Extract: your original data sources and how the data was formatted (CSV, JSON, pgAdmin 4, etc).	type of transformation needed for this data (cleaning, joining, filtering, aggregating, etc).	The type of final production database to load the data into (relational or non-relational).	The final tables or collections that will be used in the production database.
Wikipedia list of Botanical Gardens By State, Pandas df.read_html to Pandas Table. Then taking the Table and turning it into a data frame	cleaning, dropping columns, adding a state column with default value to all data frames	MongoDB (non-relational)	One collection per state or territory. Each collection has a list of botanical gardens, each garden entry is saved as a document.
USDA Native Plants by State (CSV for each state)	Cleaning Author Name out of Scientific Name and Author Column	MongoDB (non-relational)	
We wanted our data to be refrenced through a simple search. Our intent was to have all related materials available. For example, if you search for a plant by name, you can see if it's native to a state or multiple states. If you search your state, you will have similar results narrowed down by state, and a list of botanical gardens in your state. U.S. Territories have been included as 'States'.			