

CRM - Data Seeding

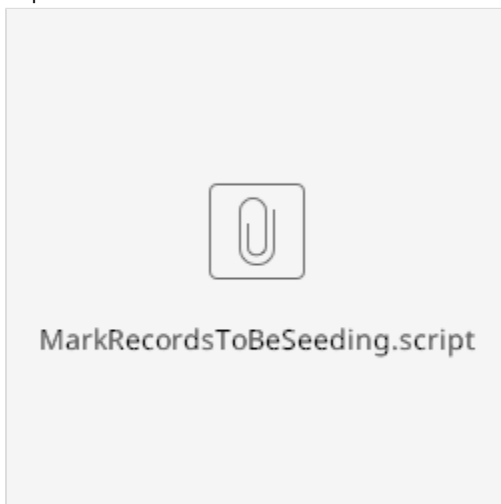
Data Seeding is the process of populating a Salesforce org (such as a sandbox) with relevant test or sample data to simulate real-world scenarios. It involves carefully selecting, preparing, and loading data sets that closely match production conditions while protecting sensitive information. By using seeded data, teams can thoroughly test features, train end users effectively, and showcase realistic demos without risking production data integrity.

Chosen approach:

- For now, the data seeding works only for Icare but it is highly recommended to use it also for Core.
- It was decided that the following objects Product, Account and Pricing Campaign have the field "Is Included in Seeding?" that indicates that these objects and its parents and children will have been moved to the new sandbox
- because the tool copies data directly from production, it is impossible to predict if the records chosen by template will not have any data errors (for example null value for required field e.t.c). Because of that during creation of template we need to keep all automatization but during sandbox seeding we need to disable all validations
- some objects like ccrz__E_Spec__c and ccrz__E_ProductCategory__c are not related with other objects via lookups or master details relations but via indirect relations (dependencies embedded in trigger or flow logic
- The production contains data that should not be accessible by developes/functionals. To avoid migrate these data to sandboxes the source of data seeding should be UAT env where the data is anonimized

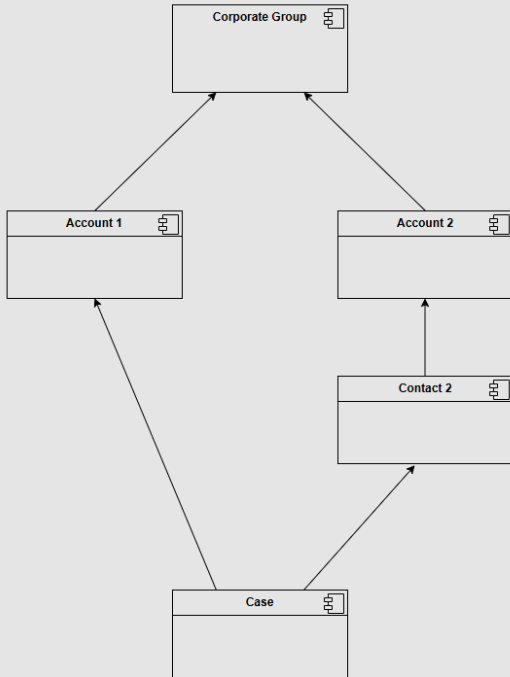
Preparation of Production/UAT (source of data seeding)

- Records to be included in seeding (field "Is Included in Seeding?" = true) should be chosen by people who knows business and technology (Dina, Ann, Julien, e.t.c). on production and then migated to UAT when the UAT sandbox is refreshed. However for now the script has been prepared that selects these records based on data sheet that used to be used to data seeding https://aodocs.altirnao.com/?locale=en_US&aodocs-domain=solvay.com#Menu_viewDoc/LibraryId_QaWbTYEtjpk7b24jVh/DocumentId_TRPZCIs0iRDCvcEAlu/ViewId_QaWbVwAysJMabBxB56 the scrip can be found here:

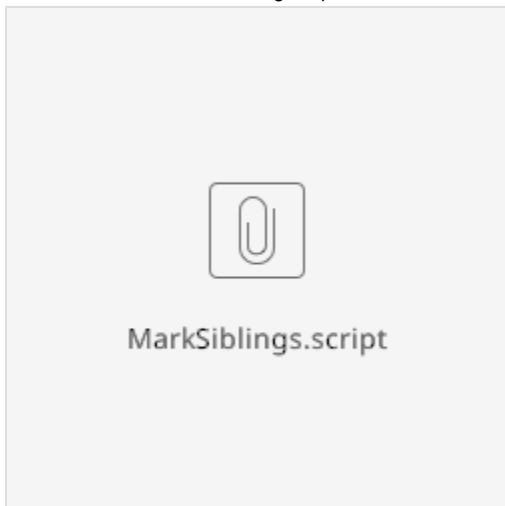


- The script mentioned above should be used only in situation when right records were not marked on production
- There is another data model constraint. Cases can be assigned to contracts of the accounts that belong to the same corporate group as the account to which the case is assigned

VALID SCENARIO

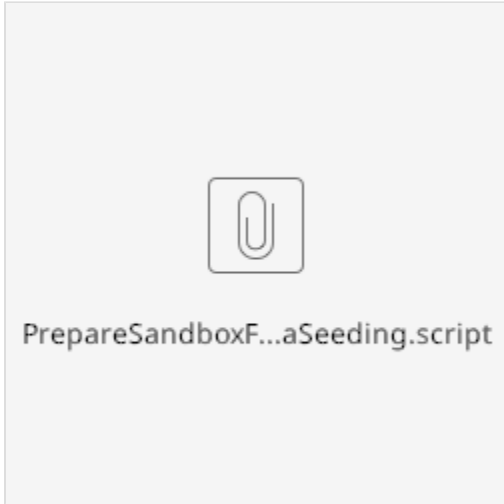


Because of this constaring when we need to mark all accounts from a given corporate group to avoid data inconsistencies when we are inserting cases. To achieve that following script was introduced that marks them.



Preparation of sandbox to which we will insert records (data seeding destination)

Because products are copied to the sandbox by salesforce itself when the new sandbox is created and own data seeding template also copies products duplicates occur. Unfortunately there is a custom mechanism that prevents of creation duplicates the entire data seeding fails. To fix this problem the script that prepares sandbox for dataseeding has been created. It needs to be executed at the destination sandbox before we run the data seeding . This script changes the material group and material code of products that were copied during sandbox creation.

