching external references as well.

Build Datasets Button

The next tab allows you to enter the name of a new Dataset and then click the Build Datasets button at right to start the download process. If you select an existing Dataset name from the menu then that Dataset will be replaced. All of the download results will be listed in the window pane at lower right.

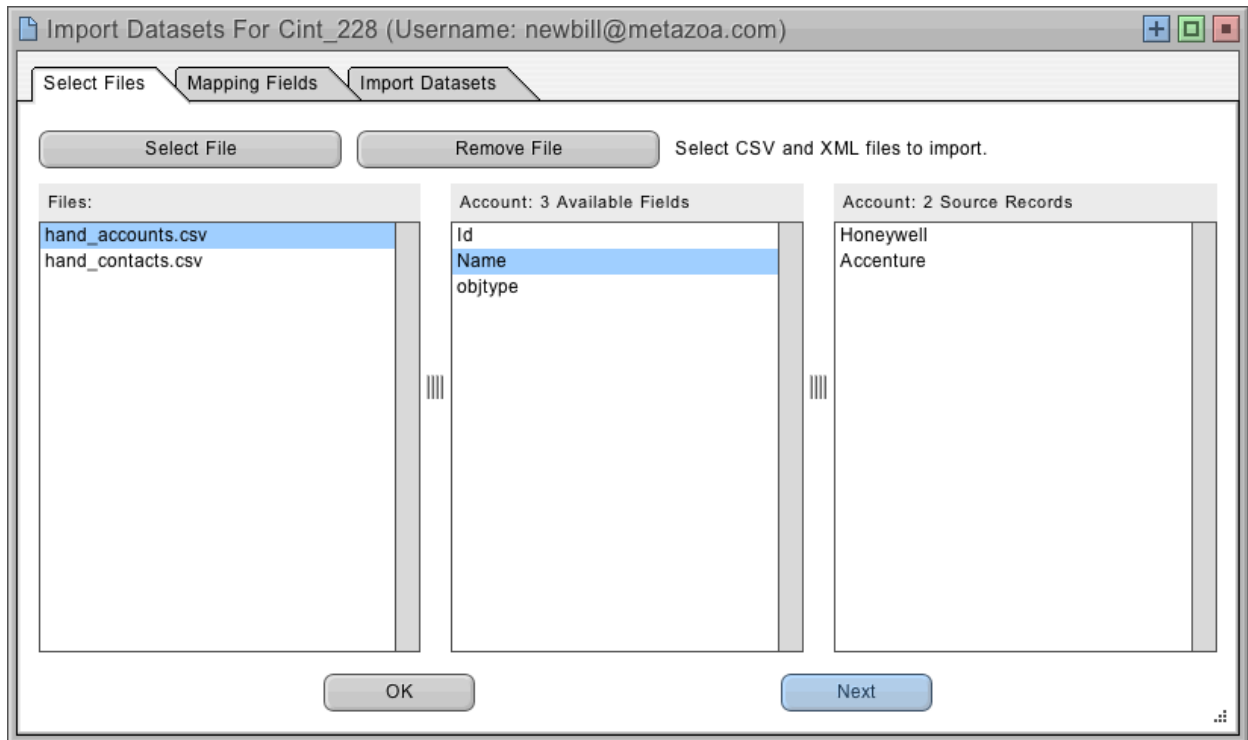


The XML files with your Dataset information will be saved in the "datasets" folder next to the "workspaces" folder in the Snapshot file system. Datasets are globally available for migration to any destination Snapshot. In other words, any source Snapshot can be used to build a Dataset, and any Dataset can be migrated to the destination Snapshot.

The last tab allows you to Build a Dataset at a specific time in the future or as a recurring event. When a scheduled Dataset is created, it will automatically replace the current Dataset by name. Unlike Snapshots, Datasets are not maintained in a time series.

Import Datasets Dialog

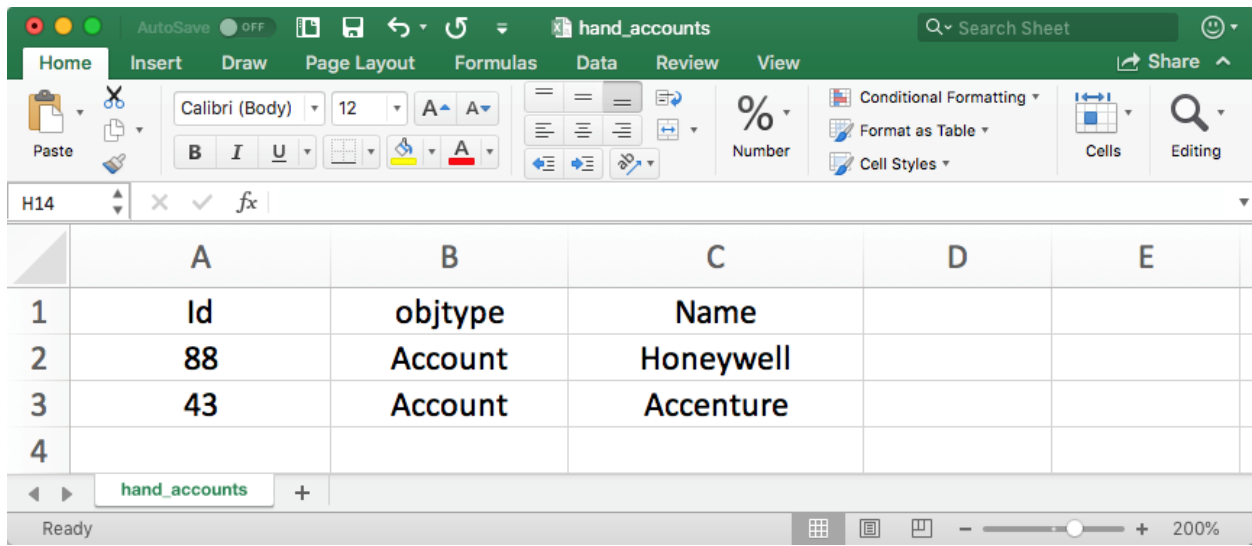
There is another way to create a Dataset. The Import Dataset command is right under Build Datasets when you right-click a Deployment Arrow. When you Import a Dataset, you can select any number of CSV or XML files and add them to the list at left. The interface will show the imported fields and source records in the window panes at right.



