

Inteligencia de Negocios

EsSalud

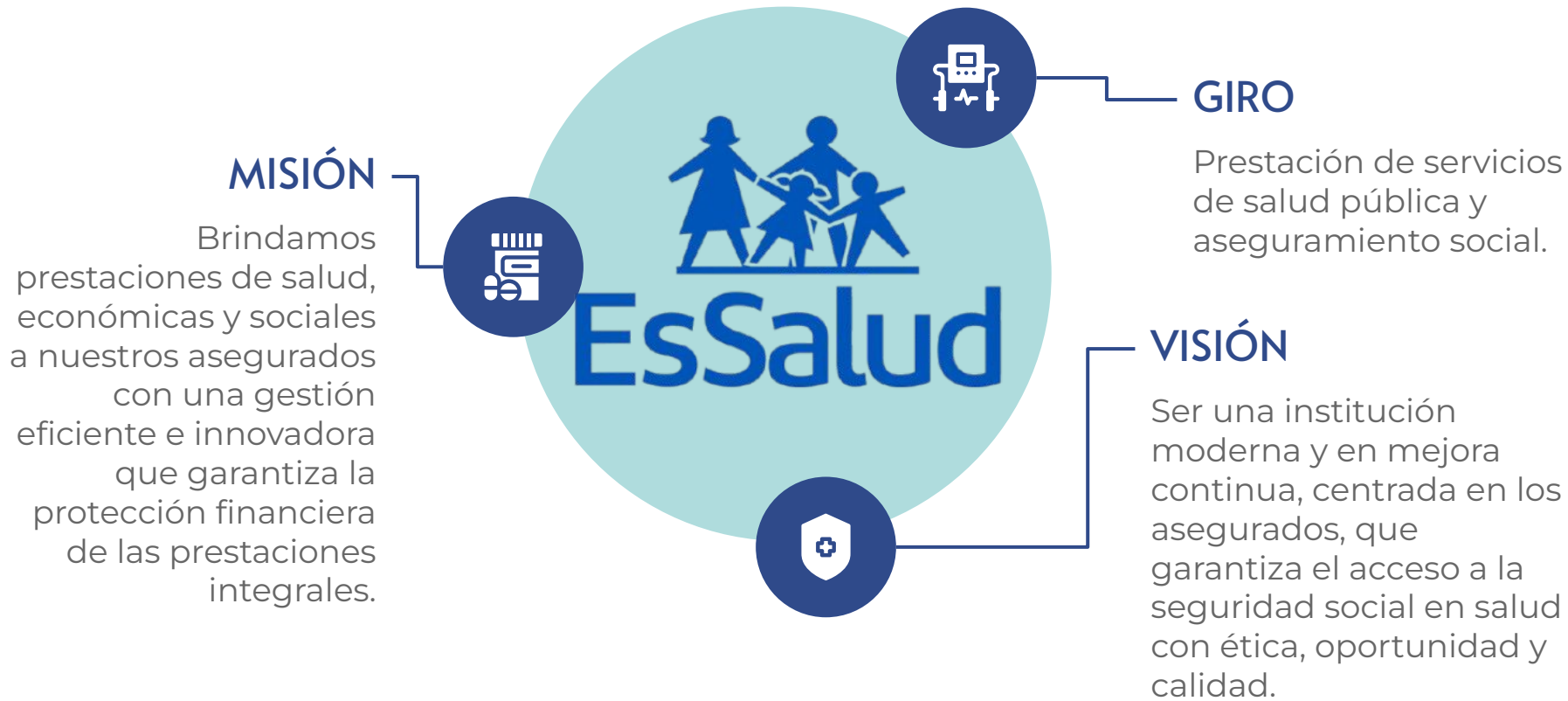


EsSalud



SEGURO SOCIAL DE SALUD

RUC: 20131257750

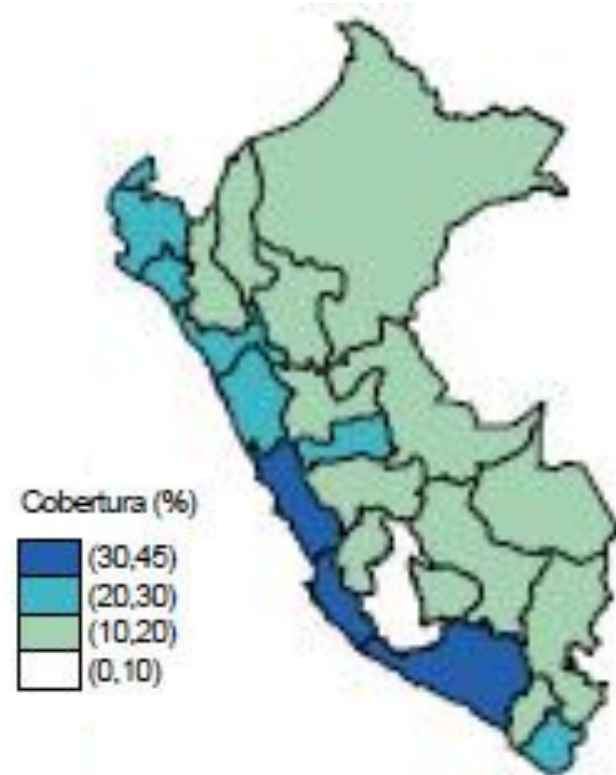


UBICACIÓN



Sede Principal

Av. Arenales, Lima



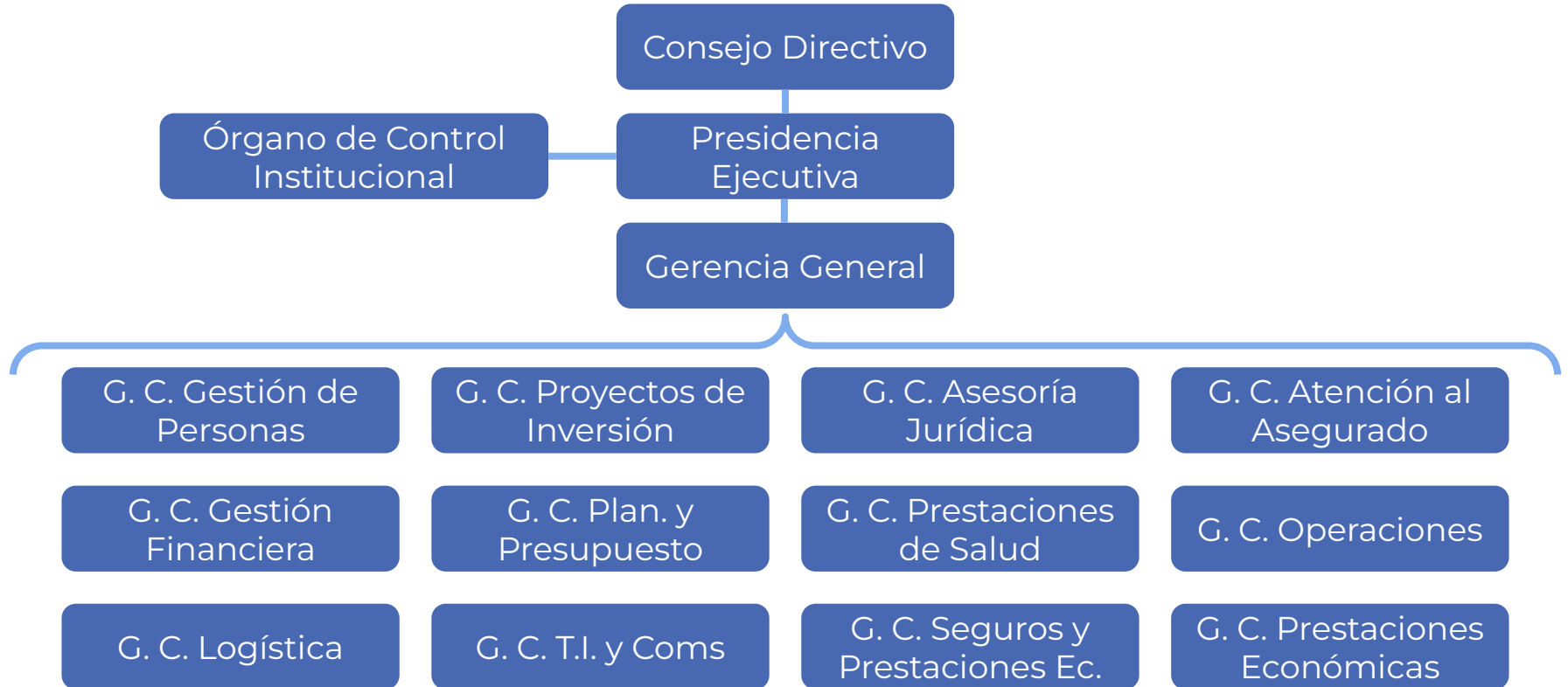
VALORES

Solidaridad
Universalidad
Igualdad
Integralidad
Ética institucional

OBJETIVOS

Mejorar la calidad de los servicios
Garantizar el acceso efectivo
Transformación institucional
Eficiencia y modernización de la gestión
Gestión ética e integridad institucional
Innovación tecnológica en salud

ORGANIGRAMA



CADENA DE VALOR

Actividades Primarias:

1. Administración de Recursos

2. Medición del Riesgo

