

Inteligencia de Negocios

EsSalud



EsSalud



SEGURO SOCIAL DE SALUD



SEDES

Sede Principal

Av. Arenales, Lima



PRODUCTOS y CLIENTES

PRODUCTOS

1 Servicios Hospitalarios

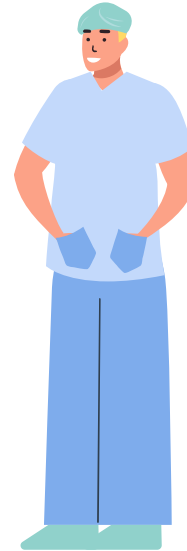
2 Servicio de Prevención de Enfermedades



CLIENTES

Asegurados 1

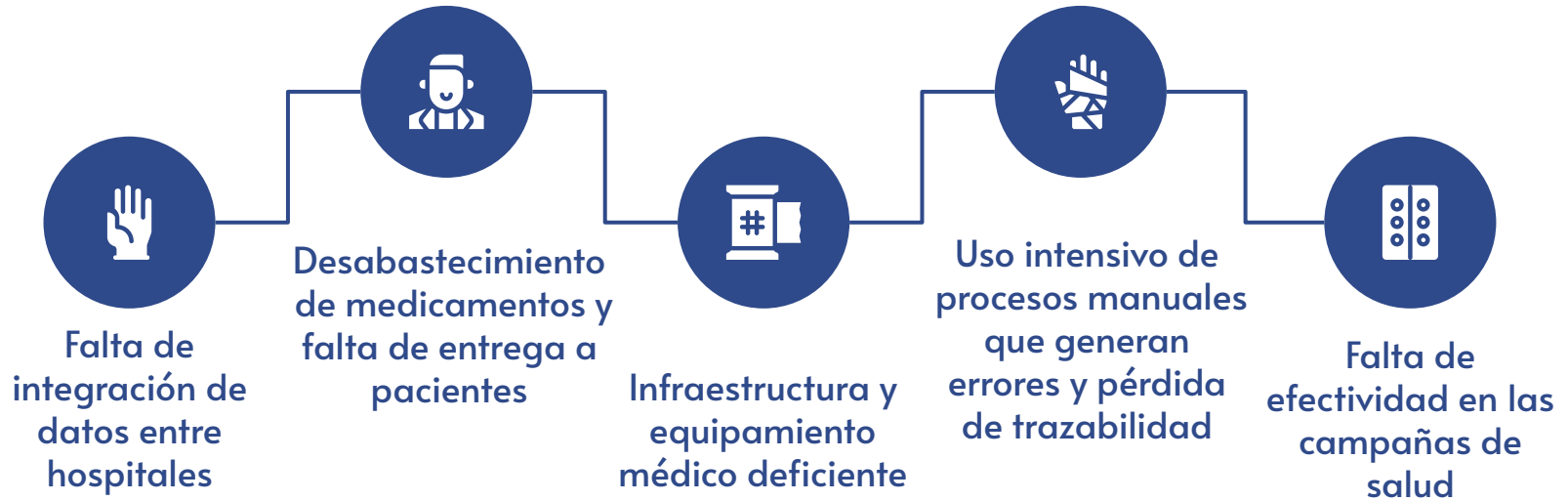
Población Regional 2



ORGANIGRAMA



Problemas



Necesidades de Información decisiones críticas

Nivel	Tipo de decision	Necesidad de información
Estratégico	Definir políticas de salud, asignar presupuesto nacional	Indicadores consolidados de incidencia y prevalencia, análisis de costos, cobertura de asegurados, efectividad de campañas de prevención
Táctico	Planificación regional y asignación de recursos	Reportes comparativos por hospital/región, ocupación de camas, stock de medicamentos, tiempo de espera promedio, indicadores de desempeño regional
Operativo	Gestión diaria y respuesta a emergencias	Datos en tiempo real: atenciones nuevas por día, resultados de laboratorio, disponibilidad de UCI, personal disponible, trazabilidad de pacientes

Problema elegido y justificación

Dificultad para anticipar la ocurrencia de enfermedades Diabetes, Hipertensión y Obesidad en ubicaciones específicas, limitando la planificación y preparación de recursos en EsSalud



Disponibilidad de data confiable para su análisis

Existen bases oficiales de EsSalud y del Estado que permiten construir modelos predictivos y anticipar la demanda de salud.



