



# ESPE

UNIVERSIDAD DE LAS FUERZAS ARMADAS  
INNOVACIÓN PARA LA EXCELENCIA

VICERRECTORADO DE INVESTIGACIÓN, INNOVACIÓN Y TRANSFERENCIA DE  
TECNOLOGÍA

CENTRO DE POSGRADOS

**SOLUCIÓN ÁGIL DE INTELIGENCIA DE NEGOCIOS UTILIZANDO UN GESTOR DE BASE DE  
DATOS COLUMNAR Y SOFTWARE OPEN SOURCE – CASO DE ESTUDIO: DIRECCIÓN DEL  
REGISTRO INTERCONECTADO DE PROGRAMAS SOCIALES DE LA SECRETARÍA NACIONAL  
DE PLANIFICACIÓN Y DESARROLLO**

Autor: Ing. Gabriela Villagómez  
Director: Ph.D YOO, SANG GUUN  
Sangolquí, Diciembre 2019



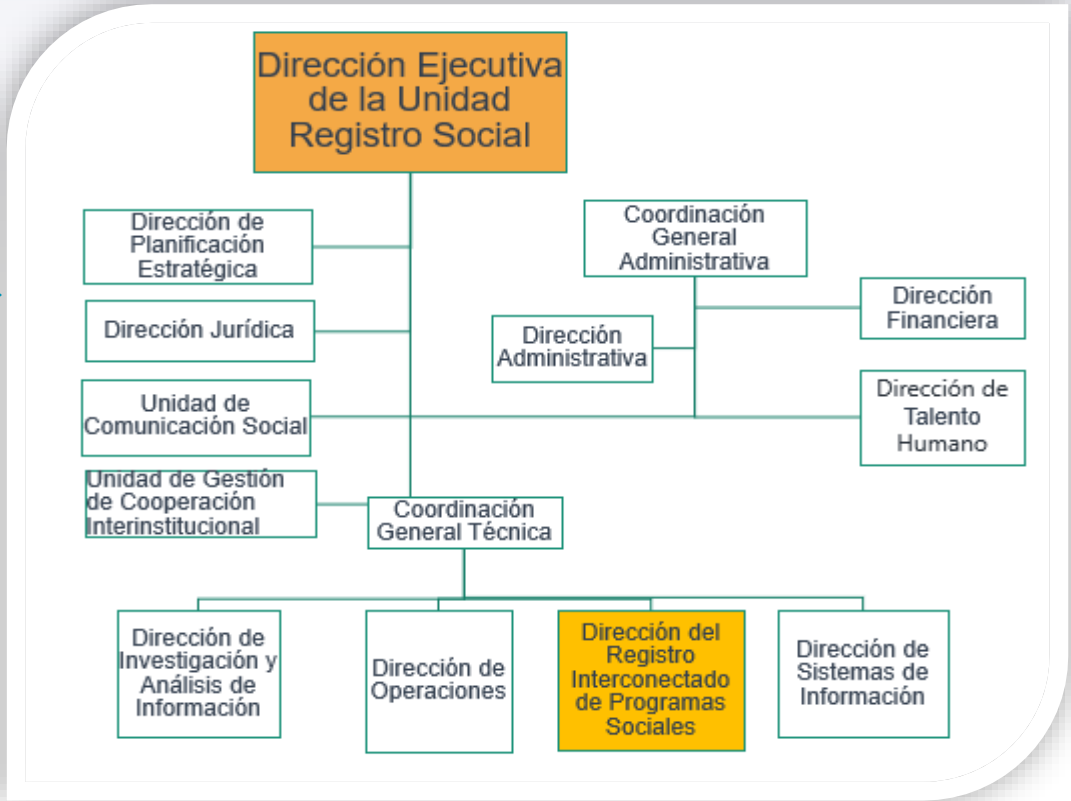
# AGENDA

- ▶ Contexto del Problema
- ▶ Planteamiento del Problema
- ▶ Objetivos
- ▶ Hipótesis
- ▶ Metodología de la Solución
- ▶ Herramientas Seleccionadas
- ▶ Arquitectura de Solución
- ▶ Resultados Obtenidos
- ▶ Conclusiones y Recomendaciones



# Contexto del Problema

1



# Planteamiento del Problema

2

Datos de carácter social



Archivos planos  
Fuentes de datos



Interna y Externa

Unidad del Registro Social



Almacenada

Procesos ETL



+

Falta de soporte técnico



Herramientas para la gestión e integración de datos

+

No seguir una metodología establecida.