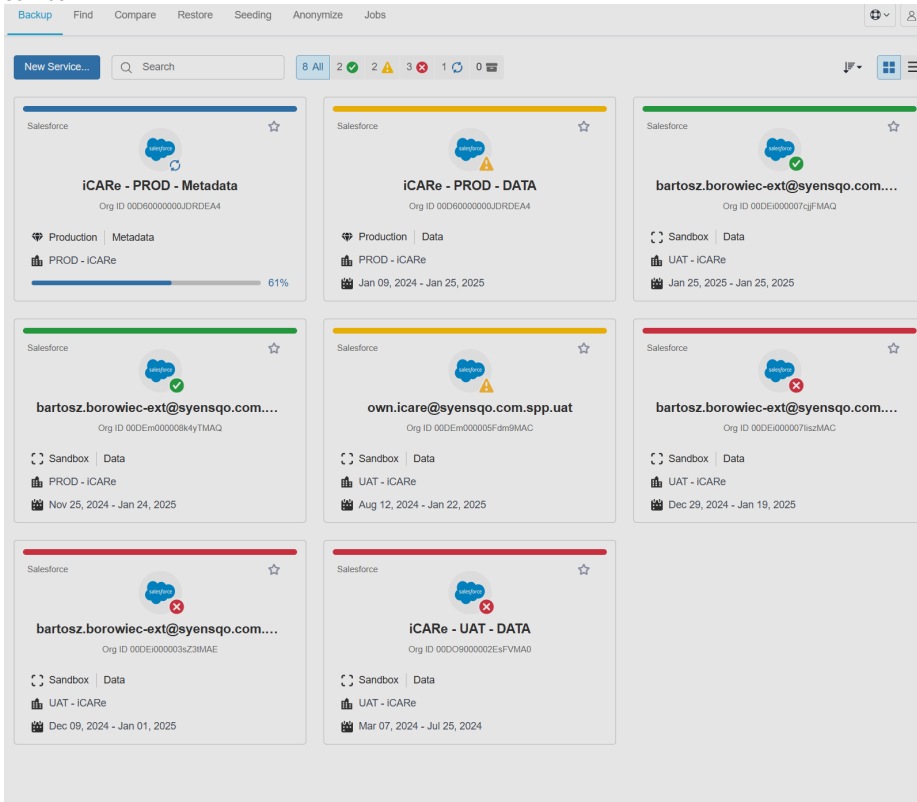


Template creation/Modification

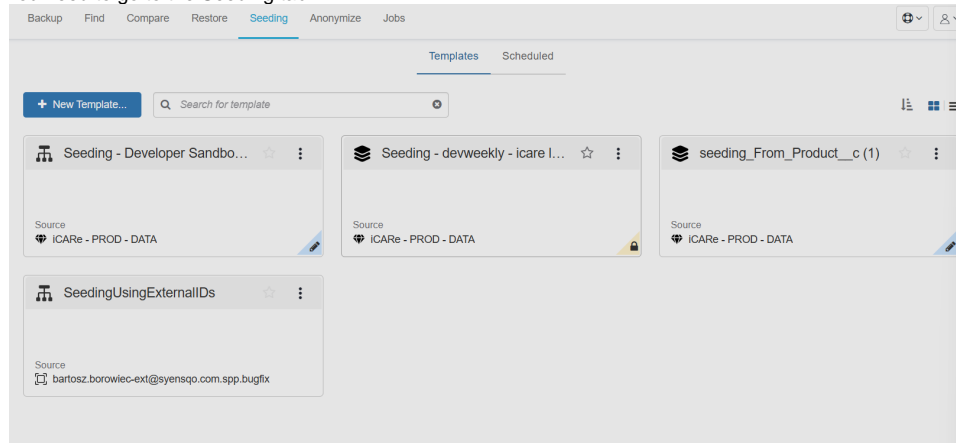
Creation of the Template for data seeding creation is intuitive but there are some hints that needs to be taken into consideration

- log to <https://launcher.myapps.microsoft.com/api/signin/6113285c-a00a-45af-82f5-201b36dce74d?tenantId=5e60f7a7-d410-4d38-b875-93ca12adc30e> (single sign on is enabled)
- You need to have "own backup" admin privileges. The privileges of data seeder allow you to "press the button" to data seed, but does not allow you to create a template

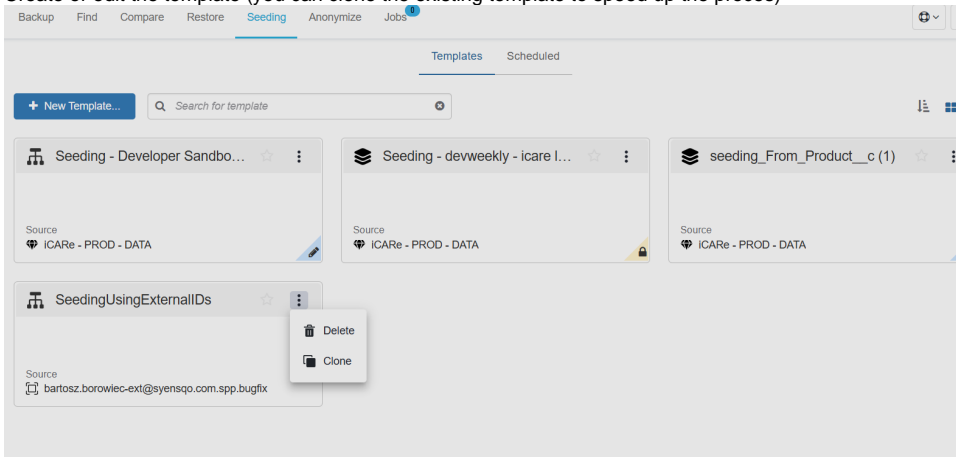
- Ensure that the destination sandbox is added as a backup service. Own backup does not allow you to seed Sanbox that is not added as a backup service



