**Estimation Methodology for Solution Development Using Rule and Control Books**

**Abstract**

This paper develops a structured methodology for estimating the time required to create solutions involving rule books and control books for data adjustment processes. The estimation framework considers inputs from domain experts, the complexity of the process, and detailed operations required for data transformation and validation. The methodology provides a systematic approach to time estimation by breaking down the process into manageable components and assigning time estimates based on defined complexity parameters.

**1. Introduction**

In data processing and adjustment workflows, creating solutions for automating manual processes involves detailed rule definitions and control frameworks. These solutions typically comprise a rule book that outlines data transformation rules and control books to validate data integrity at both pre- and post-transformation stages. Estimating the time required to design and implement such solutions is a critical step to plan resources and manage stakeholder expectations effectively. This paper outlines a methodology to estimate development time based on the complexity and nature of tasks involved.

**2. Objective**

The objective of this paper is to define a methodology for estimating the time required to:

1. Gather requirements and document the process.
2. Write the rules in the rule book, including data enrichment, filtering, aggregation, and calculations.
3. Develop pre- and post-control books for data validation and reconciliation.
4. Account for supporting activities such as testing, defect resolution, deployment, user training, and documentation.
5. Quantify the impact of key complexity parameters on the time estimation process.

**3. Methodology**

**3.1 Process Overview**

1. **Requirement Gathering**:
   * Domain experts manually performing the data adjustment process will provide detailed explanations of the workflow.
   * Inputs include:
     + Input files used (transaction data, mapping files, filters, etc.).
     + Specific rules for copying, enriching, filtering, and calculating data.
     + Allocation and aggregation rules to create the final output.
2. **Rule Book Creation**:
   * Documenting data processing rules, such as:
     + Identifying input columns for copying.
     + Lookup operations for data enrichment.
     + Filters applied for specific conditions.
     + Allocation and proportion rules.
     + Aggregation and transformation steps.
3. **Control Book Creation**:
   * **Pre-Control Book**:
     + Validations on raw input data (e.g., null value checks, specific mappings, or range validations).
   * **Post-Control Book**:
     + Validations and reconciliations on output data, including:
       - Comparing aggregated outputs with input data.
       - Verifying consolidated values.
4. **Additional Activities**:
   * Testing the solution using test cases.
   * Addressing any defects identified during testing.
   * Deploying the final solution.
   * Conducting user training sessions.
   * Creating documentation, including user manuals and process flow documents.

**3.2 Complexity Parameters**

The following parameters are used to estimate the complexity of the process:

1. **Number of Data Enrichment Files**:
   * The number of input files used for lookup, filtering, or enrichment.
2. **Allocation Rules**:
   * Whether allocation or proportion logic is required.
3. **Aggregation Requirements**:
   * Steps needed to consolidate data for the final output.
4. **Control Requirements**:
   * Number and type of validations/reconciliations in pre- and post-control books.
5. **Mapping Tables**:
   * The number of mapping tables used for lookups and enrichment.
6. **Pivoting and Unpivoting Columns**:
   * The requirement to pivot or unpivot data columns based on specific categorical columns.
   * Time estimation: Each column pivoted or unpivoted will take approximately **15 minutes**.

**3.3 Time Estimation Framework**

Each activity is assigned a base time estimate for the technical work (e.g., column copying, lookups, etc.). Additionally, a percentage-based allocation is applied to account for requirement gathering, defect resolution, deployment, user training, testing, and documentation. These supporting activities are distributed as a proportion of the core technical work, ensuring they are reflected in the final time estimate.

