



Data sampling for mock/sandbox migrations helps to ensure data accuracy, identify potential issues, and optimize performance before the actual migration takes place.

This document guides team members on data sampling when migrating data from a legacy CRM to a **Partial Copy Sandbox** for a mock data migration.

A partial sample **should not be taken** when the client has a Full Copy sandbox. Instead, a full data load based on the data filter criteria should be completed when a Full Copy sandbox is available.

Data Sample Criteria

1. Begin with the list of all objects to migrate
2. The following 'Master Data' type records should be taken in full:
 - a. Users
 - b. Products
 - c. Service Territories/Locations/Other Territory type objects
 - d. Pricebooks
 - e. Pricebook Entries
 - f. Campaigns
 - g. Any Custom Object where transactional records (such as Opportunities/Deals would look up to)
3. Account Sample: if there are fewer than 10,000 Account records, Accounts should be taken in full. If there are more than 10,000 Account records, then to sample for 10000, execute the following:
 - a. Export all records to CSV/Excel/Google Sheets
 - b. Summarize records by created year through pivot table/countif formulas
 - c. Create a table like the following:

Year	Number of Account Records	% of total	Representative amount to sample (10,000 records)
2021	10,000	10 %	1,000
2022	30,000	30%	3,000
2023	60,000	60%	6,000
Total	100,000	100%	10,000



By doing this analysis, we know we need to bring 1000 Account records created in 2021, 3,000 created in 2022 and 6,000 created in 2023. This will give a representative sample of the types of data issues one can expect to see.

4. For Related records such as Contacts and Opportunities, take ALL records related to the above 10,000 Accounts and migrate them as well. You can do this through Vlookups or SQL statements in the legacy system
5. For Leads, do a similar analysis as the Account object:
 - a. Export all records to CSV/Excel/Google Sheets
 - b. Summarize records by created year through pivot table/countif formulas
 - c. Create a table like the following:

Year	Number of Lead Records	% of total	Representative amount to sample (10,000 records)
2021	15,000	15%	1,500
2022	35,000	35%	3,500
2023	50,000	50%	5,000
Total	100,000	100%	10,000

6. When taking a sample of Account and Lead records in this way, it's important to add a layer of 'randomness.'
 - a. Split out all the data into separate spreadsheets by the created year
 - b. Create a new column called 'Random Number'
 - c. Use the formula =Randbetween(1,10000000)
 - d. Copy the column, and paste it into the same column with 'Paste Values'
 - e. Sort the data by the column 'Ascending'
 - f. Take the first 'N' number of records (based on the table created) for that sheet/year and add it to the 'data to migrate' file which includes all the Account records to migrate as part of the sample
7. Prior to migrating, do a final sanity check:
 - a. Ensure all 'record types' are represented
 - b. Ensure all 'dependent' records are migrated (ex. All Opportunities/deals related to the migrated Accounts)
 - c. Ensure there are no missing dependencies (ex. if you load Opportunities that look up to a contact that is not related to an account that's migrated)
 - i. If this happens, update the 'Accounts to migrate' file with those Accounts as well, and load the related Contacts
8. Ensure you document any data loading issues and do a pre-load validation based on these issues when you do the 'production data load'

