

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

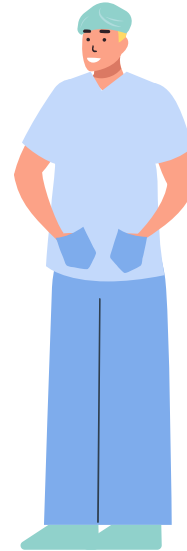
3 Aseguramiento de Salud



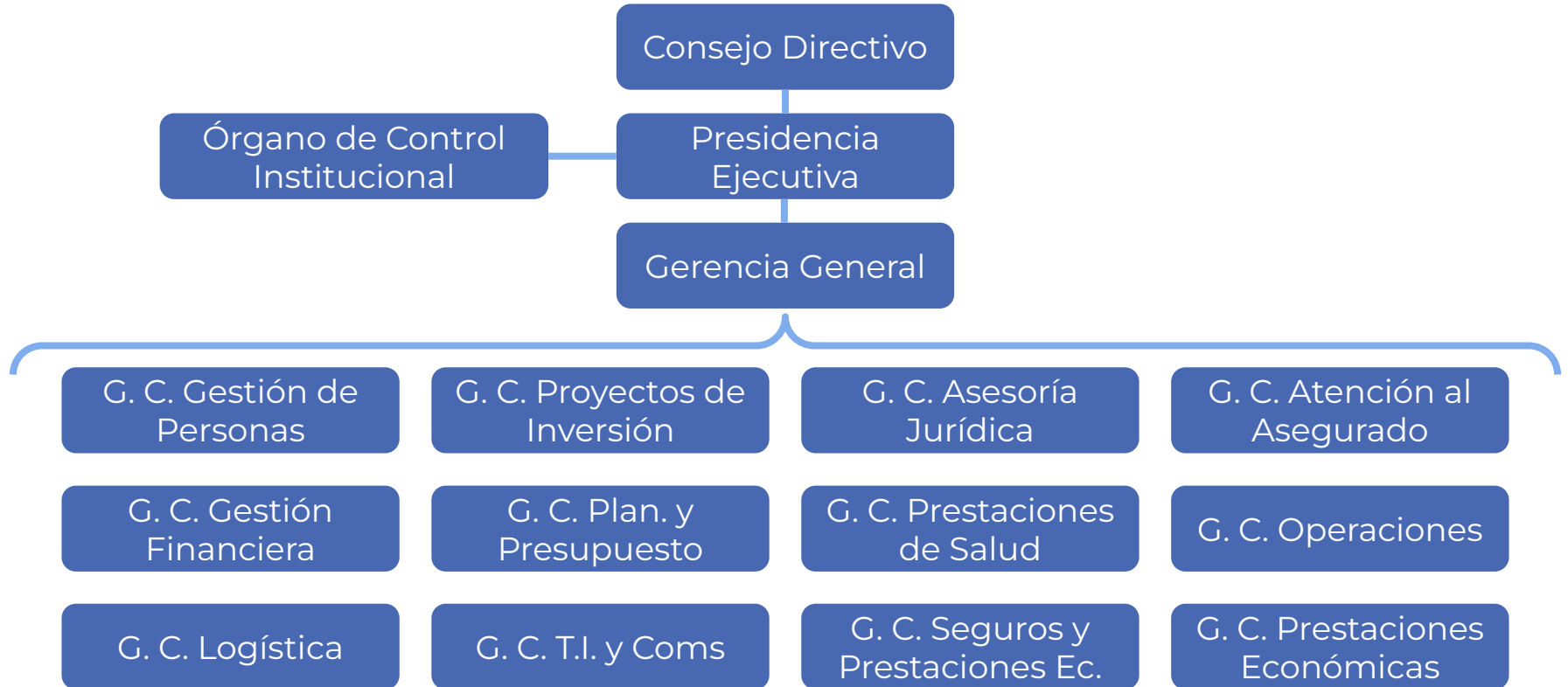
CLIENTES

Asegurados 1

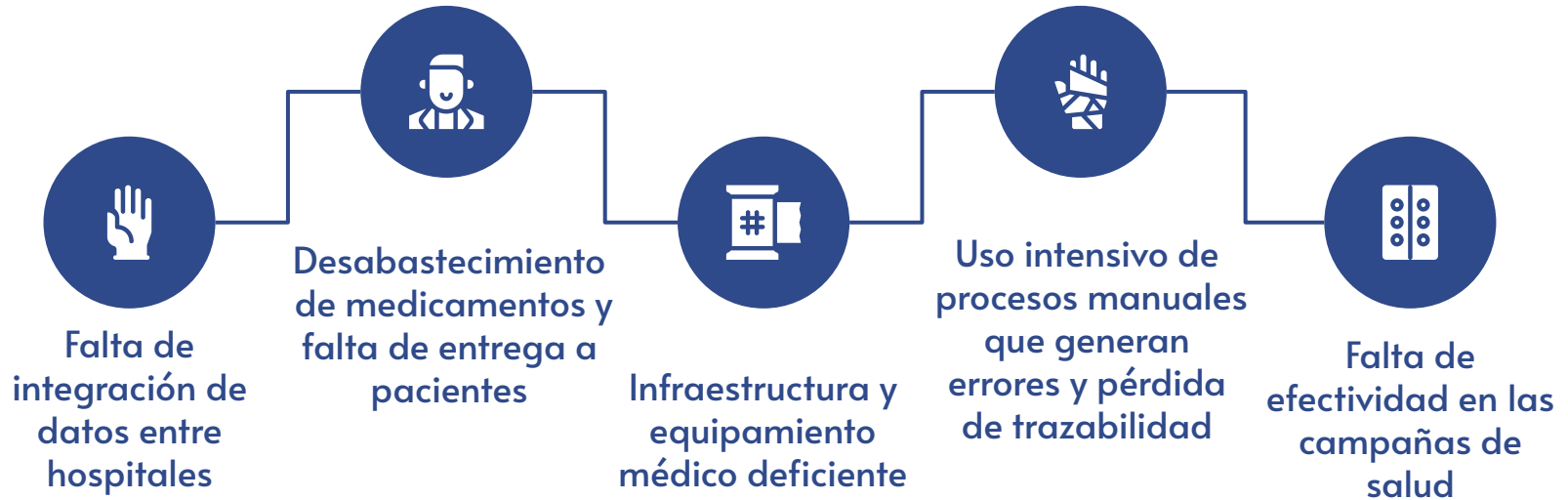
Población Regional 2



ORGANIGRAMA



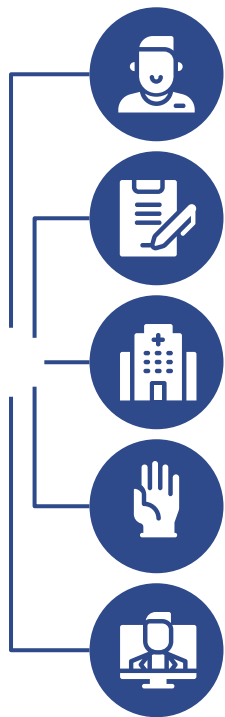
Problemas



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



Problema

Representa un punto crítico donde convergen la gestión de datos, toma de decisiones

Disponibilidad data

Datos abiertos del gobierno del Perú, se encuentran la base de datos relacionadas con campañas

Impacto directo a la salud pública

Una campaña de salud inefectiva se traduce en menor cobertura de población, falta de prevención, mas costos

Relación con otros problemas

Falta de integración de datos, reportes tardíos y ausencia de indicadores son factores para la ineficacia de campañas

