

Trabajo Nociones de Arquitectura de la Información

Versión: 2021-05-29 13:00

Observación: Cada vez que agregue nuevos elementos al documento, o que modifique algún componente del informe, revise la coherencia y consistencia con los otros elementos que hacen parte del mismo.

RESPONSABLES

Nombre Completo – Documento de Identificación
1. Santiago Rincón Perez
2. Julian Saldarriaga Blandon
3. Angie Velasquez Velez
4. Harvey Rodriguez Gil
REPO EN GITHUB: https://github.com/hrodriguezgi-un/AGD_TrabajoGrupal

Realiza este trabajo considerando los datos que generan los sistemas transaccionales e información no estructurada de tu dominio (si trabajas por ejemplo para TCC tu dominio es la mensajería; también puedes explorar en la página <https://www.kaggle.com/datasets> o <https://arxiv.org/>). Considera tener acceso a esta información, de al menos 10 MB (puede ser uno o varios archivos de texto), y **tener al menos cuatro clases conceptuales. Este documento también debe almacenarse en el REPO. Plazo Máximo de Entrega 23 de Mayo, NO SE recibirá por correo electrónico, envío por <https://forms.gle/h7ty3yZykaUq5m7y6>**

1 COMPRENSIÓN DEL NEGOCIO

1.1 DESCRIPCIÓN DEL CONTEXTO DEL NEGOCIO.

Describe en máximo 250 palabras el contexto en el cual se generan los datos y cuál es el proceso que los genera.

El contexto de los datos está soportado en la Resolución 3374 de 2000, por la cual se reglamentan los datos básicos que deben ser reportados por los prestadores de servicios de salud y las entidades Administradoras de Planes de Beneficios, la información del Registro Individual de Prestaciones de

Servicios de Salud - RIPS, constituyen una fuente de datos prioritaria para los procesos de dirección, regulación y control del Sistema General de Seguridad Social en Salud – SGSSS.

El RIPS provee los datos mínimos y básicos, y facilita las relaciones entre las entidades administradoras de planes de beneficios (pagadores), las instituciones prestadoras de servicios de salud y los profesionales independientes (prestadores) apoyando la generación de planes, programas, estrategias y operaciones, para la toma decisiones gerenciales, estratégicas, administrativas y asistenciales.

Se define analizar los registros individuales para el servicio de hospitalización en el municipio de Medellín, entre los años 2009 a 2017.

1.2 IDENTIFICACIÓN DEL PROBLEMA:

Delimite en máximo 150 palabras la problemática, así como identificar los requisitos, supuestos, restricciones y beneficios de la solución de este.

Conocer la frecuencia de utilización de los servicios de hospitalización en el municipio de Medellín, la morbilidad y su caracterización según variables sociodemográficas, con el fin de determinar la demanda de los servicios de hospitalización y la destinación de recursos en el marco del fortalecimiento de la red hospitalaria (generar políticas en salud)

Requisitos: Toda atención prestada a un paciente en las instituciones de Salud (IPS) deriva en una factura la cual debe contener unos datos básicos como: datos de identificación, datos del servicio, datos del motivo de la atención y datos del valor. La información resultante de cada factura se debe consolidar en un archivo plano, pasarlos por un validador dispuesto por el Ministerio de Salud y realizar envío de la información una vez validada a las EAPB para finalmente colocarlas a disposición del ente territorial.

Supuestos: Información veraz

Restricciones: pese a que los distintos generadores y usuarios de los datos reconocen su valor e importancia, en la actualidad se han identificado falencias tanto en la calidad del registro primario como el no uso para la gestión en las IPS y las EAPB, así como para la formulación de la política pública en salud.

Beneficios de la solución: Definir políticas públicas de morbilidad y su caracterización según variables sociodemográficas, así mismo contribuir a generar guías de práctica clínica para optimizar el cuidado del paciente, disminuir la variabilidad innecesaria en la práctica clínica.

