**UNIVERSIDAD DE TARAPACÁ**

**FACULTAD DE INGENIERÍA**

**INGENIERÍA CIVIL EN COMPUTACIÓN E INFORMÁTICA**

**ARICA – CHILE**



**Informe de avance 2 “Herramienta de automatización de validación de consistencia e integridad de campos”**

 **Equipo de Desarrollo:**

**- Sebastian Lukich Aste**

**Empresa o Unidad:**

**Digital Social Change SPA**

**Curso: Proyecto IV ICCI**

**Profesor: Diego Aracena Pizarro**

Arica, 24 de octubre del 2022

Contenido

[**1.** **Definición del proyecto** 5](#_Toc117515743)

[1.1. Contexto 5](#_Toc117515744)

[1.2. Problema 5](#_Toc117515745)

[1.3. Solución 6](#_Toc117515746)

[**2.** **Requisitos del sistema** 6](#_Toc117515747)

[2.1. Requisitos funcionales 6](#_Toc117515748)

[2.2. Requisitos no funcionales 8](#_Toc117515749)

[**3.** **Alcances del ambiente de software** 8](#_Toc117515750)

[**4.** **Subsistemas** 8](#_Toc117515751)

[**5.** **Objetivos** 9](#_Toc117515752)

[5.1. Objetivo general 9](#_Toc117515753)

[5.2. Objetivos específicos 9](#_Toc117515754)

[**6.** **Alcance** 9](#_Toc117515755)

[**7.** **Lista de actividades** 10](#_Toc117515756)

[7.1. Etapas y actividades dentro del desarrollo del proyecto 10](#_Toc117515757)

[7.2. Carta Gantt 11](#_Toc117515758)

[**8.** **Diagramas del proyecto** 12](#_Toc117515759)

[8.1. Diagrama de sistema del proyecto (alto nivel) 12](#_Toc117515760)

[8.2. Diagrama de contexto: sistema y subsistemas 12](#_Toc117515761)

[**9.** **Herramientas a utilizar** 12](#_Toc117515762)

[9.1. Pentaho Data Integration (PDI) 12](#_Toc117515763)

[9.2. Saiku Analytics 13](#_Toc117515764)

[9.3. Pentaho CDE 13](#_Toc117515765)

[**10.** **Aspectos iniciales del front-end** 13](#_Toc117515766)

[10.1. Ejemplo de gráfico de torta mostrando el porcentaje de error de cada campo respecto al total de errores en toda la “[Tabla 1]” 13](#_Toc117515767)

[10.2. Ejemplo de gráfico de barras mostrando el total de errores encontrados por sistema en decremento con el aumento de iteraciones 14](#_Toc117515768)

[10.3. Ejemplo de gráficos de barras donde la cantidad de errores disminuye con las iteraciones por tabla dentro de un “[Sistema 1]” 14](#_Toc117515769)

[**11.** **Business Process Management (BPM)** 15](#_Toc117515770)

[**12.** **Alcance de acuerdo a las herramientas** 15](#_Toc117515771)

[**13.** **Repositorio de Github** 16](#_Toc117515772)

[**14.** **Rep** 17](#_Toc117515773)

[**15.** **Almacenamiento y modelamiento de los datos** 17](#_Toc117515774)

[**16.** **Conclusiones** 19](#_Toc117515775)

[**17.** **Bibliografía** 20](#_Toc117515776)

**Introducción**

La empresa Digital Social Change Spa se dedica a la transformación digital de empresas principalmente en la asesoría e implementación de automatización relacionadas al procesamiento masivo de datos.

Digital Social Change Spa requiere implementar una herramienta de apoyo/complementaria para el proceso de validación de migración de datos. Esta herramienta es requerida puesto que al realizar una migración la cantidad de campos que se deben validar es muy alta para poder realizar una validación manual. Dado lo anterior se requiere diseñar y construir 2 modulos:

1. Proceso de Captura de Información (Back-End 20%).

Capturar el detalle de los errores.

