UNIVERSIDAD NACIONAL DE INGENIERÍA

Facultad de Ingeniería Industrial y Sistemas
Escuela Profesional de Ingeniería de Sistemas

SEGUNDA CALIFICADA SISTEMA DE INTELIGENCIA DE NEGOCIOS (SI 807-U)

PROFESOR: Dr. Ing. ARADIEL CASTANEDA, HILARIO
AYUDANTE DE CATEDRA: García, Atuncar, Fernando
CICLO: 2025 B Fecha: 29-09-25

GITHUB: David1712uni/Contralor-a-SIN

INTEGRANTES:

Código UNI	Apellidos y Nombres	Correo Electrónico	Tareas realizadas
20220008E	Andrade Saavedra, Navhi Giordano	navhi.andrade.s@uni.pe	- Preprocesamiento de datos - Carga de datos en Hive - Comprobación de querys en Hive
20220122B	Caruzo Cieza, David	david.caruzo.c@uni.pe	- Preguntas del negocio - KPI's Definidos
20200298H	Carhuas Romero Jhon Jesus	jhon.carhuas.r@uni.pe	- Diagrama estrella - Actualizar la fuente de datos

OBJETIVOS

• Aplicar la metodología Hefesto en la fase de análisis de requerimientos. • Identificar preguntas de negocio y traducirlas en KPIs.
• Elaborar el modelo conceptual preliminar (hechos y dimensiones). • Construir el inventario de fuentes OLTP.
• Implementar la primera ingesta de datos en Hortonworks (HDFS y Hive).

ALCANCE

Esta práctica se centra en el análisis funcional y en el inicio de la implementación técnica:
- Levantamiento de preguntas de negocio.
- Definición de KPIs clave y sus fichas técnicas.

- Diseño de modelo conceptual (Star Schema).
- Inventario de fuentes de datos OLTP.
- Ingesta de datasets en HDFS y creación de tablas en Hive (zona raw). - Consulta básica de control en Hive.

1. DESARROLLO

1.1. Preguntas del Negocio

Área	Rol del Usuario	Pregunta de Negocio	Nivel de Prioridad	Fuente de Datos Actual
Dirección Estratégica	Alta Dirección de EsSalud	¿Cuál es la tasa de diagnósticos de enfermedades priorizadas (diabetes, hipertensión, obesidad) por cada 1000 habitantes?	Alta	Datos Abiertos EsSalud (consultas externas), CIE10
Dirección Estratégica	Presidencia Ejecutiva	¿Cómo está evolucionando la tendencia de diagnósticos en el tiempo y en qué zonas se proyecta mayor riesgo?	Alta	Datos Abiertos EsSalud, INEI
Gestión Táctica	Gerencias Regionales	¿Qué regiones presentan mayor concentración de diagnósticos y cómo se distribuyen los casos por sexo y edad?	Alta	Datos Abiertos EsSalud, Ubigeo INEI
Gestión Táctica	Dirección de Prevención y Promoción	¿Cuál es la edad promedio al momento del diagnóstico y cómo puede orientar las campañas preventivas?	Media	Datos Abiertos EsSalud
Gestión Táctica	Coordinación de Control Médico	¿Cuánto tiempo pasa, en promedio, entre el diagnóstico inicial y el primer control posterior?	Alta	Registros hospitalarios, Datos Abiertos EsSalud

Gestión Operativa	Jefatura Hospitalaria	¿Cuál es la cobertura de diagnósticos lograda por cada red hospitalaria respecto a su población asegurada?	Alta	Datos Abiertos EsSalud, RENIPRESS
Gestión Operativa	Unidad de Recursos Humanos	¿Qué nivel de disponibilidad de especialistas se tiene en relación con los pacientes diagnosticados?	Media	RRHH EsSalud, RENIPRESS