Imported XML files use the same format that is saved in the “datasets” folder in the Snapshot file system. You can select XML files from any existing Dataset for import. However, CSV is a more popular file format, because CSV files can be brought in from other systems or created with a spreadsheet editor. The next section talks about the format that Snapshot expects for CSV files.

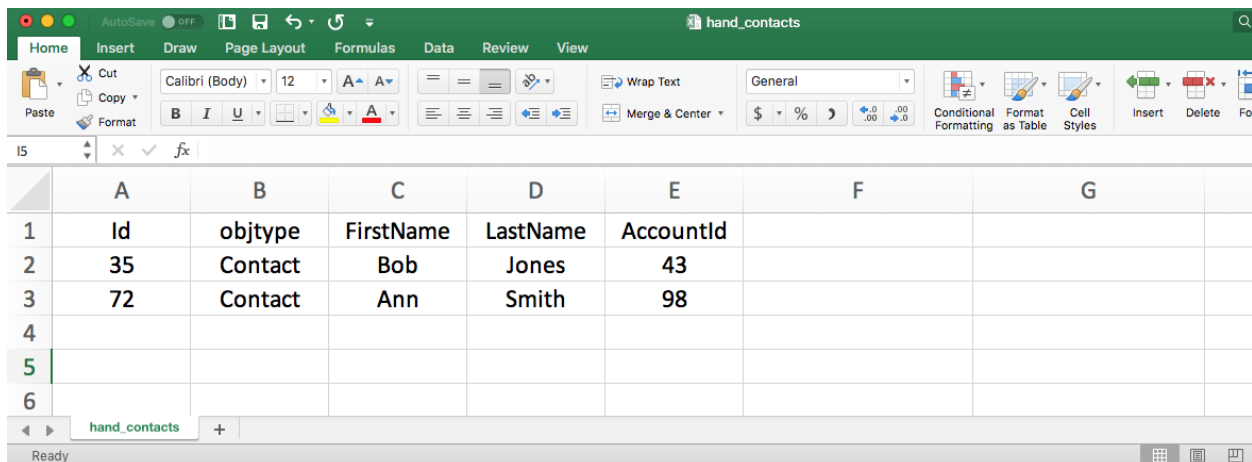
CSV File Format

Snapshot's CSV file format is capable of preserving all internal relationships and reconnecting all external relationships when the Dataset is migrated. This file format expects the first row of CSV data to be field names, followed by additional rows for each record. The columns must include the field name "Id" to specify a unique record Id, and the field name "objtype" to specify the object type. Here is an example of the CSV file format with two Account records:



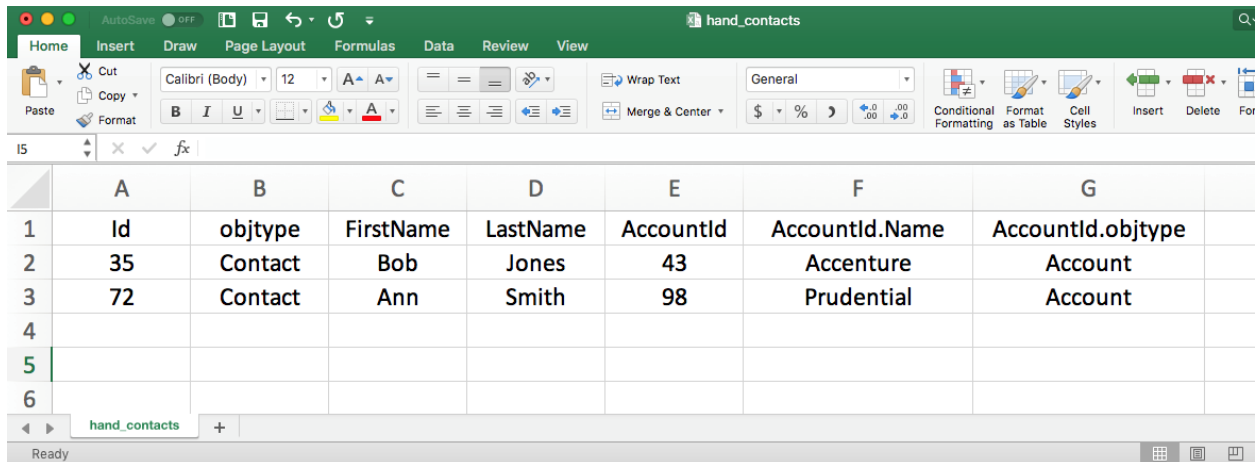
| | A | B | C | D | E |
|---|----|---------|-----------|---|---|
| 1 | Id | objtype | Name | | |
| 2 | 88 | Account | Honeywell | | |
| 3 | 43 | Account | Accenture | | |
| 4 | | | | | |