3. Servicios y Calidad al asegurado

Administración de Fondos de Salud

Administración y Monitoreo

Fondos para ejecución operativa y financiera

Prestadora de Salud

Económicos y Sociales

Satisfacción del Asegurado y Ciudadano

Margen

Margen

Procesos de Soporte:

Planificación Estratégica y Presupuesto

Administración de Riesgo / Análisis de Portafolio

Administración de la Gestión Financiera

Registro Contable

Mejora Continua de la entidad (GOP)

CEABE y Logística

Gestión de Tecnología de Información y Comunicación

Administración de la Gestión de las Personas

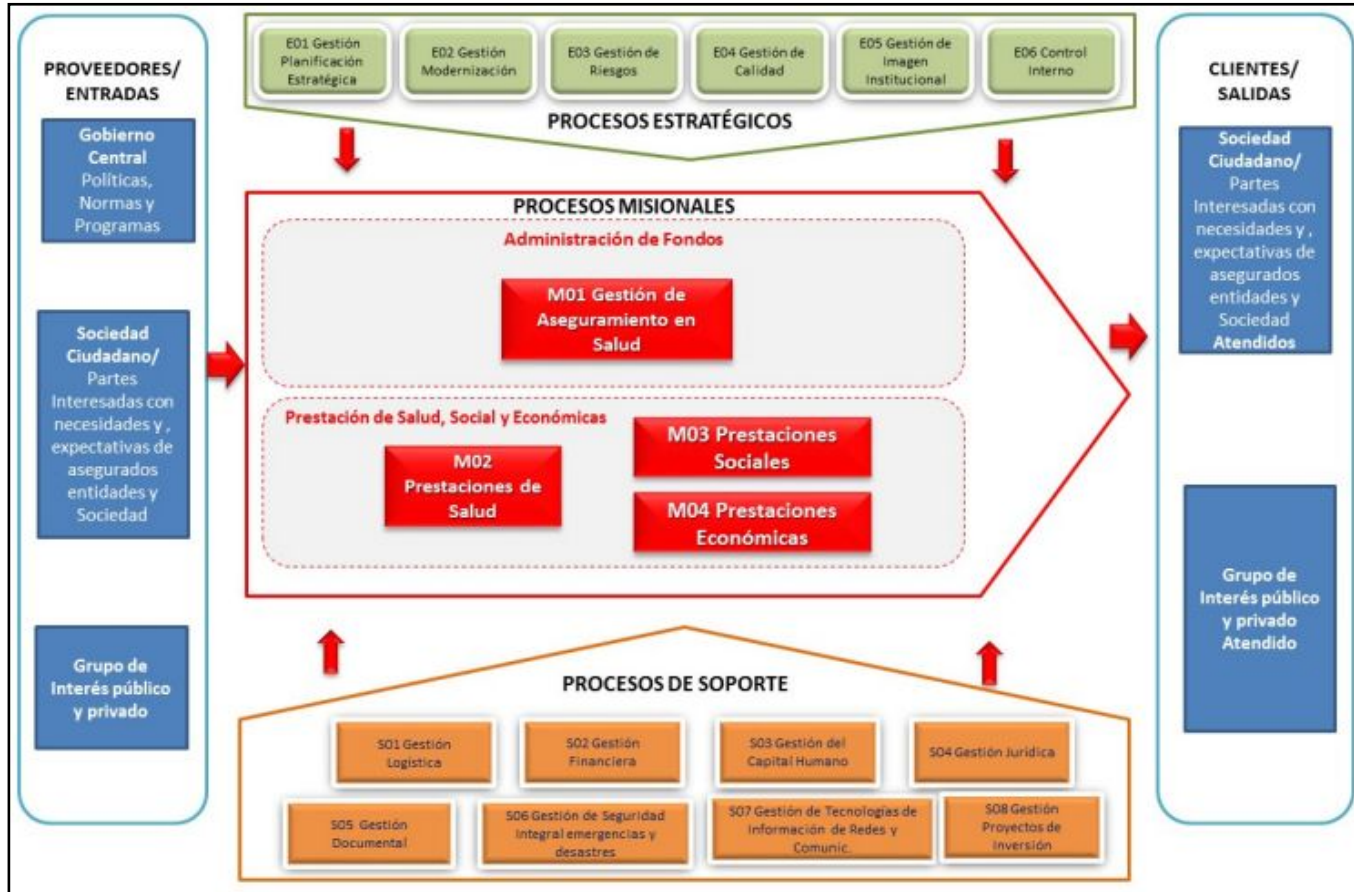
Administración de la Gestión de Asesoría Jurídica