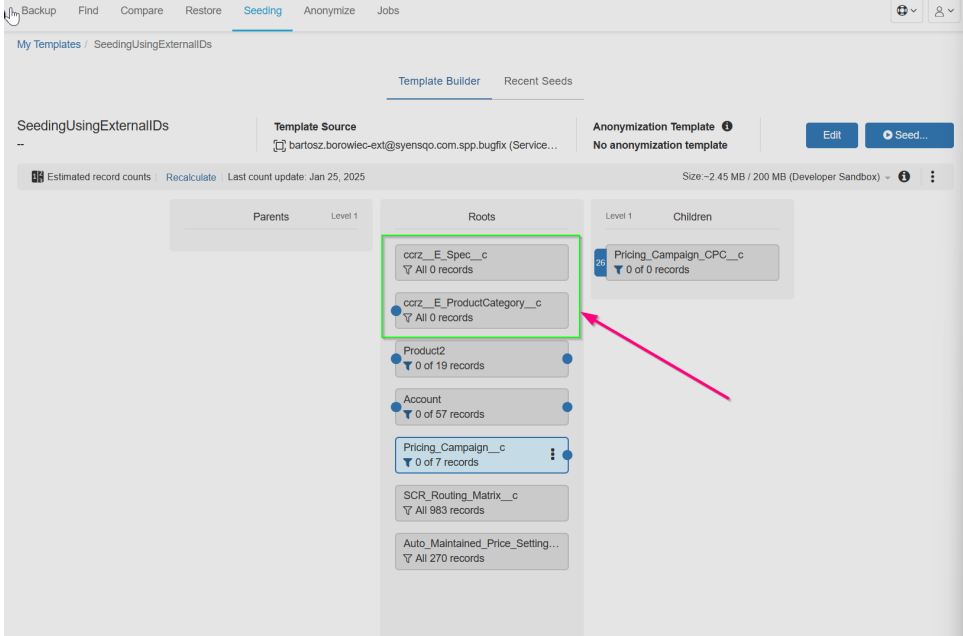
- You need to go to the Seeding tab



- Create or edit the template (you can clone the existing template to speed up the proces)



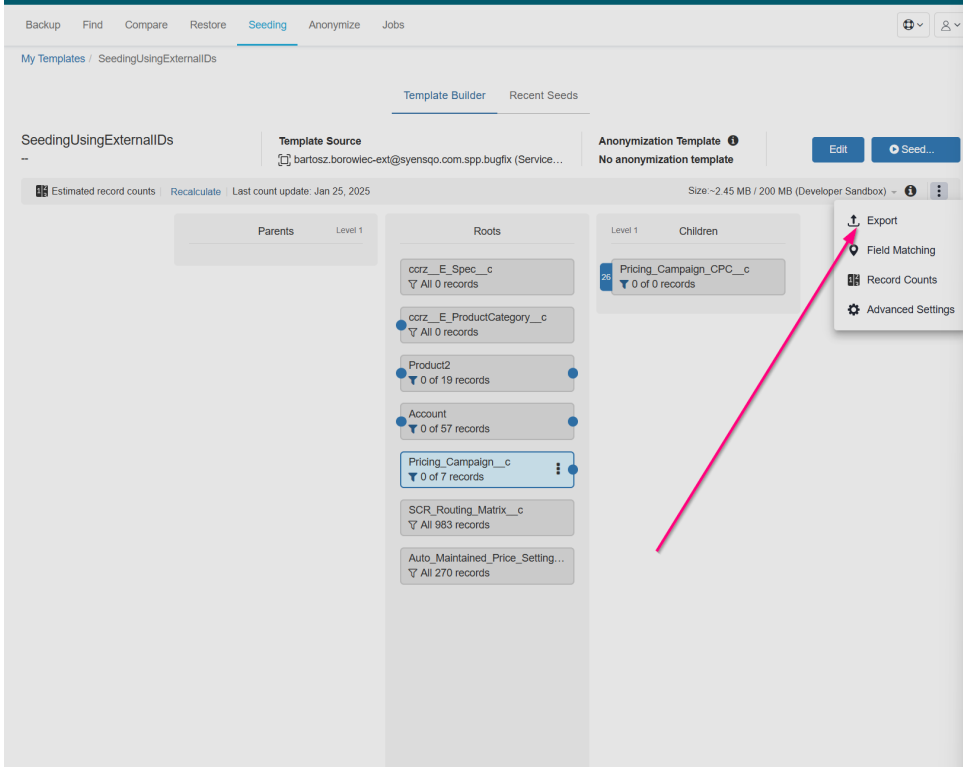
- Roots are objects that you as data seeder choose to be inserted to the sandbox. As you can see some of these objects have to be added as roots because these records are indirectly related to records that are roots and cannot be automatically found by own backup tool