1. Proceso de Seguimiento (Front-End 80%)

Visualización por iteraciones de los errores.

Para lo anterior se utilizará tecnologías de inteligencia de negocios como los cubos multidimensionales y los visores OLAP.

## **Definición del proyecto**

### Contexto

La captura de información es el proceso en el cual se sacan los datos de error de una tabla que ha sido migrada de un sistema a otro. Por ejemplo, si así fuere que la tabla tiene un campo “Sexo”, que después de la migración terminó con todos sus registros en ese campo con los valores “M”, “F” y “5”, entonces el detalle de ese error sería que en el campo “Sexo” de la tabla en cuestión hay un valor “5” fuera del conjunto válido de valores. Este detalle se registraría en otro conjunto de datos que suele estar actualizándose periódicamente por iteración.

Cuando el proceso no está automatizado, todos los recursos humanos disponibles en el área deben dedicarse a hacer consultas SQL a mano sobre cada campo de la tabla migrada. Lo que asciende a varios días de trabajo en el proyecto y más el hecho de que hay varias tablas por sistema y que se produce un informe de los errores que se encontraron por tabla puede derivar en que la tarea neta se haga inviable.

El seguimiento de la información es el proceso que se trata de, mediante gráficos, tablas, diagramas y ayudas visuales; dar seguimiento a las estadísticas de error que dejó como resultado la captura de información de error. Las estadísticas evolucionan con el avance de cada iteración que tiene el proyecto. Y las visualizaciones siguen un orden de navegación jerárquico, es decir, cada entidad mayor permite visualizar en mayor detalle sus estadísticas pasando a la visualización más específica de una de sus sub- entidades. Por ejemplo, la visualización de un sistema pasa a la visualización de uno de sus subsistemas que lo conforman.

### Problema

La empresa Digital Social Change SPA no puede hacerse cargo, simplemente con recursos humanos, del trabajo de captura de la información para todos los campos del total de tablas de cada sistema migrado. Necesita de herramientas automatizadas, que apoyen en los procesos tanto de captura de información de error, en lo que respecta a consistencia e integridad de los campos de sus tablas migradas, así como de seguimiento de la información de error de estos procesos de captura.

### Solución

Implementar a través de software de inteligencia de negocios una herramienta con dos módulos. Uno que automatice el proceso de captura de información de error, en lo que respecta a consistencia e integridad de los campos de las tablas migradas, y otro que, como complemento al anterior, haga un seguimiento que se pueda visualizar de la evolución por iteraciones de la información capturada por el proceso anterior.