1.3 DETERMINACIÓN DE OBJETIVOS:

Describe en máximo 150 palabras las metas a lograr al proponer una solución basada en un modelo de datos o de analítica (cómo y qué tipo de ventaja competitiva se ganará).

Contar con la información a un nivel de detalle tal que permita conocer la demanda de los servicios de atención en salud de hospitalización para aportar a las estadísticas de morbilidad en Colombia, definición de políticas públicas

1.4 EVALUACIÓN DE LA SITUACIÓN ACTUAL:

Describe en máximo 150 palabras el estado actual antes de implementar la solución de analítica, a fin de tener un punto de comparación que permita medir el grado de éxito de la solución.

2 COMPRENSIÓN DE LOS DATOS

2.1 RECOLECCIÓN DE DATOS

Describe en máximo 150 palabras los datos a utilizar identificando las fuentes, las técnicas empleadas en su recolección, los problemas encontrados en su obtención y la forma como se resolvieron los mismos. Además, adjunte los datos (archivos de texto, etc.) agréguelos en el github (**REPO EN GITHUB**) en un solo archivo, por favor comprímalo(s). Llame el archivo T1.2.1.Datos.zip

Link del archivo:

https://github.com/hrodriguezgi-un/AGD_TrabajoGrupal/blob/main/T1.2.1.Datos.zip

Se tomaron los datos desde la página MEDATA, sitio de datos abiertos de la Alcaldía de Medellín, correspondiente al Registro de prestaciones de Servicios de Salud RIPS para el servicio de hospitalización, el conjunto de datos es suministrado por la Secretaria de Salud de Medellín para consulta pública, la misma puede ser consultada en el link: <http://medata.gov.co/dataset/atenciones-en-hospitalizaci%C3%B3n> para el entendimiento de la información se hace necesario referenciarse en la Resolución 3374 de 2000, la cual define la estructura básica de datos y las tablas de referencia que complementan el conjunto de datos, las cuales pueden ser consultadas en la siguiente página del Ministerio de Salud: <https://web.sispro.gov.co/WebPublico/Consultas/ConsultarReferenciaBasica.aspx>, la información puede ser descargada en formato CSV con un total de 979995 y 189.28 MB.

2.2 DESCRIPCIÓN DE DATOS (DICCIONARIO):

Diligencia la siguiente tabla, puede agregar otra columna si lo considera necesario.

