

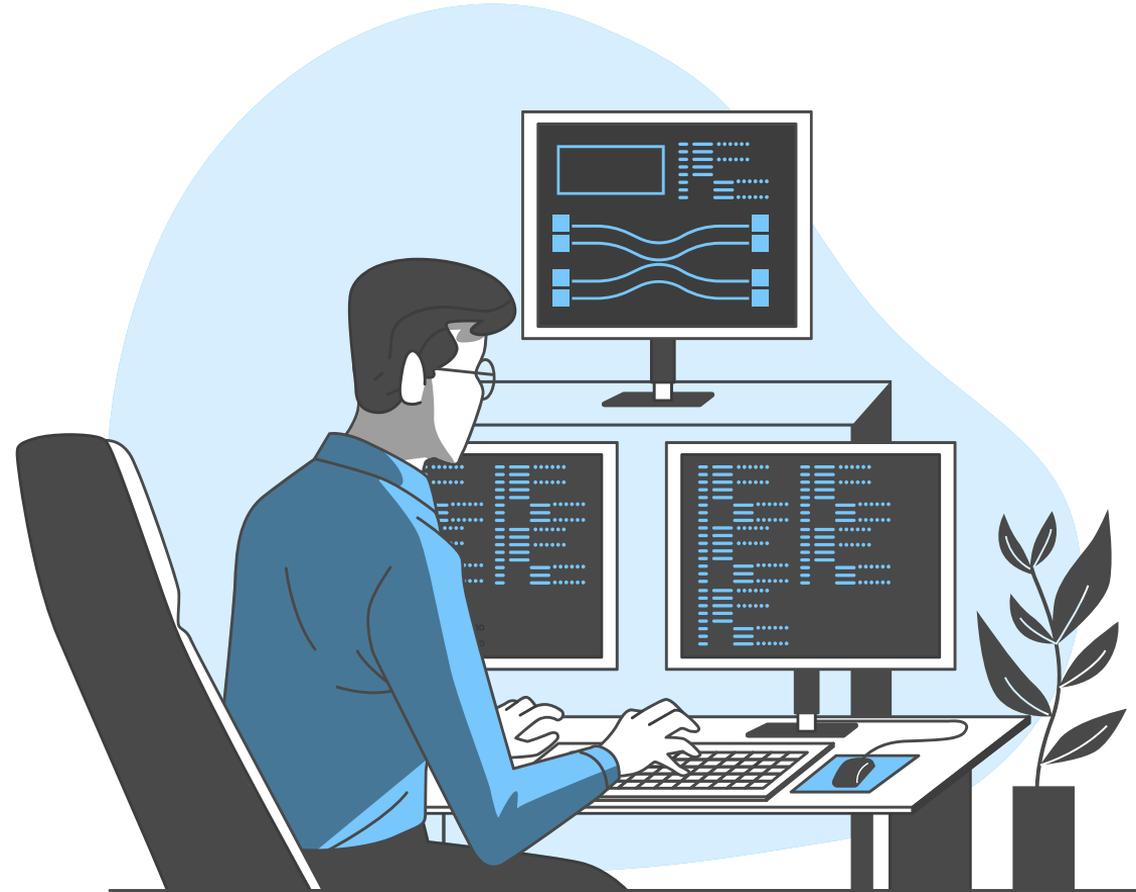
07/11/22 ...

...



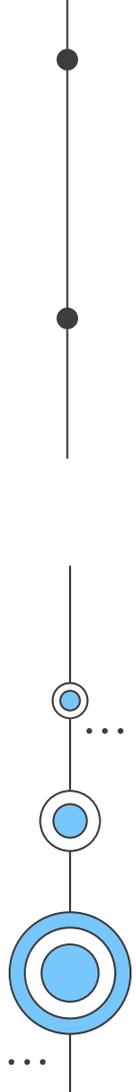
Herramienta de automatización de validación de consistencia e integridad de campos

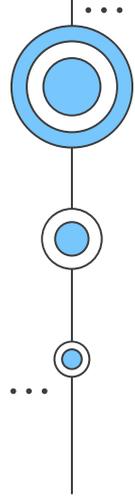
E. Desarrollo: Sebastian Lukich
Empresa o Unidad: Digital Social Change SPA
Profesor: Diego Arcena





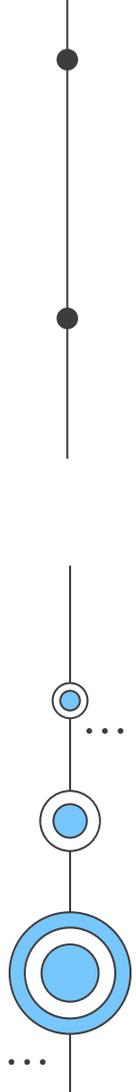
INTRODUCCIÓN



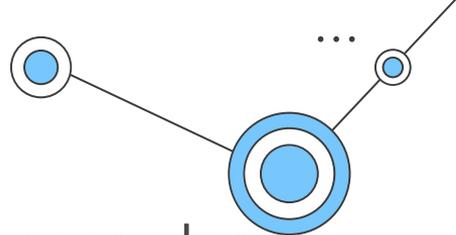


01

Definición del proyecto



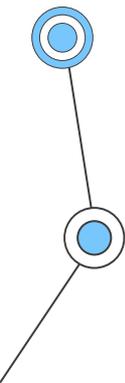
Contexto



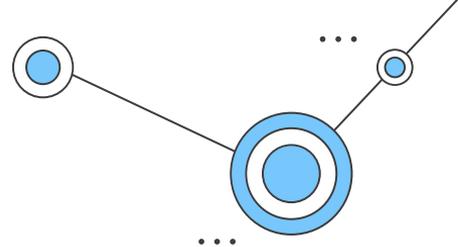
La captura de información es el proceso en el cual se sacan los datos de error de una tabla que ha sido migrada de un sistema a otro. (Ejemplo.)

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.

El seguimiento de la información es el proceso que se trata de dar seguimiento a las estadísticas de error que dejó como resultado la captura de información de error.



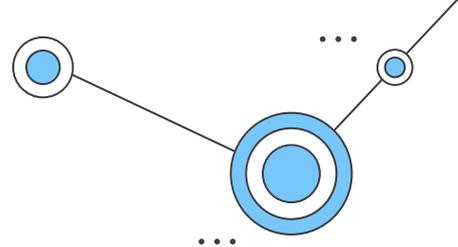
Problema



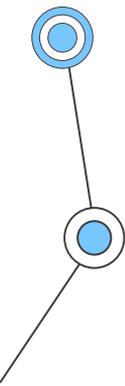
La empresa Digital Social Change SPA no puede hacerse cargo, simplemente con recursos humanos, del trabajo de captura de la información para todos los campos del total de tablas de cada sistema migrado. Necesita de herramientas automatizadas, que apoyen en los procesos tanto de captura de información 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.





