|  |  |
| --- | --- |
| FSS Logo |  |

**Dự án R&D**

**Công cụ để trích xuất thông tin mapping từ tài liệu**

**TÀI LIỆU ĐẶC TẢ YÊU CẦU**

**Hà Nội, 08 – 2024**MỤC LỤC

[1 GIỚI THIỆU 1](#_Toc175564602)

[1.1 Mục đích tài liệu 1](#_Toc175564603)

[1.2 Phạm vi tài liệu 1](#_Toc175564604)

[1.3 Khái niệm, thuật ngữ, các từ viết tắt 1](#_Toc175564605)

[2 CÔNG CỤ TRÍCH XUẤT THÔNG TIN MAPPING TỪ TÀI LIỆU 2](#_Toc175564606)

[2.1 Đặt vấn đề 2](#_Toc175564607)

[2.2 Công nghệ sử dụng 2](#_Toc175564608)

[2.3 Mô tả yêu cầu 3](#_Toc175564609)

[2.3.1 Phase 1 3](#_Toc175564610)

# GIỚI THIỆU

## Mục đích tài liệu

Tài liệu này được dùng làm tài liệu đặc tả phân tích yêu cầu công cụ cần phát triển.

## Phạm vi tài liệu

Tài liệu bao gồm:

* Mục đích của công cụ
* Đặc tả yêu cầu của công cụ

## Khái niệm, thuật ngữ, các từ viết tắt

|  |  |
| --- | --- |
| Từ viết tắt | Từ viết đầy đủ |
| FSS | Công ty Giải pháp phần mềm Tài chính |
| DWH | Data warehouse |
| TLPTYC | Tài liệu phân tích yêu cầu |
| Job Data Stage | Các mã nguồn job được xuất ra từ công cụ Data Stage |
|  |  |

# CÔNG CỤ TRÍCH XUẤT THÔNG TIN MAPPING TỪ TÀI LIỆU

## Đặt vấn đề

Trong quá trình triển khai các dự án về dữ liệu, có một số công việc sau đòi hỏi nhiều nguồn lực/thời để rà soát như:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Đầu vào** | | **Mục đích** | | | | **Công việc thực hiện** | |
| **Loại công việc** | Đặc tả mapping | Mã nguồn | Đánh giá khối lượng công việc cần thực hiện | Đánh giá rủi ro tác động hệ thống khi triển khai | Chuẩn bị môi trường dữ liệu đầy đủ | Hiểu ý nghĩa của các bảng sử dụng | Rà soát các job ảnh hưởng khi thay đổi bảng, nguồn | Thống kê các job liên quan |
| Dự án chuyển đổi nguồn dữ liệu | x | x | x | x |  |  | x | x |
| CR đổi nguồn dữ liệu, bảng dữ liệu | x | x | x | x |  |  |  |  |
| Phân tích yêu cầu | x |  |  |  | x | x | x |  |
|  |  |  |  |  |  |  |  |  |

Để giảm thiểu thời gian rà soát, mong muốn phát triển một công cụ để có thể trích xuất nhanh được việc này.

Công cụ cho phép trích xuất thông tin mapping (tối thiểu là tên bảng, tên cột) từ các tài liệu như: tài liệu word đặc tả yêu cầu, tài liệu mapping excel, Mã nguồn chương trình (ví dụ job DataStage)

## Công nghệ sử dụng

* Sử dụng Python
* Lý do: cài đặt đơn giản, dễ dàng triển khai lên môi trường của đối tác nếu cần

## Mô tả yêu cầu

Chia làm hai phase, trong phase đầu tiên yêu cầu như sau:

### Phase 1

|  |  |  |
| --- | --- | --- |
| **STT** | **Đầu vào** | **Đầu ra** |
| 1 | * File work mô tả mapping của khách hàng, trong đó có các câu lệnh SQL | * Danh sách bảng sử dụng * Danh sách bảng/cột sử dụng * Danh sách bảng/cột sử dụng và đầu mục tương ứng   Mỗi mục thành 1 file excel output |
| 2 | * Bộ job Data Stage khách hàng cung cấp (mỗi job thành 1 file) | * Danh sách bảng sử dụng * Danh sách bảng/cột sử dụng * Danh sách bảng/cột sử dụng và job tương ứng |