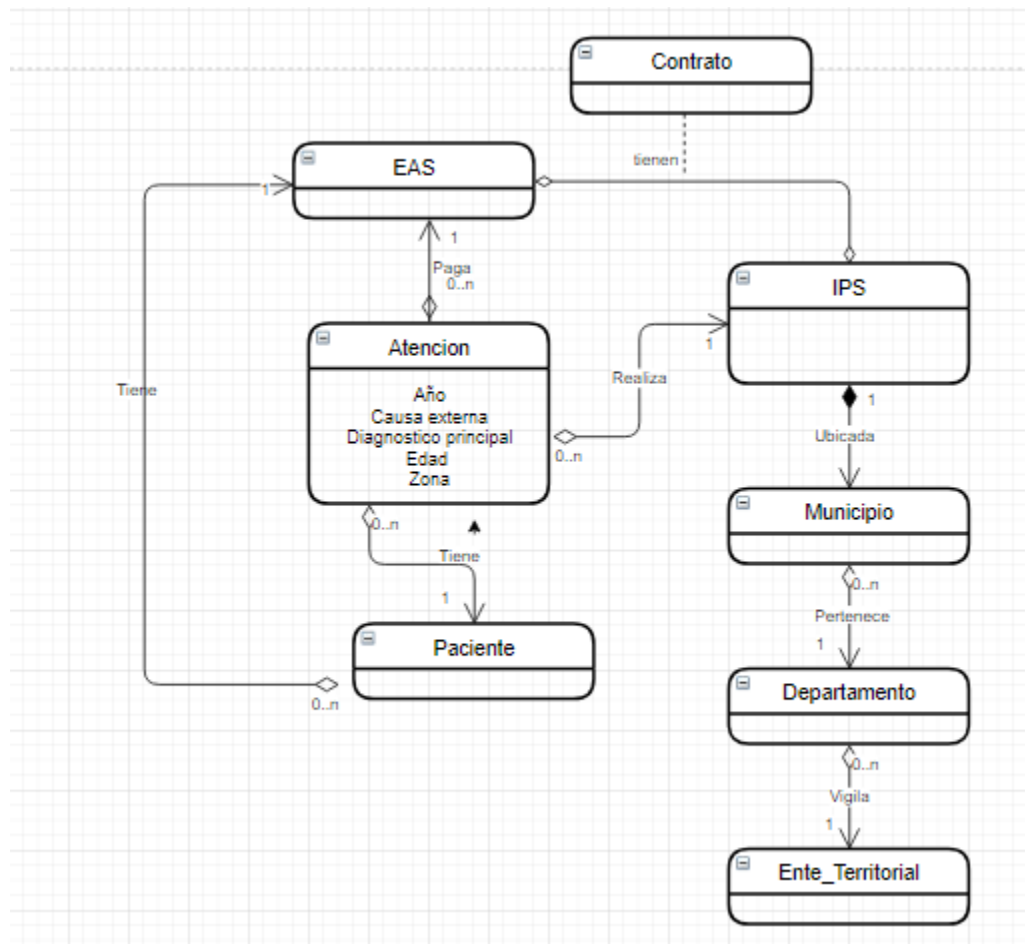
Tabla	Nombre del atributo / variable	Formato o Tipo de Dato	Descripción
departamentos	cod	varchar(3)	Código DANE (Departamento Administrativo Nacional de Estadística) para departamento en Colombia
departamentos	departamento	varchar(100)	Nombre del departamento en Colombia
municipios	cod	varchar(6)	Código DANE para municipios para Colombia
municipios	descripcion	varchar(50)	Nombre del municipio en Colombia

causa_externa	cod	int(4)	Código causa externa definidos en la Resolución 3374 de 2000 RIPS
causa_externa	causa_externa	varchar(100)	Descripción de causa externa la cual corresponde al daño intencional o no intencional a una persona descritos en los RIPS
diagnostico	cod	varchar(4)	Código CIE10 (Clasificación internacional de enfermedades en su 10 versión) diagnósticos en salud
diagnostico	diagnostico	varchar(500)	Descripción código CIE10 diagnósticos en salud
eas	cod	varchar(6)	Código de Entidad Administradora de Salud
eas	nombre_eas	varchar(200)	Descripción de Entidad Administradora de Salud
ips	cod	varchar(12)	Código único de habilitación definido desde el Registro especial de prestadores de servicios de salud REPS
ips	nombre_ips	varchar(500)	Descripción Prestador de Servicios de Salud definido desde el Registro especial de prestadores de servicios de salud REPS
tipo_edad	cod	int(4)	Código de tipo edad (años =1, meses =2, días =3) definidos en la Resolución 3374 de 2000 RIPS
tipo_edad	tipo_edad	varchar(100)	Descripción de tipo de edad (años =1, meses =2, días =3) definidos en la Resolución 3374 de 2000 RIPS
tipo_usuario	cod	int(4)	Código de tipo usuario corresponde al régimen de afiliación en salud asignado desde los RIPS
tipo_usuario	tipo_usuario	varchar(100)	Descripción de tipo usuario, que es la descripción del régimen de afiliación en salud
atencion	consecutivo	int(4)	Consecutivo de registro en la base de datos
atencion	anio	int(4)	Corresponde el año de la prestación en salud para el servicio de hospitalización
atencion	cod_eas	varchar(6)	Código de Entidad Administradora de Salud
atencion	cod_tipo_usuario	int(4)	Código de tipo usuario corresponde al régimen de afiliación en salud asignado desde los RIPS
atencion	cod_ips	varchar(12)	Código único de habilitación definido desde el Registro especial de prestadores de servicios de salud REPS
atencion	cod_municipio	varchar(6)	Código DANE para municipios para Colombia
atencion	cod_departamento	varchar(3)	Código DANE (Departamento Administrativo Nacional de Estadística) para departamento en Colombia
atencion	cod_causa_externa	int(4)	Código causa externa definidos en la Resolución 3374 de 2000 RIPS
atencion	cod_dx_ppal_egreso	varchar(4)	Código CIE10 (Clasificación internacional de enfermedades en su 10 versión) diagnóstico principal al egreso del servicio de hospitalización
atencion	cod_tipo_edad	int(4)	Código de tipo edad (años =1, meses =2, días =3) definidos en la Resolución 3374 de 2000 RIPS