This example has a row of field names followed by two records, as well as the required "objtype" and "Id" columns. The rest of the columns are used for other fields like the Account Name. In this example, the Id field is from some other system, because these are obviously not 18-character Salesforce Ids. The Ids can be in any format, but they must be self-consistent in order for all of the internal relationships to be maintained. Now let's take a look at another imported CSV file with Contact data that refers to the Account records above:



| | A | B | C | D | E | F | G |
|---|----|---------|-----------|----------|-----------|---|---|
| 1 | Id | objtype | FirstName | LastName | AccountId | | |
| 2 | 35 | Contact | Bob | Jones | 43 | | |
| 3 | 72 | Contact | Ann | Smith | 98 | | |
| 4 | | | | | | | |
| 5 | | | | | | | |
| 6 | | | | | | | |

The Contact data contains a reference to a parent Account record. If you only include the AccountId field, then Snapshot will connect each Contact to any Account with that Id in the destination org. In our example, the Contact record for Bob Jones will be properly connected to the Account record for Accenture because they both have the same Id = 43. In order for this to work, sure that the Account CSV file is processed before the Contact CSV file. You can right-click an imported file in the list to adjust the order of operations.



| | A | B | C | D | E | F | G |
|---|----|---------|-----------|----------|-----------|----------------|-------------------|
| 1 | Id | objtype | FirstName | LastName | AccountId | AccountId.Name | AccountId.objtype |
| 2 | 35 | Contact | Bob | Jones | 43 | Accenture | Account |
| 3 | 72 | Contact | Ann | Smith | 98 | Prudential | Account |
| 4 | | | | | | | |
| 5 | | | | | | | |
| 6 | | | | | | | |

Here is another version of the Contact CSV file that includes better matching information for the parent Account records. If you include the AccountId.objtype and AccountId.Name fields, then these fields will be available in the Migrate Datasets dialog to match destination objects with the same object type and name. In our example, one Contact will connect to Accenture by Id, and the other Contact will connect to Prudential by Name, if such a record exists in the destination org.

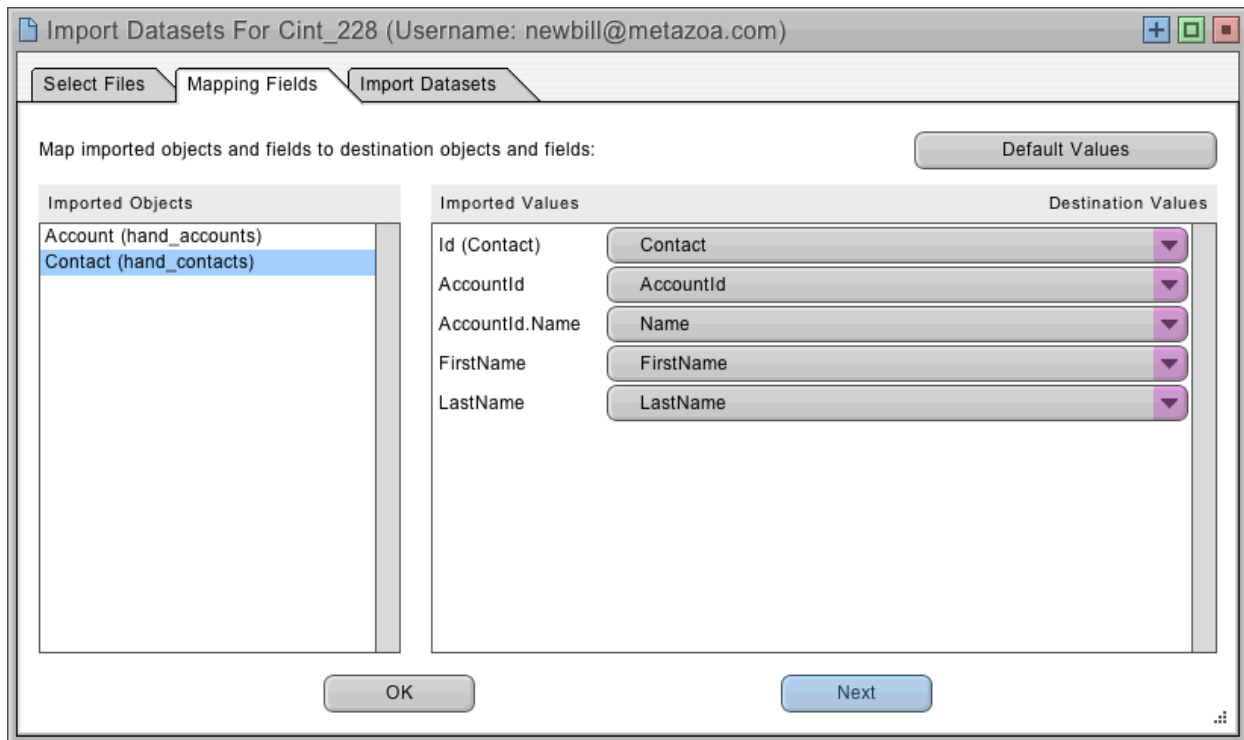
In this manner, Snapshot will use the imported CSV data to match both internal and external references. The required reference information is usually easy to include in the CSV file. For example, you could use the Salesforce SOQL Query below to capture all of the information needed to create our example CSV data for Accounts and Contacts:

```
SELECT Id,Name FROM Account
```

```
SELECT Id,FirstName,LastName,AccountId,Account.Name FROM Contact
```

Mapping Fields

The imported CSV data could be from any external system or database. The Mapping Fields tab provides an easy way to make sure that the imported object and field names match up to the object and field names in the destination org. First, select the desired destination object name at top. Then go down the list and select matching field names. The same technique applies to matching external references and fields.



