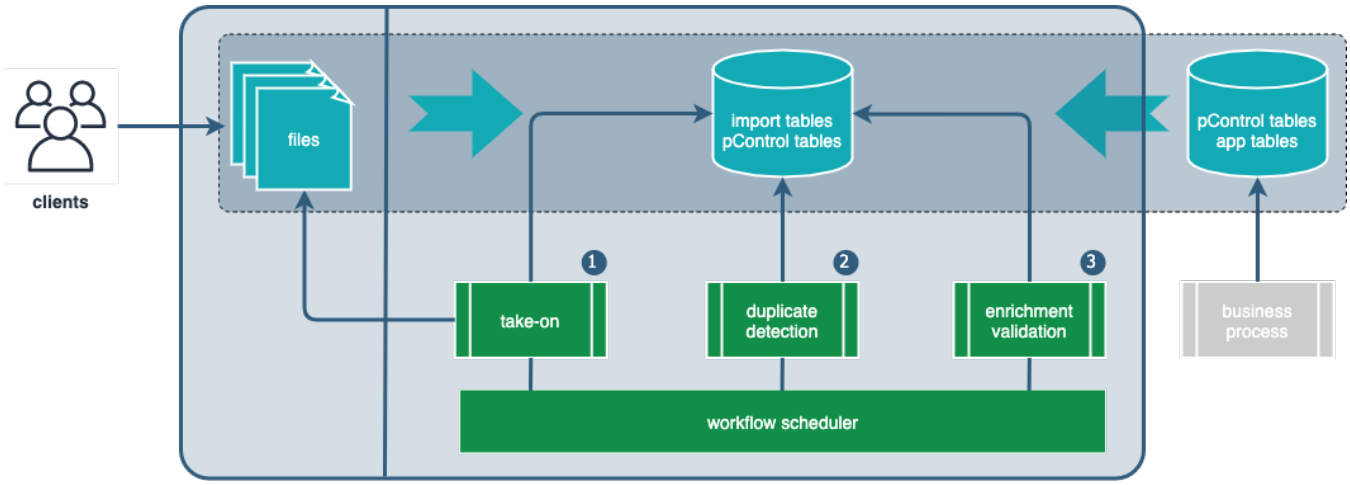
Processed Data

> *... describing common procedures on processed data (data management)*

workflow management system

- workflow management system builds, monitors, and manages the ETL workflow (Data Manager)
    1. clients upload raw data into an external layer (Landing Zone), which are data usually in the CSV and XLS formats
    2. workflow scheduler (Data Poller) schedules individual pipelines of the ETL workflow
    3. rule-based system (Rules Engine) performs transform operations, which are domain specific calculations that ensure data integrity requirements
    4. the result of ETL workflow are data ready for an export by business processes (NAV Validation, etc.)
    5. this process runs again if client uploads new raw data (this can happen many times during the day)
- the status of ETL workflow changes usually with the below order (p_status data field in import tables)
    1. ? (Take on is not completed)
    2. L (Loaded)
    3. I, U (manually inserted or updated)
    4. P (in progress)
    5. Y (when there is error)
    6. G (pending error)
    7. V (validated) and X (fail / user decision to exclude) or Z (fail / user decision to mark data as error)
        - X and Z status is decided by user, who can mark it in a checkbox in UI (predefined attribute of 'set to exclude records that fail validation')
- results of ETL workflow are stored in the core database, which has import and pControl tables



1. **take-on**
    - the system checks for the arrival of client data (transport type: file) and triggers the take-on process
    - the system executes the take-on process to load data into the designated staging tables (import tables)
    - data records are loaded into import tables and are assigned a status, either via automated process or manual process
        - take-on process is an automated process, which results into loaded records ("L" status)
        - user process is a manual process (users create or update rows), which results into inserted records ("I" status) or updated records ("U" status)
            - import table has an attribute to allow the manual process by users
            - this attribute is named as "editable table column" (if false, user cannot manually insert or update data)
    - as the result, the import table is updated with a status result (p_status with 'V' for validated or 'X' for fail)
        - the process of loading data is described in pControl table (P135)
        - the "p135_import_register" table tracks updates and invokes the "ss_import_register" workflow
    - **fund valuation oversight** has the below *import* tables for the take-on process
        - u100_investment_holdings
        - u101_external_unit_prices
        - u102_cash_asset
        - u103_pending_cash
        - u104_trial_balance
        - u107_price_adjustment
        - u108_units_on_issue
        - u109_cash_flow
        - u110_cashflows_and_orders
        - u113_pending_settlements
        - u175_security_prices
        - u180_security
        - u200_asset_transactions
        - U210_corporate_action

        - u224_corporate_actions
        - u280_benchmark
2. **duplicate detection**
    - the system identifies data duplications in staging tables (import tables)
    - the system marks duplicated records as duplicates (status: "D" for duplicated records instead of "L" for loaded records)
        - system can also delete the identified duplicated records (row is removed from the import tables)
        - the actions (and actions for specific rows) after duplicates are found can be set via attribute setting
    - data records are loaded into core tables from staging tables with such statuses (still the same import table, but with updated statuses)
3. **enrichment validation**
    - the system triggers the "enrichment validation" process for the selected data
        - this process include an enrichment of data and validation of data content
    - enrichment validations are defined in static tables
        - l281_enrich_validate_details
        - l282_enrichment_details
    - data records that have been validated are updated in core tables (status: "V" for validated records)
    - the results of ETL workflow are described in import tables and pControl tables (P135 and P183)
        - the "p183_load_monitor" table stores all the ETL results
            - it is used to find out the details of ETL results in import tables
            - it can be also used for monitoring purposes, e.g. we can see this monitoring data in UI / Swing menu with the ASM (Attention Status Monitor)
    - the changes in the core and staging database (import tables) are stored in the "t950_audit_log" table
        - we are assuming that such audit history can be used to reconstruct the data lineage for changes in import tables
    - business processes are invoked when data are populated in the p183 table
        - p183 table is used to wake funds up if Data Manager is waiting for data (using the "ss_up_load_data" workflow)