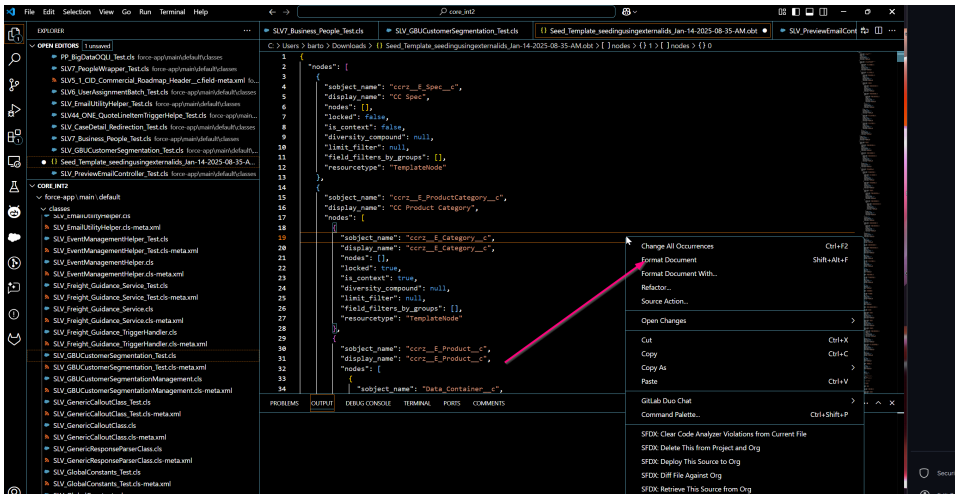
- Own Backup only lets you add new root objects at the end of the existing roots list, which can create problems if a newly introduced object (like X__c) must be inserted before object that is already a root object (for example Product2, Account e.t.c) . In that case, Product2 depends on X__c and cannot be seeded correctly unless X__c is inserted to the sandbox earlier(it means that is placed in the roots list before Product2). To work around this limitation, you can add X__c at the end of the root list, export the template as a JSON file, manually reorder the objects in the JSON, and then reimport it into Own Backup.

Lets break this solution into small steps:

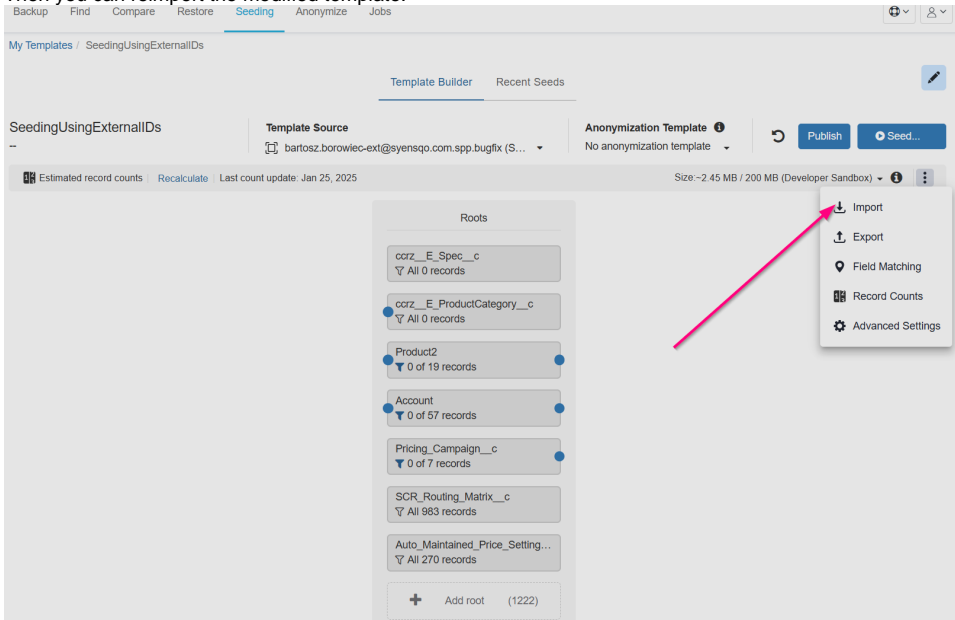
Firstly export the existing template. Own backup exports templates as JSON files.



