

Bases de Datos No Relacionales

Informe del proyecto de fin de curso

Andoni Alvarellos, Jorge Triñanes

04/07/2022

1.. Introducción

El proyecto tiene por objetivo facilitar el análisis de los datos de las compras estatales de Uruguay publicados por la Agencia Reguladora de Compras Estatales (ARCE) [2]. Éste tipo de análisis es una vía para mejorar la eficiencia en las compras y combatir la corrupción.

La información está organizada por año y está disponible para el período 2002 a 2021. Los datos están en formato json, consistente con el estándar OCDS (Open Contracting Data Standard) [10] promovido por Open Contracting Partnership (OCP) [5]. A partir de una revisión de los datos publicados por ARCE, se detectó que estos contienen un subconjunto de los campos definidos en el esquema de OCDS [9]. Los principales datos publicados son: entidad compradora, proveedor, artículo, importe, estado de la compra, fecha.

La información acerca de las compras puede ser usada con dos propósitos distintos:

- a. evaluar la eficiencia en el proceso de compras y
- b. identificar casos sospechosos de corrupción.

Para el primero se tomó como referencia la publicación “Indicators to diagnose the performance of a procurement market” [6]. Para el segundo se consideró el material “RED FLAGS for integrity: Giving the green light to open data solutions” [11]. En ambos casos se tomó en cuenta que los datos disponibles contienen un subconjunto de los datos del estándar OCDS por lo que solo resulta posible considerar el subconjunto de información efectivamente disponible.

Dados los problemas a resolver, en principio pareció más adecuado trabajar sobre una base de grafos para facilitar el análisis de las conexiones entre compradores y proveedores, que es uno de los temas de interés [11, 16]. Sin embargo, debido a la estructura de la información disponible resulta difícil cargarlas directamente en una base como Neo4j, por lo que se optó por cargar primero la información en MongoDB para facilitar su manipulación. Surgió entonces la pregunta acerca de si efectivamente sería mejor encarar el proyecto cargando los datos en una base de grafos, dado que habría un paso adicional para pasar de MongoDB a Neo4j. Se decidió encarar el proyecto utilizando MongoDB para realizar el análisis y depuración de los datos y llevar a cabo

buena parte de las consultas de interés. Adicionalmente se planteó cargar información en Neo4j apuntando a evaluar si esta base de datos de grafos efectivamente presenta ventajas significativas para el análisis de un grafo bi-partito conformado por compradores y proveedores en que las únicas relaciones de interés son de tipo comprador-proveedor. Para el proyecto se planteó utilizar cuentas gratuitas de MongoDB Atlas y Neo4j Aura. La versión gratuita de Atlas tiene una restricción de 500 Mb de espacio y la de Aura puede soportar como máximo 50.000 nodos y 175.000 relaciones. Estas restricciones se consideraron adecuadas para almacenar varios años de compras. Se decidió cargar tantos años como resulte posible, partiendo de los más recientes. Uno de los principales riesgos identificados en la definición del proyecto es el referido a la calidad de los datos por lo que una de las primeras actividades consistió en evaluar la calidad de la información disponible.

A continuación se presenta la estructura del informe. La sección 2 presenta un análisis de los datos disponibles, realizado a partir de la carga de los datos en MongoDB. Se describe la estructura de los datos y se contrasta esta con el estándar OCDS. Asimismo se presenta una evaluación de la calidad de los datos y las acciones que se llevaron a cabo para contemplar los problemas detectados. La sección 3 describe un conjunto de transformaciones de los datos realizadas antes de llevar a cabo las consultas. La sección 4 presenta las consultas implementadas. Estas surgen de filtrar las consultas de interés tomando en cuenta las que se pueden implementar a partir de la información disponible. La sección 5 presenta el análisis de la relación comprador-proveedor implementada en Neo4j. En la sección se analizan distintas alternativas respecto a los datos a cargar en la base de grafos y la elección llevada a cabo. Se describe la carga de los datos en Neo4j lo que incluye una descripción del modelo correspondiente y de los programas elaborados para llevar a cabo la carga de la información y las consultas elaboradas tanto en Neo4j para analizar las relaciones comprador-proveedor. La sección 6 describe la forma en que se llevó a cabo el proyecto y la sección 7 presenta una evaluación de los resultados obtenidos y de la ejecución del proyecto y propone posibles trabajos futuros.