**Base Time Estimates for Core Technical Work**

The base time estimates for core activities are:

|  |  |  |  |
| --- | --- | --- | --- |
| **Activity** | **Unit of Work** | **Base Time (minutes)** | **Description** |
| Column Copying | Per column | 10 | Simple data copying from one file to another. |
| Lookup for Enrichment | Per column | 20 | Includes mapping table lookups. |
| Filtering | Per condition | 15 | Filtering data based on specific criteria. |
| Allocation | Per rule | 25 | Allocation or proportion calculations. |
| Aggregation for Uploader | Per dataset | 40 | Consolidation of data into final format. |
| Reconciliation (Post-Control) | Per control point | 30 | Data comparison and validation. |
| Pivoting/Unpivoting Columns | Per column | 15 | Reformatting data structure. |

**Percentage Allocation for Supporting Activities**

The percentages below represent the proportion of each supporting activity as part of the total time for a given core technical task:

| **Supporting Activity** | **Percentage of Total Time** |
| --- | --- |
| Requirement Gathering | 10% |
| Testing | 15% |
| Defect Resolution | 10% |
| Deployment | 5% |
| User Training | 5% |
| Documentation | 10% |

**Example Calculation**

Let’s calculate the total time for a single **lookup for enrichment** activity, using the base time and percentages:

1. **Base Time for Lookup for Enrichment** = 20 minutes.
2. Supporting activity percentages applied:
   * Requirement Gathering = ( 20 \* 10% = 2 , \text{minutes} ).
   * Testing = ( 20 \* 15% = 3 , \text{minutes} ).
   * Defect Resolution = ( 20 \* 10% = 2 , \text{minutes} ).
   * Deployment = ( 20 \* 5% = 1 , \text{minute} ).
   * User Training = ( 20 \* 5% = 1 , \text{minute} ).
   * Documentation = ( 20 \* 10% = 2 , \text{minutes} ).

**Total Time for Lookup for Enrichment** = ( 20 + 2 + 3 + 2 + 1 + 1 + 2 = 31 , \text{minutes} ).

**3.4 Consolidated Time Table**

Using the base time estimates and the revised **55% allocation** for supporting activities, the total time for each core activity is recalculated as follows:

| **Activity** | **Base Time (minutes)** | **Supporting Activities (Additional 55%)** | **Total Time per Unit (minutes)** |
| --- | --- | --- | --- |
| Column Copying | 10 | 10×55%=5.510 \times 55% = 5.5 | 15.5 |
| Lookup for Enrichment | 20 | 20×55%=1120 \times 55% = 11 | 31 |
| Filtering | 15 | 15×55%=8.2515 \times 55% = 8.25 | 23.25 |
| Allocation | 25 | 25×55%=13.7525 \times 55% = 13.75 | 38.75 |
| Aggregation for Uploader | 40 | 40×55%=2240 \times 55% = 22 | 62 |
| Reconciliation (Post-Control) | 30 | 30×55%=16.530 \times 55% = 16.5 | 46.5 |
| Pivoting/Unpivoting Columns | 15 | 15×55%=8.2515 \times 55% = 8.25 | 23.25 |

**Explanation**

The supporting activities now account for 55% of the total time per unit. The **Total Time per Unit** is calculated as follows:

Total Time per Unit=Base Time+(Base Time×55%)\text{Total Time per Unit} = \text{Base Time} + (\text{Base Time} \times 55\%)

For example:

* **Column Copying**: 10+(10×0.55)=15.5 minutes10 + (10 \times 0.55) = 15.5 \, \text{minutes}
* **Lookup for Enrichment**: 20+(20×0.55)=31 minutes20 + (20 \times 0.55) = 31 \, \text{minutes}

**4. Assumptions**

1. Percentages are based on past project experience and can be adjusted for specific scenarios.
2. The core technical work estimates are accurate and represent typical complexity levels.
3. Supporting activities are distributed proportionally and are essential for successful delivery.

**Benefits of Percentage-Based Allocation**

1. Ensures that all supporting activities (e.g., testing, documentation) are accounted for without separately estimating them.
2. Provides a unified metric for each type of work (e.g., "per column" or "per rule"), simplifying overall calculations.
3. Allows scalability: the percentages can be adjusted dynamically based on project size or complexity.