1.2.KPI's definidos

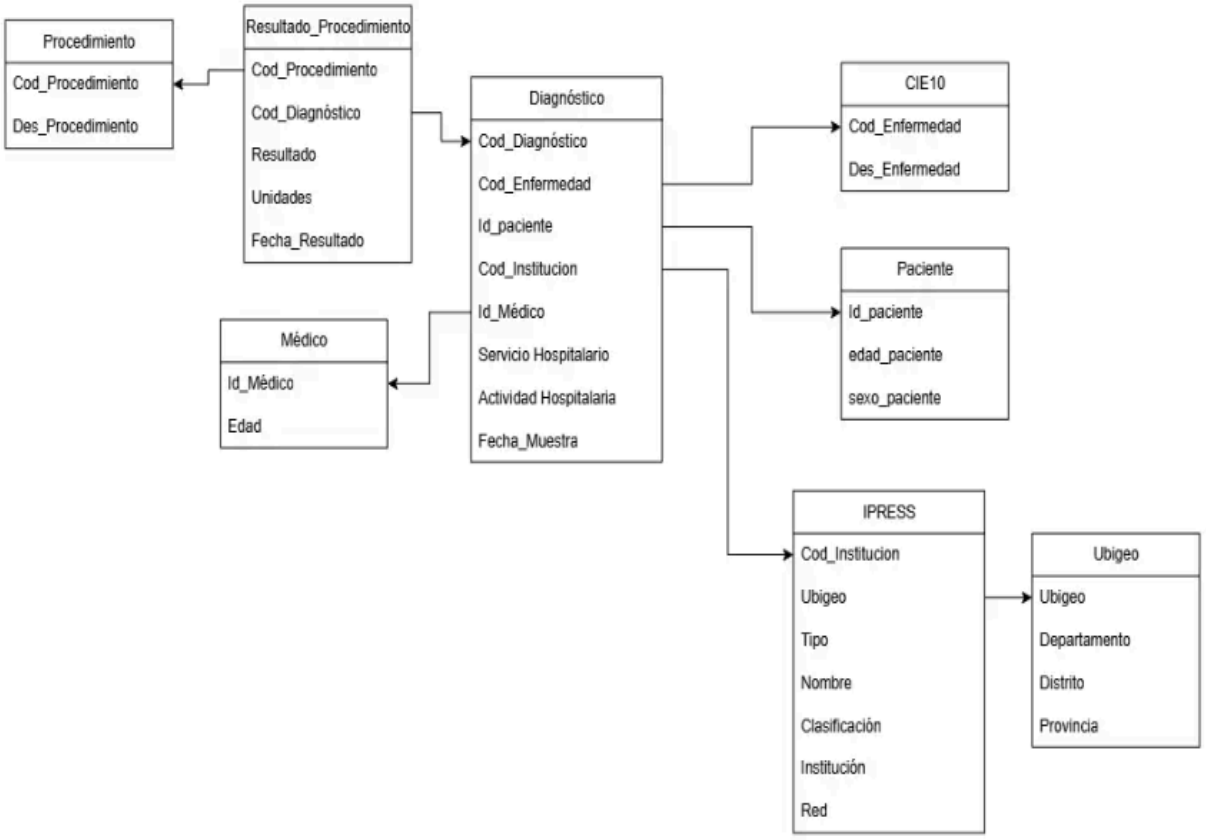
Nombre KPI	Descripción	Fórmula	Unidad	Periodicidad	Fuente de Datos	Responsable
Tasa de diagnósticos por 1000 habitantes	Mide la frecuencia de diagnósticos en la población general.	(N° diagnósticos ÷ Población total) × 1000	Casos por 1000 hab.	Mensual	Datos Abiertos EsSalud, INEI	Dirección Estratégica
Tendencia de crecimiento de casos	Variación porcentual de diagnósticos en el tiempo para anticipar zonas de riesgo.	((Casos periodo actual – Casos periodo anterior) ÷ Casos periodo anterior) × 100	%	Mensual	Datos Abiertos EsSalud	Coordinación Epidemiológica
Concentración	Porcentaje	(Casos	%	Trimestral	Datos	Gerencias

geográfica de diagnósticos	de casos concentrados en cada región.	en región ÷ Total casos) × 100			Abiertos EsSalud, Ubigeo INEI	Regionales
Edad promedio de diagnóstico	Edad media de los pacientes al momento del diagnóstico.	Σ (Edad de diagnóstico) ÷ N pacientes	Años	Trimestral	Datos Abiertos EsSalud	Dirección de Prevención
Tiempo promedio entre diagnóstico y control	Días transcurridos entre el diagnóstico inicial y el primer control.	Σ (Fecha control – Fecha diagnóstico) ÷ N pacientes	Días	Semestral	Datos hospitalarios	Coordinación de Control Médico
Distribución por sexo de diagnósticos	Diferencia en proporción de diagnósticos por sexo.	(Casos sexo ÷ Total casos) × 100	%	Trimestral	Datos Abiertos EsSalud	Unidad de Estadística
Índice de diagnósticos en población joven (<40 años)	Casos de hipertensión, diabetes y obesidad en menores de 40 años.	(Casos <40 años ÷ Total casos) × 100	%	Trimestral	Datos Abiertos EsSalud	Dirección de Epidemiología
Disponibilidad de especialistas por paciente	Relación médicos especialistas/pacientes diagnosticados.	N Médicos ÷ N pacientes atendidos	Ratio	Trimestral	RRHH EsSalud, RENIPRESS	Jefatura de Recursos Humanos
Cobertura de diagnósticos por red hospitalaria	Porcentaje de pacientes diagnosticados sobre	(Pacientes diagnosticados ÷ Población	%	Trimestral	RENIPRESS, Datos Abiertos	Gerencias de Red