2.. Datos disponibles

Esta sección describe las características de los datos de las compras estatales del estado uruguay publicadas como datos abiertos por ARCE [2]. Al momento de llevar a cabo la definición del proyecto [2] la página Web con los datos del año 2020, que fue el que se analizó con cierto detalle, presentaba un único archivo json que contenía un único documento con del orden de 5000 objetos correspondientes a otros tantos eventos de compra. Sin embargo, a partir del comienzo del trabajo efectivo en el proyecto, en la misma página aparecen para ese mismo año 24 archivos json, cada uno de los cuales contiene un único documento con del orden de 5000 objetos correspondientes a eventos de compra. Esto implica que los volúmenes de información se multiplicaron por 24, por lo que el alcance del proyecto se acotó a considerar tan solo los últimos cuatro años disponibles, del 2018 al 2021.

2.1. Descripción de los datos disponibles

Los datos de las compras de cada año se presentan como un conjunto de 24 archivos. Estos archivos tienen nombres de la forma x-MM-AAAA.json donde x toma uno de los valores a o l, MM identifica al mes y AAAA al año. Los archivos cuyos nombres comienzan con una a contienen datos de las adjudicaciones (awards) y los que comienzan con una l contienen los datos de los llamados (tender). En la estructura 1 se presenta una aproximación del esquema general de los datos contenidos en estos archivos con un formato basado en json.

Listado 1: Compras publicadas en un año-mes

```
{
  "_id" : object ,
  "version": str
  "publishedDate": str ,
  "publisher": object ,
  "uri": str ,
  "license": str ,
  "publicationPolicy": str ,
  "releases":
  [
    {
      "initiationType": str ,
      "parties":
      [
        {
          "id": str ,
          "roles": [str],
          "name": str ,
          "contactPoint":
          {
          }
        }
      ]
    }
  ]
  "tag": [str],
  "date": str .
  "ocid": str ,
  "id": str ,
  "tender":
  "id": str ,
  "hasEnquiries": bool ,
  "procurementMethodDetails": str ,
  "procurementMethod": str ,
```

```
"title": str ,
"description": str ,
"procuringEntity":
  {
    "id": str ,
    "name": str
  }
"submissionMethodDetails": str ,
"buyer":
  {
    "id": str ,
    "name": str ,
  }
"awards":
  [
    {
      "id": str ,
      "items"
      [
        {
          "classification":
            {
              "description": str ,
              "id": str ,
              "scheme": str
            }
          "unit":
            {
              "id": str ,
              "name": str ,
              "value":
                {
                  "amount": str ,
                  "currency": str
                }
            }
        }
      ]
    }
  ]
"status": str ,
"date": str ,
"value":
  {
    "amount": str ,
```

```
        "currency": str
    }
    "documents":
    [
        {
            "id": str,
            "documentType": str,
            "datePublished": str,
            "url": str,
            "language": str,
            "format": str
        }
    ]
}
]
```

2.2. Carga en MongoDB

Para cargar los datos en los documentos se utilizó la herramienta `mongoimport` presente en el paquete `mongodb-database-tools`^[3] provisto por MongoDB. Para poder analizar posibles diferencias entre los esquemas de los datos de los distintos archivos (uno por año) presentes en el conjunto de datos, se decidió cargar cada uno en una colección diferente. Se realizó mediante el siguiente script `zsh`:

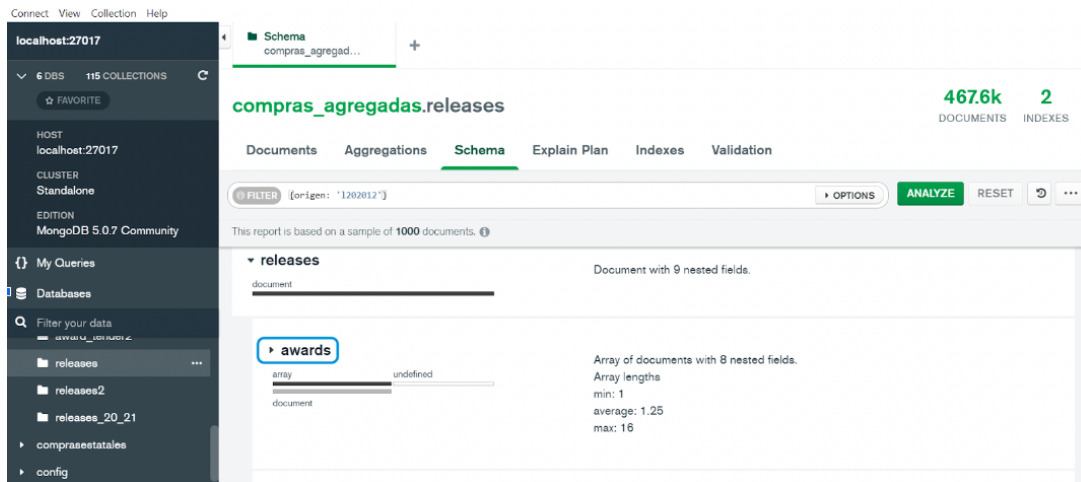
```
for folder in ./**;do
  for file in $folder/*(N.);do
    mongoimport --collection=compras$(basename $folder)
    "mongodb+srv://BDNRT01:BDNRT01@cluster0.kj6ip.mongodb.net/
    comprasestatales" $file;
  done;
done;
```

