

⚠ Make a copy of this document to add your own notes and comments.

ARR Rollup Challenge

Your mission, should you choose to accept it, is to design and implement an ARR (🚩) Rollup solution for Acme, Inc. This solution should be repeatable and reliable.

Note: ARR stands for Annual Recurring Revenue.

Acme, Inc. needs to know how much revenue Acme is receiving from each account, and how much revenue Acme receives from an entire account hierarchy.

Tables of the Database

Below, you will find three CSV files which will act as your database.

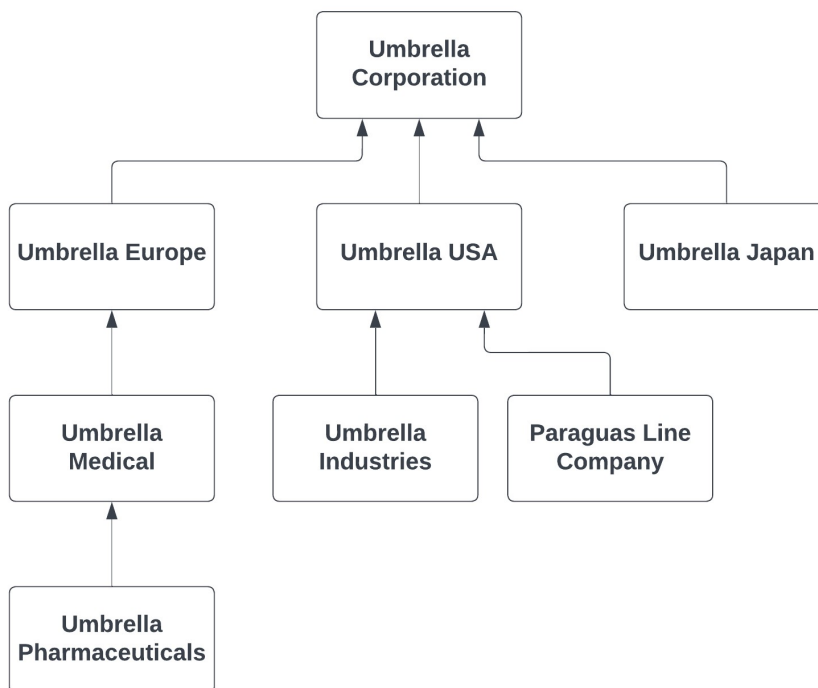
File	Description
accounts.csv	<p>This file contains a list of accounts. Think of an account as a company, for example: Umbrella Corporation, or Wayne Industries.</p> <p>Note: Accounts can live in a hierarchy. Look to the next section for more info.</p>
subscriptions.csv	<p>This file contains a list of subscriptions. A subscription is a group of products that Acme is selling to the account.</p> <p>For example, if you buy the Disney+ Bundle, it will come with Disney+, Hulu, and ESPN+. To keep track of that, there will be a single row in the subscriptions file tying the purchase to the account. Then the subscription items file will have three rows, one for Disney+, one for Hulu, and one for ESPN+.</p> <p>Each subscription will be tied to an account.</p>
subscription_items.csv	<p>This file contains a list of subscription items. A subscription item defines the product being sold in the overall subscription. It will also store information on the price of the subscription item.</p> <p>Each subscription item will be tied to a subscription.</p>

Account Hierarchy

Account hierarchies are important to keep track of as companies can form complex structures. In Acme's case, they will sell to any legal entity. So if multiple companies in the same account hierarchy would like to buy Acme's product, then that is allowed.

Note: Some companies will only sell to the ultimate parent company so even they need to understand a company's hierarchy so they are selling to the right people.

Let's use the Umbrella Corporation as an example of a company structure.



Let's establish some key terms using the diagram above:

- **Ultimate Parent**
 - The ultimate parent is the account at the very top of the hierarchy.
 - In the diagram above, **Umbrella Corporation** is the ultimate parent. No other account in that hierarchy is the ultimate parent. There can only be one ultimate parent in a hierarchy.
 - If an account is by itself with no parent or child accounts, then the account will be its own ultimate parent account.

- **Parent**
 - A parent is an account that has one or more child accounts.
 - In the diagram above, **Umbrella USA**, **Umbrella Europe**, and **Umbrella Medical** are parent accounts.
- **Child**
 - A child is an account that has a parent account, and has no child accounts.
 - In the diagram above, **Umbrella Pharmaceuticals**, **Umbrella Industries**, **Paraguas Line Company**, and **Umbrella Japan** are child accounts.

There is no limit to how many child companies a parent company can have.

Once an account hierarchy is set up, you can calculate a key data point which is the ARR of the account hierarchy. Let's say Umbrella USA is paying Acme \$30,000, and Umbrella Medical is paying Acme \$15,000. The ARR of the account hierarchy is \$45,000.

Requirements

Your task is to populate three columns in *accounts.csv*:

- **ultimate_parent_id**
 - This column should be populated with the id of the ultimate parent account.
- **arr**
 - This column should be populated with Acme's revenue for the account.
- **hierarchy_arr**
 - This column should be populated with Acme's revenue for the entire account hierarchy.

You can use any programming language to complete this task.

The script should be repeatable, and maintainable. In lieu of a real database, treat the CSV files as tables in a database. The CSV files are large files to allow you to find ways to optimize your script.

We will ask for your output so that we can verify it.

For determining the ARR of an account, you will need to look at the subscriptions and subscription items. Take a look at the next section to see what the columns mean, and to determine how you can relate the data.

Only active subscriptions, and active subscription items should be considered for the ARR. An active subscription or subscription item is where the Start Date is today or earlier, and the End Date is today or later. Use the table below for reference:

Start Date	End Date	Active?
5/30/2022	5/29/2023	No
1/1/2023	12/31/2023	Yes
12/1/2023	11/30/2024	No
5/31/2023	5/30/2024	Yes

To determine the revenue for each subscription item, there are three data points provided:

- **quantity**
 - The number of items that are being sold.
- **list_price**
 - The annual list price of one item.
- **discount**
 - The discount for the items. The format in the file will be 0.07 which is a 7% discount.

For an example, if someone from Umbrella Corporation buys 7 licenses that are \$15 a year with a 10% discount, the ARR for that would be \$94.50.

Once you have the ARR of each account, then you can calculate the ARR of the account's hierarchy which will be the sum of the ARR of the accounts inside the hierarchy.

Columns of each Table

Here are the columns you will see in **accounts.csv**

Column	Description
id	A unique identifier for the account.
parent_id	The id of the account that is the parent of this account.
name	Name of the account.

ultimate_parent_id *	The id of the ultimate parent account.
arr *	The sum of the ARR that this account is paying.
hierarchy_arr *	The sum of the ARR that this account's hierarchy is paying.

** denotes the columns your script will be populating*

Here are the columns you will see in **subscriptions.csv**

Column	Description
id	A unique identifier for the subscription.
account_id	The id of the account that this subscription belongs to.
start_date	The start date of this subscription.
end_date	The end date of this subscription.

Here are the columns you will see in **subscription_items.csv**

Column	Description
id	A unique identifier for the subscription item.
subscription_id	The id of the subscription that this subscription item belongs to.
product_name	The name of the product that this subscription item is for.
quantity	The number of products being sold.
list_price	The annual list price of this subscription item.
discount	The discount for this subscription item.
start_date	The start date of this subscription item.
end_date	The end date of this subscription item.