Alta relevancia social

La campañas son la cara visible frente a la ciudadanía

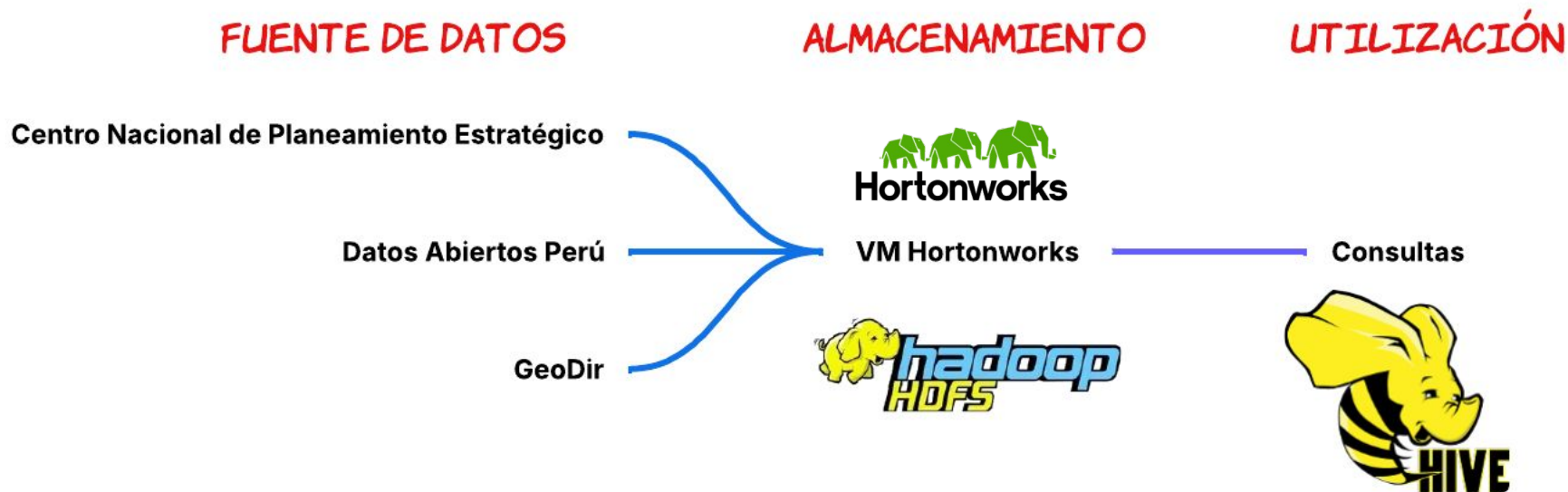
KPI's Iniciales

Nombre del KPI	Descripción	Formula	Unidad de medida	Frecuencia
Tasa de ocupación hospitalaria	Mide el nivel de utilización de camas en hospitales	$(\text{Camas ocupadas} \div \text{Camas disponibles}) \times 100$	%	Diario
Tiempo promedio de espera en consulta	Tiempo promedio desde la solicitud hasta la atención	$\Sigma (\text{hora atención} - \text{hora solicitud}) \div N \text{ pacientes}$	Horas	Diario
Cobertura de campañas preventivas	Porcentaje de población objetivo alcanzada en campañas	$(\text{Población atendida} \div \text{Población objetivo}) \times 100$	%	Mensual
Oportunidad de reportes epidemiológicos	Rapidez en la entrega de reportes desde los establecimientos	$(\text{Reportes entregados a tiempo} \div \text{Total de reportes}) \times 100$	%	Semanal
Incidencia de enfermedades priorizadas	Casos nuevos detectados de enfermedades críticas	$(\text{Nuevos casos} \div \text{Población total}) \times 1000$	Casos por 1000 hab.	Mensual
Tasa de errores en registros de campaña de salud	Proporción de inconsistencias en registros de campañas	$(\text{Errores detectados} \div \text{Registros revisados}) \times 100$	%	Mensual