2.3. Estructura de los datos

A continuación se utilizó la herramienta “Analyze Schema” de MongoDB `compass`, para tener una visión general de los datos presentados. El mismo toma una muestra de los documentos en la colección y presenta la estructura de los datos analizados, mostrando el porcentaje de documentos en los que se encuentra cada campo. La figura 1 contiene un ejemplo del resultado de ésta función.

De este análisis se obtuvo que no hay una diferencia significativa en la estructura general de las publicaciones (*releases*) para los distintos años.

Figura 1: Ejemplo output Analyze Schema



2.4. Contenido y comparación con el estándar OCDS

Como se había identificado durante la etapa de definición del proyecto, en términos generales los *releases* conforman con la estructura definida por el estándar de OCDS de datos[9], aunque no se presentan todos los datos del estándar. Sin embargo, estas *releases* no se encuentran agrupadas en registros (*record*) de acuerdo al esquema propuesto por OCDS[7]. Éstos registros agrupan toda la información publicada a través del tiempo para un mismo proceso de compras. Considerando que la información de empaquetado y publicación no era necesaria para éste proyecto, se extrajo la información de todos los releases y se aplicó el proceso de combinación definido en el estándar[8].

Antes de implementar el algoritmo fue necesario extraer todas las releases, en los distintos paquetes, durante esta etapa también se combinaron los datos que inicialmente se habían cargado en colecciones distintas. Para eso se aplicó el siguiente pipeline a todas las colecciones:

```
[{
  $unwind: '$releases'
}, {
  $replaceRoot: {
    newRoot: '$releases'
  }
}, {
  $merge: 'releases'
}]
```

Las particularidades del algoritmo de combinación complican su implementación como un pipeline de mongodb. Sin embargo, la Open Contracting Partnership disponibiliza un paquete de python con un conjunto de operaciones definidas en OCDS: ocdskit. En

particular, el comando *compile??* realiza este proceso sobre un conjunto de releases en formato json. Aprovechando el carácter de base documental y la compatibilidad de modelos, se realizó un dump mediante la herramienta mongoexport (también incluida en mongodb-database-tools), se procesó el archivo resultante con ocdskit y se volvió a cargar el resultado (releases de tipo *compiled*) a una nueva colección en la que todos los releases contienen la última información disponibilizada al momento de la consulta de los datos.

```
mongoexport --db=comprasestatales --collection=releases
--out=releases.json;
cat releases.json | ocdskit compile > compiledReleases.json;
mongoimport --db=comprasestatales
--collection=compiledReleases compiledReleases.json
```

2.5. Calidad de los datos

En esta etapa se analizó los datos de los releases mediante el uso de las funciones de búsqueda y agregación de mongodd, de cara al conjunto de consultas propuestas por el estándar [11, 6]. El principal objetivo era identificar si la integración de datos de distintos releases formaban records coherentes (pertenecientes al mismo proceso), para eso se aplicó filtros mediante el uso del campo id (interno a la entidad informante) y efectivamente para cada release resultante todos los datos contienen datos del mismo proceso de compras. Si bien el conjunto de datos resultantes contiene la estructura adecuada para el procesamiento, se identificaron debilidades en la completitud de los datos que limitan la validez del estudio de muchas de ellas.

- Aproximadamente un 50% de los procesos no poseen información del llamado (*tender*), sino solamente de la adjudicación final.

Consecuencia: los resultados de las consultas que analizan información referente al llamado (como tipo de proceso) sólo analizan el subconjunto de procesos con esta información.

- No hay información de montos (estimados) en los llamados

Consecuencia: no se puede evaluar algunos indicadores de eficiencia que refieren a la planificación del gasto.

- El esquema usado para la clasificación de artículos no está entre los sugeridos por OCDS[15], sino que es propio de ARCE.

Consecuencia: el análisis de las métricas referentes a los mercados de artículos depende de esta clasificación.

- Los valores totales para una adjudicación (*award.value*) no están informados para todos procesos, sin embargo, los que no tienen este valor presentan la información de valor para cada artículo.

Consecuencia: Fue necesaria una transformación previa de los datos para incluir este valor (ver sección 3.).