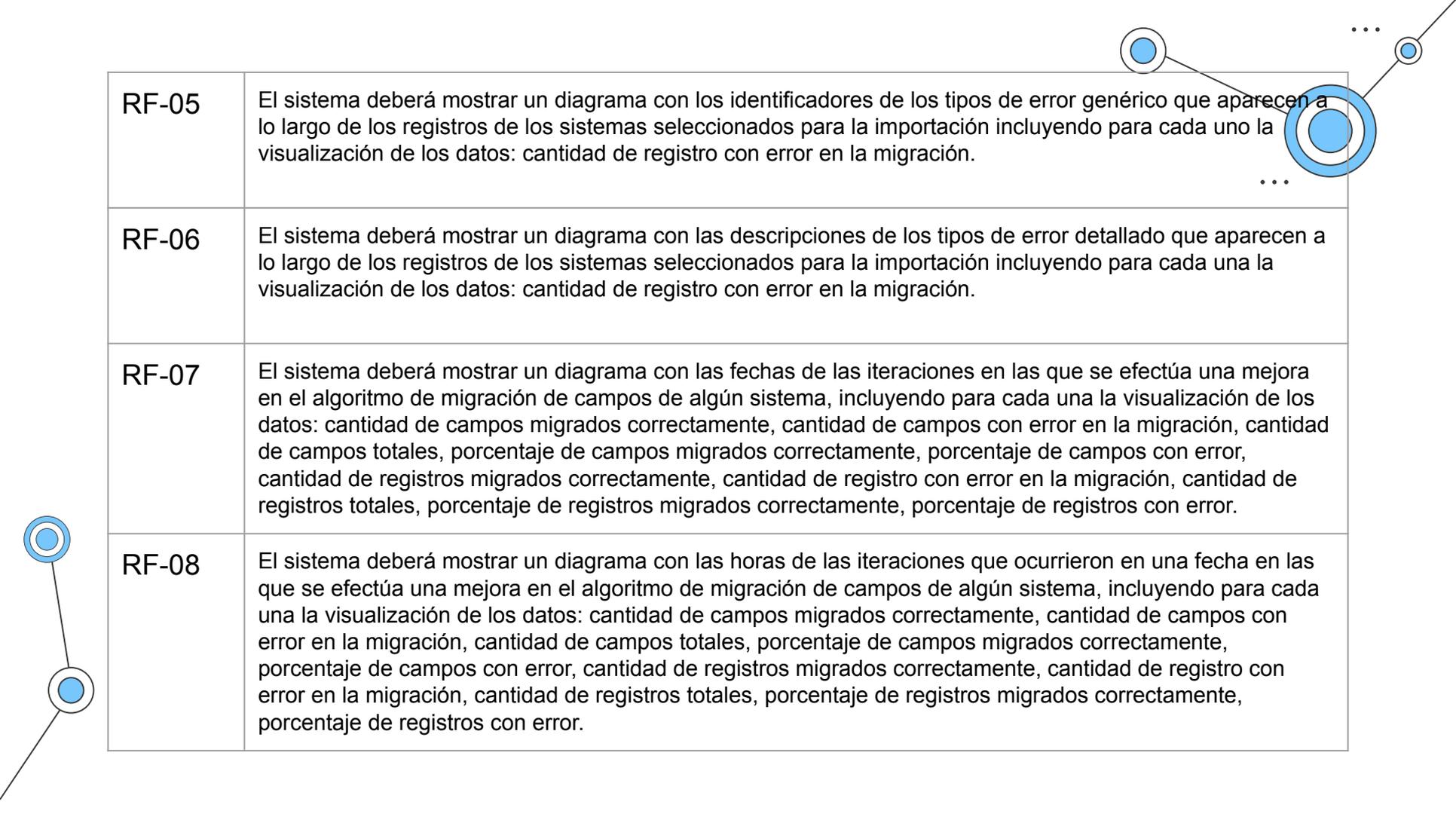
02

Requisitos del sistema

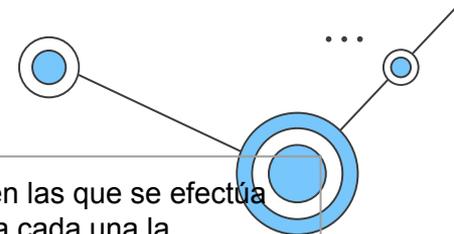


Requisitos funcionales

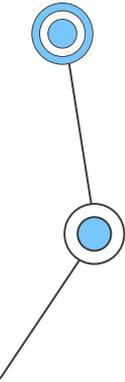
Código	Descripción
RF-01	El sistema deberá mostrar un diagrama con los nombres de los sistemas seleccionados para la importación incluyendo para cada uno la visualización de los datos: cantidad de campos migrados correctamente, cantidad de campos con error en la migración, cantidad de campos totales, 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.

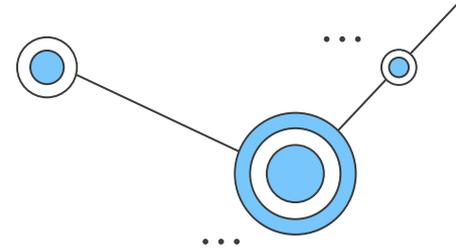


RF-05	El sistema deberá mostrar un diagrama con los identificadores de los tipos de error genérico que aparecen a lo largo de los registros de los sistemas seleccionados para la importación incluyendo para cada uno la visualización de los datos: cantidad de registro con error en la migración. ...
RF-06	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.





Requisitos no funcionales

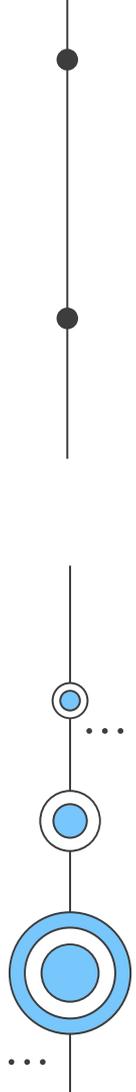
Código	Descripción
RF-01	El sistema deberá utilizar una base de datos para almacenar los datos sobre los errores de la validación de registros de sistemas destinados a migración.
RF-02	El sistema deberá utilizar una herramienta de visor OLAP del grupo Pentaho Saiku o STPivot de Stratebi.

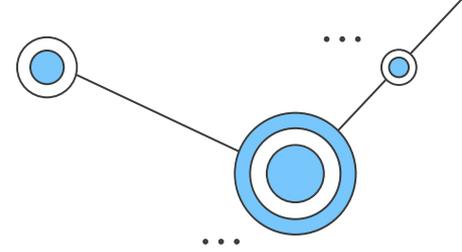




03

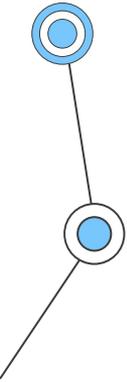
Objetivos



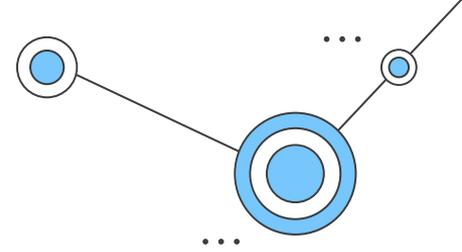


Objetivo general

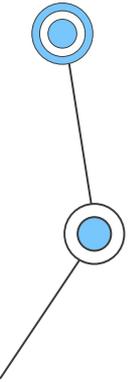
Producir un módulo para un proceso automático de captura de información de errores de migración y otro que sirva de tablero para la visualización del seguimiento de los errores capturados en las migraciones de sistemas que ejecuta la empresa Digital Social Change.



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.



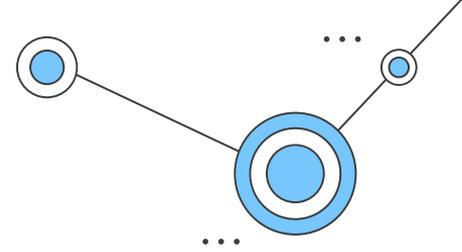


04

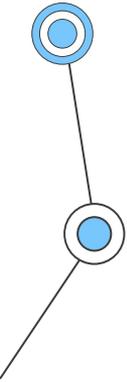
Alcance



Alcance



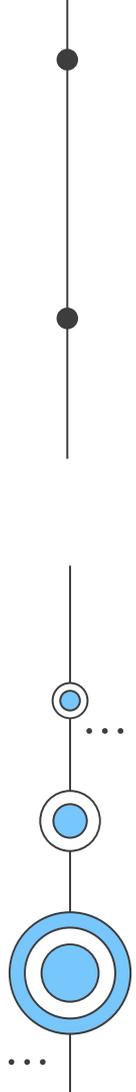
Proveer a la empresa Digital Social Change Spa de un software que automatice las tareas de hacer las validaciones a mano campo por campo para las tablas de los sistemas que pasan por migración en la empresa, y de un software que muestre el avance gráfica y dinámicamente de estas validaciones conforme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





05

Lista de
actividades



Etapas y actividades

- 1) 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".
- 2) 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".
- 3) 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".
- 4) Pruebas: "Refinamiento sucesivo de módulo de Pentaho", "Refinamiento sucesivo de módulo de Saiku".
- 5) Cierre del proyecto: "Carta de aprobación del cliente", "Exposición de resultados".



06

Diagramas del proyecto



Diagrama de sistema del proyecto (alto nivel)