atencion	edad	int(4)	Edad según el tipo de edad definido, ya sean años, meses o días.
atencion	sexo	varchar(2)	Sexo del paciente
atencion	zona	varchar(2)	Zona de residencia habitual ya sea urbana = U o rural R
atencion	servicio	varchar(50)	Corresponde al servicio de hospitalización
atencion	total_atenciones	int(4)	Total de atenciones de acuerdo a las variables anteriores

2.3 MODELO DEL DOMINIO

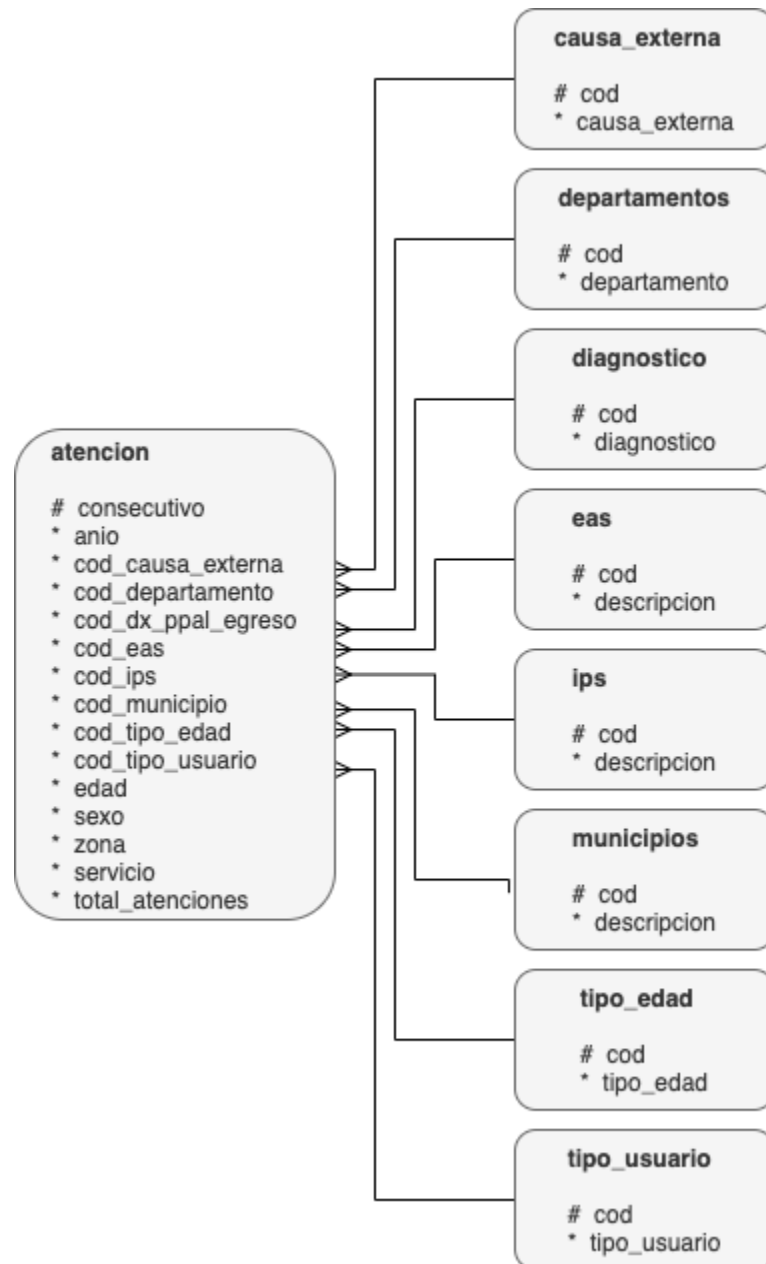
Observación: Incluya el gráfico del modelo del dominio que representa la estructura de datos de su problema.