- Algunos procesos no contienen la información para Proveedores, Comprador, o Artículos.

Consecuencia: A la hora de calcular algunos datos en los que esta información era relevante se decidió utilizar el valor “Not informed” para los procesos que no la disponibilizan.

3.. Transformaciones previas

Antes de realizar las consultas se realizaron algunas alteraciones a los datos para facilitar el análisis dentro del alcance de este proyecto.

En primer lugar se decidió pasar todos los montos a moneda nacional, para poder compararlos. Se eligió tomar como estimación el promedio de la cotización del año y se ejecutaron los siguientes comandos:

```
db.compiledReleases.update({
  "awards": {
    $exists: true
  }
},
{
  $mul: {
    "awards.$[award].items.$[item].unit.value.amount":
    <cotizacion_dolar>
  },
  $set: {
    "awards.$[award].items.$[item].unit.value.currency": "UYU"
  }
},
{
  multi: true,
  arrayFilters: [
    {
      "award.items": {
        $exists: true
      }
    },
    {
      "item.unit.value.currency": "USD"
    }
  ]
}
```



```
});
db.compiledReleases.update(
{
  "awards": {
    $exists: true
  }
},
{
  $mul: {
    "awards.$[award].value.amount": <cotizacion_dolar>
  },
  $set: {
    "awards.$[award].value.currency": "UYU"
  }
},
{
  multi: true,
  arrayFilters: [
    {
      "award.value.currency": "USD"
    }
  ]
});
```

Además, como se señaló en la sección anterior, se decidió calcular los valores totales de una adjudicación a partir de los valores de los artículos cuando fue posible:

```
db.compiledReleases.update({
  "awards": {
    $exists: true
  },
  "awards.value": {
    $exists: false
  }
},
[
  {
    $addFields: {
      "awards.value": {
        currency: "UYU",
        amount: {
          $reduce: {
```

```

input: "$awards.items",
initialValue: 0,
in: {
  $add: [
    "$$value",
    {
      $sum: {
        $map: {
          input: "$$this",
          as: "item",
          in: {
            $multiply: [
              "$$item.unit.value.amount",
              "$$item.quantity"
            ]
          }
        }
      }
    }
  ]
}
],
{
  multi: true
})

```

4.. Consultas

En esta sección se presentan las consultas realizadas sobre la base de datos documental, con el objetivo de medir indicadores que alerten sobre la posibilidad de dos problemáticas: ineficiencia en el gasto y corrupción. La selección de las consultas se realizó a partir de una revisión de los documentos publicados por la Open Contracting Partnership *Indicators to diagnose the performance of a procurement market*[6] y *Red flags for integrity*[11] respectivamente y las planillas asociadas [14], [13] y [12] con la correspondiente información para medirlos con los datos del estándar OCDS.

En base al análisis detallado en la sección 2. se seleccionaron las siguientes consultas:

1. Cantidad e importes totales de llamados y adjudicaciones por año.
2. Cantidad de anulaciones / cancelaciones.
3. Cantidad de entidades compradoras.
4. Cantidad e importes por tipo de compra.
5. Cantidad de proveedores en el período.
6. Proporción de llamados abiertos (competitivos) respecto al total (en cantidad y en importes).
7. Proporción de llamados con un único oferente en los llamados abiertos respecto al total (en cantidad y en importes).
8. Promedio y mediana de cantidad de oferentes en llamados competitivos
9. Índice de Herfindahl-Hirschman: Es la suma de los cuadrados del porcentaje del valor de mercado que fue asignado a un proveedor, para cada artículo. El índice varía entre 0 y 10000. Valores mayores a 4000 indican alta concentración. Resultan de interés los ítems con valores mayores a 4000, y para estos poder contar con los importes asociados a cada uno.
10. Ratio número de adjudicaciones/número de proveedores para un ítem.
11. Información por entidad compradora: Cantidad de proveedores, Cantidad de adjudicaciones, Importe adjudicado, Top 10 por importe adjudicado
12. Cantidad y monto de adjudicaciones por proveedor

En el anexo **A** se encuentran las consultas elaboradas en MongoDB.