	asegurados en la red.	n asegurada en red) × 100			EsSalud	
Variabilidad regional de diagnósticos	Diferencia entre la región con mayor y menor tasa de diagnósticos.	Máx(tasa diagnósticos) – Mín(tasa diagnósticos)	%	Trimestral	Datos Abiertos EsSalud, Ubigeo INEI	Dirección Estratégica

1.3. Modelo conceptual preliminar

Inserte aquí el diagrama conceptual (Star Schema) mostrando hechos y dimensiones.



1.4. Inventario de fuentes OLTP:

Sistema	Área usuaria	Tipo	Tecnología	Frecuencia actualización	Observaciones
Sistema de Planeamiento Estratégico – CEPLAN	Planeamiento Estratégico y Estadística	OLTP(maestro de referencia geográfica)	Archivo XLS	Anual / Semestral (cuando se actualizan	Incluye descripciones y metadatos de cada ubigeo; sirve como

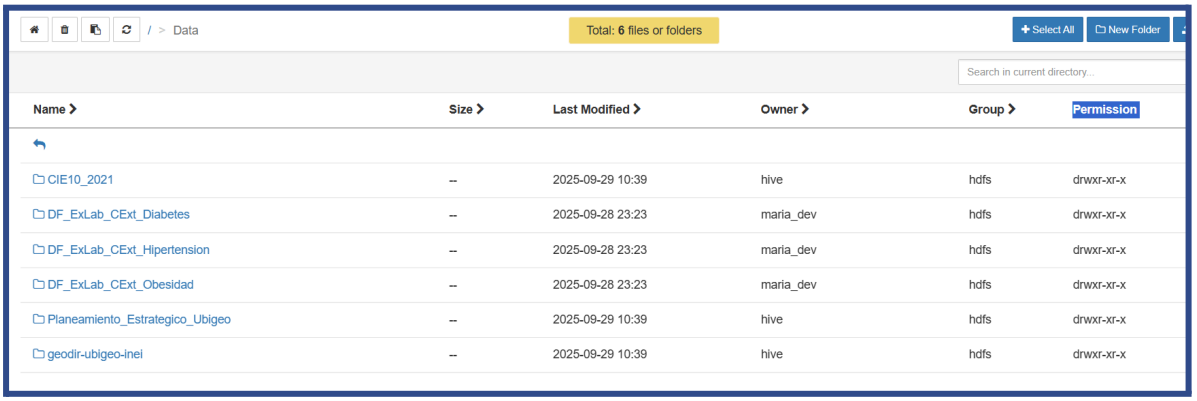
				planes y ubigeos)	insumo referencial para consolidar datos de salud y otras áreas del Estado.
Sistema de Consultas Externas – Datos Abiertos EsSalud	Red hospitalaria / Consulta externa	OLTP (registros médicos de atenciones y diagnósticos)	Archivos CSV	Diario	Incluye registros de exámenes de laboratorio en consultas externas con diagnóstico de diabetes, hipertensión y obesidad (2020–2024).
Sistema de Codificación Geográfica – Ubigeo (INEI)	Planeamiento / Estadística	OLTP (catálogo maestro geográfico)	Archivo CSV	Eventual / según necesidad (referencia en reportes, cruces de información y georreferencia ción)	Contiene códigos y descripciones de distrito, provincia y departamento. Es la tabla de referencia estándar para enlazar información de salud con ubicación geográfica oficial.
Sistema de Clasificación Internacional de Enfermedades – CIE10	Áreas médicas y de estadística en EsSalud / SuSalud	OLTP (catálogo maestro de enfermedade s)	Archivo CSV	Eventual / según necesidad (referencia en diagnósticos y reportes)	Contiene el código y descripción de enfermedades según la Clasificación Internacional de Enfermedades (CIE10).
RENIPRESS – Registro Nacional de Instituciones Prestadoras de Servicios de Salud	SuSalud / Gestión de infraestructur a hospitalaria	OLTP (maestro de establecimie ntos de salud)	Archivo XLS	Eventual / según necesidad (planeamiento, auditoría y gestión de red)	Contiene información oficial de todos los hospitales, clínicas y centros de salud registrados en el Perú.