3 MODELO ENTIDAD-RELACIÓN

3.1 TOMA DE PANTALLA DEL MODELO E-R

Observación: lo que se pide, puede usar <https://draw.io> o Microsoft Visio® y modele usando la notación de Barker.



3.2 SENTENCIA O CONSULTA DE CREACIÓN DEL TABLA(S)

Observación: Escriba el código en el Sistema de Gestión de Bases de Datos Relacionales de su elección (se recomienda SQLite por simplicidad, mediante <https://sqlitebrowser.org/>) para crear

las tablas que corresponda con su conjunto de datos específico. Almacene en el repositorio **(REPO EN GITHUB)** el script con el nombre de T1.3.2.Creacion_Tablas.sql

Link del archivo:

https://github.com/hrodriguezgi-un/AGD_TrabajoGrupal/blob/main/df_files/T1.3.2.Creacion_Tablas.sql

3.3 SENTENCIAS PARA INSERTAR DATOS

Observación: Escriba el código para insertar los datos en cada una de las tablas creadas. Almacene en el repositorio **(REPO EN GITHUB)** el script con el nombre de T1.3.3.Insertar_Datos.sql

Link del archivo:

https://github.com/hrodriguezgi-un/AGD_TrabajoGrupal/blob/main/df_files/T1.3.3.Insertar_Datos.sql

3.4 SENTENCIA DE CONSULTA

Observación: realice la exploración básica de los datos, conteos totales y por categorías, máximos, promedio y mínimos. Es decir, aplique estadística descriptiva con el fin de conocer las propiedades de los datos y entenderlos lo mejor posible. Use solamente sentencias SQL. Anexe las tomas de pantalla donde evidencie la sentencia SQL y su correspondiente ejecución. Además, Almacene en el repositorio **(REPO EN GITHUB)** el script con el nombre de T1.3.4.Consultar_Datos.sql

Link del archivo:

https://github.com/hrodriguezgi-un/AGD_TrabajoGrupal/blob/main/scripts/T1.3.4.Consultar_Datos.sql

```
-- 1. TOTAL DE REGISTROS POR AÑO
SELECT
    ANIO
    ,COUNT(1)
FROM
    atencion
GROUP BY
    ANIO;
```

123 ANIO	123 COUNT(1)
2,015	111,411
2,016	112,357
2,017	114,948

```
-- 2. TOTAL DE REGISTROS POR ENTIDAD ADMINISTRADORA DE SALUD EAS
SELECT
    E.descripcion AS NOMBRE_EAS
    ,SUM(A.total_atenciones) AS TOTAL_ATENCIONES
FROM
    atencion AS A
    LEFT JOIN eas AS E ON
        A.cod_eas = E.cod
GROUP BY
```

```

        E.descripcion
ORDER BY
        SUM(A.total_atenciones) DESC
LIMIT 10;