## **Requisitos del sistema**

### 2.1. Requisitos funcionales

|  |
| --- |
| Requisitos funcionales |
| Código | Descripción |
| RF-01 | El sistema deberá mostrar un diagrama con los nombres de los sistemas seleccionados para la importación incluyendo para cada uno la visualización de los datos: cantidad de campos migrados correctamente, cantidad de campos con error en la migración, cantidad de campos totales, porcentaje de campos migrados correctamente, porcentaje de campos con error. |
| RF-02 | El sistema deberá mostrar un diagrama con los nombres de los subsistemas seleccionados para la importación incluyendo para cada uno la visualización de los datos: cantidad de campos migrados correctamente, cantidad de campos con error en la migración, cantidad de campos totales, porcentaje de campos migrados correctamente, porcentaje de campos con error. |
| RF-03 | El sistema deberá mostrar un diagrama con los nombres de las tablas seleccionadas para la importación incluyendo para cada una la visualización de los datos: cantidad de campos migrados correctamente, cantidad de campos con error en la migración, cantidad de campos totales, porcentaje de campos migrados correctamente, porcentaje de campos con error, cantidad de registros migrados correctamente, cantidad de registro con error en la migración, cantidad de registros totales, porcentaje de registros migrados correctamente, porcentaje de registros con error. |
| RF-04 | El sistema deberá mostrar un diagrama con los nombres de los campos seleccionados para la importación incluyendo para cada una la visualización de los datos: cantidad de registros migrados correctamente, cantidad de registro con error en la migración, cantidad de registros totales, porcentaje de registros migrados correctamente, porcentaje de registros con error. |
| RF-05 | El sistema deberá mostrar un diagrama con los identificadores de los tipos de error genérico que aparecen a lo largo de los registros de los sistemas seleccionados para la importación incluyendo para cada uno la visualización de los datos: cantidad de registro con error en la migración. |
| RF-06 | El sistema deberá mostrar un diagrama con las descripciones de los tipos de error detallado que aparecen a lo largo de los registros de los sistemas seleccionados para la importación incluyendo para cada una la visualización de los datos: cantidad de registro con error en la migración. |
| RF-07 | El sistema deberá mostrar un diagrama con las fechas de las iteraciones en las que se efectúa una mejora en el algoritmo de migración de campos de algún sistema, incluyendo para cada una la visualización de los datos: cantidad de campos migrados correctamente, cantidad de campos con error en la migración, cantidad de campos totales, porcentaje de campos migrados correctamente, porcentaje de campos con error, cantidad de registros migrados correctamente, cantidad de registro con error en la migración, cantidad de registros totales, porcentaje de registros migrados correctamente, porcentaje de registros con error. |
| RF-08 | El sistema deberá mostrar un diagrama con las horas de las iteraciones que ocurrieron en una fecha en las que se efectúa una mejora en el algoritmo de migración de campos de algún sistema, incluyendo para cada una la visualización de los datos: cantidad de campos migrados correctamente, cantidad de campos con error en la migración, cantidad de campos totales, porcentaje de campos migrados correctamente, porcentaje de campos con error, cantidad de registros migrados correctamente, cantidad de registro con error en la migración, cantidad de registros totales, porcentaje de registros migrados correctamente, porcentaje de registros con error. |
| RF-09 | El sistema deberá mostrar un diagrama con los números de orden de las iteraciones en las que se efectúa una mejora en el algoritmo de migración de campos de algún sistema, incluyendo para cada una la visualización de los datos: cantidad de campos migrados correctamente, cantidad de campos con error en la migración, cantidad de campos totales, porcentaje de campos migrados correctamente, porcentaje de campos con error, cantidad de registros migrados correctamente, cantidad de registro con error en la migración, cantidad de registros totales, porcentaje de registros migrados correctamente, porcentaje de registros con error. |
| RF-10 | El sistema permitirá traslados de un diagrama a otro, siempre que cumplan con la relación de estar en un nivel de abstracción inferior a la categoría principal del primer diagrama. Como pasar del diagrama sobre “sistemas” al diagrama de la categoría inferior “subsistemas”, a partir de la elección del usuario de qué sistema del conjunto se desea observar en más detalle. |
| RF-11 | El sistema permitirá filtros en los diagramas de los requisitos funcionales de 1 al 10 que permitirán la personalización en la navegación del usuario por las visualizaciones de los datos que dispone el sistema. |
| RF-12 | El sistema automatizará la captura de la información de errores contenida en las tablas que pasaron por un proceso de migración. |

### 2.2. Requisitos no funcionales

|  |
| --- |
| Requisitos no funcionales |
| Código | Descripción |
| RNF-01 | El sistema deberá utilizar una base de datos para almacenar los datos sobre los errores de la validación de registros de sistemas destinados a migración. |
| RNF-02 | El sistema deberá utilizar una herramienta de visor OLAP del grupo Pentaho Saiku o STPivot de Stratebi. |

## **Alcances del ambiente de software**

* Automatización del proceso de captura del detalle de los errores.
* Visibilidad de estadísticas de errores de migración en distinto nivel de detalle.
* Visibilidad de estadísticas de errores de migración en distinta representación gráfica.