Infraestructura y Gestión en Proyectos de Inversión

Cooperación Internacional

Gestión Documental

MAPA DE PROCESOS



PRODUCTOS y CLIENTES

PRODUCTOS

1 Servicios Hospitalarios

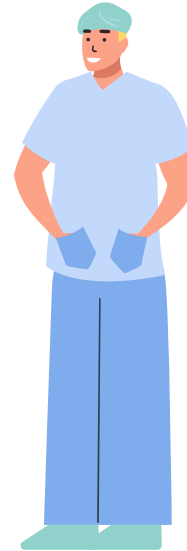
2 Servicio de Prevención de Enfermedades



CLIENTES

Asegurados 1

Población Regional 2



ANALISIS FODA

FORTALEZAS	OPORTUNIDADES
<p>Datos históricos para modelos predictivos.</p> <p>Experiencia previa en predicción de demanda.</p> <p>Procesos normados de distribución de camas.</p>	<p>Uso de técnicas de series temporales y machine learning.</p> <p>Cooperación con entidades académicas.</p> <p>Impulso gubernamental a la transformación digital.</p>
DEBILIDADES	AMENAZAS
<p>Falta de integración de datos asistenciales.</p> <p>Cultura reactiva, poca planificación predictiva.</p> <p>Procesos de camas no orientados a anticipación.</p>	<p>Epidemias o desastres que elevan la demanda.</p> <p>Saturación en emergencias (caso Almenara).</p> <p>Brecha digital que limita el acceso equitativo.</p>

Problemas



PRIORIZACIÓN DE PROBLEMAS

PROBLEMA	IMPACTO EN MISIÓN	ESFUERZO PARA RESOLVER	PRIORIDAD
Incapacidad para anticipar demanda de emergencias / hospitalizaciones	Muy alto	Alto	Alta
Falta de capacidad de reserva para picos	Alto	Medio-Alto	Alta
Programación de personal no ajustada	Medio-Alto	Medio	Media-Alta
Planificación de insumos con márgenes insuficientes	Medio	Medio	Media
Sistemas de información fragmentados	Medio	Alto	Media

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.

Necesidades de Información decisiones críticas

Nivel	Tipo de decision	Necesidad de información
Estratégico	<ul style="list-style-type: none">- Definir políticas y reservas de capacidad.- Aprobar inversiones en infraestructura y tecnología predictiva.- Establecer criterios de distribución regional de recursos.	<ul style="list-style-type: none">- Proyecciones nacionales/regionales de demanda.- Modelos predictivos y escenarios.- Indicadores históricos y análisis costo-beneficio.
Táctico	<ul style="list-style-type: none">- Planificar recursos por hospital o red.- Ajustar personal y camas.- Activar reservas o medidas de contingencia.	<ul style="list-style-type: none">- Pronósticos semanales/mensuales de ocupación.- Datos de personal, camas e insumos.- Alertas de saturación y rotación de camas.
Operativo	<ul style="list-style-type: none">- Asignación diaria de camas y turnos.- Derivaciones y gestión del flujo de pacientes.- Monitoreo en tiempo real de recursos	<ul style="list-style-type: none">- Dashboards en tiempo real.- Listas de espera y alertas de congestión.- Estado de insumos y personal disponible.

KPI's Definidos

Nombre del KPI	Definición	Fórmula	Unidad	Frecuencia	Fuente de datos	Responsable
Promedio general de resultado por año	Mide la tendencia anual de los valores promedio de resultado.	Promedio(resultado de exámenes) agrupado por año.	Valor promedio por año	Anual	Datos Abiertos EsSalud, INEI	Unidad de Estadística Médica
Tasa de diagnósticos por grupo etario	Indica qué grupo etario concentra más diagnósticos.	$(\text{Cantidad de diagnósticos} / \text{Total de diagnósticos}) \times 100$	% del total de casos	Mes	Datos Abiertos EsSalud	Coordinación Epidemiológica
Departamento con mayor carga de diagnósticos	Identifica la región más afectada.	Suma(cantidad de diagnósticos) por región	Nº de diagnósticos	Mes	Datos Abiertos EsSalud, Ubigeo INEI	Gerencias Regionales

Nombre del KPI	Definición	Fórmula	Unidad	Frecuencia	Fuente de datos	Responsable
Variación porcentual anual del promedio de resultados	Mide si los niveles promedio de la enfermedad mejoran o empeoran cada año.	$(\text{Promedio año actual} - \text{Promedio año anterior}) / \text{Promedio año anterior} \times 100$	% de variación	Mes	Datos Abiertos EsSalud	Dirección de Prevención
Promedio de resultado por tipo de diabetes	Permite detectar si ciertos tipos presentan peores valores.	Promedio de resultado por enfermedad	Valor promedio	Mes	Datos hospitalarios	Coordinación de Control Médico
Distribución mensual de diagnósticos	Detecta estacionalidad o picos en diagnósticos.	Total (cantidad de diagnósticos)	Nº de diagnósticos por mes	Mes	Datos Abiertos EsSalud	Unidad de Estadística

Nombre del KPI	Definición	Fórmula	Unidad	Frec	Fuente	Responsabl e
Promedio ponderado de resultados por cantidad de diagnósticos	Integra la gravedad promedio considerando el peso de cada registro.	$\Sigma(\text{promedio resultado} \times \text{cantidad diagnósticos}) / \Sigma (\text{cantidad diagnósticos})$	Valor ponderado	Mes	Datos Abiertos EsSalud	Dirección de Epidemiología
Departamento con mejor control promedio	Identifica la región con mejores valores promedio de control.	Mínimo(promedio resultado)	Valor más bajo promedio	Mes	RRHH EsSalud, RENIPRESS	Jefatura de Recursos Humanos
Proporción de casos con complicaciones	Evalúa la gravedad de los diagnósticos según si hay complicaciones.	$(\text{Casos con complicación} / \text{Total de casos}) \times 100$	% de casos complicados	Mes	Datos Abiertos EsSalud	Gerencias de Red
Índice de concentración por grupo etario	Mide si los diagnósticos están concentrados en un grupo etario o distribuidos	$\Sigma((\text{cantidad diagnósticos grupo} / \text{total diagnósticos})^2)$	Índice (0 = disperso, 1 = concentrado)	Mes	Datos Abiertos EsSalud, Ubigeo INEI	Dirección Estratégica