Back-end



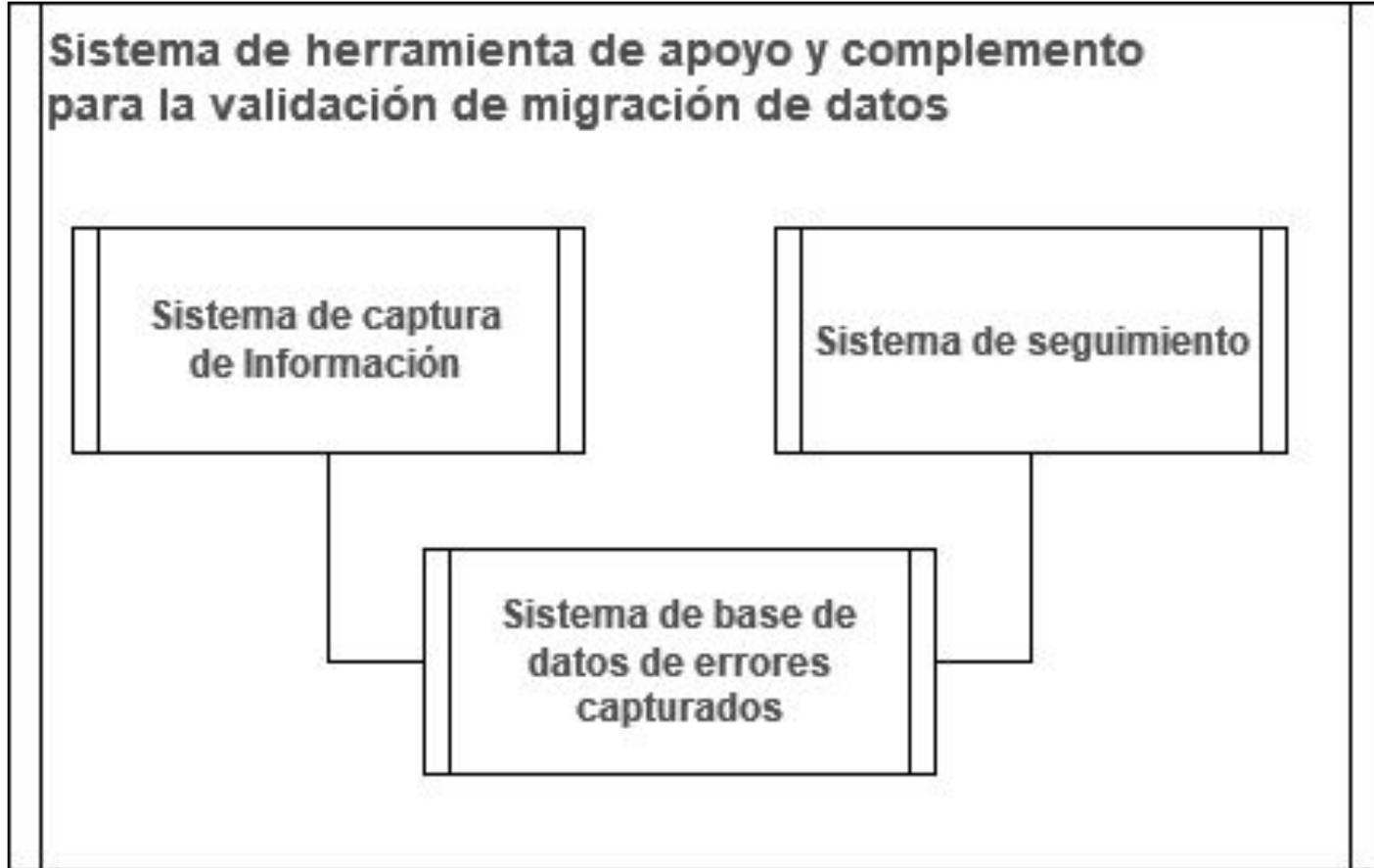
Proceso
automatizado de
captura de datos

Front-end



Dashboard
(Panel de
estadísticas)

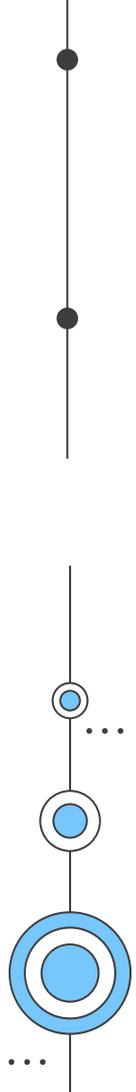
Diagrama de contexto (sistema y subsistemas)



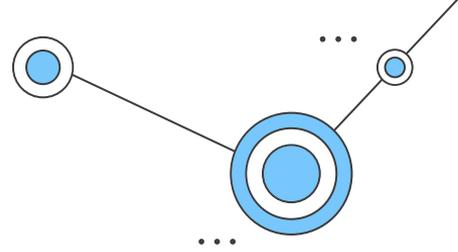


07

Herramientas
a utilizar



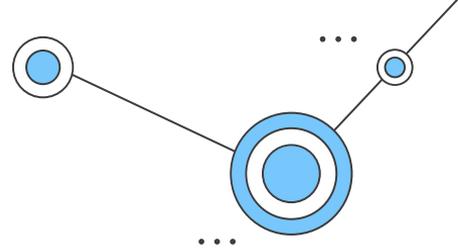
Pentaho Data Integration



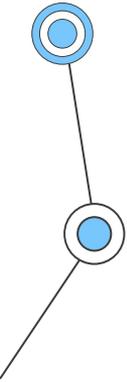
Es una de las herramientas o componentes de Pentaho Suite que permite que se utilicen técnicas ETL, es decir, poder implementar procesos de extracción, transformación y carga de datos. PDI ofrece también datos analíticos muy precisos, eliminando las complejidades involucradas en la codificación al proporcionar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

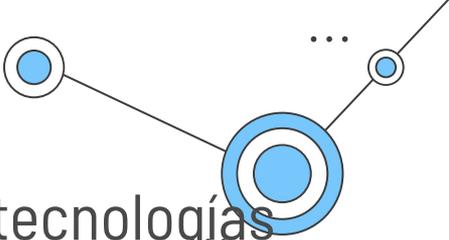


Saiku Analytics



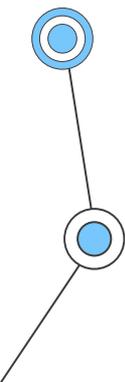
Saiku permite a los usuarios comerciales explorar fuentes de datos complejas, utilizando una interfaz familiar de arrastrar y soltar y una terminología comercial fácil de entender, todo dentro de un navegador.

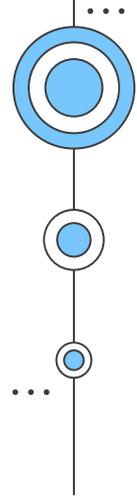




Pentaho CDE

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.*



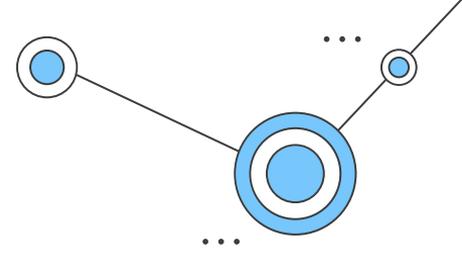


08

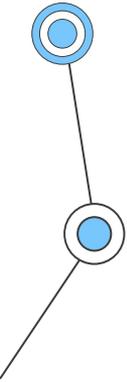
Alcance de acuerdo a la herramientas

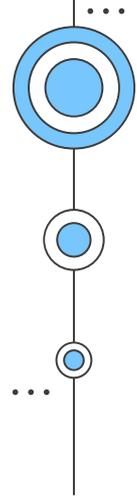


Alcance de acuerdo a la herramientas



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.