Proyecto de inteligencia de negocio.

Modelos de datos multidimensional



Tiempo considerable de acceso a los datos

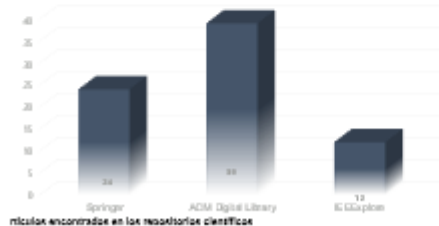
# Objetivos

3



1.

Realizar una revisión de literatura para determinar las características de herramientas de gestión de datos columnar y prácticas más adecuadas, para la implementación de proyectos de inteligencia de negocios que contribuirá en el planteamiento de la propuesta de mejora.



Artículos encontrados en los repositorios científicos



2.

Implementar una nueva solución de inteligencia de negocios enfocada a la información de carácter social en el sector gubernamental, con el fin de modernizar la gestión de datos existente en la DRIPS.



3.

Evaluar los resultados mediante el uso de herramientas de integración y gestión de datos, con el propósito de comprobar el tiempo de respuesta de consultas aplicado al modelo multidimensional en un gestor de datos relacional y columnar



# Metodología de Solución

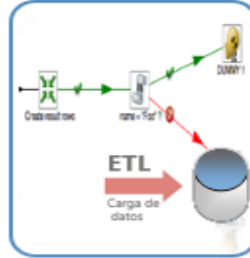
4



- Requisitos de datos
- Recopilación de requerimientos
- Selección de herramientas tecnológicas.



- Análisis y determinación de esquemas.
- Análisis y determinación de componentes del diseño de datos dimensional
- Diseñar el modelamiento dimensional.



- Construcción de procesos y pruebas ETL.
- Calendarización y tiempo de procesamiento



- Construcción de reporte de datos.
- Bitácora de componentes.



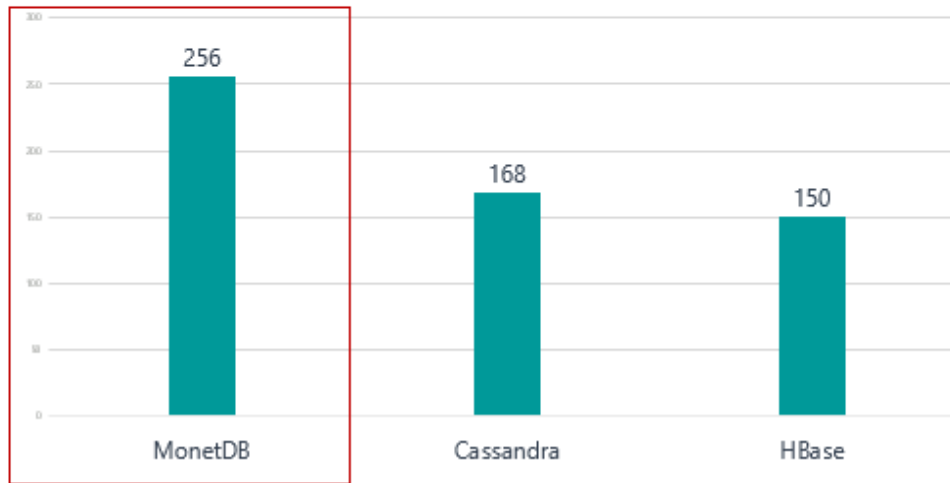
- Mantenimiento de base de datos y herramientas de software.
- Generación de respaldos.