2. EVIDENCIA TÉCNICA

2.1. Implementación de Hortonworks

- Archivos cargados en HDFS (/data/raw/).

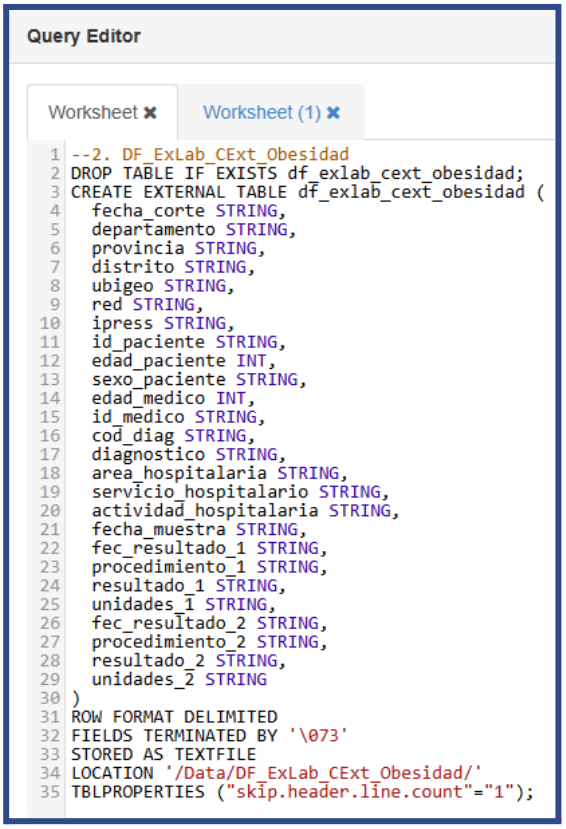
- Datos cargados en Ambari:



- Datos cargados, vista desde la consola:

```
[maria_dev@sandbox-hdp ~]$ hdfs dfs -ls /Data/
Found 6 items
drwxr-xr-x   - hive      hdfs          0 2025-09-29 15:39 /Data/CIE10_2021
drwxr-xr-x   - maria_dev hdfs          0 2025-09-29 04:23 /Data/DF_ExLab_CExt_Diabetes
drwxr-xr-x   - maria_dev hdfs          0 2025-09-29 04:23 /Data/DF_ExLab_CExt_Hipertension
drwxr-xr-x   - maria_dev hdfs          0 2025-09-29 04:23 /Data/DF_ExLab_CExt_Obesidad
drwxr-xr-x   - hive      hdfs          0 2025-09-29 15:39 /Data/Planeamiento_Estrategico_Ubigeo
drwxr-xr-x   - hive      hdfs          0 2025-09-29 15:39 /Data/geodir-ubigeo-inei
[maria_dev@sandbox-hdp ~]$ hdfs dfs -ls /Data/CIE10_2021/
Found 1 items
-rw-r--r--   1 maria_dev hdfs     913690 2025-09-29 03:26 /Data/CIE10_2021/CIE10_2021.csv
```