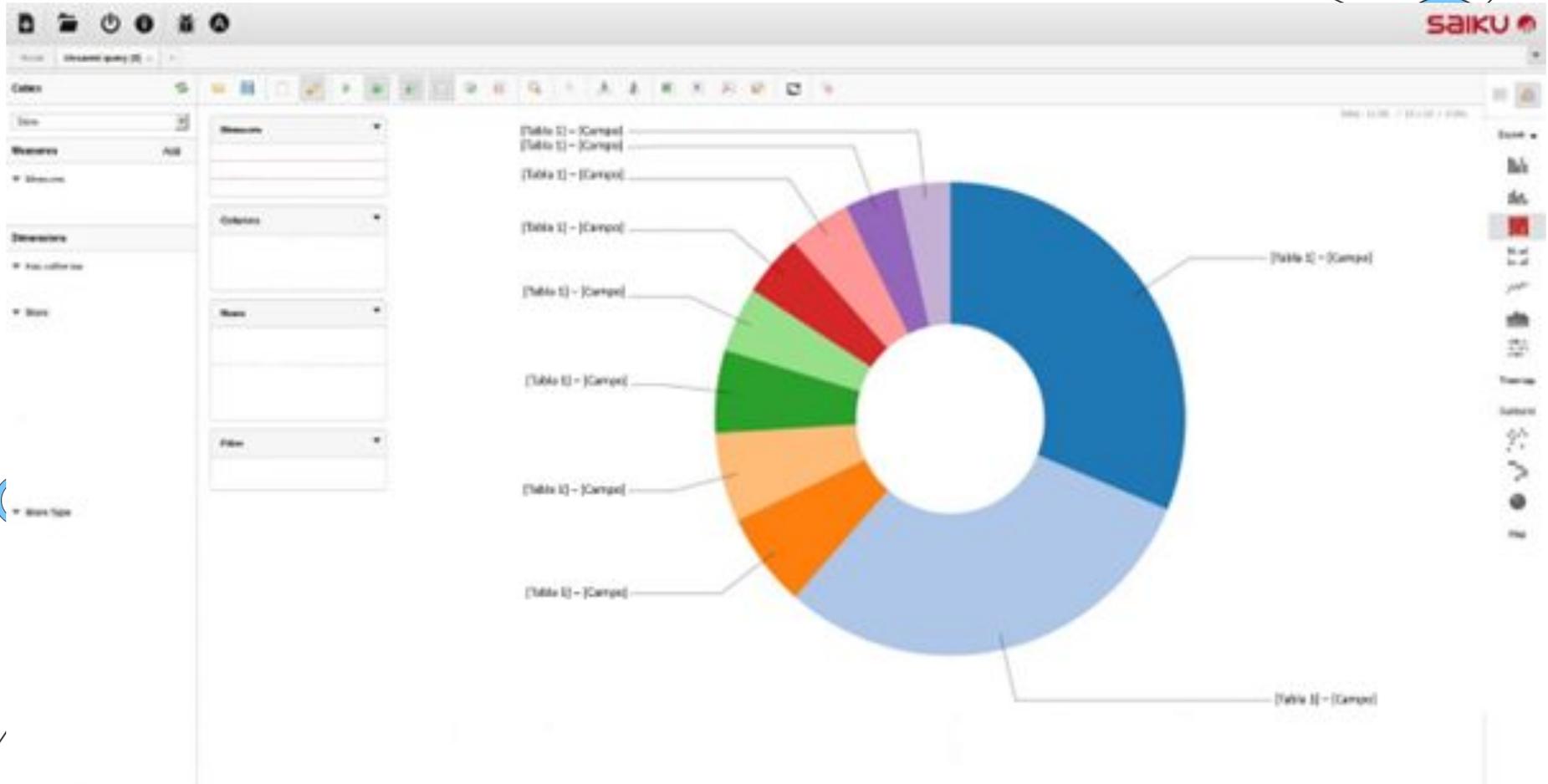


09

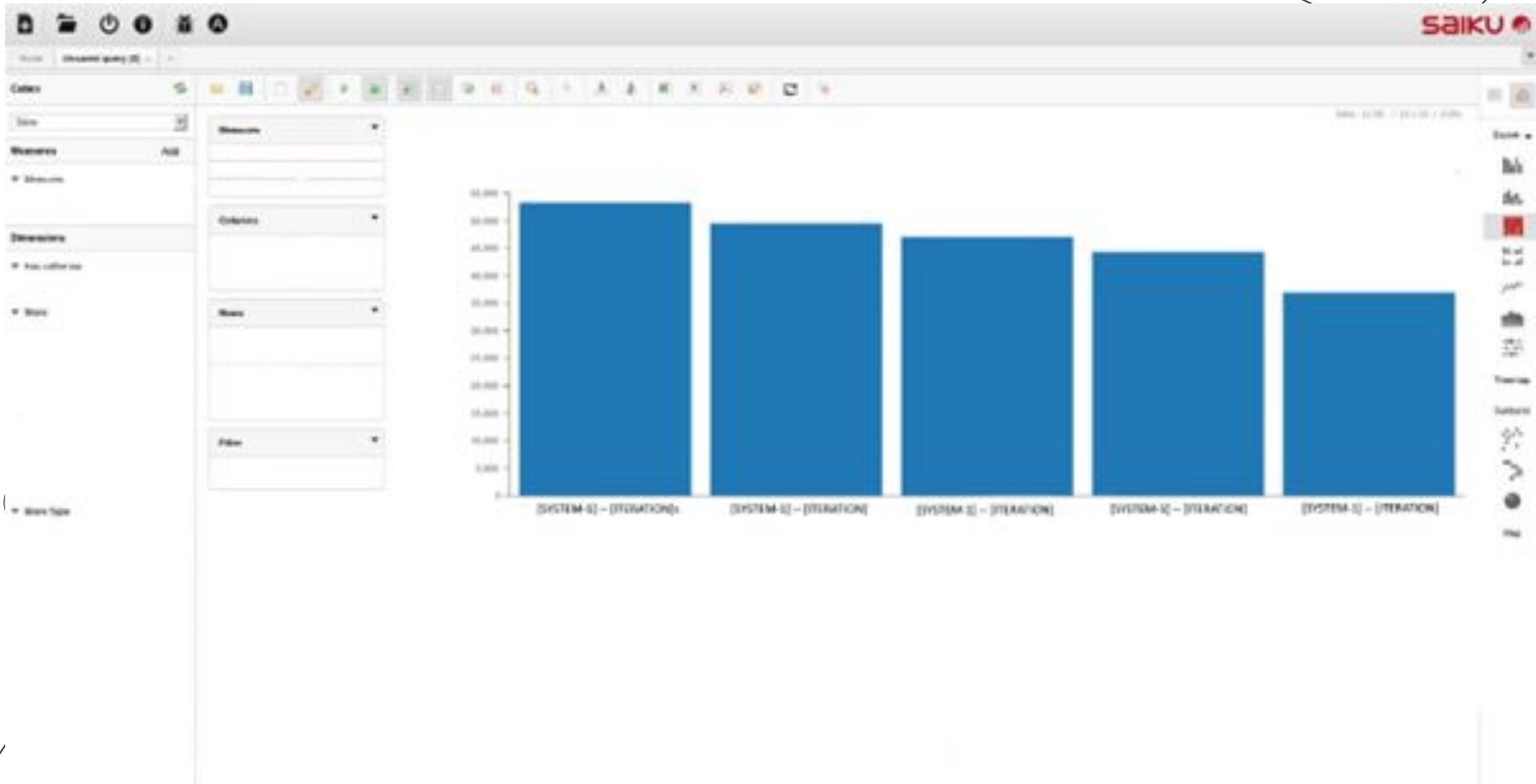
Aspectos iniciales del Front-end



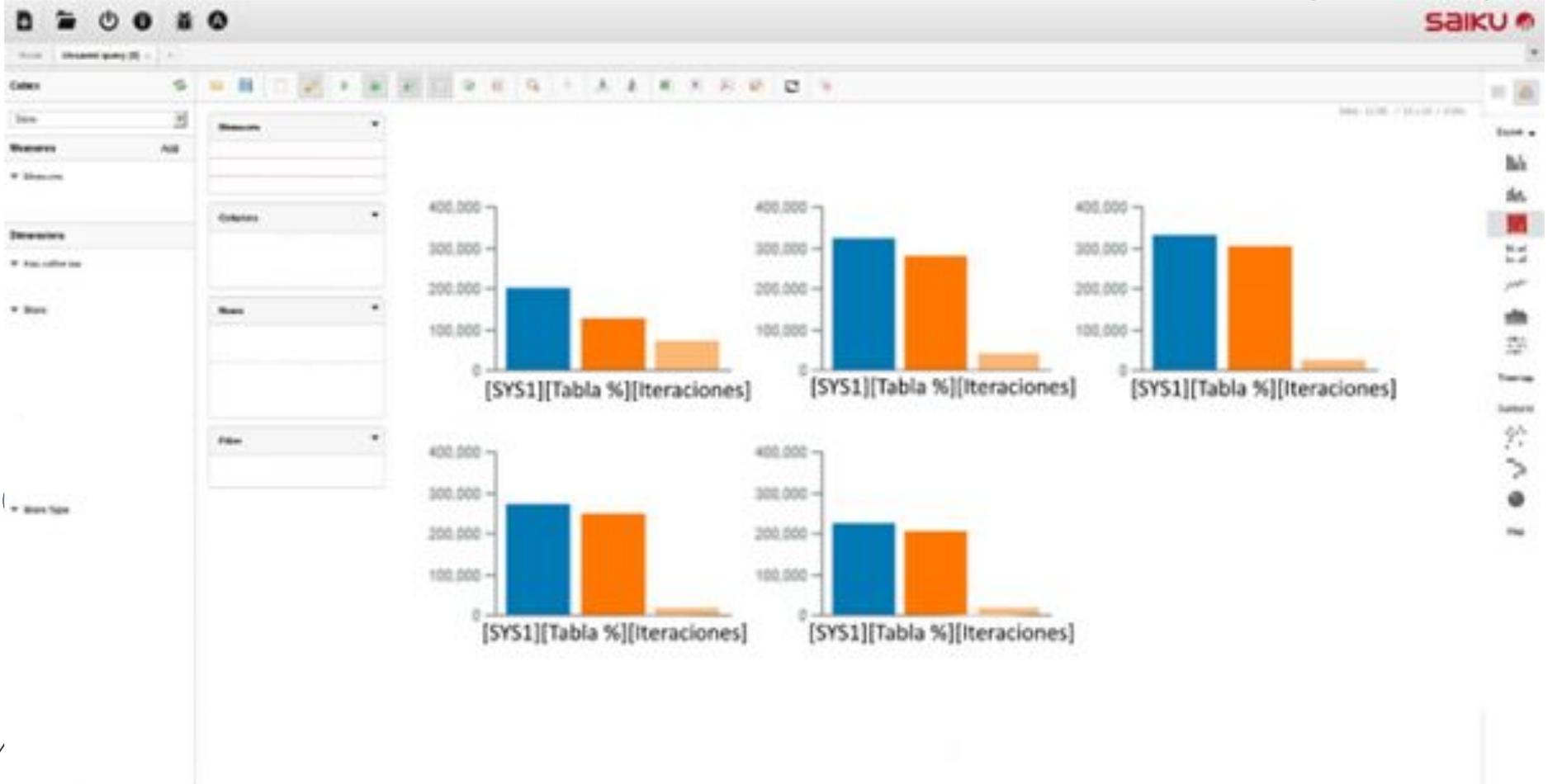
Ejemplo de gráfico de torta mostrando el porcentaje de error de cada campo respecto al total de errores en toda la "[Tabla 1]"

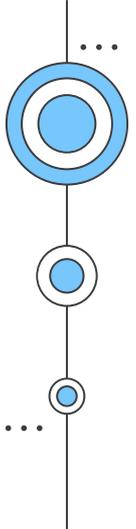


Ejemplo de gráfico de barras mostrando el total de errores encontrados por sistema en decremento con el aumento de iteraciones