# Herramientas Seleccionadas – Gestión de Datos

5

Características	Cassandra	Hbase	MonetDB
<b>Almacenamiento columnar</b>	Si Gestor de base de datos orientada en columnas basados en ideas de BigTable and DynamoDB	Si Gestor de base de datos orientado en columnas basados en ideas Apache Hadoop y en conceptos de BigTable.	Si Gestor de base de datos multimodelo: i) Relacional, y ii) Orientado en columnas.
<b>Licencia Open Source</b>	Si	Si	Si
<b>Multiplataforma</b>	Si BSD, Linux, OS X, Windows	Si Linux, Unix, Windows (Usando Cygwin)	Si FreeBSD, Linux, OS X, Solaris, Windows
<b>Soporte SQL</b>	Si SQL-like SELECT, DML and DDL statements (CQL)	No	Si
<b>Foreign Keys</b>	No	No	Si
<b>Procedimientos almacenados</b>	No	Si	Si (En Sql, C,R)
<b>APIs y otros métodos de acceso</b>	Si Protocolo propietario Thrift	Java API, RESTful, HTTP API, Thrift	Si Native C Library, JDBC, ODBC
<b>Triggers</b>	Si	Si	Si
<b>Concurrencia</b>	Si	Si	Si
<b>Durabilidad</b>	Si	Si	Si
<b>Concepto de usuarios</b>	Si Acceso para los usuarios definido por objeto	Si Listas de control de acceso (ACL) para RBAC, integración con Apache Ranger para RBAC y ABAC	Si Acceso específico según el estándar SQL
<b>Soporta conceptos de transacciones</b>	No	Si ACID de una fila (en millones de columnas)	Si ACID
<b>Aplicación</b>	Internet de las cosas (IoT), aplicaciones de detección de fraude, motores de recomendación, catálogos de productos y listas de reproducción y aplicaciones de mensajería.	Proyectos de gran procesamiento, almacenamiento y análisis de datos (Big data), informes de analítica de datos.	Base de datos analítica, Data warehouse, OLAP, Data mining, Scientific database

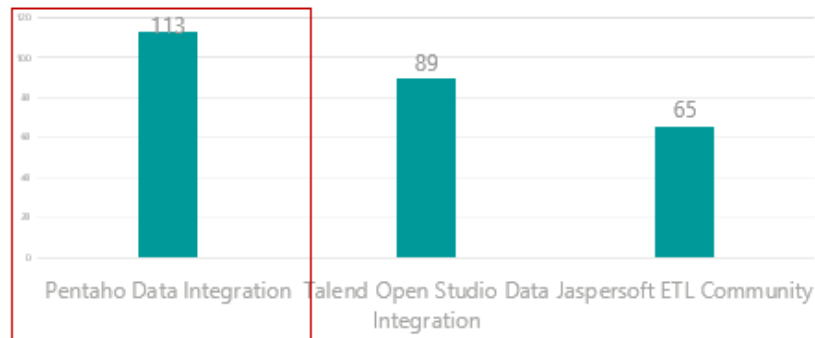
Rank			DBMS	Database Model	Score	
Oct 2019	Sep 2019	Oct 2018			Oct 2019	Sep 2019
1.	1.	1.	Cassandra	Wide column	123.22	-0.18
2.	2.	2.	HBase	Wide column	54.84	-0.88
54.	57.	53.	MonetDB	Relational, Multi-model	1.88	+0.22