Las primeras cinco consultas tienen por objetivo poder encuadrar el análisis de las compras. Tener idea de los importes totales, la cantidad de compras totales y por tipo de compra, la cantidad de compradores y de proveedores. Las otras consultas brindan información acerca de la competencia en el mercado. Las consultas 7 y 8 no se pudieron implementar porque no está disponible la información acerca de las ofertas. De acuerdo a nuestra comprensión del estándar OCDS la disponibilidad de esta información es muy relevante para asegurar la transparencia e integridad del proceso de compras. Un indicador muy interesante es el índice de Herfindahl-Hirschman que da una idea de cuánta competencia hay en un mercado en particular de cierto tipo de artículos. En caso de no contar con información de tipo, se puede calcular artículo por artículo. En este caso se presenta el pipeline para el cálculo incluyendo además el monto total adjudicado en el mercado de ese artículo y la cantidad de adjudicaciones. Estos campos ayudan a analizar la situación de los mercados con alta concentración. A modo de ejemplo se agrega al final una etapa de \$match adicional que realiza un corte en particular sobre los datos. Además, en esta consulta, se intenta inferir el valor de cada artículo para los casos en el que el único valor presente es el total de la adjudicación.

4.1. Resultados

A continuación se presenta un análisis inicial sobre los indicadores señalados en el punto anterior. El resultado de la ejecución de las consultas puede encontrarse en el anexo B

1. Cantidad e importes totales de llamados y adjudicaciones por año.

Lo primero que se puede notar al analizar esta consulta es que existe información publicada para años anteriores a los que corresponde el paquete de datos. En principio, no sería problemático, ya que sería posible que existieran actualizaciones de procesos anteriores, pero registros con fecha de más de 30 años anterior al período parecerían indicar otro problema de calidad de datos no detectado en el análisis del esquema. Por otra parte, se constata que la diferencia entre adjudicaciones y llamados publicados a través de los años de publicación.

2. Cantidad de anulaciones / cancelaciones.

De acuerdo a la guía de Open Contracting[14], un alto número de anulaciones o cancelaciones puede ser un indicador de ineficiencia. Dentro del conjunto de datos reportados, hay 27.227 procesos que no fueron adjudicados sobre un total de 1.110.699. Al rededor de un 2.45 %.

3. Cantidad de entidades compradoras.

El número de entidades compradoras es un indicador del tamaño del mercado e informa posibles conclusiones para el resto de los indicadores presentados. En este conjunto de datos se incluyen un total de 395 entidades compradoras diferentes.

4. Cantidad e importes por tipo de compra.

Open contracting, en su guía de casos de uso, identifica como un riesgo a la integridad (sospecha de corrupción) un alto porcentaje de procesos que se resuelven por procesos directos frente a procesos competitivos. En los datos presentados, la gran mayoría de los procesos corresponden a llamados de licitación abierta (tanto en cantidad como en monto total), sin embargo corresponde señalar el alto número de procesos para los que no se cuenta con esta información (más de 700.000).

5. Cantidad de proveedores en el período. Un mayor número de proveedores frente a compradores es un indicador de mayor competitividad (y por lo tanto integridad) en el mercado. En este caso se identificaron 29.384 proveedores diferentes.

6. Proporción de llamados abiertos (competitivos) respecto al total (en cantidad y en importes).

Constuyendo sobre el indicador 4, podemos ver que el 66.7 % de los procesos incluido carece de información respecto al procedimiento empleado en la adjudicación. Aquí queda claro el impacto del problema de calidad de datos identificado al

comienzo del proyecto y constituye un problema de transparencia ya que dificulta la elaboración de un identificador de integridad. Analizando la porción de datos para los que conocemos el procedimiento, más del 95 % de los procesos de compras (y del monto total asociado) corresponde a procesos abiertos, lo cual sería un indicador muy positivo pero este dato carece de solidez debido al problema anteriormente mencionado.

7. Proporción de llamados con un único oferente en los llamados abiertos respecto al total (en cantidad y en importes).

El estándar señala que un mayor número de procesos que se resuelven con único oferente es una posible señal de llamados diseñados y compras directas encubiertas [14] [13]. Sin embargo, con los datos incluidos en este dataset no es posible determinar este indicador debido a que sólo se presenta la información de la oferta ganadora. Esto conspira una vez más contra la transparencia y el valor de los datos publicados.

8. Promedio y mediana de cantidad de oferentes en llamados competitivos

Se repite la situación del punto anterior.

9. Índice de Herfindahl-Hirschman

Este indicador permite determinar la concentración de los mercados para cada artículo presente en el conjunto de datos. Tras aplicar la métrica se encontró un gran número de artículos con un índice de 4.000, valor sugerido para identificar posibles mercados concentrados [12]. Para aproximar un análisis inicial, agregamos algunos filtros que buscan obtener los mercados más significativos, tanto en cantidad como en el total de dinero que se asignó a los distintos proveedores. Filtrando aquellos que tienen una cantidad significativa de procesos y ordenando por el monto total asignado, podemos ver que los principales artículos con mercados con baja competitividad corresponden al rubro de medicina. Si bien este indicador señala la baja competitividad como un posible indicador de faltas de integridad, es posible que los números obtenidos respondan a la realidad del sector en cuestión. La interpretación más detallada de estos valores escapa al alcance definido para este proyecto. Además de esa situación, la ejecución de la consulta da como resultado otra cantidad de artículos con mercados concentrados en variados rubros (papelería, transporte, software, etc.)