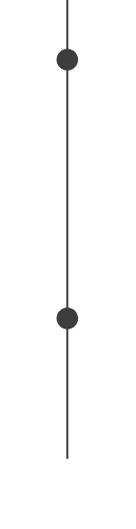
Ejemplo de gráficos de barras donde la cantidad de errores disminuye con las iteraciones por tabla dentro de un "[Sistema 1]"

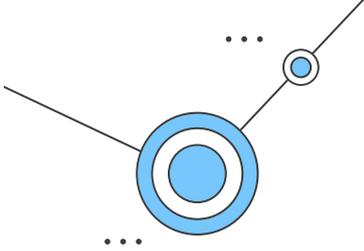
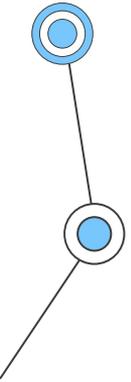
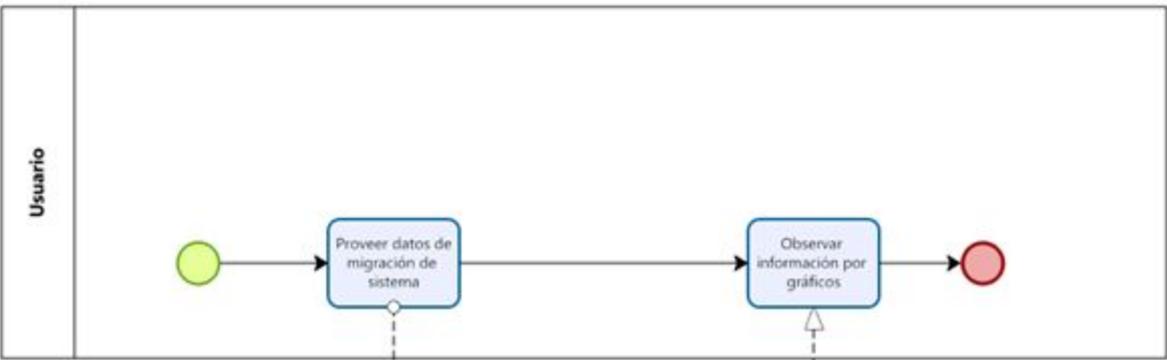
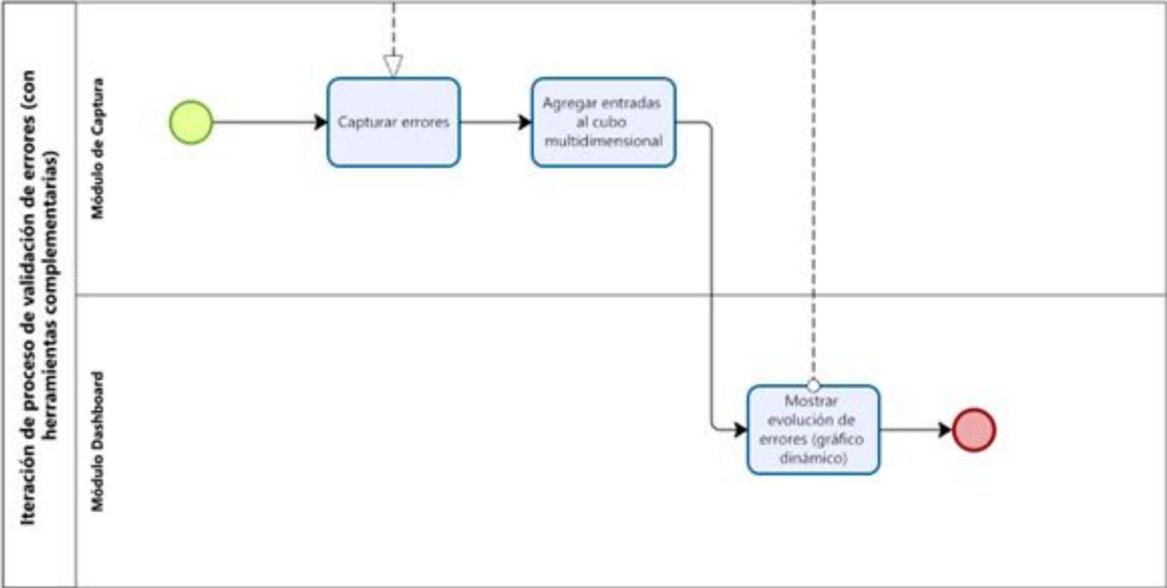




10

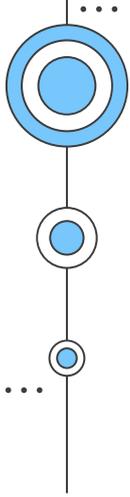
BPM





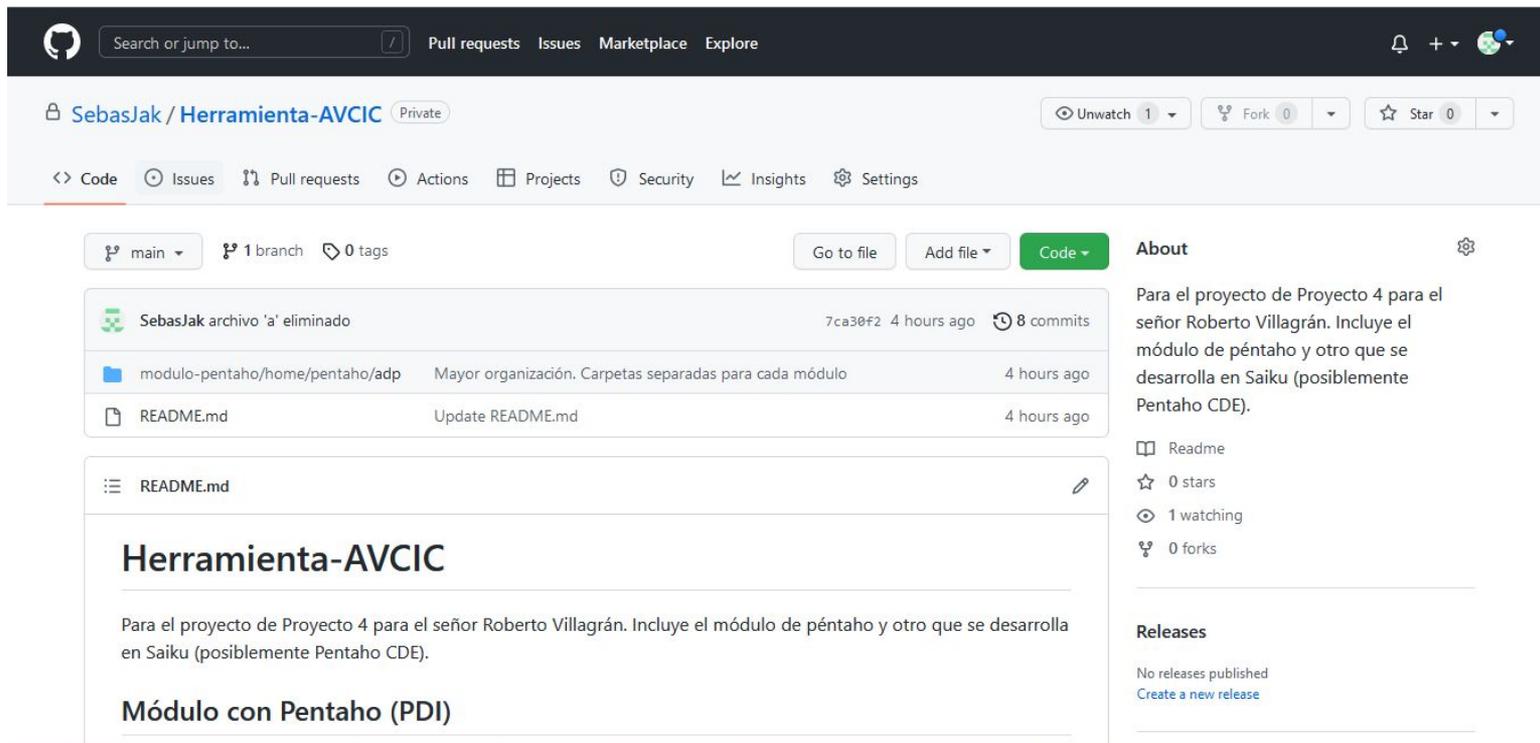
11

Repositorio de GitHub



Página principal del repositorio

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



Search or jump to...

Pull requests Issues Marketplace Explore

SebasJak / Herramienta-AVCIC Private

Unwatch 1 Fork 0 Star 0

Code Issues Pull requests Actions Projects Security Insights Settings

main 1 branch 0 tags

Go to file Add file Code

Commit	Message	Time
SebasJak	archivo 'a' eliminado	7ca30f2 4 hours ago 8 commits
modulo-pentaho/home/pentaho/adp	Mayor organización. Carpetas separadas para cada módulo	4 hours ago
README.md	Update README.md	4 hours ago

README.md

Herramienta-AVCIC

Para el proyecto de Proyecto 4 para el señor Roberto Villagrán. Incluye el módulo de pentaho y otro que se desarrolla en Saiku (posiblemente Pentaho CDE).