```

RBC NOMBRE_EAS	TOTAL_ATENCIONES
ALIANZA MEDELLIN ANTIOQUIA SAVIA SALUD SUBSIDIADO	123,214
EPS SURA	66,531
EPS NUEVA EPS	53,378
CCF COMFAMA DE ANTIOQUIA	50,150
EPS COOMEVA	38,811
SIN ADMINISTRADORA	26,162
ASEG SURAMERICANA DE SEGUROS DE VIDA S.A	20,340
EPS SALUD TOTAL S.A.	17,518
DSSA ANTIOQUIA	14,670
EPS CAFESALUD S.A	13,782

```

-- 3. TOTAL ATENCIONES POR IPS Y SERVICIO PARA HOMBRES MAYORES DE 50 AÑOS
SELECT
        I.descripcion AS NOMBRE_IPS
        ,A.servicio AS SERVICIO
        ,SUM(A.total_atenciones) AS TOTAL_ATENCIONES
FROM
        atencion AS A
        LEFT JOIN ips AS I ON
                A.cod_ips = I.cod
WHERE
        A.sexo = 'H'
        AND A.edad > 50
GROUP BY
        I.descripcion, A.servicio
ORDER BY
        SUM(A.total_atenciones) DESC
LIMIT 10;

```

RBC NOMBRE_IPS	RBC SERVICIO	TOTAL_ATENCIONES
IPS UNIVERSIDAD DE ANTIOQUIA IPS UNIVERSITARIA	HOSPITALI	11,689
FUNDACION HOSPITALARIA SAN VICENTE DE PAUL	HOSPITALI	6,583
EMPRESA SOCIAL DEL ESTADO METROSALUD	HOSPITALI	5,763
EMPRESA SOCIAL DEL ESTADO HOSPITAL GENERAL DE MEDELLIN LUZ CASTRO DE GUTIERREZ	HOSPITALI	5,657
FUNDACION INSTITUTO NEUROLOGICO DE ANTIOQUIA	HOSPITALI	4,615
HOSPITAL PABLO TOBON URIBE	HOSPITALI	4,570
COMUNIDAD DE HERMANAS DOMINICAS DE LA PRESENTACION DE LA SANTISIMA VIRGEN DE TOURS PROVINCIA DE MEDELLIN	HOSPITALI	4,317
NUEVA CLINICA SAGRADO CORAZON S.A.S	HOSPITALI	2,712
PROMOTORA MEDICA LAS AMERICAS SA	HOSPITALI	2,418
E.S.E. HOSPITAL LA MARIA	HOSPITALI	2,137

```

-- 4. 2 PRIMEROS DIAGNÓSTICO Y CAUSA EXTERNA CON MAYOR FRECUENCIA
SELECT
        D.diagnostico
        ,CE.causa_externa
        ,SUM(A.total_atenciones) AS total_atenciones
FROM
        atencion as A
        LEFT JOIN diagnostico AS D ON
                A.cod_dx_ppal_egreso = D.cod
        LEFT JOIN causa_externa AS CE ON
                A.cod_causa_externa = CE.cod
GROUP BY
        1,2
ORDER BY
        3 DESC
LIMIT 2;

```


diagnostico	causa_externa	total_atenciones
PARTO UNICO ESPONTANEO PRESENTACION CEFALICA DE VERTICE	Enfermedad general	19,392
INFECCION DE VIAS URINARIAS SITIO NO ESPECIFICADO	Enfermedad general	16,503

4 MONGODB

4.1 SENTENCIA O CONSULTA DE CREACIÓN DEL DOCUMENTO(S)

Observación: Escriba el código en MongoDB para crear al menos 20 documentos que correspondan a su conjunto de datos específico. Almacene en el repositorio **(REPO EN GITHUB)** el script con el nombre de T1.4.1.Creacion_Documentos.sql

Link del archivo:

https://github.com/hrodriguezgi-un/AGD_TrabajoGrupal/blob/main/scripts/T1.4.1.Creacion_Documentos.mongodb