- Scripts CREATE EXTERNAL TABLE en Hive.
  - Creando la tabla de Obesidad

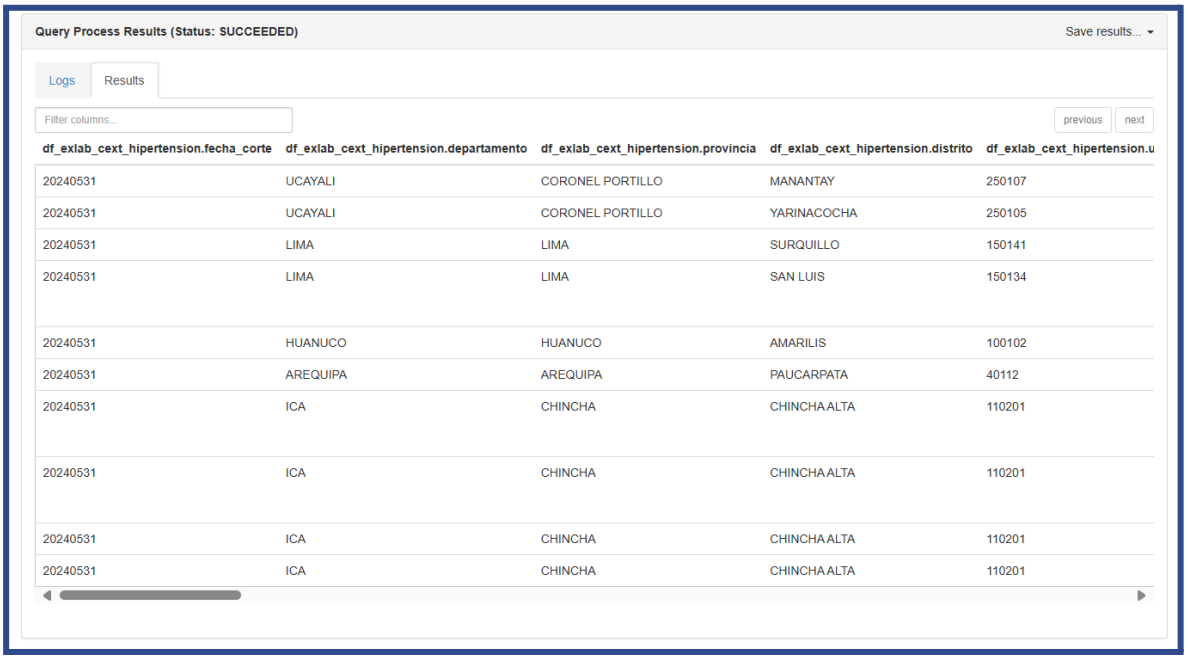


- Resultados de consulta simple
  - Consulta

Worksheet

```
1 SELECT *
2 FROM df_exlab_cext_hipertension
3 LIMIT 10;
```

- Resultado:



Query Process Results (Status: SUCCEEDED) — Save results...

Logs | Results

Filter columns... — previous | next

df_exlab_cext_hipertension.fecha_corte	df_exlab_cext_hipertension.departamento	df_exlab_cext_hipertension.provincia	df_exlab_cext_hipertension.distrito	df_exlab_cext_hipertension.u
20240531	UCAYALI	CORONEL PORTILLO	MANANTAY	250107
20240531	UCAYALI	CORONEL PORTILLO	YARINACOCHA	250105
20240531	LIMA	LIMA	SURQUILLO	150141
20240531	LIMA	LIMA	SAN LUIS	150134
20240531	HUANUCO	HUANUCO	AMARILIS	100102
20240531	AREQUIPA	AREQUIPA	PAUCARPATA	40112
20240531	ICA	CHINCHA	CHINCHA ALTA	110201
20240531	ICA	CHINCHA	CHINCHA ALTA	110201
20240531	ICA	CHINCHA	CHINCHA ALTA	110201
20240531	ICA	CHINCHA	CHINCHA ALTA	110201

ENTREGABLES:
INFORME DETALLADO DEL PRACTICA: con cada producto entregable

RÚBRICA DE CALIFICACIÓN

Preguntas de negocio	Relevantes, claras y alineadas con los objetivos; bien justificadas.	Preguntas adecuadas, aunque faltan detalles menores.	Preguntas poco claras o incompletas; no todas alineadas al negocio.
Definición de KPIs	Fichas técnicas completas, con fórmula, frecuencia y responsable definidos.	KPIs bien definidos, con pequeñas fallas en ficha técnica.	KPIs incompletos o poco consistentes.
Modelo conceptual	Diagrama estrella completo, con hechos y dimensiones bien definidos.	Modelo adecuado, con fallas menores en jerarquías o atributos.	Modelo incompleto, con dimensiones o hechos poco claros.
Inventario OLTP	Cobertura completa de sistemas, tablas y campos clave; bien documentado.	Inventario adecuado, con algunos vacíos menores.	Inventario incompleto o con errores de documentación.
Evidencia técnica Hortonworks	Ingesta en HDFS + tablas Hive (raw) correctas + consulta validada + Carga en GitHub.	Ingesta y tablas funcionales, con fallas menores + Carga en GitHub.	Evidencia incompleta o sin validación+ Carga en GitHub.
Presentación y redacción	Documento claro, ordenado, con redacción y ortografía correctas, cargado en el repositorio GitHub.	Documento entendible, con errores menores de forma, cargado en el repositorio GitHub.	Documento poco claro o con fallas frecuentes de redacción, cargado en el repositorio GitHub.