Otro dato a mencionar es que una vez más podemos ver una gran cantidad de adjudicaciones que no tienen datos sobre artículos o proveedores. Si bien esto se evidencia por el problema de calidad ya mencionado anteriormente, los datos que no poseen información parecen carecer a la vez de ambos campos, lo que resulta en una altísima concentración del artículo "Not informed".

10. Ratio número de adjudicaciones/número de proveedores para un ítem.

Al igual que el anterior, en ésta consulta obtenemos información sobre concentración de mercados. Una ratio mayor indica una alta concentración de las adjudicaciones a un grupo reducido de proveedores[14]. Los datos resultantes parecen indicar una vez más que existen muchos mercados de artículos que tienen una alta concentración en el conjunto de datos reportados. Al igual que en el anterior punto, el área de productos y servicios médicos cuenta con los mayores ratios de concentración.

11. Información por entidad compradora: Cantidad de proveedores, Cantidad de adjudicaciones, Importe adjudicado, Top 10 por importe adjudicado

En este caso, recolectamos información sobre las entidades con mayor monto de compras reportado dentro del conjunto de datos. El análisis de los datos obtenidos no cae dentro del alcance del proyecto.

12. Cantidad y monto de adjudicaciones por proveedor Proveedores que concentran una mayor proporción de los montos adjudicados pueden constituir una amenaza para la competitividad[13]. Se presentan los datos de algunos de los adjudicatarios con un mayor monto acumulado en el período informado.

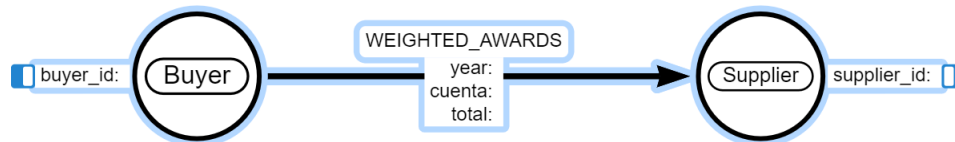
5.. Análisis de las relaciones comprador-proveedor

Esta sección describe el análisis llevado a cabo de la relevancia de las relaciones comprador-proveedor. El primer paso consiste en definir el modelo adecuado para realizar ese análisis en Neo4j. Una vez definida la estructura de la base de datos dada por los nodos, relaciones y propiedades, se debe implementar la programación necesaria.

5.1. Modelo

Se consideraron diferentes alternativas de modelado. La primera consiste en cargar la mayor parte de la información almacenada en MongoDB. Esto implica contar con la información de los compradores, proveedores, llamados, adjudicaciones, incluyendo el detalle de los ítems e importes. La segunda alternativa considerada consiste en tener toda la información de las compras pero abierta por ítem. Esta alternativa es interesante para realizar análisis por ítem además de por importe. Tiene como inconveniente el gran volumen de datos requerido y la gran cantidad de relaciones que involucra. La tercera alternativa consiste en tener solo la información de los importes por cada compra. La cuarta alternativa considerada consiste en tener solo los datos de los totales de importes vendidos por un proveedor a un comprador en un año. Este es el modelo más simple que permite analizar las relaciones entre compradores y proveedores. Los principales indicadores de la relevancia de la relación entre un comprador y un proveedor en un período están dados por el monto total comprado por el comprador a ese proveedor y la cantidad de compras a ese proveedor. La figura 2 presenta ese modelo.