## **Subsistemas**

* Sistema de captura de información
* Sistema de seguimiento
* Sistema de base de datos de errores capturados

## **Objetivos**

### Objetivo general

Producir un módulo para un proceso automático de captura de información de errores de migración y otro que sirva de tablero para la visualización del seguimiento de los errores capturados en las migraciones de sistemas que ejecuta la empresa Digital Social Change.

### Objetivos específicos

1. Implementar un software de captura automatizada de datos de errores producidos en las migraciones de sistemas migrados por Digital Social Change Spa.
2. Conocer la estructura de almacenamiento de los datos de los errores registrados por el proceso automático de validación de campos de tablas migradas con la que funcionará el tablero de estadísticas de evolución de estos mismos errores.
3. Escoger el diseño de las vistas con las que se visualizará la evolución de los errores en el tablero de estadísticas de evolución de errores generados por las validaciones de los campos de las tablas de los distintos sistemas y subsistemas sometidas a proceso de migración.
4. Implementar por software un tablero de estadísticas que permita ver la evolución de los errores generados por las validaciones de los campos de las tablas de los distintos sistemas y subsistemas que fueron sometidas a proceso de migración.

## **Alcance**

Proveer a la empresa Digital Social Change Spa de un software que automatice las tareas de hacer las validaciones a mano campo por campo para las tablas de los sistemas que pasan por migración en la empresa, y de un software que muestre el avance gráfica y dinámicamente de estas validaciones conforme va mejorando el método de migración de sistemas.

## **Lista de actividades**

### Etapas y actividades dentro del desarrollo del proyecto

El proyecto constará de las siguientes etapas que a su vez se subdividen en actividades dentro de la planificación:

**Inauguración del proyecto:**

“Establecer la metodología de trabajo (acuerdo con cliente)”, “Elegir horarios de reuniones (acuerdo con cliente)“, “Entender el trasfondo de la empresa y su necesidad”.

**Diseño del proyecto:**

“Buscar trabajos relacionados y reportar experiencias anteriores (de la empresa)”, “Establecer listado de requisitos (acuerdo firmado)”, “Hacer diagrama en general del sistema”, “Hacer modelo estrella”, “Hacer diagrama de contexto”, “Hacer diagrama de comunicación/ de secuencia”, “Hacer diagrama colaborativo”.

**Implementación:**

“Hacer pantallazos de primer prototipo”, “Llenado de tablas (estructura reportada por la empresa)”, “Boceto de concepto con Excel”, “Levantar servidor con Pentaho más base de datos”, “Desarrollar primer módulo con Pentaho”, “Desarrollar segundo módulo con Saiku”.

**Pruebas:**

“Refinamiento sucesivo de módulo de Pentaho”, “Refinamiento sucesivo de módulo de Saiku”.

**Cierre del proyecto:**

“Carta de aprobación del cliente”, “Exposición de resultados”.

### Carta Gantt



*Figura 1.*

## **Diagramas del proyecto**

### Diagrama de sistema del proyecto (alto nivel)



*Figura 2.*

### Diagrama de contexto: sistema y subsistemas



*Figura 3.*

## **Herramientas a utilizar**

### Pentaho Data Integration (PDI)

Es una de las herramientas o componentes de Pentaho Suite que permite que se utilicen técnicas ETL, es decir, poder implementar procesos deextracción, transformación y carga de datos. PDI ofrece también datos analíticos muy precisos, eliminando las complejidades involucradas en la codificación al proporcionar bibliotecas en profundidad para el mismo.

### Saiku Analytics

Saiku permite a los usuarios comerciales explorar fuentes de datos complejas, utilizando una interfaz familiar de arrastrar y soltar y una terminología comercial fácil de entender, todo dentro de un navegador.

