**Overview of the Monthly Flow**

Each month, you gather:

1. **Oracle USD** and **Oracle CAD** files
2. **CKs** file containing “NIS Details by Basin - US” and “NIS Details by Basin - CA”
3. **Various Tableau pulls** (multiple workbooks/views)
4. **Existing “Reference File”** (the monthly reference repository you update each cycle)

You then:

1. **Extract, transform, and pivot** data from Oracle and CKs to **append** to the **monthly reference file**.
2. **Pull additional data** (GL transactions, unassigned data, activity metrics, etc.) from Tableau.
3. **Combine** the updated reference file data + the new monthly Tableau data to do your analysis:
   * Compare CKs vs. GL for completeness.
   * Identify unassigned transactions at the district level and distribute them based on activity metrics.
   * Incorporate idle-day cost (GAP), standard estimates (for standard cost), and forecast data (for forecast cost).
4. **Build the “Main2”** table that contains the final breakout by project (actual + standard + forecast comparisons).
5. **Append** the “Main2” table into a **master file**. The master file contains a long-term record of every month’s Main2 data.

**Files in Detail**

**1. Reference File (Monthly Reference)**

* **Purpose**: Holds the **live reference data** for the month, combining:
  + Updated Oracle & CKs pivot tables
  + Ongoing FX references (so you know the monthly FX rates)
  + Possibly other reference tabs (like unassigned distributions from prior months, if you carry them forward)
* **Lifecycle**:
  + Start with last month’s reference file (because it may have historical tables, e.g., past month’s CKs pivot, past month’s FX).
  + Append/overwrite any data for the new month (CKs pivot, new FX rate, etc.).
  + Save as the “latest reference file” so it’s up-to-date.

**2. Master File**

* **Purpose**: Long-term repository of the **final “Main2”** table (and possibly other final analyses if you want to keep them).
* **Lifecycle**:
  1. Each month, once you have your final “Main2” built, you open the master file, append the new rows for the current month, and save.
  2. Over time, this master file grows to hold the full monthly history of final numbers.

**High-Level Step-by-Step**

Below is a more **detailed** monthly cycle:

1. **User picks**:
   * Oracle USD (for month M)
   * Oracle CAD (for month M)
   * CKs file (for month M)
2. **Open Last Month’s Reference File** (e.g. RefData.xlsx)
   * Contains an FX sheet with prior months’ rates, a CKs\_Pivot sheet with prior months’ pivot data, maybe other reference tabs for overhead distributions, etc.
3. **Process the 3 monthly files** (USD, CAD, CKs):
   * **Merge Oracle** to find new data for “USD” and “CA” (the Oracle IS19 sheets).
   * **Pivot CKs** to produce a monthly pivot of the “NIS Details…” data.
   * **Calculate this month’s FX** (comparing Oracle USD vs. CAD).
   * **Append** new pivot rows to the CKs\_Pivot sheet in RefData.xlsx.
   * **Append** or **overwrite** the new FX rate in the FX sheet in RefData.xlsx.
   * Save the updated RefData.xlsx (it now has the new month’s reference data included).
4. **Pull from Tableau**:
   * **GL transactions** (by project)
   * **Unassigned data** (district-level transactions)
   * **Activity metrics** (like job counts, hours, crews, etc.)
   * **Other references** (maybe standard rates, forecast data, etc.)
   * Place these into your **in-memory** structure (or directly into a “Tableau Pull” sheet in RefData.xlsx, if you prefer to keep them in the reference file for easy access).
5. **Do the Analysis**:
   * **Compare CKs pivot** (from RefData.xlsx) with **GL transactions** (from Tableau) for completeness.
   * **Create** unassigned distributions, using the newly pulled activity metrics. Distribute them at a project level.
   * **Incorporate** GAP (idle days cost) from whichever source if needed.
   * **Calculate** “Standard” cost (estimated rate \* actual activity).
   * **Calculate** “Forecast” cost (estimated rate \* forecasted activity).
   * Build the final combined table with columns for:
     + **Actual** cost (CKs or GL, after unassigned reclass)
     + **Standard** cost
     + **Forecast** cost
     + **GAP**
     + …plus any additional fields you need.  
       Let’s call this final combined table **Main2** (for month M).
6. **Append Main2** (month M) to the **Master File**:
   * Open MasterFile.xlsx.
   * Read the existing “Main2” sheet (which has prior months).
   * Append the new rows for month M.
   * Save MasterFile.xlsx.
7. **Output**:
   * You have an updated RefData.xlsx for next month’s baseline.
   * You have an updated MasterFile.xlsx that includes the newly created Main2 rows.