KPI's Iniciales

Nombre del KPI	Descripción	Formula	Unidad de medida	Frecuencia
Efectividad de alertas tempranas.	Porcentaje de alertas tempranas correctas	$(\text{Alertas correctas} \div \text{Alertas totales}) \times 100$	%	Mensual
Disponibilidad de medicamentos críticos	Stock real frente a lo planificado	$(\text{Stock disponible} \div \text{Stock planificado}) \times 100$	%	Semanal
Ratio de personal por paciente	Relación entre personal médico y pacientes atendidos	$(\text{N médicos} \div \text{N pacientes atendidos})$	Ratio	Diario
Nivel de integración de datos de campañas	Grado de centralización de datos de campañas.	$(\text{Campañas con datos integrados} \div \text{Total de campañas}) \times 100$	%	Trimestral

Arquitectura de datos



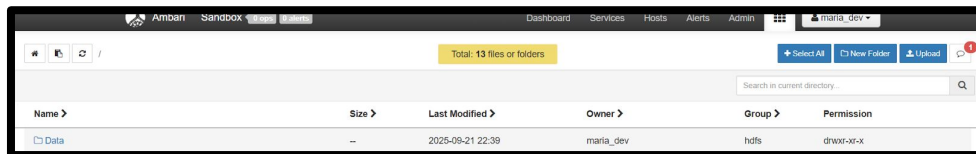
Fuente de datos

Repositorio	Archivo	Descripción
Centro Nacional de Planeamiento Estratégico	Datos-planeamiento-estrategico.xlsx	Incluye descripciones de cada ubigeo
Datos Abiertos Perú	DF_CExterna_2015_2022.csv	Contiene los datos de todas las consultas externas realizadas entre 2015 y 2022
	DF_ExLab_CExt_Diabetes.csv	Contiene todas las consultas externas con diagnósticos de diabetes y sus detalles
	DF_ExLab_CExt_EnfermedadRenal.csv	Contiene todas las consultas externas con diagnósticos de enfermedades renales y sus detalles
	DF_ExLab_CExt_Hiperlipidemia.csv	Contiene todas las consultas externas con diagnósticos de hiperlipidemia y sus detalles
	DF_ExLab_CExt_Hipertension.csv	Contiene todas las consultas externas con diagnósticos de hipertensión y sus detalles
	DF_ExLab_CExt_Obesidad.csv	Contiene todas las consultas externas con diagnósticos de obesidad y sus detalles
GeoDir	Geodir_Ubigeo_Inei.xlsx	Contiene detalle de todos los ubigeos con sus condiciones económicas y demográficas

