Story UFC-102: Transaction Import Duplication Detection

The following document will go over the design of the Transaction Import Duplication Detection feature, as outlined in Story ticket UFC-102.

Acceptance Criteria

For reference, the following are the acceptance criteria for this ticket:

- When I move to the 'Adjust Transactions' step of the import process, I can see that some transactions have been marked as duplicates.

- I can visually see that a transaction has been marked as a potential duplicate (e.g. a colored background or icon or something).

- I can toggle whether or not the transaction is actually included in the final import.

- By default, any transactions that are marked as duplicates are marked to not be included in the final import.

- I can choose to hide any transactions that have been marked as duplicates while I edit the other transactions.

- The criteria for deciding if a transaction is a duplicate is dependent on the date and amount of the transaction (i.e. if a transaction already exists with the given date/amount, then it should be marked as a potential duplicate).

- If I continue on to the summary step with transactions that have been marked as duplicates, then they should not show up in the summary, and should not be imported.

Design (v1)

At a high level, this feature should work roughly as follows:

i. During the `parseCsvIntoTransactions` saga, the process will back check all existing transactions to see if a to-be-imported transaction is a potential duplicate.

    – This might seem like overkill, but what if we added an index to the `transactionsIndex` for indexing the transactions by date? That would help speedup the duplicate checking process, plus enable things like the charts (which group over time/day) to be a bit more efficient as well.

        • Additionally, it would help with other date-related filtering tasks in the future (could even help with the existing `month` picker on the `Transactions` page).

ii. If it is a potential duplicate, it should set a flag to indicate as much.

   – This flag could be in one of (mainly) two places: either on the `Transaction` object itself, or as a separate piece of state in the `transactionsImport` slice (e.g., a map of ID → flag).

   – If we put it on the `Transaction` object, then we should follow advice to split out a new subclass (e.g. `ImportableTransaction`) and move the `targetAccount` property along with it.

      • One advantage of this is that the flag is then colocated with the other transaction data, meaning that the processes for updating/displaying the value are very similar as for the rest of the transaction properties.

      • One downside is that we now have specialized `Transaction` classes, for which I am hesitant since, ya know, "OOP sucks".

   – If we instead put it in the `transactionsImport` slice, then we'll have to somehow make the `TransactionsImportTable` (and subsequently, the `TransactionsTableCard`) have to know how to use this piece of state.

      • But then we wouldn't have multiple `Transaction` classes.

   – I think I'm leaning more towards the `ImportableTransaction` object, just to keep the simplicity of how transaction properties are updated and such.

   – Regardless, I think it'd be a smart idea to move a bit more of the 'import related' functionality of the `TransactionsTableCard` (e.g. `useTargetAccount`) into `TransactionsTableCard`. This would entail moving things over like specific components, operating on a separate copy of the `columns`, and composing the original table component.

iii. Once the user is in the `Adjust Transactions` step, they'll see that the table has a new column with a checkbox for each row. Obviously, 'checked' would be 'duplicate'.

iv. It would also be nice to add a background color to the rows of duplicate transactions. I'm thinking something orange? ala 'warning'?

v. Once the user moves to the `Complete Import` step, the `cleanTransactionsForSummary` saga should now remove any duplicate transactions.

vi. The table in the `Complete Import` step definitely shouldn't be showing the 'Duplicate' column.

Design (v2)

This design takes into account some updated acceptance criteria and ideas that were generated -- in particular, instead of having a 'duplicated' column, instead just have an 'Import?' (i.e. 'Include?') column that users can use to toggle whether or not they want to import certain transactions.

Then, duplication detection would just toggle this inclusion off (by default) and add a little warning icon to mark the transaction as a potential duplicate.

- HOLD UP, let's consider for a second... is it actually better to use a row background color instead of an icon for highlighting duplicates? Let's consider the design decisions that were just made below, specifically of putting the warning message into a toast like box. If we used a yellow background on the duplicate rows, I think it would have better semantics than just an icon. Additionally, and this is the real kicker, I think it would improve the UX since, as a user scans down the table on the right side (where most of their work is, just categorizing), they'll be able to more quickly determine which transactions are duplicates, than having to scan back to the left side of the table to find an icon. I think I'll try out both these designs and just see which looks better. But it would seem like a colored background would have some theoretical advantages over using icons.

The following are just refinements to the design details in v1 above.

i.   Keep the same.

ii. I think I've settled on using a separate `ImportableTransaction` class. But now that we'll have a toggleable 'Include' column, that means we'll need to add two new fields to this `ImportableTransaction`:

   1. `isDuplicate`

   2. `includeInImport`

iii. The `Adjust Transactions` step table will now have a column titled `Import?`. Next to the checkboxes in each row, a warning icon will be displayed if the transaction has been marked as a potential duplicate.

   – Hmm... after having looked at some designs for tables with checkboxes, it seems like almost **no one uses a label in the column header** for the checkboxes (when the checkboxes are the far-left column and used for selection); instead they just put another checkbox for selecting all/none. Which makes sense... this is a pretty common pattern (ala gmail). While this is obviously the more popular design, and **I think we should use it**, I just hope users will understand what the checkboxes are for in the case where it's maybe their first time using the import process, having no duplicate transactions, and wondering what these checkboxes are for. Like, it *should* be obvious, but we all know users...

iv. Since we'll have the warning icons, I don't think we need the colored background anymore.

  – OR NOT??

v. Keep the same.

vi. Keep the same.

  Now for some new details.

vii. There should be a message above the table indicating what the warning icons represent.

  – This message should only be present when any duplicate transactions have been detected.

  – The message should be something along the following:

    • "It looks like some transactions could be possible duplicates. These have been marked with <warning icon|colored background> and excluded from being imported. You can choose to include these transactions for import if you wish."

  – The message itself should be styled like a warning toast inside the Step card: yellow background, square rounded corners, yada yada, to make itself more obvious that this is something notable.

viii. There should also be a tooltip on the icons indicating 'Possible Duplicate'.

ix. There should be a checkbox above the table to hide all of the duplicate transactions.

  – The `ImportTransactionsTable` would then have to perform the filtering before passing the transactions down to `TransactionsTableCard`.

  – Maybe this checkbox should go inside the card of the warning message?

Tasks

UFC-140

• Create the `ImportableTransaction` model.

• Modify the `parseCsvIntoTransactions` saga to do the duplication detection and use the `ImportableTransaction` model.

UFC-141

- Move the import-related things from `TransactionsTableCard` to `ImportTransactionsTable`.

- Add the checkbox column to `ImportTransactionsTable` when in editing mode.

- Modify the rows to have a background color when they the transaction is marked as a duplicate.

- Add the warning message + "hide duplicates" checkbox to `StepAdjustTransactions`.

- Modify `ImportTransactionsTable` to have a prop to toggle hiding duplicate transactions.

- Modify the `canGotoNextStep` logic so that it only checks 'marked for import' transactions to make sure they're valid.

- Modify the `cleanTransactionsForSummary` to exclude all transactions that have been marked to not be imported.