Then edit this JSON file in Visual Studio code or any other json editor. VSC is quite good because it allows you to format JSON files to the format that are easy to edit



In VSC Move the node with object X_c before the node representing the Product2 and save the modified file
Then you can reimport the modified template:

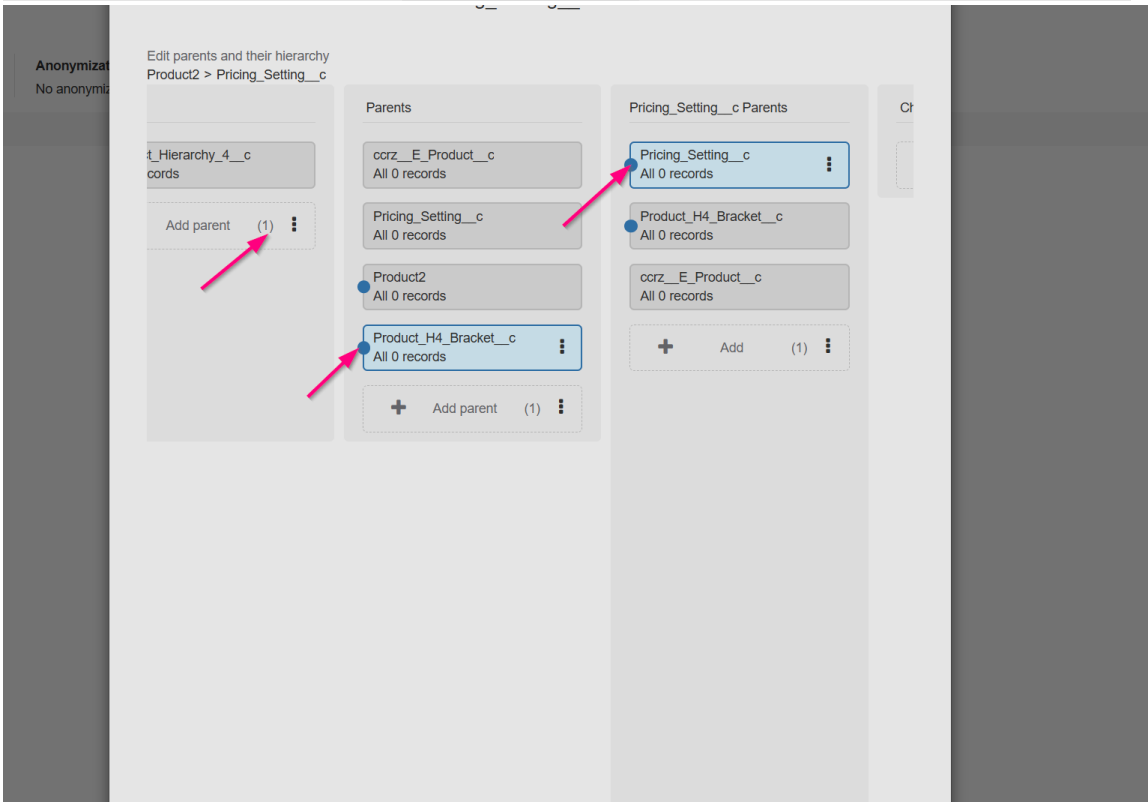
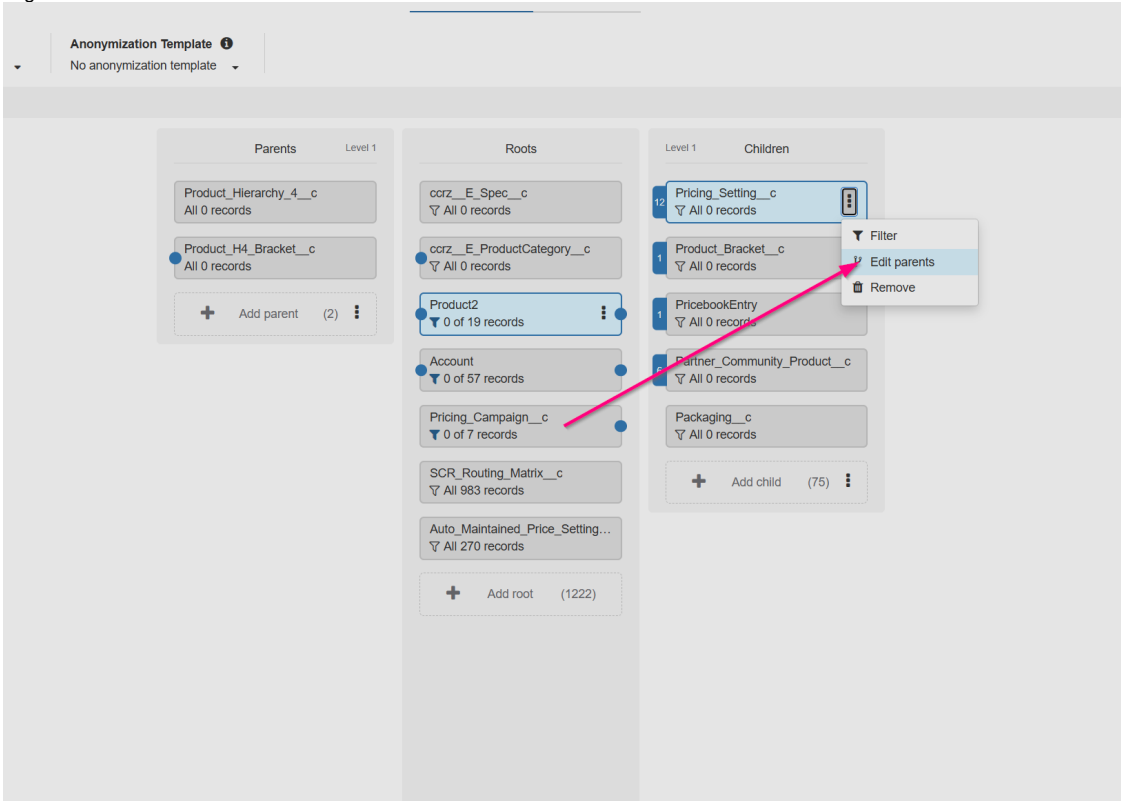