Módulo con Pentaho (PDI)

About

Para el proyecto de Proyecto 4 para el señor Roberto Villagrán. Incluye el módulo de pentaho y otro que se desarrolla en Saiku (posiblemente Pentaho CDE).

Readme 0 stars 1 watching 0 forks

Releases

No releases published [Create a new release](#)

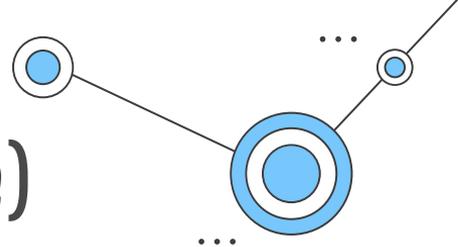
Vista de los archivos/carpetas (mod-pentaho)

main ▾ Herramienta-AVCIC / modulo-pentaho / home / pentaho / adp / Northwind /

Go to file Add file ▾ ...

 SebasJak Mayor organización. Carpetas separadas para cada módulo 43c2daa 4 hours ago  History
..
 Categories Mayor organización. Carpetas separadas para cada módulo 4 hours ago
 Customers Mayor organización. Carpetas separadas para cada módulo 4 hours ago
 EmployeeTerritories Mayor organización. Carpetas separadas para cada módulo 4 hours ago
 Employees Mayor organización. Carpetas separadas para cada módulo 4 hours ago
 OrderDetails Mayor organización. Carpetas separadas para cada módulo 4 hours ago
 Orders Mayor organización. Carpetas separadas para cada módulo 4 hours ago
 Products Mayor organización. Carpetas separadas para cada módulo 4 hours ago
 Region Mayor organización. Carpetas separadas para cada módulo 4 hours ago
 Shippers Mayor organización. Carpetas separadas para cada módulo 4 hours ago
 Suppliers Mayor organización. Carpetas separadas para cada módulo 4 hours ago
 Territories Mayor organización. Carpetas separadas para cada módulo 4 hours ago
 base Mayor organización. Carpetas separadas para cada módulo 4 hours ago
 macro_5.xlsm Mayor organización. Carpetas separadas para cada módulo 4 hours ago

Vista de los archivos/carpetas (mod-pentaho, más profundamente)



main ▾ Herramienta-AVCIC / modulo-pentaho / home / pentaho / adp / Northwind / Customers /

Go to file Add file ▾ ...

SebasJak Mayor organización. Carpetas separadas para cada módulo 43c2daa 4 hours ago History

..

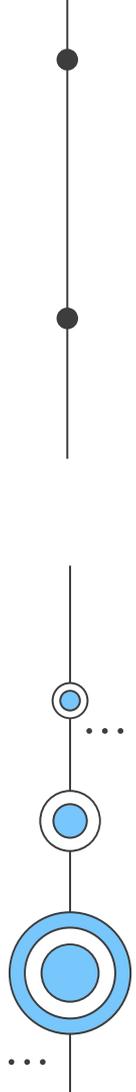
Job_Northwind_Customers.kjb	Mayor organización. Carpetas separadas para cada módulo	4 hours ago
Job_Northwind_Customers_MAIN.kjb	Mayor organización. Carpetas separadas para cada módulo	4 hours ago
TRX_Northwind_Customers_CustomerID.ktr	Mayor organización. Carpetas separadas para cada módulo	4 hours ago
TRX_Northwind_Customers_Phone.ktr	Mayor organización. Carpetas separadas para cada módulo	4 hours ago
TRX_Northwind_Customers_RESPALDO.ktr	Mayor organización. Carpetas separadas para cada módulo	4 hours ago
TRX_Northwind_Customers_RESUMEN.ktr	Mayor organización. Carpetas separadas para cada módulo	4 hours ago



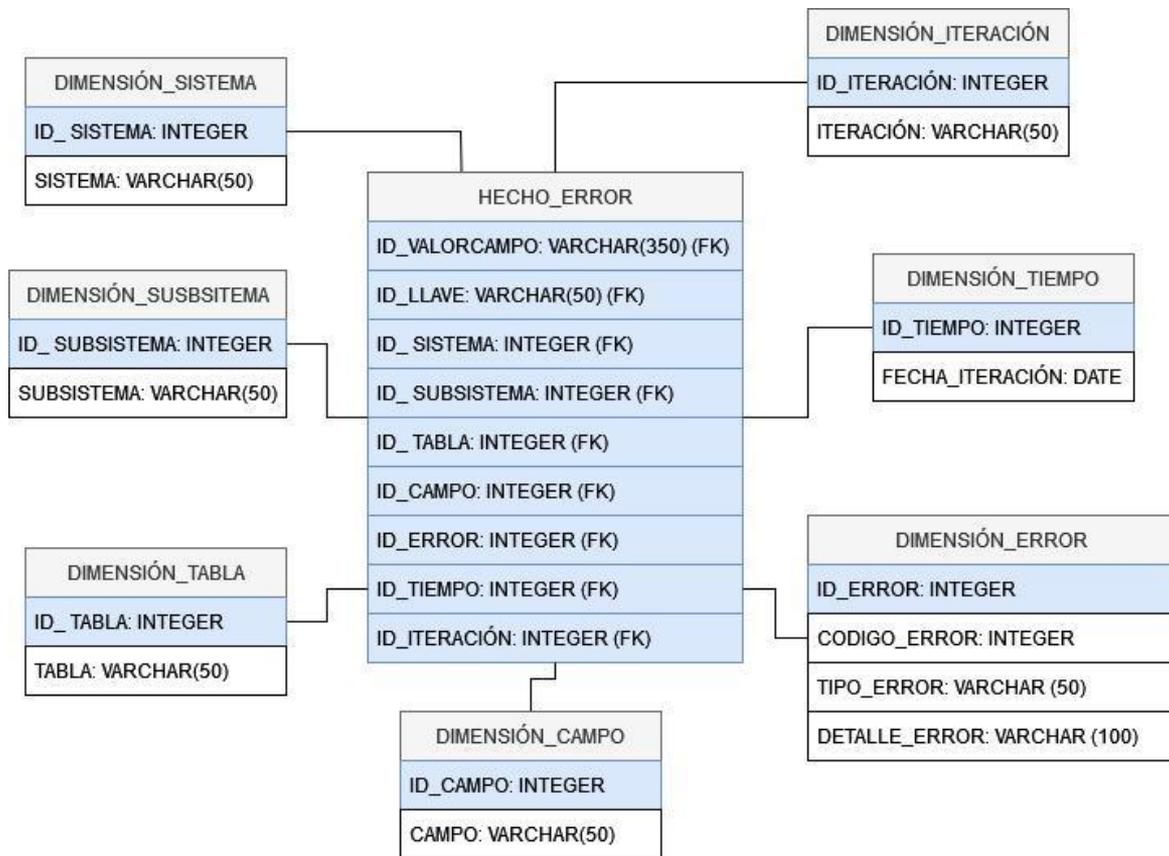


12

Almacenamiento y modelamiento de datos



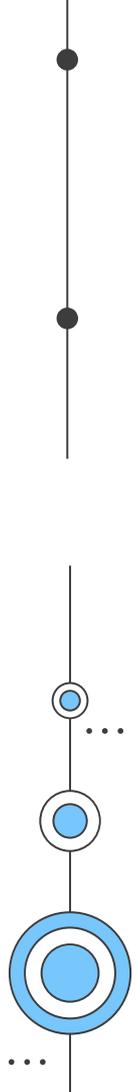
Modelo estrella para datamart

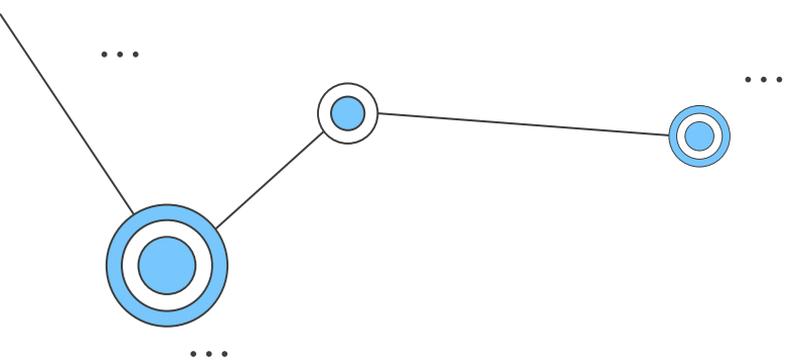




13

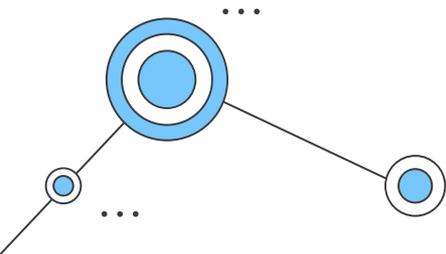
Implementación del sistema (pantallazos/vivo)





Módulo con Pentaho

Para mostrar y hacer pruebas de su funcionamiento se utilizó la base de datos de prueba Northwind.