Figura 2: Esquema de la base Neo4j



5.2. Carga de los datos en Neo4j

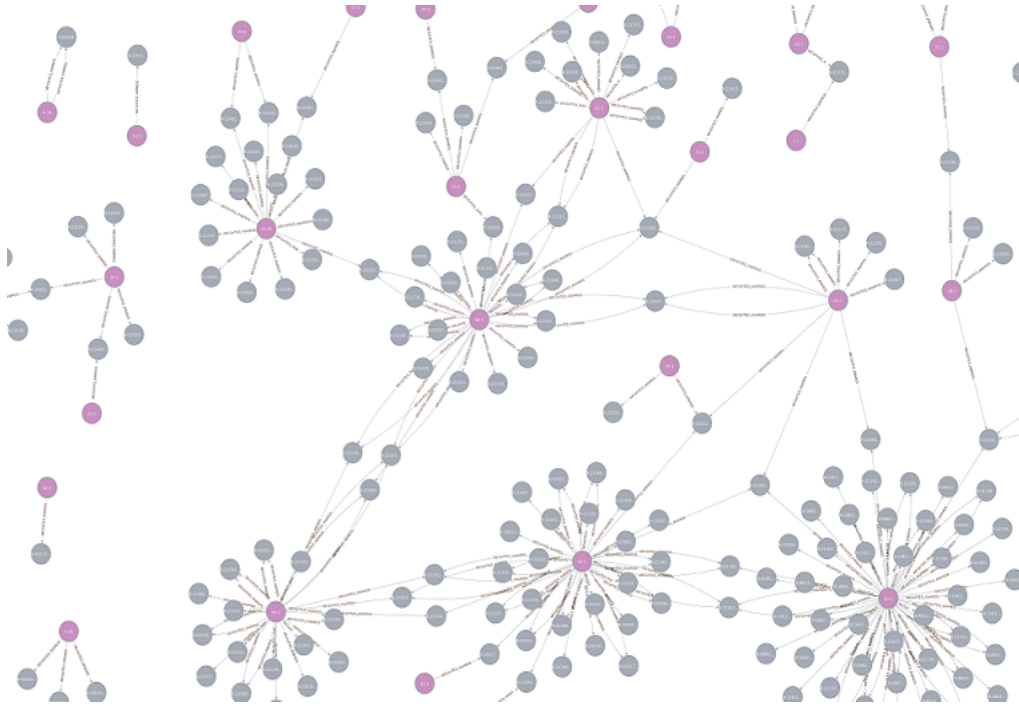
La carga de los datos en Neo4j se realizó a partir de archivos csv generados como salida de agregaciones de MongoDB. Un archivo con los datos de los compradores, otro con los datos de los vendedores y otro más con los datos de las relaciones. Primero se cargaron los nodos y luego las relaciones. Los dos primeros no merecen más comentarios. Por el contrario, la generación del archivo para cargar las relaciones tiene cierta complejidad. Se partió de los datos de las adjudicaciones (awards), que son las compras efectivamente realizadas. En MongoDB se seleccionaron los "releases" publicados en los años 2020 y 2021 con "tag" "award", y con "award.date" de esos años. Se dejaron afuera registros con el campo "awards.items.unit.value.amount" con valor negativo que corresponden a ventas del estado, no a compras. También se descartaron 128 registros correspondientes al año 2021 que carecen del campo "buyer_id" que identifica al comprador. La mayor parte de las compras están valoradas en pesos uruguayos (UYU) pero una parte significativa lo están en dólares USA (USD). Estos importes se convirtieron a pesos uruguayos a la cotización promedio de cada uno de los años: 42,01 para 2020 y 43,55 para 2021. De esta forma todos los importes quedaron expresados en pesos corrientes, lo que se considera adecuado dado el marco del proyecto. Se cargaron 276 compradores, 11020 proveedores y 58337 relaciones con totales de compra.

5.3. Consultas en Neo4j

La carga de los datos de las relaciones ponderadas entre compradores y proveedores permite realizar diversos análisis del grafo correspondiente. Los grafos de las compras son normalmente muy densos, pero la ponderación permite solo considerar un subconjunto de relaciones que tengan un peso significativo y den lugar a un grafo menos denso.

En la figura 3 se presenta una parte del grafo obtenido seleccionando las relaciones entre compradores y proveedores que en los años 2021 o 2022 hayan involucrado un importe total superior a \$40.000.000 (pesos uruguayos) a valores corrientes. La consulta devuelve 287 nodos (48 compradores y 239 proveedores) y 349 relaciones. En el Listado 2 se presenta el texto de la consulta correspondiente.

Figura 3: Parte del grafo de compradores y proveedores en compras de más de 40 millones de pesos en 2020 o 2021 - El color rosado identifica a compradores y el gris a proveedores



Listado 2: Consulta en Cypher para seleccionar los nodos de compradores y proveedores relacionados por compras de más de 40 millones de pesos en 2020 o 2021

```
MATCH (b: Buyer) -[a:WEIGHTED_AWARDS] -(s: Supplier)
WHERE a.total > 40000000
RETURN a, b, s
```