- Always add all parents of root objects. Otherwise, you will face issues during the insertion object. The rule of thumb is to add all parents, grandparents and great-grandparents of root objects. To add parents you need to edit the template and press on the "blue dot" on the left of the root object or the root dot of its parent:

The screenshot shows the 'SeedingUsingExternalIDs' interface. At the top, there are navigation tabs: Backup, Find, Compare, Restore, Seeding (active), Anonymize, and Jobs. Below this, the 'Template Builder' and 'Recent Seeds' sections are visible. The main area displays a hierarchical tree of objects across four levels: Level 3 (Parents), Level 2 (Parents), Level 1 (Parents), and Level 1 (Roots). Each object in the tree has a blue dot on its left side, which is highlighted by a red arrow. The objects and their record counts are as follows:

- Level 3 (Parents):
 - Product_Hierarchy_4_c (All 0 records)
 - Product_H4_Bracket_c (All 0 records) - Blue dot highlighted by red arrow
 - + Add parent (2)
- Level 2 (Parents):
 - Data_Container_c (All 0 records)
 - Product2 (All 0 records) - Blue dot highlighted by red arrow
 - ccrz_E_Attribute_c (All 0 records)
 - ccrz_E_Product_c (All 0 records)
 - + Add parent (1)
- Level 1 (Parents):
 - ccrz_E_Category_c (All 0 records)
 - ccrz_E_Product_c (All 0 records) - Blue dot highlighted by red arrow
 - + Add parent (1)
- Level 1 (Roots):
 - ccrz_E_Spec_c (All 0 records)
 - ccrz_E_ProductCategory_c (All 0 records) - Blue dot highlighted by red arrow
 - Product2 (All 0 of 10 records)
 - Account (All 0 of 57 records)
 - Pricing_Campaign_c (All 0 of 7 records)
 - SCR_Routing_Matrix_c (All 983 records)
 - Auto_Maintained_Price_Setting... (All 270 records)
 - + Add root (1222)