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	CIE10.csv	Contiene el código de todas la enfermedades y su descripción

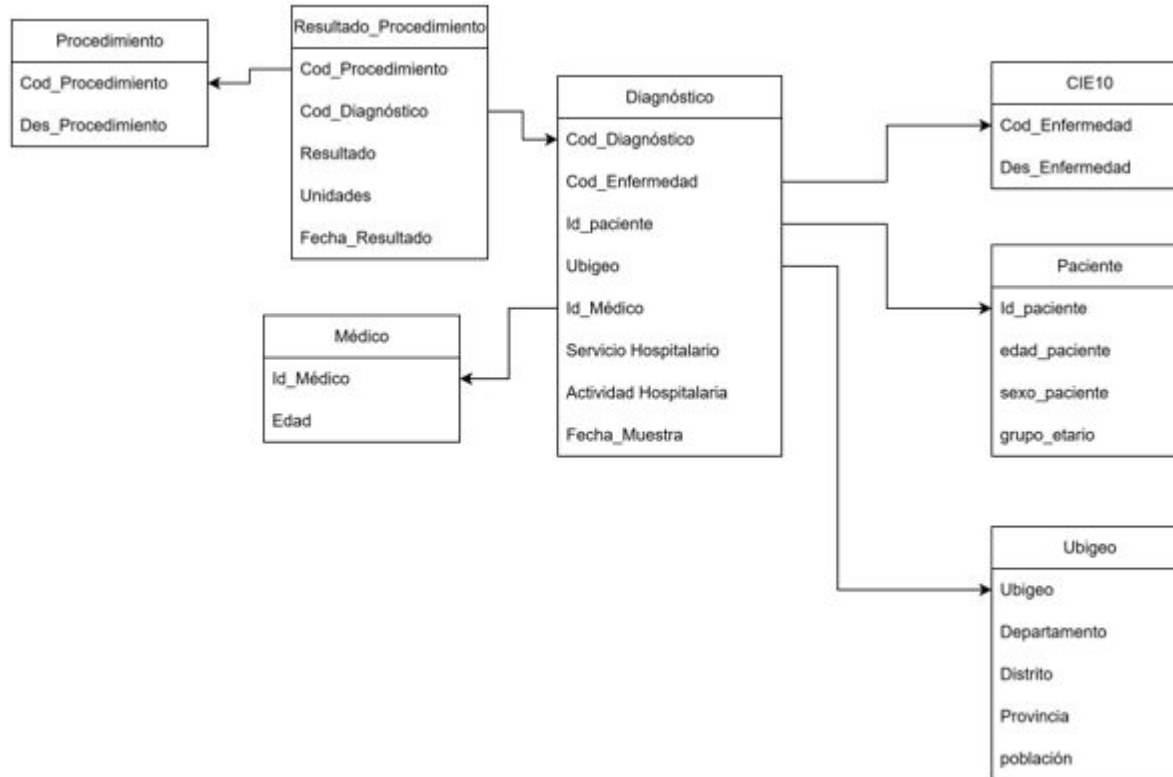
Inventario de Fuentes de datos OLTP

Sistema	Área usuaria	Tipo	Tecnología	Frecuencia de utilización	Observaciones
Sistema de Planeamiento Estratégico – CEPLAN	Planeamiento Estratégico y Estadística	OLTP (maestro de referencia geográfica)	Archivo XLS	Mensual	Incluye descripciones y metadatos de cada ubigeo; sirve como 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.
Sistema de Codificación Geográfica – Ubigeo (INEI)	Planeamiento / Estadística	OLTP (catálogo maestro geográfico)	Archivo CSV	Eventual / según necesidad	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	Área usuaria	Tipo	Tecnología	Frecuencia de utilización	Observaciones
Sistema de Clasificación Internacional de Enfermedades – CIE10	Áreas médicas y de estadística en EsSalud / SuSalud	OLTP (catálogo maestro de enfermedades)	Archivo CSV	Eventual / según necesidad	Contiene el código y descripción de enfermedades según la Clasificación Internacional de Enfermedades (CIE10).