Impacto directo en la planificación y prevención

Predecir enfermedades ayuda a asignar recursos médicos y reducir la saturación hospitalaria y los costos.



Relación con otros problemas identificados

La falta de integración y los procesos manuales limitan la calidad de los modelos predictivos.



Alta relevancia social y estratégica

Anticipar brotes fortalece la confianza ciudadana y mejora la legitimidad institucional de EsSalud.

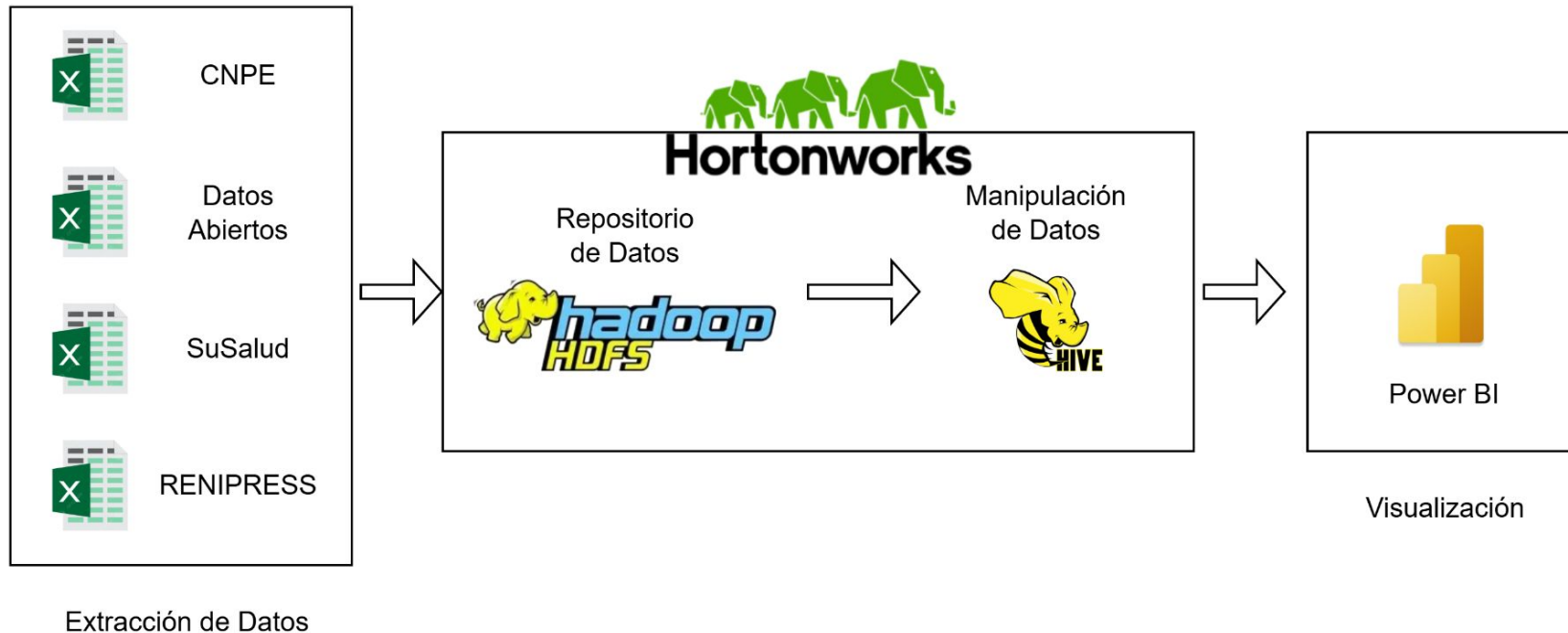
KPI's Iniciales

Nombre del KPI	Descripción	Formula	Unidad	Frecuencia
Tasa de diagnósticos por 1000 habitantes	Mide la frecuencia de diagnósticos de estas patologías en la población.	$(N^{\circ} \text{ diagnósticos} \div \text{Población total}) \times 1000$	Casos por 1000 hab.	Mensual
Tendencia de crecimiento de casos	Mide la variación porcentual de diagnósticos en el tiempo para anticipar zonas de riesgo.	$((\text{Casos periodo actual} - \text{Casos periodo anterior}) \div \text{Casos periodo anterior}) \times 100$	%	Mensual
Concentración geográfica de diagnósticos	(Casos en región \div Total de casos) \times 100	$(\text{Casos en región} \div \text{Total de casos}) \times 100$	%	Trimestral
Edad promedio de diagnóstico	Estima la edad promedio de los pacientes al momento del diagnóstico, útil para segmentar campañas.	$\Sigma (\text{Edad de diagnóstico}) \div N \text{ pacientes}$	Años	Trimestral
Tiempo promedio entre diagnóstico y control	Mide días transcurridos entre diagnóstico inicial y primer control posterior.	$\Sigma (\text{Fecha control} - \text{Fecha diagnóstico}) \div N \text{ pacientes}$	Días	Semestral
Distribución por sexo de diagnósticos	Diferencia en proporción de diagnósticos por sexo (Paciente)	$(\text{Casos sexo} \div \text{Total casos}) \times 100$	%	Trimestral

KPI's Iniciales