The screenshot shows a software window titled "Import Datasets For Cint_228 (Username: newbill@metazoa.com)". It has three tabs: "Select Files", "Mapping Fields" (which is active), and "Import Datasets".

Inside the "Mapping Fields" tab, there is a section titled "Map imported objects and fields to destination objects and fields:". To the right of this text is a button labeled "Default Values".

Below this section, there are two main columns:

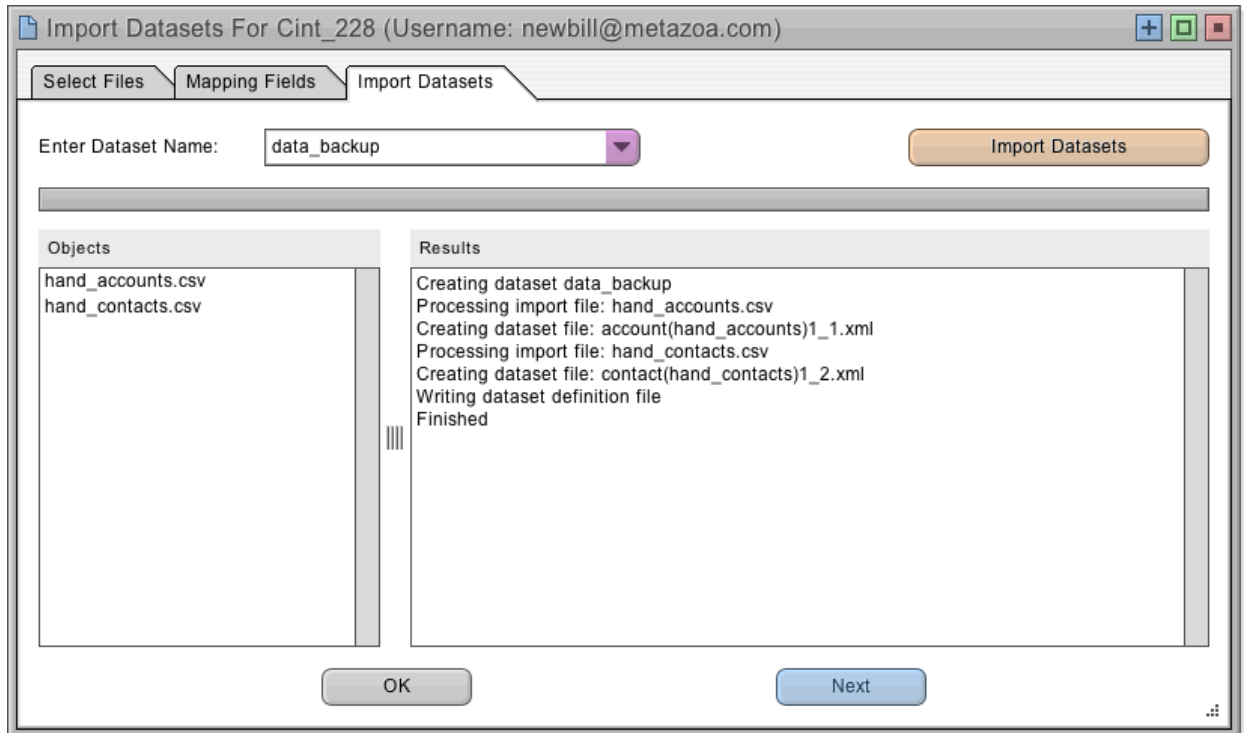
- Imported Objects:** A list box containing two items: "Account (hand_accounts)" and "Contact (hand_contacts)". The "Contact (hand_contacts)" item is currently selected and highlighted in blue.
- Imported Values:** A list of fields corresponding to the selected object. It includes:
 - Id (Contact) mapped to "Contact"
 - AccountId mapped to "AccountId"
 - AccountId.Name mapped to "Name"
 - FirstName mapped to "FirstName"
 - LastName mapped to "LastName"Each mapping is shown in a row with a dropdown arrow on the right side of the destination value.

At the bottom of the dialog, there are two buttons: "OK" and "Next".

Be aware that the Build Dataset and Import Dataset dialogs work somewhat differently. When you Build a Dataset, the records are downloaded from the source Salesforce org, but when you Import a Dataset, the records are mapped against the destination Salesforce org. So be sure that you have selected the correct Deployment Arrow and destination Snapshot when you are importing CSV data. Otherwise, you may not see the expected object and field names from the destination org in the mapping interface above.

Import Datasets Button

The last tab allows you to enter the name of a new Dataset and then click the Import Datasets button at right to start the import process. If you select an existing Dataset name from the menu then that Dataset will be replaced. All of the import results will be listed in the window pane at lower right. Remember, at this point you are only importing the CSV files and creating a new Dataset. If you want to migrate that data to an actual Salesforce org you will need to use the Migrate Dataset dialog, explained in the next section.



Import Datasets For Cint_228 (Username: newbill@metazoa.com)

Select Files Mapping Fields Import Datasets

Enter Dataset Name:

Objects

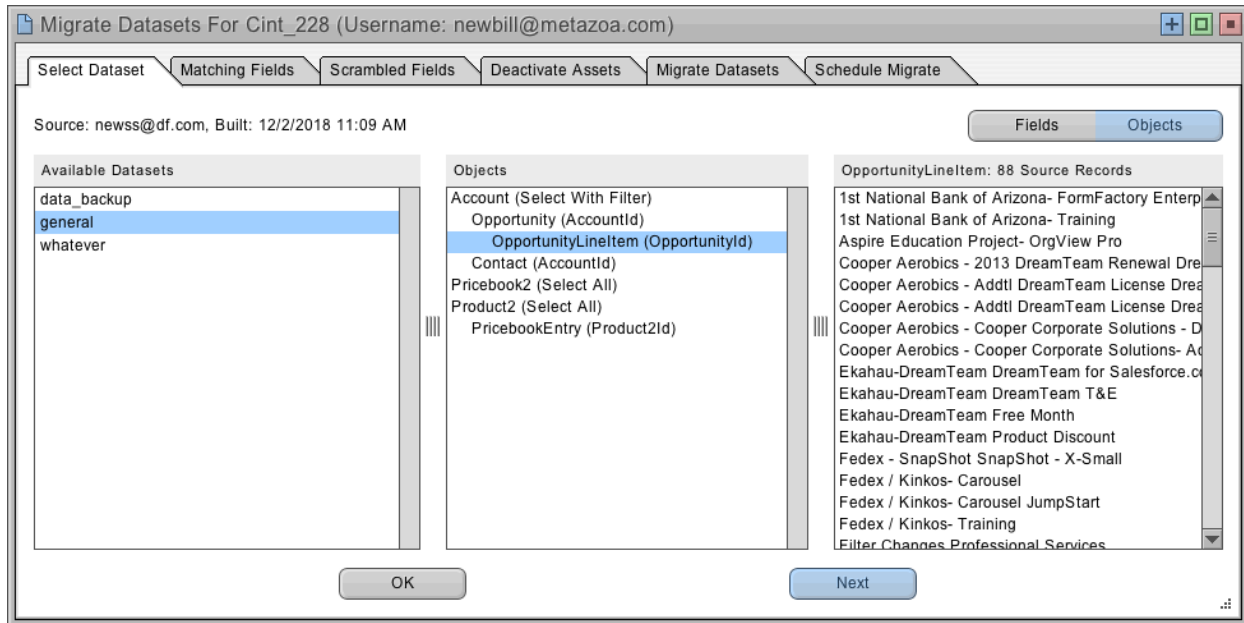
- hand_accounts.csv
- hand_contacts.csv

Results

```
Creating dataset data_backup
Processing import file: hand_accounts.csv
Creating dataset file: account(hand_accounts)1_1.xml
Processing import file: hand_contacts.csv
Creating dataset file: contact(hand_contacts)1_2.xml
Writing dataset definition file
Finished
```

Migrate Datasets Dialog

After a Dataset has been created, you are ready to migrate these records to a destination Salesforce org. Right-click a Deployment Arrow that is connected to the correct destination org and select the Migrate Datasets option to get started. The Migrate Datasets dialog allows you to select any of the global Datasets from the list at left and see the objects and fields that are available in the list at right. The next four tabs provide options for matching fields, scrambling fields, deactivating assets, and finally migrating the selected Dataset.



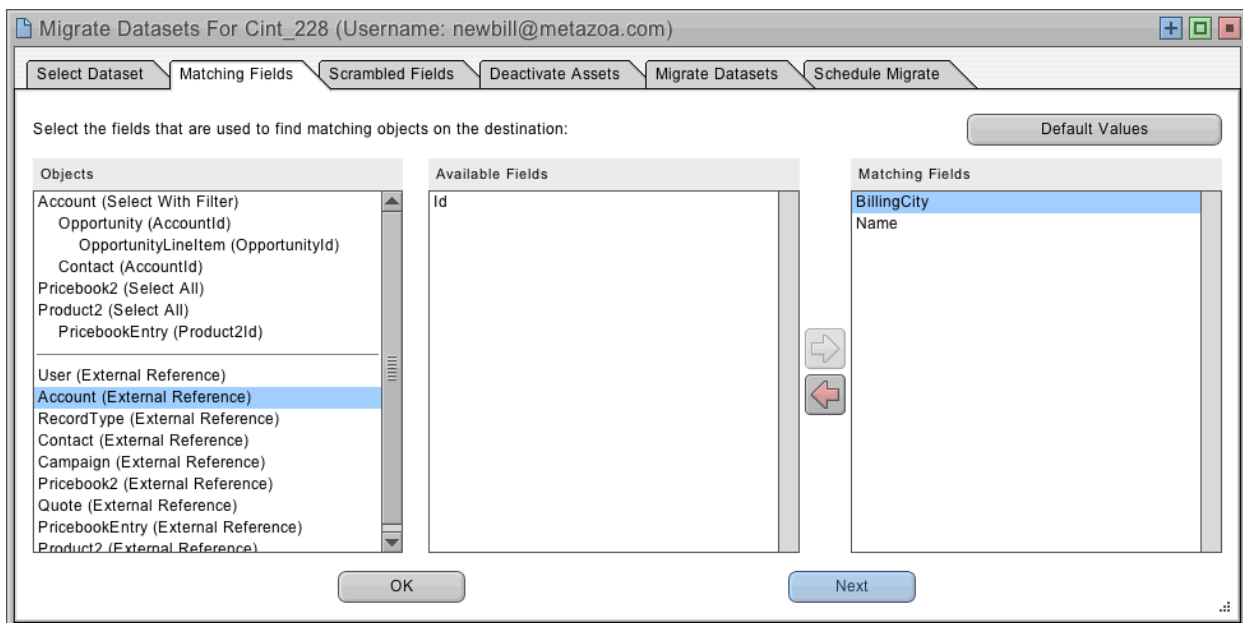
Here is a power user tip. You can right-click any of the objects in the middle window pane and export the data as an XML or CSV file. These files will be in the correct format for the Import Datasets dialog. For example, you could export a CSV file, edit the file as a spreadsheet, and then import your changes.

Matching Fields

Snapshot uses the loaded fields to identify matching records in the destination org during migration. If a source record matches a destination record, then the corresponding destination record is updated. If a source record does not match any destination record, then a new destination record is created.