4.2 SENTENCIA DE CONSULTA

Observación: Realice la exploración básica de los datos, conteos totales y por categorías, máximos, promedio y mínimos. Es decir, aplique estadística descriptiva con el fin de conocer las propiedades de los datos y entenderlos lo mejor posible. Use solamente sentencias SQL. Anexe las tomas de pantalla donde evidencie la sentencia SQL y su correspondiente ejecución. Además, Almacene en el repositorio **(REPO EN GITHUB)** el script con el nombre de T1.4.2.Consultar_Datos.sql

Link del archivo:

https://github.com/hrodriguezgi-un/AGD_TrabajoGrupal/blob/main/scripts/T1.4.2.Consultar_Datos.mongodb

```
// Documentos asociados a la IPS: 'FUNDACION HOSPITALARIA SAN VICENTE DE PAUL'
db.atenciones.find( {ips: {$eq: 'FUNDACION HOSPITALARIA SAN VICENTE DE PAUL'} }
);
```

```

Edit Document
{
  "_id": {
    "$oid": "60b2526328bbd0428fd61ee0"
  },
  "consecutivo": 641306,
  "anio": 2015,
  "eas": "SIN ADMINISTRADORA",
  "ips": "FUNDACION HOSPITALARIA SAN VICENTE DE PAUL",
  "departamento": "ANTIOQUIA",
  "municipio": "MEDELLIN",
  "causa_externa": "Otro tipo de accidente",
  "dx_ppal_egreso": "QUEMADURAS QUE AFECTAN DEL 10% AL 19% DE LA SUPERFICIE DEL CUERPO",
  "edad": 1,
  "tipo_edad": "ANIOS",
  "sexo": "M",
  "zona": "U",
  "servicio": "HOSPITALI",
  "total_atenciones": 1
},
{
  "_id": {
    "$oid": "60b2526328bbd0428fd61ee1"
  },
  "consecutivo": 641307,
  "anio": 2015,
  "eas": "SIN ADMINISTRADORA",
  "ips": "FUNDACION HOSPITALARIA SAN VICENTE DE PAUL",
  "departamento": "ANTIOQUIA",

```

```

// Cantidad de Atenciones por Hombre y Mujer menores o iguales a 20 años
const aggregation1 = [
  { $match: { tipo_edad: { $eq: 'ANIOS' }, edad: { $lte: 20 } } },
  { $group: { _id: '$sexo', atenciones: { $sum: '$total_atenciones' } } }
];
db.atenciones.aggregate(aggregation1);

```

```

[
  {
    "_id": "M",
    "atenciones": 450
  },
  {
    "_id": "H",
    "atenciones": 459
  }
]

```

5 ANÁLISIS DE LECTURA

Observación: Considerando el artículo: “The Definitive Guide to Graph Databases for the RDBMS Developer” de Neo4J. Compartido en las carpeta de lecturas recomendadas. Analice y responda cada pregunta en máximo 150 palabras:

1. ¿Cuáles son las limitaciones, que se pueden inferir de la lectura, para migrar los conjuntos de datos relacionales a NoSQL?
 - El tiempo de descarga de datos puede llegar a hacer amplio, ya que es necesario por ejemplo “Coincidir con los nodos existentes en función de las búsquedas de atributos, unir nodos y relaciones con etiquetas y atributos de los datos de la fila, configurar nuevas etiquetas y eliminar las obsoletas”

2. ¿Cuáles limitaciones adicionales que se deben considerar, a parte de las mencionadas en el artículo?
 - Se debe contar con un diseño previo para no complejizar los queries
3. ¿Cuáles son las razones (criterios) que se deben considerar para migrar un conjunto de datos relacionados a NoSQL?
 - El rendimiento del sistema: tiempos de respuesta largos para consultar los datos
 - Consultas de tipo self joins
 - Múltiples cambios sobre los esquemas
 - Ciclos de desarrollo acelerados.