Nombre del KPI	Descripción	Formula	Unidad de medida	Frecuencia
Índice de diagnósticos en población joven (<40 años)	Casos de hipertensión/diabetes/o besidad en menores de 40 años (Paciente + Diagnóstico).	$(\text{Casos } <40 \text{ años} \div \text{Total casos}) \times 100$	%	Trimestral
Disponibilidad de especialistas por paciente	Relación médicos especialistas/pacientes diagnosticados (Médico + Diagnóstico).	$N \text{ Médicos} \div N \text{ pacientes atendidos}$	Ratio	Trimestral
Cobertura de diagnósticos por red hospitalaria	Casos diagnosticados en cada red vs asegurados en esa red (IPRESS + Ubigeo).	$(\text{Pacientes diagnosticados} \div \text{Población asegurada en red}) \times 100$	%	Trimestral
Variabilidad regional de diagnósticos	Diferencia entre la región con mayor y menor tasa de diagnósticos (Ubigeo + Diagnóstico).	$\text{Máx}(\text{tasa diagnósticos}) - \text{Mín}(\text{tasa diagnósticos})$	%	Trimestral

Arquitectura de datos



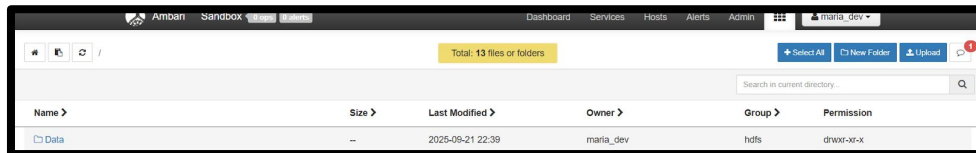
Fuente de datos

Repositorio	Archivo	Descripción
Centro Nacional de Planeamiento Estratégico	Datos-planeamiento-estrategico.xlsx	Incluye descripciones de cada ubigeo
Datos Abiertos Perú	DF_ExLab_CExt_Diabetes.csv	Contiene todas las consultas externas con diagnósticos de diabetes y sus detalles desde el 2020 hasta el 2024
	DF_ExLab_CExt_Hipertension.csv	Contiene todas las consultas externas con diagnósticos de hipertensión y sus detalles desde el 2020 hasta el 2024
	DF_ExLab_CExt_Obesidad.csv	Contiene todas las consultas externas con diagnósticos de obesidad y sus detalles desde el 2020 hasta el 2024
	Ubigeo.csv	Contiene el código de ubigeo, su distrito, provincia y departamento
SuSalud	CIEIO.csv	Contiene el código de todas la enfermedades y su descripción
RENIPRESS	USLRC20250923.xls	Contiene información de todos los hospitales del Perú