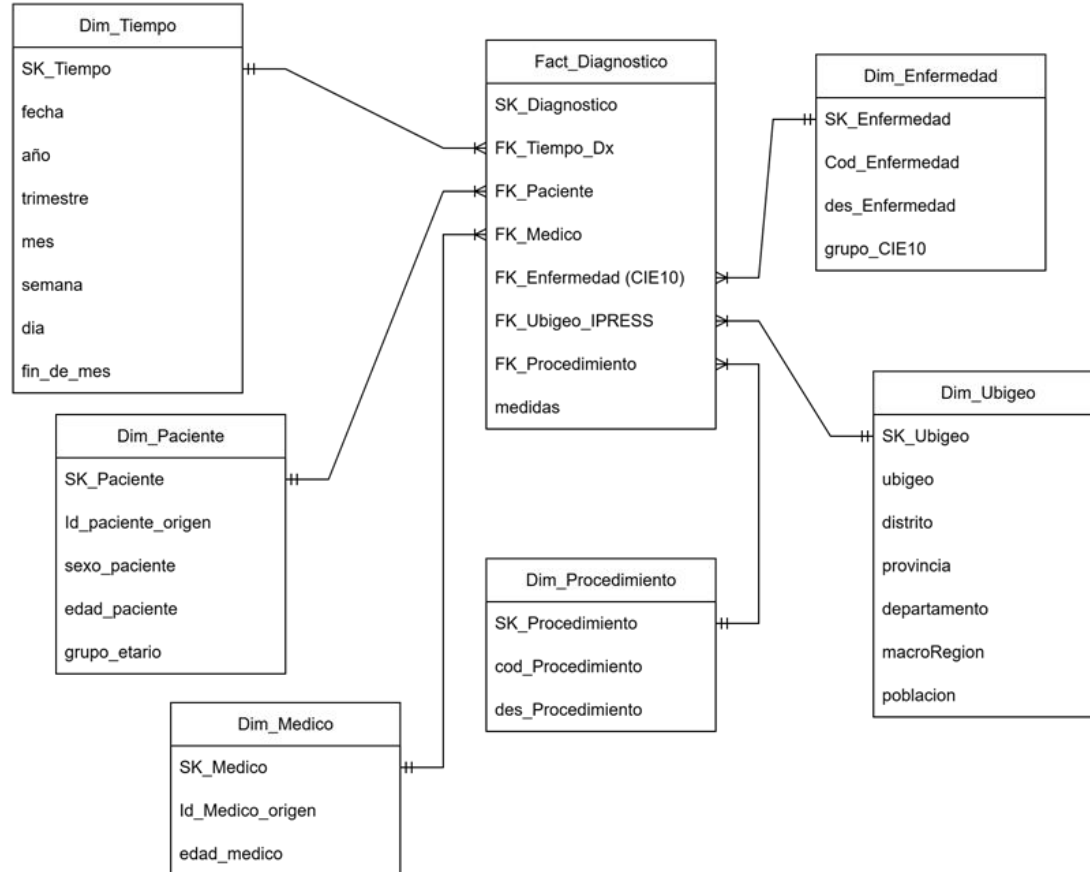
Modelado de datos

Modelo conceptual

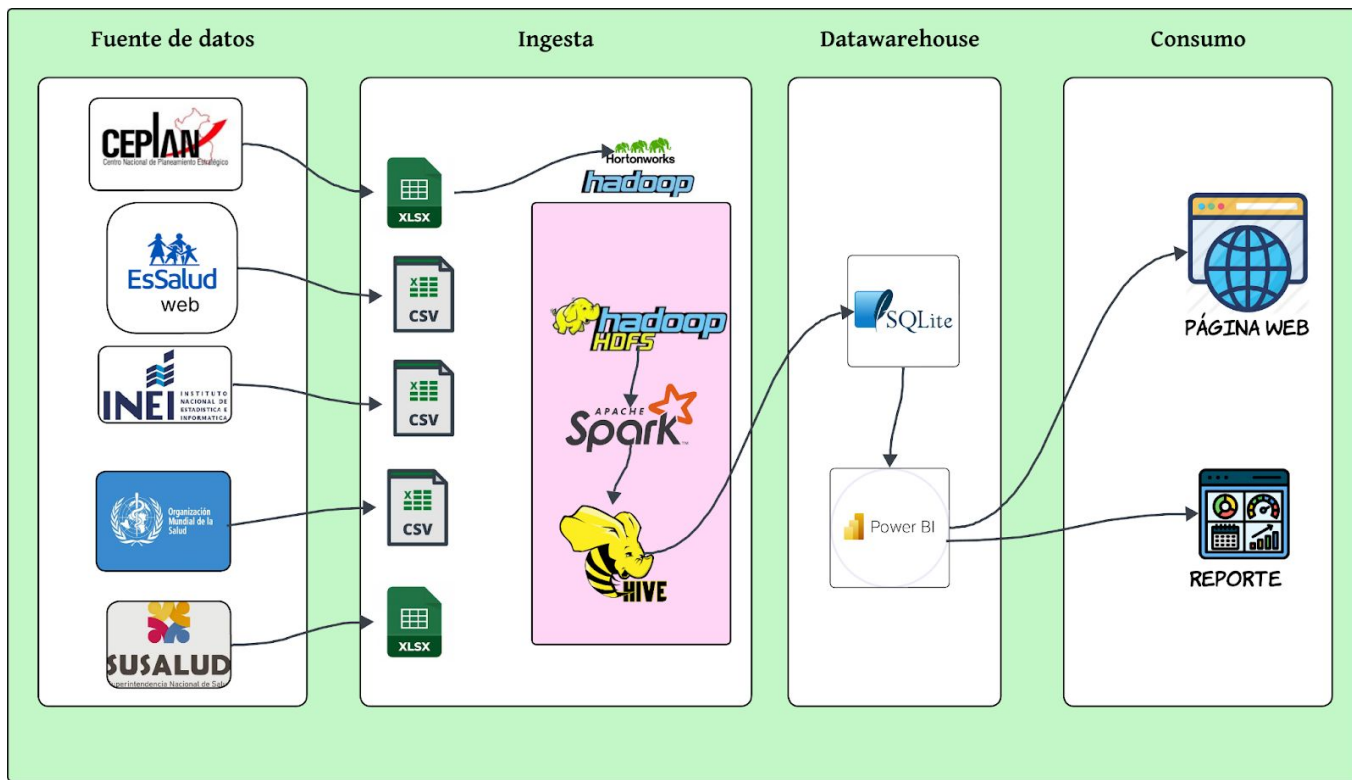


Modelado de datos

Esquema estrella

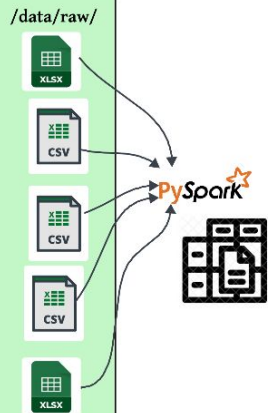


Arquitectura de datos



Proceso ETL

Extract



Transform

A. LIMPIEZA Y ESTANDARIZACIÓN

Stage

- Eliminación de cabeceras vacías y duplicados (dropDuplicates, dropna).
- Conversión de fechas (to_date(col("FECHA_MUESTRA"), "yyyyMMdd"))
- Normalización de nombres para matching textual. (regexp_replace, lower)
- Renombrado de columnas a formato estándar (withColumnRenamed).

B. CREACIÓN DE TABLAS TEMPORALES

ODS

df_CIE
Campos: cod_enfermedad, des_enfermedad

df_Paciente
Campos: id_paciente, edad_paciente, sexo_paciente

df_Medico
id_medico, edad_medico

df_Procedimiento
Se combinan procedimientos de los tres conjuntos (Diabetes, Renal, Hiperlipidemia)

df_Resultado_Procedimiento
cod_diagnostico, cod_procedimiento, resultado, unidades, fecha_resultado

C. TABLAS PRINCIPALES DE DIAGNÓSTICO

- df_Diabetes_main
- df_Renal_main
- df_Hiperlipidemia_main

```
cod_enfermedad
id_paciente
cod_institucion
id_medico
servicio_hospitalaria
actividad_hospitalaria
fecha_muestra
ubigeo
cod_diagnostico (autonomadico)
```



D. ENRIQUECIMIENTO CON CATÁLOGOS EXTERNOS

df_Geodir
Aporta los atributos geográficos por ubigeo: departamento, provincia, distrito, población

df_Ipress
Estandarización de establecimientos de salud: cod_institucion, nombre, tipo, clasificacion, institucion, red

Load

Data warehouse



Tabla	Descripción
Dim_Enfermedad (df_CIE)	Catálogo CIE10
Dim_Paciente (df_Paciente_final)	Pacientes únicos
Dim_Medico (df_Medico_final)	Médicos únicos
Dim_Procedimiento	Catálogo de
Dim_IPRESS (df_Ipress)	Establecimientos
Dim_Ubigeo (df_Geodir)	Ubicación
Fact_Diagnostico (df_diagnostico)	Registro principal de
Fact_Resultado_Procedimiento	Resultados de