# Herramientas Seleccionadas – Integración de Datos

6

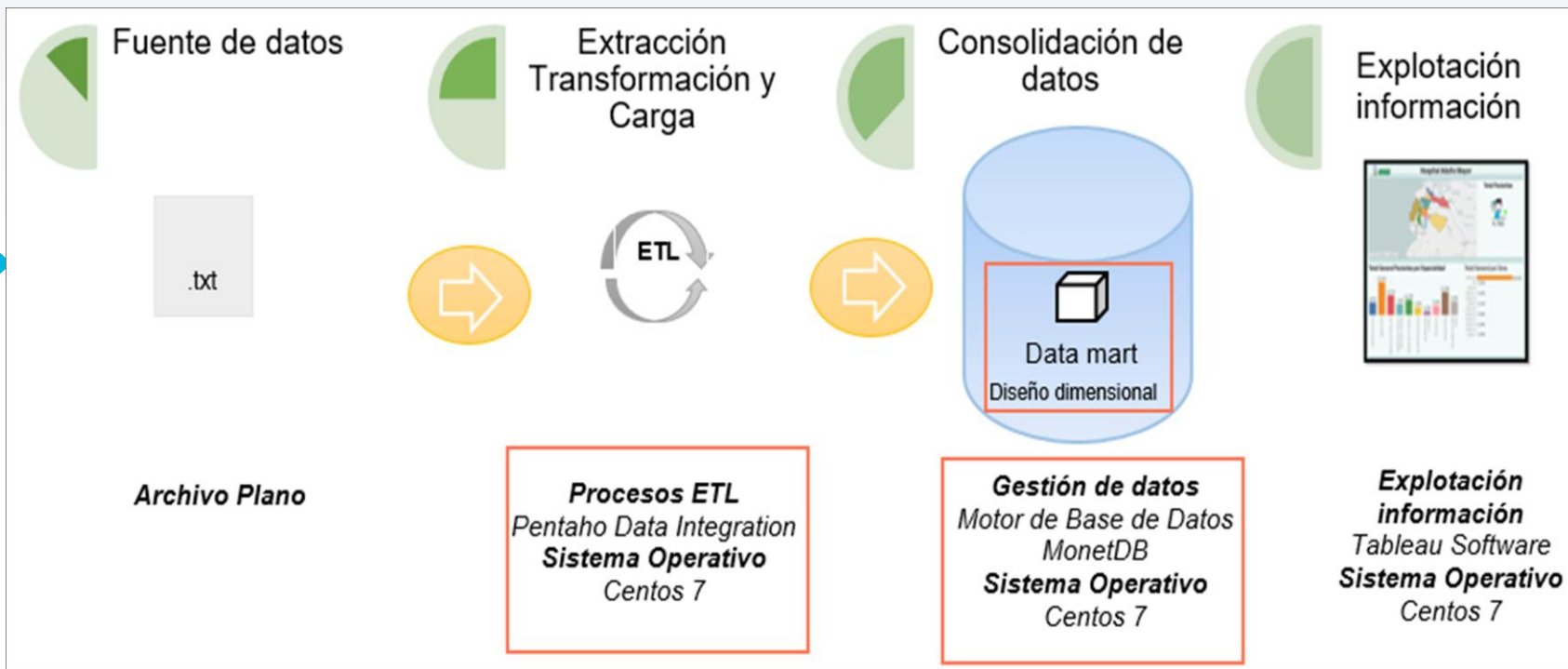
Características	Talend Open Studio Data Integration	Jaspersoft ETL Community	Pentaho Data Integration
Licencia Open Source	Si	Si	Si
Soporte Comunitario	Si	Si Soporte limitado para carga de trabajo en lotes. (Gartner Inc, 2019)	Si
Versión Estable	Versión 7.3.1M2 – Septiembre 2019	Versión 6.0.1 - released Marzo 2016	Versión 8.2 Octubre 2018
Multiplataforma	Si Windows Mac	Si Linux - versión 1.1.2 Windows Mac	Si Linux Windows Mac
Cargas de datos en lotes	Si	Si Por sí solo no puede abordar el alcance completo de las cargas de trabajo masivas / relacionadas con lotes. (Gartner Inc, 2019)	Si
Alto Rendimiento	Si	Si	Si
Conexión a múltiples fuentes de datos	Si Limitada	Si Limitada	Si Amplia