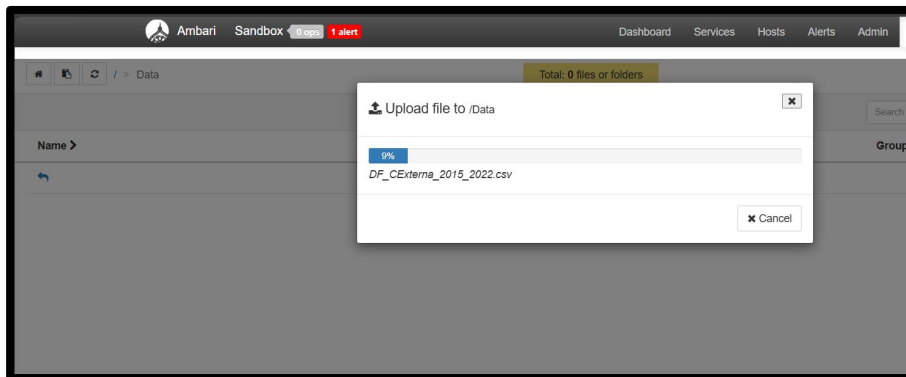
Ingesta de datos



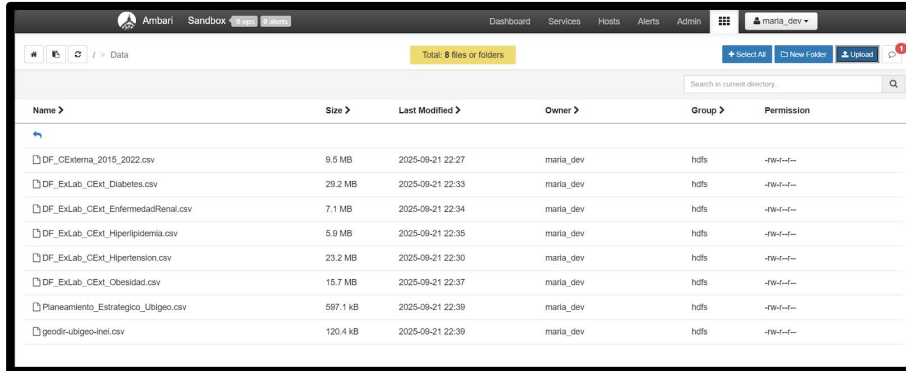
CREANDO LA CARPETA DE DATA:



SUBIENDO LOS ARCHIVOS CSV:



ARCHIVOS SUBIDOS:



Almacenamiento de datos

HDFS:

Es el sistema de almacenamiento distribuido de Hadoop, diseñado para guardar grandes volúmenes de datos dividiéndolos en bloques y replicándolos en distintos nodos, garantizando alta disponibilidad, tolerancia a fallos y acceso eficiente a la información.