### Pentaho CDE

El Community Dashboard Editor (CDE) y sus tecnologías subyacentes (CDF, ​​CDA y CCC) permiten un rápido desarrollo e implementación de los Dashboards Pentaho CTools. La herramienta CDE se creó para simplificar los procesos de creación, diseño y representación de los paneles de CTools. Es una herramienta poderosa y completa, que integra a la perfección la interfaz de usuario con fuentes de datos y componentes personalizados. *Esta herramienta se requiere investigar ya que no está en los requerimientos, pero parece el caso que permitiría un resultado más adecuado para el dashboard que se requiere.*

## **Aspectos iniciales del front-end**

### Ejemplo de gráfico de torta mostrando el porcentaje de error de cada campo respecto al total de errores en toda la “[Tabla 1]”



*Figura 4.*

### Ejemplo de gráfico de barras mostrando el total de errores encontrados por sistema en decremento con el aumento de iteraciones



*Figura 5.*

### Ejemplo de gráficos de barras donde la cantidad de errores disminuye con las iteraciones por tabla dentro de un “[Sistema 1]”



*Figura 6.*

## **Business Process Management (BPM)**

El diagrama describe el flujo donde el Módulo de Captura de la herramienta captura los errores de las tablas de migración una vez que el usuario los provee, añade las entradas respectivas al cubo multidimensional de datamart y luego muestra la evolución respectiva de los errores de migración a través del gráfico dinámico con el que interactúa el usuario y obtiene la información que buscaba.



*Figura 7.*

## **Alcance de acuerdo a las herramientas**

Automatizar la validación de campos de los sistemas de las migraciones utilizando Pentaho Data Integration. Crear un dashboard hecho en Saiku Analytics para mostrar gráfica y dinámicamente la evolución de las validaciones.

## **Repositorio de Github**

El avance de este proyecto se puede encontrar en el repositorio remoto de Github con la siguiente URL: <https://github.com/SebasJak/Herramienta-AVCIC>

Las capturas de pantalla muestran el avance que se tiene en cuanto a archivos hasta el momento además de la organización de carpetas que contiene el repositorio más su descripción.



Figura 8. Página principal del repositorio de Github.



Figura 9. Vista de los archivos/carpetas del modulo-pentaho.



Figura 10. Vista de los archivos de Pentaho (transformations, jobs).

## **Almacenamiento y modelamiento de los datos**

Como lo que se está desarrollando es este proyecto en cuanto a los datos que se van a utilizar para el dashboard y se irán actualizando es un datamart, entonces el almacén para estos datos es naturalmente un cubo multidimensional, que a continuación se describe utilizando un modelo estrella.



Figura 11. Modelo estrella del cubo multidimensional.

## **Conclusiones**

Se concluye con este avance del proyecto que:

* Los mecanismos de comunicación con el cliente como lo fueron las reuniones semanales en esta etapa desempeñan un factor clave no sólo para mostrar el avance del proyecto sino también para establecer los conocimientos en común que tienen tanto cliente como desarrollador y llegar a acuerdos de manera más formal sobre lo que se tiene que llegar a hacer con el proyecto.
* Pentaho presenta una naturaleza más manipulable de lo que muestra la fachada de su interfaz, al dar la opción de crear archivos pentaho desde fuera de la aplicación PDI, por ejemplo, mediante el uso de macros de Excel como se utilizó en este avance de proyecto para las múltiples tablas de la base de datos de ejemplo Northwind. Esto se puede explotar mediante la práctica y el dominio de estas herramientas de automatización como son los macros.
* La diversidad de administradores de bases de datos puede llegar a dificultar el trabajo de agregar nuevas bases de datos al repertorio que uno tenía, por la razón de que los lenguajes de consultas difieren en cierto grado entre sí, unos funcionan con más credenciales que otros, la instalación de uno nuevo que no se utilizaba antes puede requerir parches, etc. Como se vio en este avance con la instalación y uso de Microsoft SQLServer.

## **Bibliografía**

1. Apuntes del profesor Diego Aracena
2. Documentación de Pentaho. <https://help.hitachivantara.com/Documentation>
3. Documentación de Microsoft SQL Server. https://www.microsoft.com/en-us/sql-server/sql-server-downloads