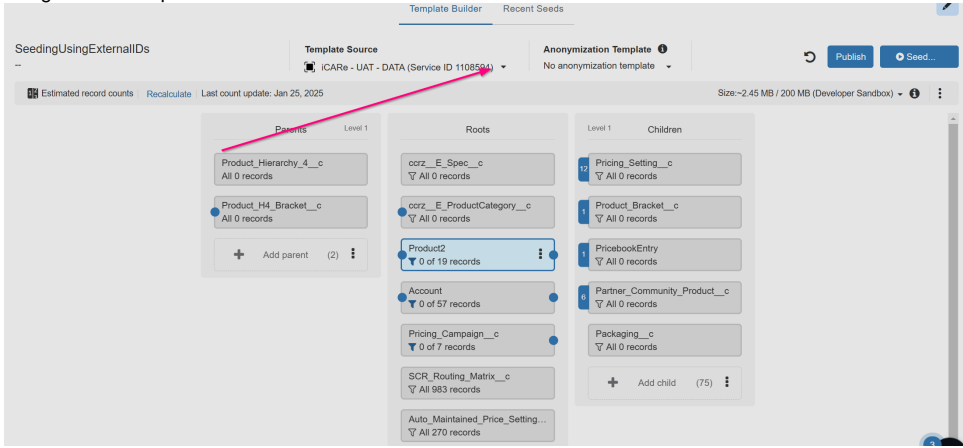
When adding children of the roots object you also need to add parents and grandparents (or even great-grandparents) of these children. Otherwise, you might face data inconsistencies:



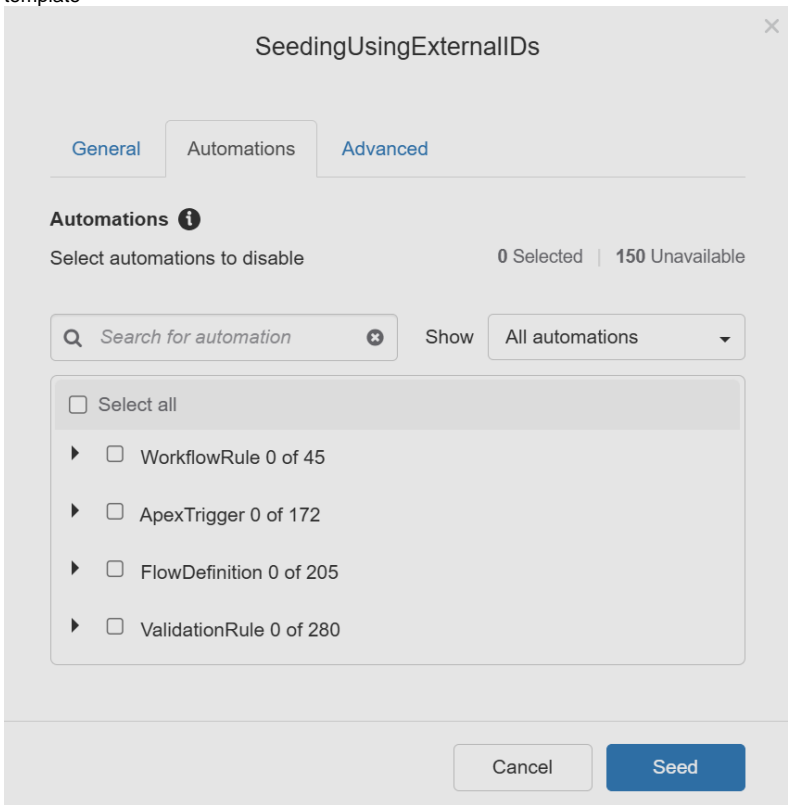
Data seeding.

When the template is ready for data seeding (test data seeding or real data seeding you need to

- Choose the data seeding source. You need to remember that the source sandbox has to be added as a data-seeding service before you are using it in the template.



- For test data seeding always enable all automation and validation rules this allows you to see all possible data inconsistencies and modify the template



- For real data seeding for a sandbox that is a part of the pipeline when the time is a key, it is recommended to disable all validation rules or even all other automation(workflows, triggers and flows) . It is better when a few inconsistent records is inserted into the sandbox than not insert them and as a consequence not inserting all records that are directly or indirectly dependent on these inconsistent records.

- When you are playing with the template you can use the following script to delete records that are already inserted



`cleanTheSandbox.script`