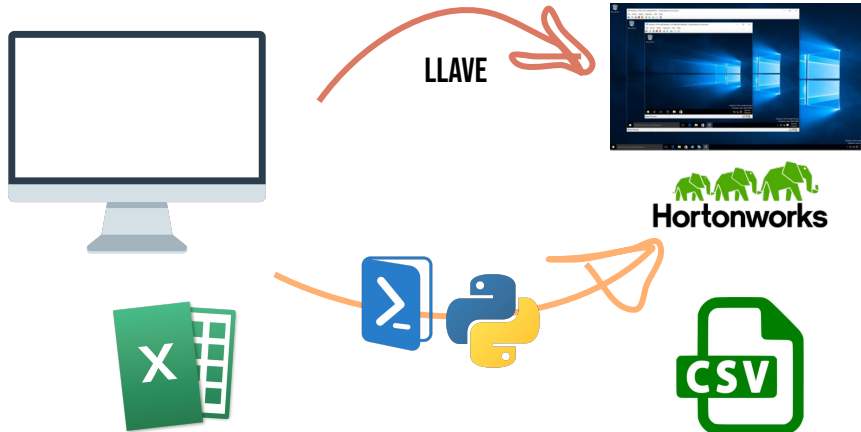
Ingesta de datos

```
[root@sandbox-hdp ~]# hdfs dfs -ls /data/raw/
Found 7 items
-rw-r--r-- 1 root hdfs      913690 2025-10-12 21:41 /data/raw/CIE10_2021.csv
-rw-r--r-- 1 root hdfs 241790497 2025-10-12 21:41 /data/raw/DF_ExLab_CExt_Diabetes.csv
-rw-r--r-- 1 root hdfs 49791254 2025-10-12 21:41 /data/raw/DF_ExLab_CExt_EnfermedadRenal.csv
-rw-r--r-- 1 root hdfs 46893990 2025-10-12 21:42 /data/raw/DF_ExLab_CExt_Hiperlipidemia.csv
-rw-r--r-- 1 root hdfs   839681 2025-10-12 21:41 /data/raw/Datos_planeamiento_estrategico_Distrital.csv
-rw-r--r-- 1 root hdfs   120413 2025-10-12 21:42 /data/raw/geodir_ubigeo_inei_ubigeo_inei.csv
-rw-r--r-- 1 root hdfs 11945413 2025-10-12 21:42 /data/raw/ipress_Listado_de_Establecimientos.csv
```

```
PS C:\Users\Lenovo> ls $env:USERPROFILE\.ssh

Directorio: C:\Users\Lenovo\.ssh

Mode                LastWriteTime         Length Name
----                -
-a-----         12/10/2025   11:47           3389 horton_id_rsa
-a-----         12/10/2025   11:47            749 horton_id_rsa.pub
```



Transformación de datos

Carga de datos



PySpark

```
%pyspark
df_CIE = spark.read.option("header", "true").csv("/data/raw/CIE10_2021.csv")
df_CIE.printSchema()
df_CIE.show(5)
```

```
root
|-- CODIGO: string (nullable = true)
|-- DESCRIPCION: string (nullable = true)
```

CODIGO	DESCRIPCION
Y21.4	Ahogamiento y sum...
Y21.5	Ahogamiento y sum...
Y21.6	Ahogamiento y sum...
Y21.7	Ahogamiento y sum...
Y21.8	Ahogamiento y sum...

only showing top 5 rows

```
%pyspark
df_Diabetes = spark.read.option("header", "true").csv("/data/raw/DF_ExLab_CExt_Diabetes.csv")
df_Diabetes.printSchema()
df_Diabetes.show(5)
```

FECHA_CORTE;DEPARTAMENTO;PROVINCIA;DISTRITO;UBIGEO;RED;IPRESS;ID_PACIENTE;EDAD_PACIENTE;SEXO_PACIENTE;EDAD_MEDICO;ID_MEDICO;COD_DIAG;DIAGNOSTICO;AREA_HOSPITALARIA;SERVICIO_HOSPITALARIO;ACTIVIDAD_HOSPITALARIA;FECHA_MUESTRA;FEC_RESULTADO_1;PROCEDIMIENTO_1;RESULTADO_1;UNIDADES_1;FEC_RESULTADO_2;PROCEDIMIENTO_2;RESULTADO_2;UNIDADES_2
20240531;UCAVALI;...
20240531;UCAVALI;...
20240531;TACNA;TA...
20240531;TACNA;TA...
20240531;SAN MART...

only showing top 5 rows

Took 6 sec: Last updated by anonymous at October 12 2025, 5:00:29 PM.

```
%pyspark
df_Enf_Renal = spark.read.option("header", "true").csv("/data/raw/DF_ExLab_CExt_EnfermedadRenal.csv")
df_Enf_Renal.printSchema()
df_Enf_Renal.show(5)
```

```
root
|-- FECHA_CORTE;DEPARTAMENTO;PROVINCIA;DISTRITO;UBIGEO;RED;IPRESS;ID_PACIENTE;EDAD_PACIENTE;SEXO_PACIENTE;EDAD_MEDICO;ID_MEDICO;COD_DIAG;DIAGNOSTICO;AREA_HOSPITALARIA;SERVICIO_HOSPITALARIO;ACTIVIDAD_HOSPITALARIA;FECHA_MUESTRA;FEC_RESULTADO_1;PROCEDIMIENTO_1;RESULTADO_1;UNIDADES_1;FEC_RESULTADO_2;PROCEDIMIENTO_2;RESULTADO_2;UNIDADES_2
+-----+
|FECHA_CORTE;DEPARTAMENTO;PROVINCIA;DISTRITO;UBIGEO;RED;IPRESS;ID_PACIENTE;EDAD_PACIENTE;SEXO_PACIENTE;EDAD_MEDICO;ID_MEDICO;COD_DIAG;DIAGNOSTICO;AREA_HOSPITALARIA;SERVICIO_HOSPITALARIO;ACTIVIDAD_HOSPITALARIA;FECHA_MUESTRA;FEC_RESULTADO_1;PROCEDIMIENTO_1;RESULTADO_1;UNIDADES_1;FEC_RESULTADO_2;PROCEDIMIENTO_2;RESULTADO_2;UNIDADES_2|
+-----+
|20240531;CALLAO;C...|
|20240531;HUANUCO;...|
|20240531;HUANUCO;...|
|20240531;HUANUCO;...|
+-----+
```

Took 1 sec: Last updated by anonymous at October 12 2025, 5:01:20 PM.