**Logical Flow Diagram**

arduino

Copy

┌─────────────────────────┐

│ Last Month's │

│ Reference File │

│ (RefData.xlsx) │

└───────┬─────────────────┘

│(open & read)

│

┌───────▼─────────────────┐

│User Provides: │

│1) Oracle USD/CAD (month) │

│2) CKs (month) │

└───────┬─────────────────┘

│(process & pivot)

▼

┌─────────────────────────┐

│ Update & Append │

│ 1) CKs pivot → RefData │

│ 2) FX → RefData │

└───────┬─────────────────┘

│(save RefData)

▼

┌─────────────────────────┐

│ Tableau Pulls │

│ - GL by Project │

│ - Unassigned, etc. │

│ - Activity Metrics │

└───────┬─────────────────┘

│

▼

┌─────────────────────────┐

│ Combine & Analyze │

│ (RefData + Tableau) │

│ Compare CKs vs. GL, │

│ create distributions, │

│ standard, forecast, etc.│

└───────┬─────────────────┘

│(produce Main2)

▼

┌─────────────────────────┐

│ Master File │

│ (MasterFile.xlsx) │

│ (contains all Main2) │

└───────┬─────────────────┘

│(append new month)

│(save MasterFile)

▼

(End of Monthly Cycle)

**Implementation Guidance**

**A. Reference File Handling**

* In your code, you’ll have functions like append\_fx\_to\_ref\_file(), append\_cks\_pivot\_to\_ref\_file(), etc.
* These read the existing data from the reference file, append the new month’s data, and rewrite.
* The reference file might also store:
  + Standard Rates
  + Project Lists
  + Crew Home Count
  + Tech, etc.  
    (Some of these you may already be pulling from Tableau. Decide if you store them in the reference file for convenience or keep them always external.)

**B. Master File Handling**

* In your code, a function like append\_main2\_to\_master(main2\_df, master\_filepath) will:
  1. Open MasterFile.xlsx.
  2. Convert the “Main2” sheet to a DataFrame.
  3. Append the new data from main2\_df (the final data for month M).
  4. Write it back out.

**C. Appendix: Potential Project Layout**

arduino

Copy

my\_project/

├─ config.ini

├─ main.py

├─ logs/

│ └─ ...

├─ modules/

│ ├─ ui.py

│ ├─ logging\_setup.py

│ ├─ data\_processing.py

│ ├─ ref\_file\_integration.py <-- new: for reading/writing the reference file

│ ├─ master\_file\_integration.py <-- new: for reading/writing the master file

│ ├─ pivot\_cks.py

│ ├─ tableau\_integration.py

│ ├─ workbook\_builder.py

│ └─ ...

└─ ...

* ref\_file\_integration.py could hold the “append CKs pivot,” “append FX,” etc.
* master\_file\_integration.py could hold “append Main2.”
* The final “analysis” steps can be in data\_processing.py or a dedicated “analysis.py.”

**Key Points to Remember**

1. **Reference File** = updated monthly, stores the data you’ll reuse or compare this month and next.
2. **Master File** = accumulates final output (Main2).
3. **Tableau + Oracle + CKs** = monthly inputs that feed into building/updating the reference file, which is then used to compile the final data.
4. **Unassigned Distribution** logic can be placed in a function that references the activity metrics from Tableau. It updates a DataFrame that merges or re-allocates costs from “Unassigned” to the correct projects.
5. **Standard** (actual activity \* standard rates) and **Forecast** (forecast activity \* forecast rates) are typically columns in the final Main2. You’ll need to store or pull standard rates and forecast data from either the reference file or Tableau each month, or both.

**Final Thoughts**

This architecture ensures:

* **Historical data** is safely stored in two places:
  1. **Ref File** for reference data that changes monthly but you may want to keep prior months for context.
  2. **Master File** for the final monthly “Main2.”
* Each month’s input files (Oracle, CKs, and Tableau pulls) feed into your reference file updates and final analysis.
* The final output is appended into a single “Master File,” so you can always see a month-by-month “Main2” in one place.

You can refine or adjust as your processes evolve. The approach is flexible enough for the complexity you’ve described (unassigned distributions, GAP, forecast vs. actual, etc.) but still keeps the flow consistent month-to-month.