# Arquitectura de la Solución

7



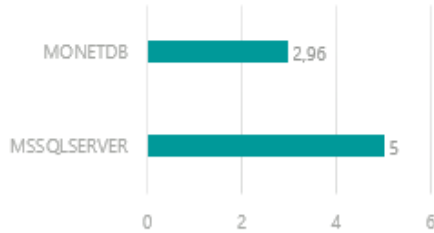
# Resultados Obtenidos



8

1. Consulta de selección sin join

TIEMPO EN SEGUNDOS



	MSSQLSERVER	MONETDB
■ Tiempo en Segundos	5	2,96

40,2%

2. Consulta de selección con un inner join

TIEMPO

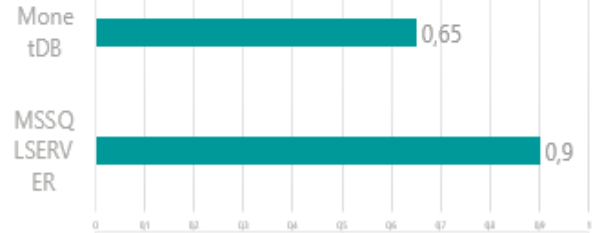


	MSSQLSERVER	MonetDB
■ Tiempo	0,9	0,65

28%

3. Consulta de selección con tres inner join

TIEMPO

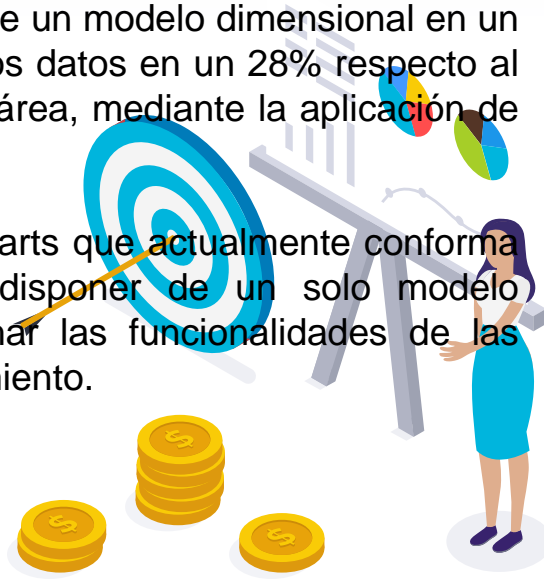


	MSSQLSERVER	MonetDB
■ Tiempo	0,9	0,65

28%

# Conclusiones

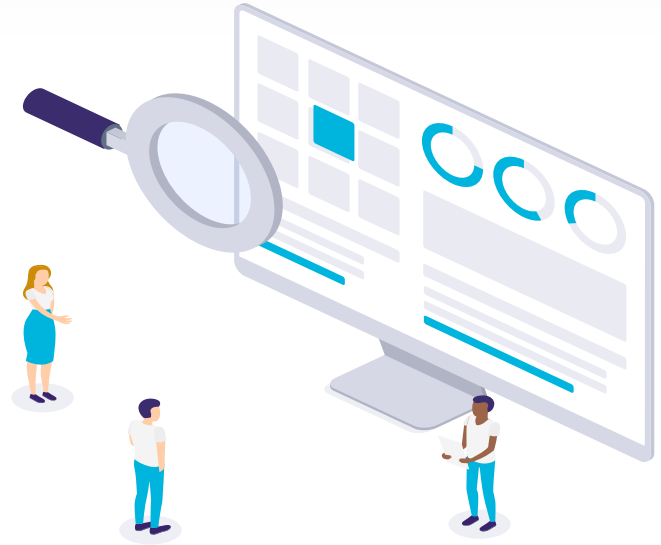
- Se realizó una revisión de literatura para determinar las características de herramientas de gestión de datos columnar y prácticas más adecuadas, para la implementación de proyectos de inteligencia de negocios que contribuyó en el planteamiento de la propuesta de mejora.
- Se evaluó, los resultados mediante el uso de herramientas de gestión de datos, con el propósito de comprobar el tiempo de respuesta de acceso a los datos aplicado al modelo multidimensional en un gestor de datos relacional y columnar.
- Como resultado del proyecto, se demostró que la implementación de un modelo dimensional en un gestor de base de datos columnar mejoró el tiempo de acceso a los datos en un 28% respecto al gestor de base de datos transaccional utilizado actualmente en el área, mediante la aplicación de una metodología de inteligencia de negocio.
- Con la solución propuesta se pretendió disminuir todos los data marts que actualmente conforma el data warehouse del caso de estudio, de tal manera de disponer de un solo modelo multidimensional que almacene toda la información, y aprovechar las funcionalidades de las herramientas de gestores de datos columnar para mejorar el rendimiento.



# Recomendaciones

- Como trabajo futuro, se podría realizar el mismo estudio, pero utilizando otros gestores de bases de datos columnar open source, para analizar los tiempos de respuesta y comparar los resultados obtenidos en este estudio.
- Se recomienda, contar con una guía metodológica para realizar el proceso de integración de datos que se adapte al entorno y reglas de negocio de cada empresa.

10



# Gracias.