FINISHED

Transformación de datos

Limpieza de datos

PySpark



ubigeo	distrito	provincia	departamento	poblacion
10101	Chachapoyas	Chachapoyas	Amazonas	29171
10102	Asuncion	Chachapoyas	Amazonas	288
10103	Balsas	Chachapoyas	Amazonas	1644
10104	Cheto	Chachapoyas	Amazonas	591
10105	Chiliquin	Chachapoyas	Amazonas	687
10106	Chuquibamba	Chachapoyas	Amazonas	2064
10107	Granada	Chachapoyas	Amazonas	379
10108	Huancas	Chachapoyas	Amazonas	1329
10109	La Jalca	Chachapoyas	Amazonas	5513
10110	Leimebamba	Chachapoyas	Amazonas	4206

id_paciente	edad_paciente	sexo_paciente
eJwzNDY1NTQ0NDczN...	64	FEMENINO
eJwzNLQ0NbcwtbAwN...	51	FEMENINO
eJwzNDE2MzG0NLAwM...	66	FEMENINO
eJwztdQ3NrS0BCIjC...	44	MASCULINO
eJwzNDYxMTGyMDI2N...	76	FEMENINO
eJwzNDY2tLAWNTE0M...	55	FEMENINO
eJwzNLY0MLAwNzA3N...	54	FEMENINO
eJwzNDIwMTM1sDQws...	48	MASCULINO
eJwzNLMwMLE0NDEwM...	47	MASCULINO
eJwzMjAyNLYwMjE0N...	38	MASCULINO

cod_diagnostico	cod_procedimiento	resultado	unidades	fecha_resultado
335027	1	99.0	mg/dL	2023-05-18
155766	1	114.98	mg/dL	2022-04-02
60665	1	144.0	mg/dL	2021-04-26
289348	1	114.2	mg/dL	2023-02-14
29517	1	91.0	mg/dL	2020-11-04
64224	1	149.0	mg/dL	2021-05-17
111879	1	250.0	mg/dL	2021-11-11
49653	1	186.0	mg/dL	2021-02-15
127578	1	97.45	mg/dL	2021-12-29
188596	1	89.0	mg/dL	2022-06-21

cod_enfermedad	id_paciente	cod_institucion	id_medico	servicio_hospitalario	actividad_hospitalaria	fecha_muestra	cod_diagnostico
E11.9	eJwzNDAwtdC0NDMxN...	CAP I MANANTAY	eJwzNjA2MzE2NLY0N...	MEDICINA GENERAL	ATENCION MEDICA ...	2020-01-02	1
E13.9	eJwzNDAwtdjAxtNUxs...	P.M. ALAMEDA	eJwzsjS1NDI2MjE2N...	MEDICINA GENERAL	ATENCION MEDICA ...	2020-01-02	2
E11.9	eJwzNDcwMLOWMDKzt...	DANIEL ALCI...	eJwztdLQwMjM1NTY3N...	ENDOCRINOLOGIA	ATENCION MEDICA ...	2020-01-02	3
E11.9	eJwzNDcwNDU3Mja1M...	CAP II OSCAR FERN...	eJwztdLS0MDE1sjQ3N...	MEDICINA FAMILIAR...	ATENCION MEDICA ...	2020-01-02	4
E11.9	eJwzNDAs0szCQ0gMDM...	P.M. SAPOSOA	eJwzNDS1MD00MDEyM...	MEDICINA GENERAL	ATENCION MEDICA ...	2020-01-02	5
E11.9	eJwzNLY0tdC1sDQ1N...	H.I LA ESPERANZA	eJwzNLYwNzcXNzQxN...	MEDICINA INTERNA	ATENCION MEDICA ...	2020-01-02	6
E11.9	eJwzNLY0MjM3NzE2M...	H.I LA ESPERANZA	eJwzNLY0NzE3MjMzM...	MEDICINA GENERAL	ATENCION MEDICA ...	2020-01-02	7
E11.9	eJwzNLY0tdjQ1NzYwN...	H.III SUAREZ-ANGAMOS	eJwzNDA0ACjJazNLU...	ENDOCRINOLOGIA	ATENCION MEDICA ...	2020-01-02	8
E11.4	eJwzNDU0NzA0MDc0t...	H.II HUANO	eJwzMjAwM7S0MLU0N...	MEDICINA INTERNA	ATENCION MEDICA ...	2020-01-02	9
E11.9	eJwzMjMrUwMjYxsJA3s...	P.M. SAPOSOA	eJwzNDS1MD00MDEyM...	MEDICINA GENERAL	ATENCION MEDICA ...	2020-01-02	10

Transformación de datos

Creación de tabla fact

PySpark 



```
df_fact = df_fact.select(
    "cod_diagnostico",
    "id_paciente",
    "sexo_paciente",
    "grupo_etario",
    "cod_institucion",
    "nombre",
    "departamento",
    "red",
    "cod_enfermedad",
    "des_enfermedad",
    "des_procedimiento",
    "resultado",
    "unidades",
    "fecha_resultado"
)
```

```
df_fact = (
    df_diagnostico
    .join(df_paciente, "id_paciente", "left")
    .join(df_resultado_procedimiento, "cod_diagnostico", "left")
    .join(df_procedimiento, "cod_procedimiento", "left")
    .join(df_ipress, "cod_institucion", "left")
    .join(df_ubigeo, "ubigeo", "left")
    .join(df_CIE, "cod_enfermedad", "left")
)
```

df_dianostico

 **RAW**

VS

CURATED

 df_fact

Subida de datos

Integración con Hive



```
%pyspark
spark.sql("CREATE DATABASE IF NOT EXISTS raw_db")
spark.sql("CREATE DATABASE IF NOT EXISTS curated_db")
```

```
%pyspark
df_fact.write.mode("overwrite").saveAsTable("curated_db.fact_diagnostico")
```

```
%pyspark
df_diagnostico.write.mode("overwrite").saveAsTable("raw_db.diagnostico")
df_paciente.write.mode("overwrite").saveAsTable("raw_db.paciente")
df_medico.write.mode("overwrite").saveAsTable("raw_db.medico")
df_resultado_procedimiento.write.mode("overwrite").saveAsTable("raw_db.resultado_procedimiento")
df_procedimiento.write.mode("overwrite").saveAsTable("raw_db.procedimiento")
df_ipress.write.mode("overwrite").saveAsTable("raw_db.ipress")
df_CIE.write.mode("overwrite").saveAsTable("raw_db.cie")
df_ubigeo.write.mode("overwrite").saveAsTable("raw_db.ubigeo")
```