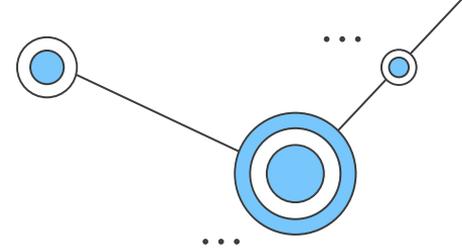


Estructura general para los archivos de la verificación de integridad de un sistema

) > Herramienta-AVCIC > modulo-pentaho > home > pentaho > adp > Northwind

Nombre	Fecha de modificación	Tipo	Tamaño
base	4/11/2022 17:49	Carpeta de archivos	
Categories	11/10/2022 04:45	Carpeta de archivos	
Customers	11/10/2022 05:14	Carpeta de archivos	
Employees	11/10/2022 05:30	Carpeta de archivos	
EmployeeTerritories	11/10/2022 05:38	Carpeta de archivos	
OrderDetails	5/10/2022 00:16	Carpeta de archivos	
Orders	11/10/2022 05:49	Carpeta de archivos	
Products	11/10/2022 05:53	Carpeta de archivos	
Region	11/10/2022 05:59	Carpeta de archivos	
Shippers	11/10/2022 06:02	Carpeta de archivos	
Suppliers	11/10/2022 06:06	Carpeta de archivos	
Territories	11/10/2022 06:11	Carpeta de archivos	

Interior de cada carpeta de tabla



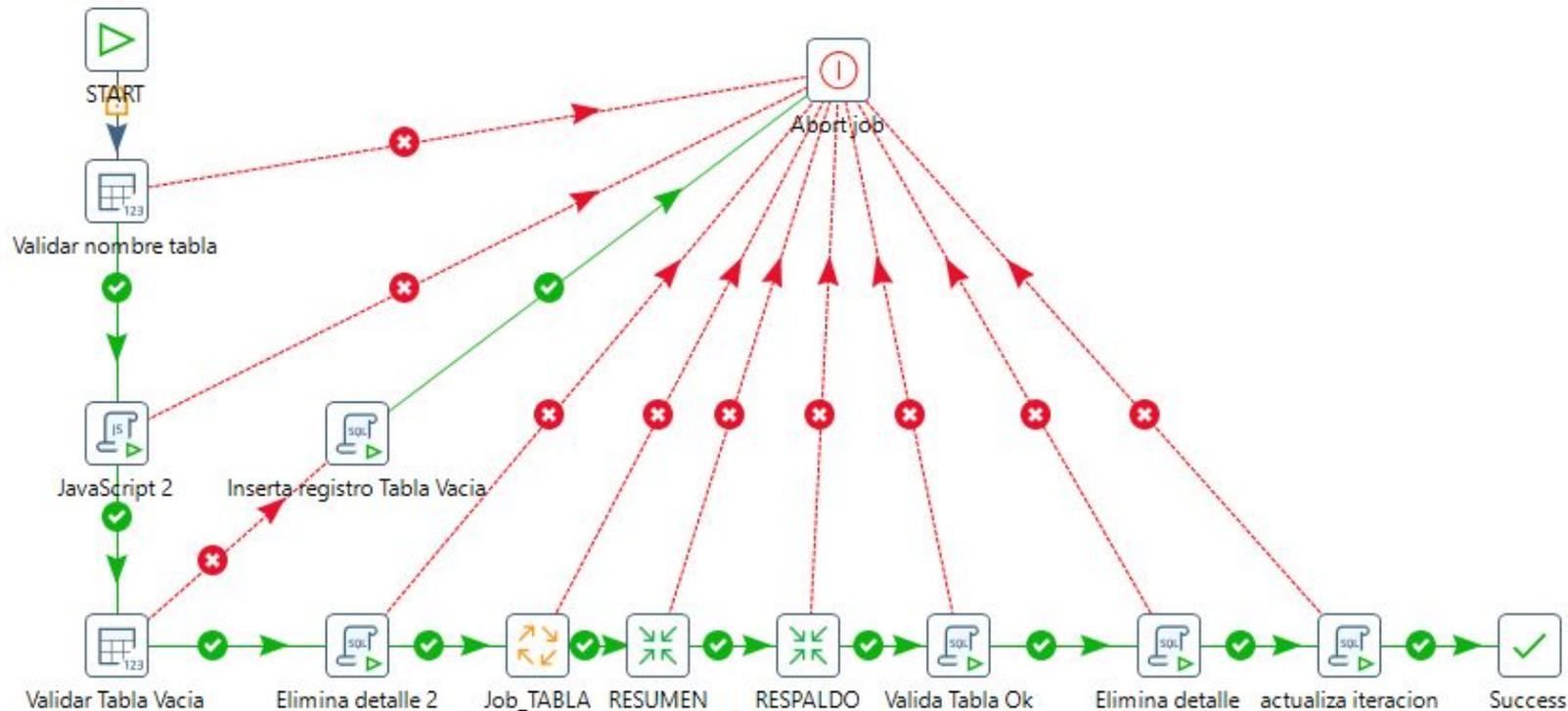
AVCIC > modulo-pentaho > home > pentaho > adp > Northwind > Products

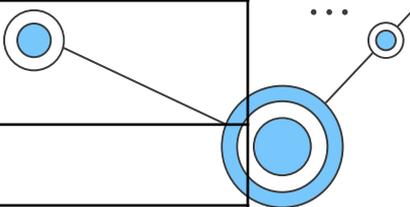
Nombre	Fecha de modificación	Tipo	Tamaño
 Job_Northwind_Products_MAIN.kjb	26/10/2022 18:12	Archivo KJB	26
 Job_Northwind_Products.kjb	26/10/2022 17:54	Archivo KJB	13
 TRX_Northwind_Products_ProductID.ktr	26/10/2022 17:52	Archivo KTR	27
 TRX_Northwind_Products_CategoryID.ktr	11/10/2022 06:45	Archivo KTR	19
 TRX_Northwind_Products_SupplierID.ktr	11/10/2022 06:41	Archivo KTR	19
 TRX_Northwind_Products_RESUMEN.ktr	11/10/2022 05:57	Archivo KTR	19
 TRX_Northwind_Products_RESPALDO.ktr	11/10/2022 05:57	Archivo KTR	18



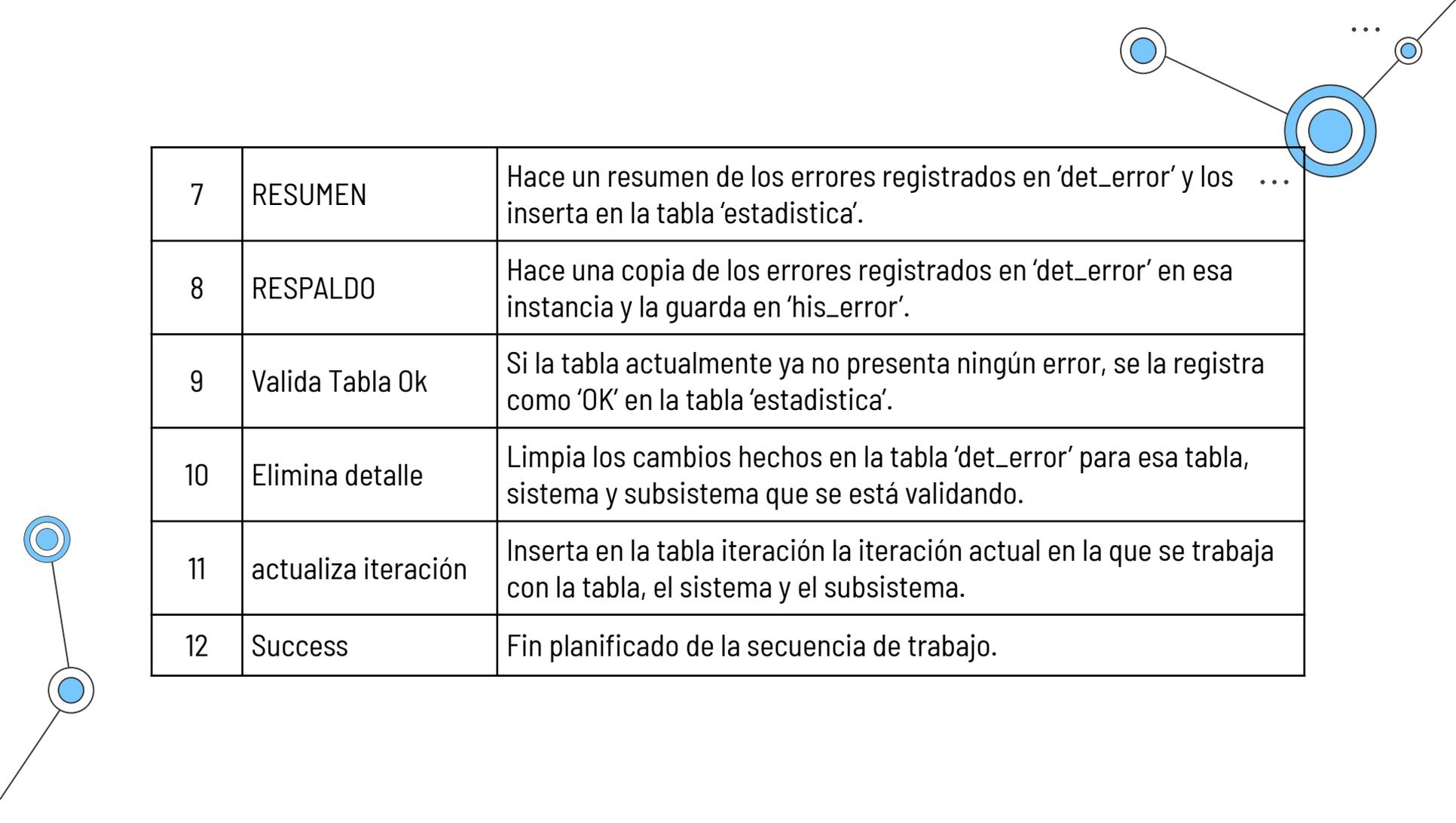
Archivo Pentaho principal (MAIN)