Snapshot automatically selects common matching fields for you. However, you can also manually select matching fields. For example, if you want to match Accounts by Name and BillingCity, then be sure that both Name and BillingCity are loaded. The selected fields create a logical AND filter for matching destination records.

One powerful way to match destination objects is with External Ids. Other common matching fields include object names, email addresses, and usernames. These fields will automatically be available to select for matching. Some Salesforce sandboxes have the same Ids as production orgs. In that case you can simply use the Id field for matching destination objects.



Migrate Datasets For Cint_228 (Username: newbill@metazoa.com)

Select Dataset Matching Fields Scrambled Fields Deactivate Assets Migrate Datasets Schedule Migrate

Select the fields that are used to find matching objects on the destination: Default Values

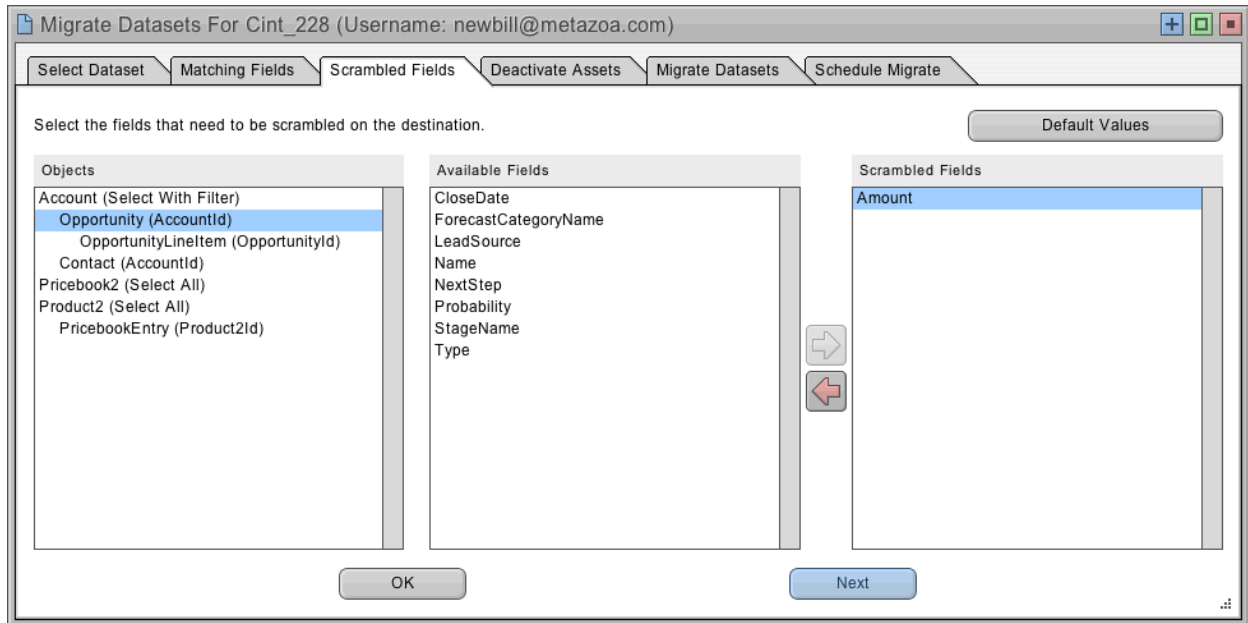
| Objects | Available Fields | Matching Fields |
|-------------------------------------|------------------|-----------------|
| Account (Select With Filter) | Id | BillingCity |
| Opportunity (AccountId) | | Name |
| OpportunityLineItem (OpportunityId) | | |
| Contact (AccountId) | | |
| Pricebook2 (Select All) | | |
| Product2 (Select All) | | |
| PricebookEntry (Product2Id) | | |
| User (External Reference) | | |
| Account (External Reference) | | |
| RecordType (External Reference) | | |
| Contact (External Reference) | | |
| Campaign (External Reference) | | |
| Pricebook2 (External Reference) | | |
| Quote (External Reference) | | |
| PricebookEntry (External Reference) | | |
| Product2 (External Reference) | | |

OK Next

Underneath the dividing line you will see “external references” to other objects in the destination org. Snapshot will automatically connect external references to matching objects on the destination org when the Dataset is migrated. You can manually specify additional fields needed for matching external references as well.

Scrambled Fields

Datasets are often used to move records into a Salesforce Sandbox or Developer Edition for testing or application development. In these situations, you may want to scramble data records that contain sensitive information. These fields might contain financial information, such as credit cards or bank accounts, or personal information, such as email addresses or Social Security numbers. The Scramble Fields tab provides an easy way to obscure fields on the destination org. Move the fields that you want to scramble over to the list at right.



The screenshot shows a web application window titled "Migrate Datasets For Cint_228 (Username: newbill@metazoa.com)". The window has several tabs: "Select Dataset", "Matching Fields", "Scrambled Fields" (which is active), "Deactivate Assets", "Migrate Datasets", and "Schedule Migrate".

Below the tabs, there is a instruction: "Select the fields that need to be scrambled on the destination." and a "Default Values" button.

The main area is divided into three columns:

- Objects:** A list of Salesforce objects with expandable sub-items. "Opportunity (AccountId)" is selected and highlighted in blue. Other objects include Account (Select With Filter), OpportunityLineItem (OpportunityId), Contact (AccountId), Pricebook2 (Select All), Product2 (Select All), and PricebookEntry (Product2Id).
- Available Fields:** A list of fields for the selected object. The fields are: CloseDate, ForecastCategoryName, LeadSource, Name, NextStep, Probability, StageName, and Type.
- Scrambled Fields:** A list of fields that have been moved to the "Scrambled Fields" category. The field "Amount" is currently in this list and is highlighted in blue.