Query Process Results (Status: SUCCEEDED)

fact_diagnostico.aod_diagnostico	fact_diagnostico.id_paciente	fact_diagnostico.sexo_paciente	fact_diagnostico.grupo_etario	fact_diagnostico.departamento	fact_diagnostico.aod_enfermedad	fact_diagnostico.des_enfermedad	fact_diagnostico.des_procedimiento	fact_diagnostico.resultado	fact_diagnostico.unidades	fact_diagnostico.fecha_resultado	fact_diagnostico.año
148	aJw2MjA1MTYwMDUwMTA0bWZlZGQhYXQ=	MASCULINO	Adulto (30-59)	null	E11.9	Diabetes mellitus no insulino dependiente, sin men...	DOSAJE DE GLUCOSA EN SANGRE, CUANTITATIVO (EXCEPTO CINTA REACTIVA)	87.0	mg/dL	2020-01-07	2020
463	aJw2NDcwMDUwMTA0bWZlZGQhYXQ=	MASCULINO	Adulto mayor (60+)	Tarma	E11.9	Diabetes mellitus no insulino dependiente, sin men...	DOSAJE DE GLUCOSA EN SANGRE, CUANTITATIVO (EXCEPTO CINTA REACTIVA)	125.0	mg/dL	2020-01-08	2020
471	aJw2NDYsTzNlZGQhYXQhYXQhYXQhYXQ=	MASCULINO	Adulto (30-59)	Puno	E11.8	Diabetes mellitus no insulino dependiente, con complicaciones no especificadas	DOSAJE DE GLUCOSA EN SANGRE, CUANTITATIVO (EXCEPTO CINTA REACTIVA)	144.0	mg/dL	2020-01-08	2020
496	aJw2NLWQwMDUwMTA0bWZlZGQhYXQ=	MASCULINO	Adulto mayor (60+)	Lima	E11.9	Diabetes mellitus no insulino dependiente, sin men...	DOSAJE DE GLUCOSA EN SANGRE, CUANTITATIVO (EXCEPTO CINTA REACTIVA)	185.0	mg/dL	2020-01-08	2020
833	aJw2NDQwNDYwNDYwNDYwNDYwNDYwNDYwNDY=	FEMENINO	Adulto mayor (60+)	Lima	E11.9	Diabetes mellitus no insulino dependiente, sin men...	DOSAJE DE GLUCOSA EN SANGRE, CUANTITATIVO (EXCEPTO CINTA REACTIVA)	92.0	mg/dL	2020-01-20	2020
1088	aJw2NLWwMDUwMTA0bWZlZGQhYXQhYXQ=	FEMENINO	Adulto (30-59)	Pasco	E11.6	Diabetes mellitus no insulino dependiente, con otras complicaciones especificadas	DOSAJE DE GLUCOSA EN SANGRE, CUANTITATIVO (EXCEPTO CINTA REACTIVA)	352.0	mg/dL	2020-01-14	2020
1238	aJw2NDYsTzNlZGQhYXQhYXQhYXQhYXQ=	MASCULINO	Adulto mayor (60+)	La Libertad	E11.9	Diabetes mellitus no insulino dependiente, sin men...	DOSAJE DE GLUCOSA EN SANGRE, CUANTITATIVO (EXCEPTO CINTA REACTIVA)	143.0	mg/dL	2020-01-14	2020

Cubo Olap

Hecho

fact _diagnostico

```
%pyspark
df_fact = df_fact.withColumn("anio", F.year("fecha_resultado"))
df_fact = df_fact.withColumn("mes", F.month("fecha_resultado"))
df_fact = df_fact.withColumn("resultado", F.col("resultado").cast("double"))
df_fact.write.mode("overwrite").saveAsTable("curated_db.fact_diagnostico")
```

```
from pyspark.sql import functions as F
```

```
df_cubo = (
    df_fact.groupBy("grupo_etario", "departamento", "des_enfermedad", "mes", "anio")
    .agg(
        F.avg("resultado").alias("promedio_resultado"),
        F.count("cod_diagnostico").alias("cantidad_diagnosticos")
    )
)
df_cubo.write.mode("overwrite").saveAsTable("curated_db.cubo_olap")
```

cubo_olap.grupo_etario	cubo_olap.departamento	cubo_olap.des_enfermedad	cubo_olap.mes	cubo_olap.anio	cubo_olap.promedio_resultado	cubo_olap.cantidad_diagnosticos
Adulto (30-59)	La Libertad	Diabetes mellitus especificada, sin mención de complicación	2	2021	203.97222222222223	38
Adulto mayor (60+)	Piura	Diabetes mellitus no insulinodependiente, sin mención de complicación	1	2022	160.01551724137934	232
Adolescencia (12-17)	La Libertad	Diabetes mellitus insulinodependiente, sin mención de complicación	2	2022	191.5	4
Adulto (30-59)	Lambayeque	Diabetes mellitus no insulinodependiente, sin mención de complicación	3	2024	187.83571929824564	285
Adulto mayor (60+)	Junin	Diabetes mellitus especificada, con complicaciones múltiples	8	2021	225.5	2
Adulto (30-59)	La Libertad	Diabetes mellitus insulinodependiente, sin mención de complicación	8	2022	224.02777777777777	22

Dimensiones

Grupo Etario

Departamento

Enfermedad

Mes - Año

Dashboard



Medidas

```
Resultado x Diagnóstico =  
DIVIDE(  
    SUMX(  
        Cubo_nuevo,  
        Cubo_nuevo[cantidad_diagnosticos] * Cubo_nuevo[promedio_resultado]  
    ),  
    [cant_diagnosticos]  
)
```

```
cant_diagnosticos = SUM(Cubo_nuevo[cantidad_diagnosticos])
```

```
tasa_diagnosticos =  
DIVIDE(  
    [cant_diagnosticos],  
    CALCULATE( [cant_diagnosticos], ALLSELECTED() ),  
    0  
) * 100
```

```
1 Casos_cn_complicacion_% =  
2 VAR CasosConComp =  
3     CALCULATE(  
4         SUM(Cubo_nuevo[cantidad_diagnosticos]),  
5         FILTER(  
6             Cubo_nuevo,  
7             NOT CONTAINSSTRING(UPPER(Cubo_nuevo[des_enfermedad]), "SIN")  
8         )  
9     )  
10 VAR TotalCasos =  
11     SUM(Cubo_nuevo[cantidad_diagnosticos])  
12  
13 RETURN  
14 DIVIDE(CasosConComp, TotalCasos, 0) * 100
```

```
ly_prom_resultado % =  
VAR AnioActual = SELECTEDVALUE ( Cubo_nuevo[anio] )  
VAR PromActual = AVERAGE ( Cubo_nuevo[promedio_resultado] )  
VAR PromAnterior =  
    CALCULATE (   
        AVERAGE ( Cubo_nuevo[promedio_resultado] ),  
        FILTER (   
            ALL ( Cubo_nuevo ),  
            Cubo_nuevo[anio] = AnioActual - 1  
        )  
    )  
RETURN  
DIVIDE ( PromActual - PromAnterior, PromAnterior )
```




Power BI

Dashboard



GRACIAS