Hace referencia a los demás de su sector



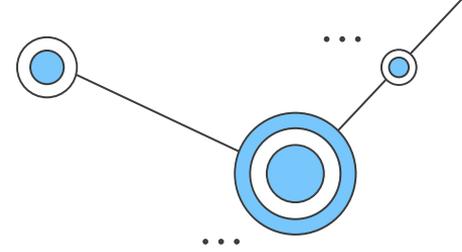


N° del paso	Nombre	Descripción
1	START	Inicio de la secuencia de trabajo.
2	Validar nombre tabla	Antes de hacer operaciones en la base de datos con el nombre de la tabla se verifica que este esté registrado en la tabla 'tablasmigracion'.
3	JavaScript 2	Seteo de la fecha y hora de la iteración con base a funciones integradas en Pentaho.
4	Validar Tabla Vacía	Verifica que de hecho hayan registros en la tabla que se desea validar. Luego, si se detecta que la tabla está vacía se inserta un registro en la tabla 'estadística' con la marca de tiempo, iteración, nombre de sistema, subsistema y tabla.
5	Elimina detalle 2	Elimina de la tabla 'det_error' todos los registros que puedan haber con el nombre de esa tabla, sistema y subsistema a la vez.
6	Job_TABLA	Se invoca al job que hace las validaciones personalizadas para cada campo de la tabla a validar.



7	RESUMEN	Hace un resumen de los errores registrados en 'det_error' y los inserta en la tabla 'estadistica'.
8	RESPALDO	Hace una copia de los errores registrados en 'det_error' en esa instancia y la guarda en 'his_error'.
9	Valida Tabla Ok	Si la tabla actualmente ya no presenta ningún error, se la registra como 'OK' en la tabla 'estadistica'.
10	Elimina detalle	Limpia los cambios hechos en la tabla 'det_error' para esa tabla, sistema y subsistema que se está validando.
11	actualiza iteración	Inserta en la tabla iteración la iteración actual en la que se trabaja con la tabla, el sistema y el subsistema.
12	Success	Fin planificado de la secuencia de trabajo.

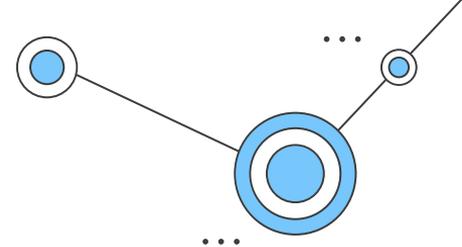
Archivo para macro Excel



Como el módulo requiere de un conjunto de archivos que comparten la misma estructura pero aplicados específicamente a cada tabla del sistema en migración, se implementó un macro de Microsoft Excel (tomando un archivo base y complementándolo). El macro genera automáticamente los archivos de Pentaho que necesita cada tabla para el sistema que se configuró previamente tras cierta configuración inicial usando un archivo de Excel. Las funciones de los grupos de celdas en el archivo Excel son las que se explican en la tabla después de la figura siguiente.



Archivo para macro Excel



	A	B	C	D	E	F	G	H	I
1	NRO_CAMPO	NOMBRE_CAMPO	SISTEMA	SUBSISTEMA	TABLA	LLAVE	ESQUEMA	MOTOR DE BD.	RESPONSABLE
2	1	Tablatest-ID	RETAIL	Northwind	TablaTest	EUS_KEY	Northwind.dbo	SQLSERVER	Melvi Sepulveda Ramos
3	2	CampoA							
4	3	CampoB							
5	4	CampoC							
6	5	CampoFinal							
7									
8									
9									
10									
11									
12									
13									
14									
15									

①

②

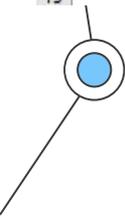
③

④

⑤

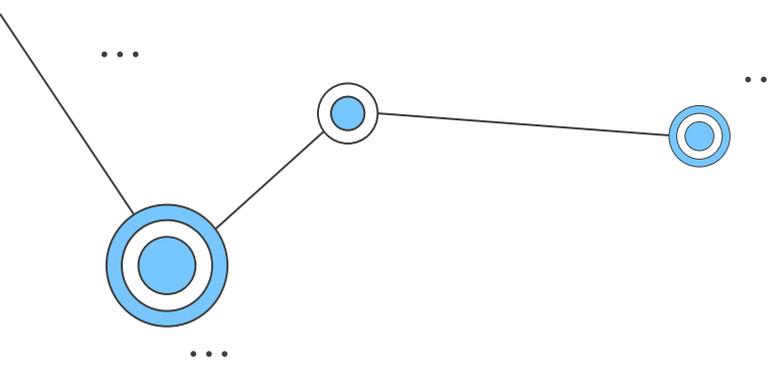
⑥

⑦

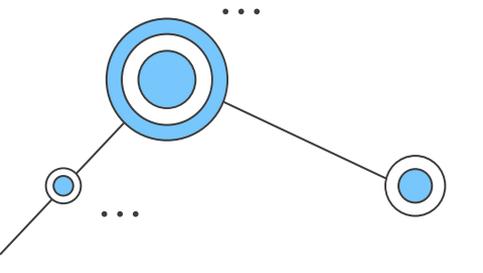


Funciones de cada zona del archivo Excel

Grupo de celdas	Descripción	Valores que se esperan ...
1	Nombres de los campos que se van a validar de la tabla.	Strings alfanuméricos sin espacios
2	Nombres del sistema y subsistema en migración. El nombre que se utiliza hasta el momento para la generación automática del nombre de los archivos es sólo el del subsistema.	String alfanumérico sin espacios
3	Nombre de la tabla a la que generar los archivos para su validación.	String alfanumérico sin espacios
4	Este dato no se usa. Era parte del archivo base.	-
5	Nombre del esquema de la base de datos que quedó como resultado de la migración.	String alfanumérico sin espacios. Acepta también punto.
6	Nombre del motor de base de datos en el que se hacen las operaciones en las tablas de la base de datos. Hasta el momento el macro está implementado sólo para SQLSERVER.	'DB2', 'SQLSERVER', 'SYBASE'
7	Este dato no se usa. Era parte del archivo base.	-



Módulo con Saiku Analytics



El avance del módulo de Saiku Analytics se vió frenado por ciertos imprevistos en el desarrollo del módulo con pentaho descrito en detalle anteriormente. Sin embargo, el servidor en donde poder utilizar la plataforma se encuentra en pie actualmente y se espera poder tener listo el avance y funcionalidad terminada para la siguiente presentación.

Plataforma Saiku Analytics levantada en el servidor

← → ↻ 146.83.109.226:9080/pentaho/Home?locale=en_US ☆

File View Tools Help

Home ▾ slukich ▾

[Browse Files](#)

[Create New](#)

[Manage Data Sources](#)

[Documentation](#)

Recents

You haven't opened anything recently. Browse your files.

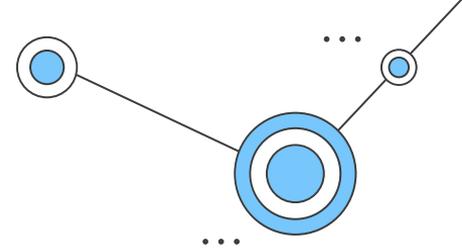
[Browse Files](#)

Pentaho Business Analytics

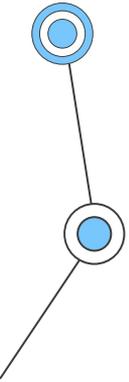
Get help and contribute with your knowledge. Find here some ways how you can do it.

[Documentation](#) [Forums](#) [Mailing Lists](#) [Blogs](#)

Conclusiones



- Uso del repositorio remoto apoya efectivamente.
- Conocimiento de herramientas de apoyo por sobre el conocimiento del lenguaje.
- Las pruebas continuas asentan el aprendizaje nuevo para el desarrollo.



07/11/22 ...

...



Herramienta de automatización de validación de consistencia e integridad de campos

E. Desarrollo: Sebastian Lukich
Empresa o Unidad: Digital Social Change SPA
Profesor: Diego Arcena