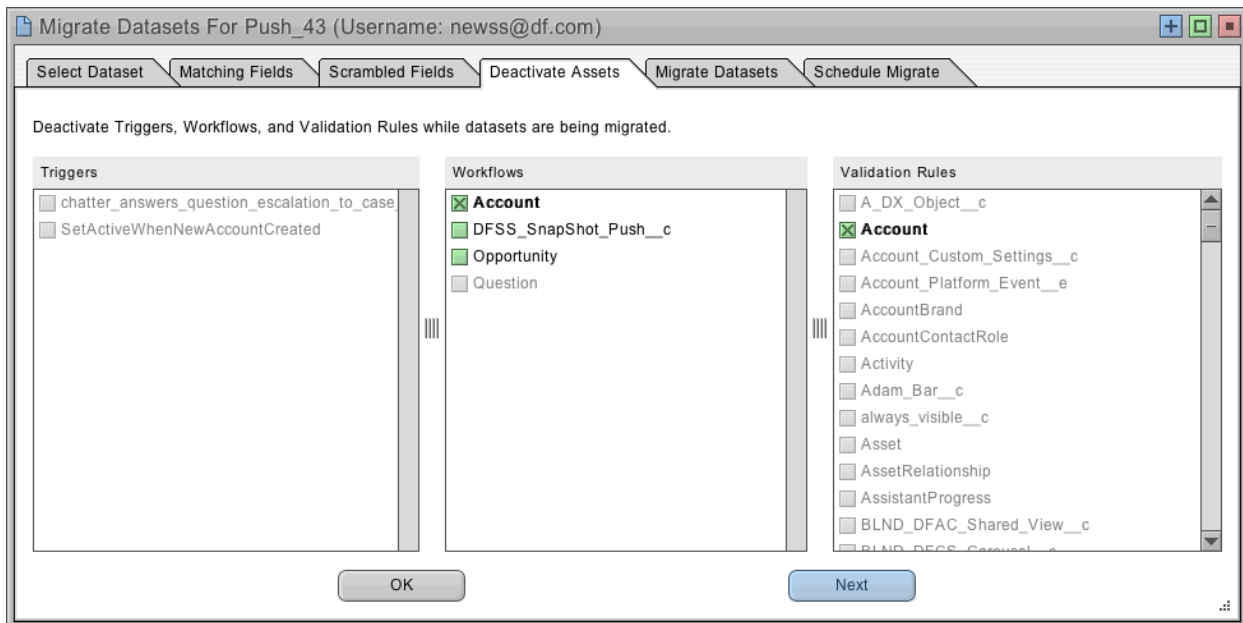
Between the "Available Fields" and "Scrambled Fields" columns are two arrow buttons: a right-pointing arrow (to move a field to the scrambled list) and a left-pointing arrow (to move a field back to the available list).

At the bottom of the window are two buttons: "OK" and "Next".

Deactivate Assets

When a Salesforce record is inserted or updated, various Apex Triggers, Workflow Rules, and Validation Rules might be invoked. Apex Triggers perform custom actions before or after records are changed. Workflow Rules can also be invoked when records change, at which point they will perform automated actions. Validation Rules verify that the record data meets some kind of criteria before being inserted, updated, or deleted.

All of these automated behaviors can cause potentially undesirable effects during data migration. For example, thousands of emails might be sent out, or some records might not be updated. The Deactivate Assets tab provides an easy way to deactivate Apex Triggers, Workflow Rules, and Validation Rules in the destination org before data migration is attempted. After migration, the deactivated triggers and rules will be turned back on.



Migrate Datasets For Push_43 (Username: newss@df.com)

Select Dataset Matching Fields Scrambled Fields Deactivate Assets Migrate Datasets Schedule Migrate

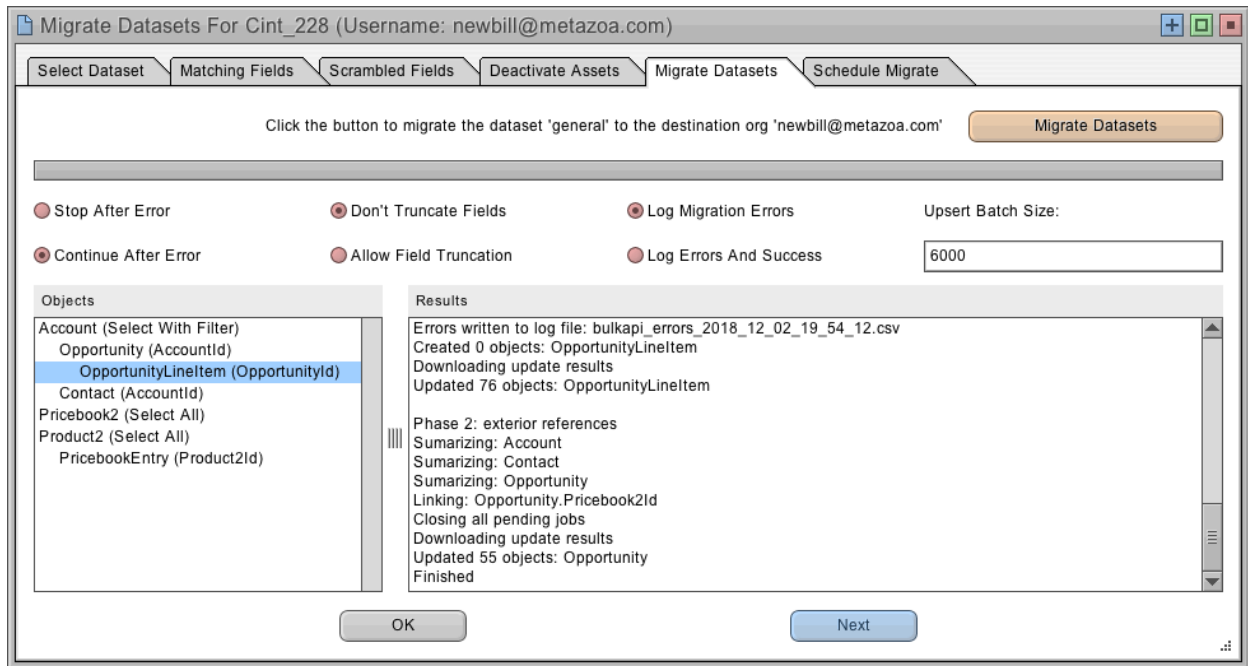
Deactivate Triggers, Workflows, and Validation Rules while datasets are being migrated.

