2 - Data Migration Step-by-step - Before Loading

Introduction

You're going to have to map data from various sources into Salesforce. **IT'S THAT BIG MIGRATION TIME.**

Well let's make sure you don't have to do it again in two days because data is missing, or delete production data.

> Salesforce does not back up your data.

If you delete your data, and the amount deleted is bigger than what is in the recycle bin, if will be deleted forever. You could try restoring it via Workbench, praying that the automated Salesforce jobs haven't wiped your data yet.
If you update data, the moment the update hits the database (the DML) is done, the old data is lost. Forever.

If you don't have a backup, you could try seeing if you turned on field history.

If worst comes to worst you can pay 10 000€ (not joking, see here) to Salesforce to restore your data. Did I mention that Salesforce would give you a CSV extract of the data you had in Salesforce ? Yeah they don't restore the org for you. You'd still need to restore it table per table with a data loading tool.

But let's try to avoid these situations, by following these steps. These steps apply to any massive data load, but especially in case of deletions.

GENERAL DATA OPERATIONS STUFF

Tools

>

> Do not use Data Loader if you can avoid it. If you tried doing a full data migration with Dataloader, you will not be helped. By this I mean I will laugh at you and go back to drinking coffee. Dataloader is a BAD tool.

Amaxa is awesome and handles objects that are related to one another. It's free and awesome.

Jitterbit is like Dataloader but better. It's free. It's getting old though, and some of the newer stuff won't work like Time fields.

Talend requires some tinkering but knowing it will allow you to migrate from almost anything, to almost anything.
Hell you can even use SFDX to do data migrations.

But yeah don't use dataloader. Even Dataloader.io is better, and that's a paid solution. Yes I would recommend you literally pay rather than use Dataloader.

If you MUST use dataloader, EXPORT THE MAPPINGS YOU ARE DOING. You can find how to do so in the data loader user guide: https://developer.salesforce.com/docs/atlas.en-us.dataLoader.meta/dataLoader/data_loader.htm

> Even if you think you will do a data load only once, the reality is you will do it multiple times. Plus, for documentation, having the mapping file is best practice anyway. Always export the mapping, or make sure it is reusable without rebuilding it, whatever the tool you use.

Volume

If you are loading a big amount of data or the org is mature, read this document entirely before doing anything. LDV starts at a few million records in general, or several gigabytes of data. Even if you don't need this right now, reading it should be best practice in general.

Yes, read the whole thing. The success of the project depends on it, and the document is quite short.

Deletions

> If you delete data in prod without a backup, this is bad.
> If the data backup was not checked, this is bad.
> If you did not check automations before deleting, this is also bad.

Seriously, before deleting ANYTHING, EVER:

- get backup
- check automations
- check backup is valid.

Data Mapping

For Admins or Consultants: you should avoid mapping the data yourself. Any data mapping you do should be with someone from the end-user's who can understand you are saying. If no one like this is available, spend time with a business operative so you can do the mapping and make them sign off on it.

The client signing off on the mapping is drastically important, as this will impact the success of the data load, AND what happens if you do not successfully load it - or if the client realizes they forgot something.

Basic operations for a data mapping are as follow:

- study Source and target data model
- establish mapping from table to table, field to field, or both if necessary.
- for each table and field, establish Source of Truth, meaning which data should take priority if conflicts exist
- establish an ExternalId from all systems to ensure data mapping is correct
- define which users can see what data. Update permissions if needed.

Data retrieval

Data needs to be extracted from source system. This can be via API, an ETL, a simple CSV extract, etc. Note that in general it is better if storing data as CSV can be avoided - ideally you should do a point-to-point load which simply transforms the data - but as most clients can only extract csv, the following best practices apply:

- Verify Data Format
  - Date format yyyy-mm-dd
  - DateTime format yyyy-mm-ddT00:00:00z
  - Emails not longer than 80 char
  - Text containing carriage returns is qualified by "
  - Other field-specific verifications re. length and separators for text, numbers, etc.
- Verify Table integrity
  - Check that all tables have basic data for records:
    - LastName, Account for Contact

    - Name for Account
    - Any other system mandatory fields
  - Check that all records have the agreed-upon external Ids
- Verify Parsing
- Do a dummy load to ensure that the full data extracted can be mapped and parsed by the selected automation tool

Data Matching

You should already have created External Ids on every table, if you are upserting data.
If not, do so now.
DO NOT match the data in excel.

Yes, INDEX(MATCH()) is a beautiful tool. No, no one wants you to spend hours doing that when you could be doing other stuff, like drinking a cold beer.

> If you're using VLOOKUP() in Excel, stop. Read up on how to use INDEX(MATCH()). You will save time, the results will be better, and you will thank yourself later. Only thing to remember is to always add "0" as a third parameter to "MATCH" so it forces exact results.

Store IDs of the external system in the target tables, in the ExternalId field. Then use that when recreating lookup relationships to find the records.

This saves time, avoids you doing a wrong matching, and best of all, if the source data changes, you can just run the data load operation again on the new file, without spending hours matching IDs.

Revision #2
Created 2019-07-07 18:06:22 UTC by Windyo
Updated 2020-03-10 10:56:10 UTC by thejamesjames