HDFS Disk Usage

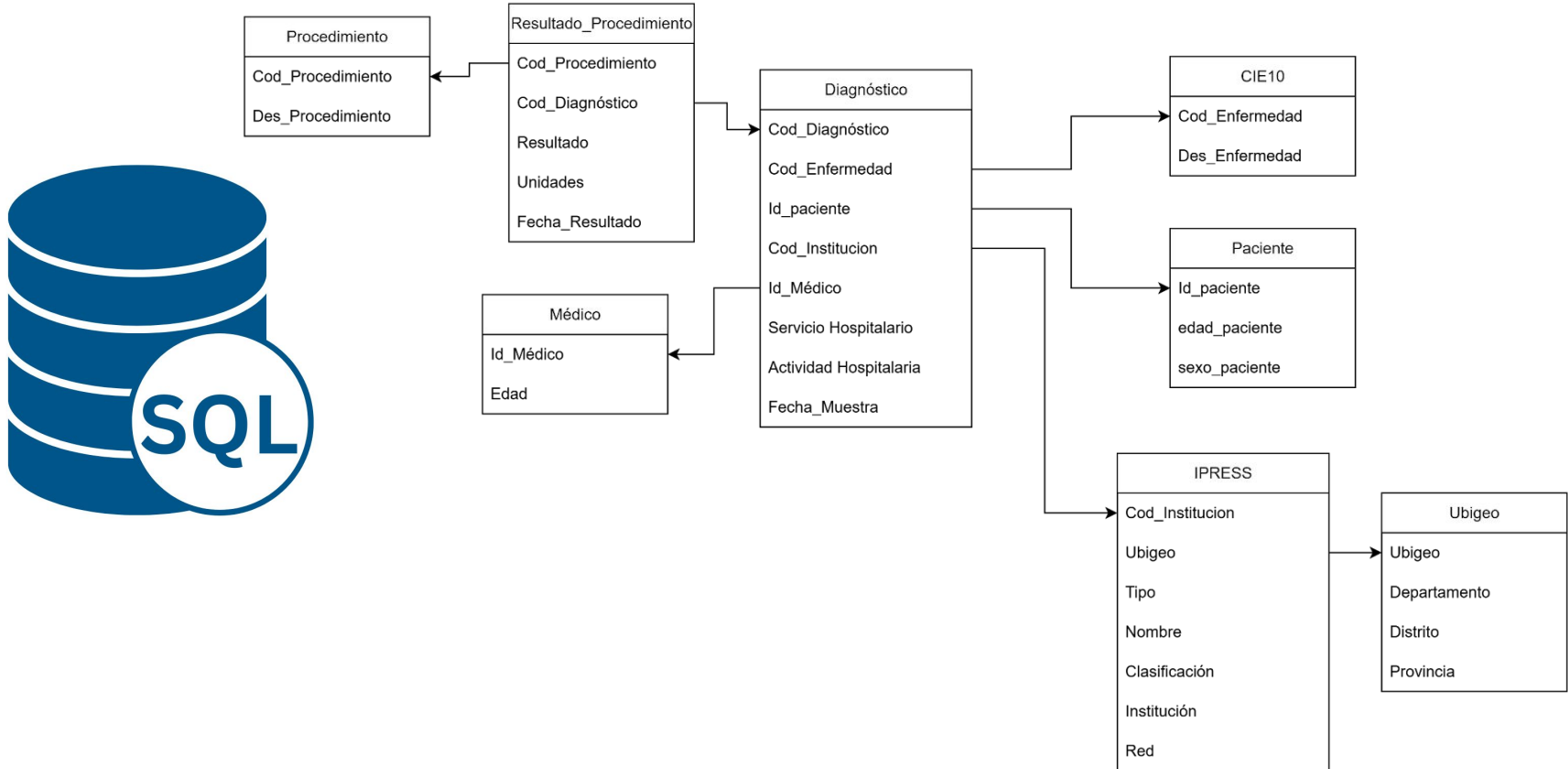


HDFS Disk Usage



DFS used
2.2 GB (2.05%)
non DFS used
20.0 GB (18.85%)
remaining
78.4 GB (73.96%)

Transformación de datos



Utilización de datos

Hive:

- Es una herramienta de *data warehouse* que permite consultar y analizar datos almacenados en Hadoop mediante un lenguaje similar a SQL (HiveQL).
- Facilita a los analistas trabajar con datos masivos sin necesidad de programar en Java o MapReduce.



Worksheet

```
1 USE essalud_db;
2
3 CREATE EXTERNAL TABLE IF NOT EXISTS consulta_externa (
4     PERIODO STRING,
5     SES_DESRED STRING,
6     SES_DESCAS STRING,
7     DES_GRU STRING,
8     DES_ACT STRING,
9     DES_SER STRING,
10    CANTIDAD STRING,
11    DEPARTAMENTO STRING,
12    PROVINCIA STRING,
13    DISTRITO STRING,
14    UBIGEO STRING,
15    FECHA_CORTE STRING
16 )
17 ROW FORMAT DELIMITED
18 FIELDS TERMINATED BY ','
19 STORED AS TEXTFILE
20 LOCATION '/Data/DF_CExterna_2015_2022'
```

Execute

Explain

Upload

Save as...

Query Process Results (Status: SUCCEEDED)

GRACIAS