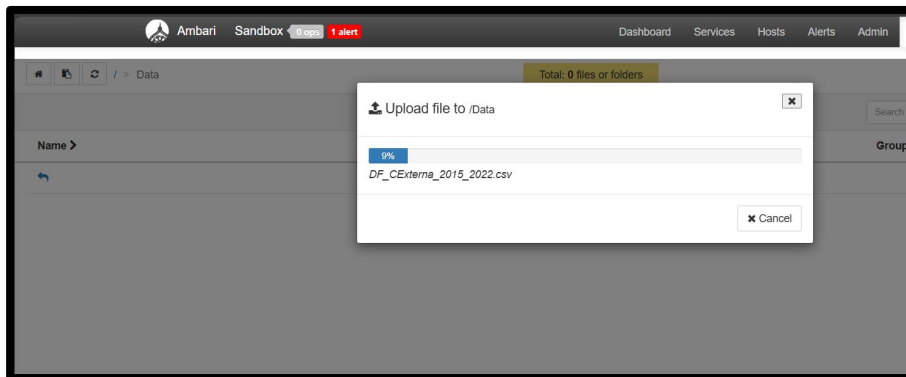
Ingesta de datos



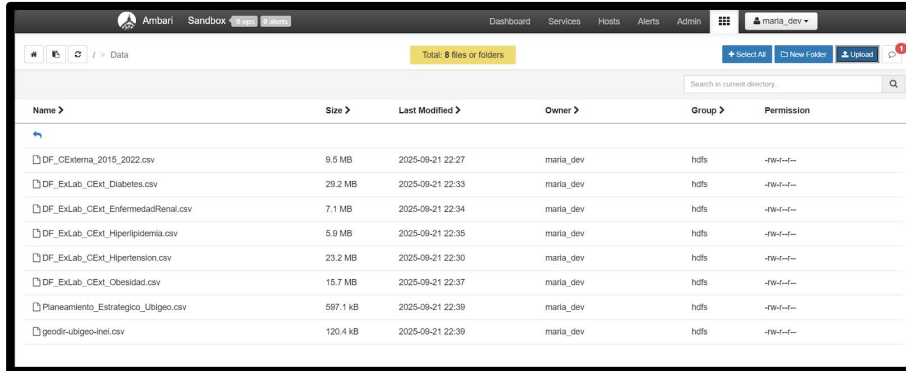
CREANDO LA CARPETA DE DATA:



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.



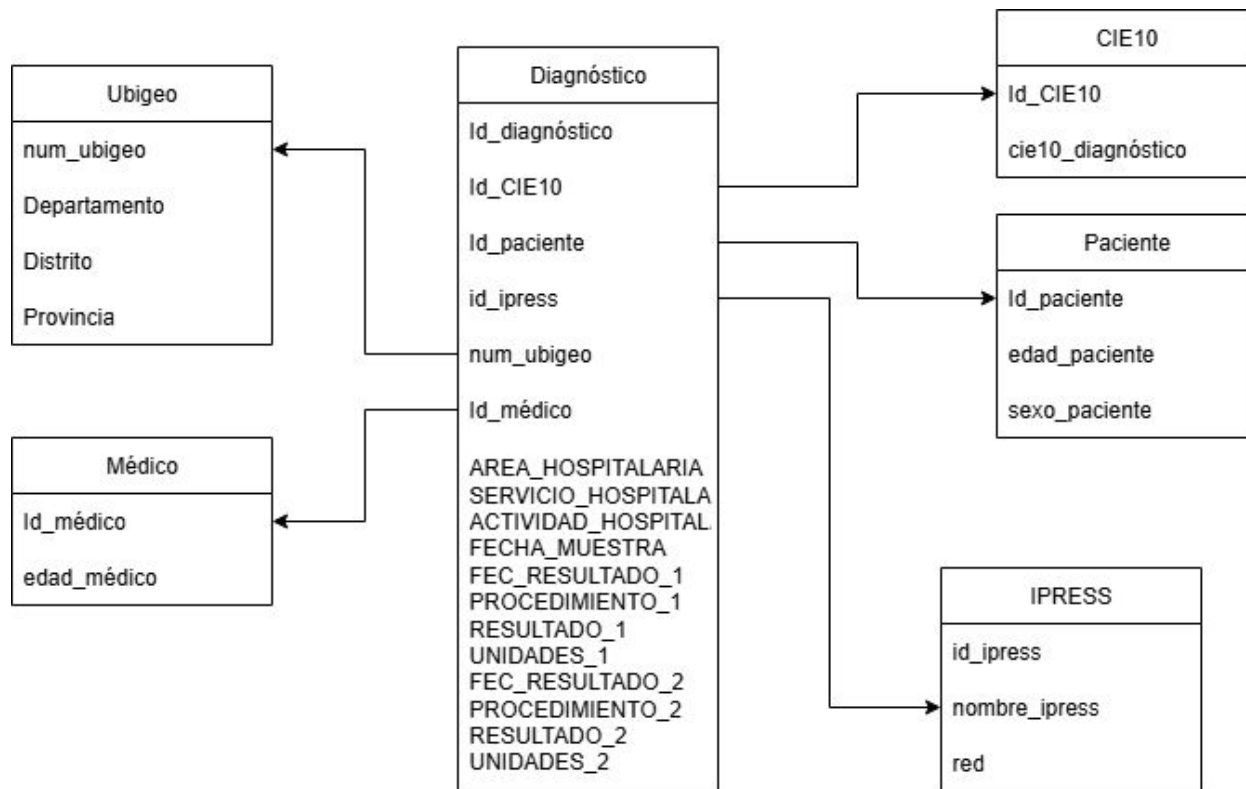
HDFS Disk Usage



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