| Triggers | Workflows | Validation Rules |
|--|---|---|
| <input type="checkbox"/> chatter_answers_question_escalation_to_case | <input checked="" type="checkbox"/> Account | <input type="checkbox"/> A_DX_Object__c |
| <input type="checkbox"/> SetActiveWhenNewAccountCreated | <input checked="" type="checkbox"/> DFSS_SnapShot_Push__c | <input type="checkbox"/> Account_Custom_Settings__c |
| | <input checked="" type="checkbox"/> Opportunity | <input type="checkbox"/> Account_Platform_Event__e |
| | <input type="checkbox"/> Question | <input type="checkbox"/> AccountBrand |
| | | <input type="checkbox"/> AccountContactRole |
| | | <input type="checkbox"/> Activity |
| | | <input type="checkbox"/> Adam_Bar__c |
| | | <input type="checkbox"/> always_visible__c |
| | | <input type="checkbox"/> Asset |
| | | <input type="checkbox"/> AssetRelationship |
| | | <input type="checkbox"/> AssistantProgress |
| | | <input type="checkbox"/> BLND_DFAC_Shared_View__c |
| | | <input type="checkbox"/> BLND_DFAC_Shared_View__c |

OK Next

Migrate Datasets Button

The next tab has the main interface for migrating datasets to the destination org. First, make sure that the migration options are set correctly, these are discussed in more detail below. Then click the Migrate Datasets button to get started. All of the details of the migration will be written to the window pane at lower right.



The screenshot shows a window titled "Migrate Datasets For Cint_228 (Username: newbill@metazoa.com)". The window has several tabs: "Select Dataset", "Matching Fields", "Scrambled Fields", "Deactivate Assets", "Migrate Datasets" (which is the active tab), and "Schedule Migrate".

Below the tabs, there is a button labeled "Migrate Datasets" and a message: "Click the button to migrate the dataset 'general' to the destination org 'newbill@metazoa.com'".

Below this, there are several radio button options for migration settings:

- ☐ Stop After Error
- ☒ Don't Truncate Fields
- ☒ Log Migration Errors
- Upsert Batch Size:
- ☒ Continue After Error
- ☐ Allow Field Truncation
- ☐ Log Errors And Success

Below the options, there are two main sections:

- Objects:** A list of objects to migrate, including "Account (Select With Filter)", "Opportunity (AccountId)", "OpportunityLineItem (OpportunityId)" (which is highlighted), "Contact (AccountId)", "Pricebook2 (Select All)", "Product2 (Select All)", and "PricebookEntry (Product2Id)".
- Results:** A text area showing the progress of the migration. It includes messages such as "Errors written to log file: bulkapi_errors_2018_12_02_19_54_12.csv", "Created 0 objects: OpportunityLineItem", "Downloading update results", "Updated 76 objects: OpportunityLineItem", "Phase 2: exterior references", "Sumarizing: Account", "Sumarizing: Contact", "Sumarizing: Opportunity", "Linking: Opportunity.Pricebook2Id", "Closing all pending jobs", "Downloading update results", "Updated 55 objects: Opportunity", and "Finished".

At the bottom of the window, there are two buttons: "OK" and "Next".

Migration Options

There are various options for Data Migration, including:

- Stop After Error
- Continue After Error

If an error occurs, then Snapshot will either stop processing additional files with the Bulk API or continue. All errors are written to the log files. Common errors include too many duplicate records, email addresses in the wrong format, field data in the wrong format, etc.

- Don't Truncate Fields
- Allow Field Truncation

One common problem when moving data between different types of orgs is text strings that are too long for the destination field. Select this option to automatically truncate fields to the correct length or otherwise report an error.

- Log Migration Errors
- Log Errors and Success

All errors are written to the Log Files located in the “datasets” folder in the Snapshot file system. Optionally you can also log successful migrations. The log file contains the source and destination Ids as well as any error message.

The Upsert Batch Size field is used for specifying smaller batch sizes. There is a limit of 12 duplicate records per batch, so a smaller batch size may be helpful in avoiding this error.

Conclusion

This whitepaper has discussed the best practices for data migration using Snapshot on the Salesforce platform. The Snapshot product from Metazoa provides a best-of-breed solution for continuous integration with a highly flexible toolset.

support@metazoa.com

1-833-METAZOA

1-833-638-2962

<https://www.metazoa.com>

Twitter: @metazoa4sf

Facebook: <https://www.facebook.com/metazoa4sf>

LinkedIn: <https://www.linkedin.com/company/18493594/>