Figura 4: Detalle de compradores y proveedores en compras de más de 40 millones de pesos en 2020 o 2021

b.buyer_id	s.supplier_id	a.awards_year	a.total
80-1	R-010127250015	2021	67.317.413
81-1	R-020113000019	2020	62.216.790
61-1	R-020124300019	2020	132.458.974
61-1	R-020124300019	2021	68.034.477
66-1	R-020288140014	2020	139.314.588
66-1	R-020288140014	2021	112.576.233
4-26	R-020384400013	2021	180.507.420
98-1	R-020406690019	2020	56.444.609
4-26	R-020415520015	2021	66.967.571
29-7	R-040013090010	2020	133.808.400
61-1	R-050067690013	2020	80.560.438
61-1	R-050067690013	2021	258.206.916
61-1	R-070012680010	2020	190.063.192
61-1	R-070012680010	2021	160.553.731
4-26	R-080194250014	2021	78.319.680
81-1	R-080221950018	2020	121.123.144
93-1	R-100161220017	2021	47.755.555
81-1	R-110394010012	2020	118.291.237
29-68	R-110400500012	2021	52.800.000
61-1	R-140098390010	2021	47.119.250
25-1	R-150106330015	2020	41.638.808
61-1	R-160117370010	2020	59.068.964
10-4	R-160175570019	2020	332.412.595
61-1	R-160175570019	2020	85.125.199
4-1	R-170226020018	2021	44.931.220
25-1	R-190228630012	2020	66.982.316
4-1	R-190228630012	2021	44.695.370

En la figura 4 se muestra una parte de los datos correspondientes al grafo de la figura 3, ordenados por "supplier_id" para facilitar la identificación de casos en que un proveedor lo es de más de un comprador. El análisis del grafo de la figura 3 permite identificar algunas comunidades de forma directa ya que existen algunas zonas del grafo que están desconectadas del resto. Para las áreas que están conectadas, puede resultar un tanto trabajoso. La consulta del listado 3 devuelve una lista ordenada de parejas de compradores para ayudar en la identificación de comunidades. Esa consulta, partiendo de cada comprador que podría pertenecer al grafo, devuelve todos los compradores que están conectados por un camino de largo 2. Un cambio en el primer elemento de la pareja en el arreglo de parejas del resultado indica que comienza una nueva comunidad candidata. La consulta identifica comunidades candidatas que no son disjuntas. Este

enfoque deja afuera las comunidades compuestas por un único comprador aislado, las que deben ser consideradas por separado. De esta forma es posible obtener una primera versión de la parte correspondiente a los compradores de esas comunidades. Una vez que se tiene la lista de todos los compradores de una comunidad, obtener los proveedores que la conforman es directo ya que todo proveedor la integra por estar relacionado con algún comprador.

Una mejor solución para el problema de la detección de comunidades consiste en utilizar algoritmos específicos disponibles en Neo4j Data Science Library [4] presentados en el apartado Graph algorithms/Community detection, pero esto excede el alcance del proyecto.

Listado 3: Consulta en Cypher para brindar una primera aproximación a la identificación de comunidades en el grafo de la figura 2

```
MATCH (b:Buyer)
WITH COLLECT(b) AS Coll_b
UNWIND Coll_b AS b
MATCH (b:Buyer)-[a1:WEIGHTED_AWARDS]-()
-[a2:WEIGHTED_AWARDS]-(m:Buyer)
WHERE a1.total > 40000000 AND a2.total > 40000000 AND
NOT b.buyer_id = m.buyer_id
RETURN COLLECT(DISTINCT[b.buyer_id , m.buyer_id]) AS Coll_bm
```

6.. Ejecución del proyecto

En esta sección se describe la forma en la que se llevó a cabo el proyecto, las principales actividades y desviaciones respecto al plan original que fue presentado en el documento de definición del proyecto [1].

En el cuadro 1 se presenta el cronograma real de ejecución del proyecto. En términos generales el proyecto se ejecutó de acuerdo al plan inicial. Las principales desviaciones estuvieron relacionadas con la calidad de los datos, la comprensión de los mismos y el estudio de la documentación del estándar OCDS. En el plan inicial estaba prevista tan solo una tarea de evaluación de calidad al comienzo del proyecto. Sin embargo, si bien se detectaron algunos problemas de calidad al inicio, la detección de nuevos problemas continuó a lo largo del proyecto por lo que fue necesario tomar decisiones y acciones para superarlos. Por otra parte, la comprensión del contenido de los datos llevó más tiempo del planificado y a medida que avanzaba el proyecto se detectó documentación adicional relevante.

Actividad	5			6			7	
	16	23	30	06	13	20	27	04
Preparación del proyecto								
Definición del proyecto	X							
Evaluar calidad de los datos		X						
Primer plan de trabajo	X							
Exposición del plan		X						
Ajustar objetivos y alcance		X						
Ejecución								
Obtener datos								
Cargar en MongoDB	X	X	X			X	X	
Evaluar volúmenes		X	X			X		
Identificar recursos de necesarios		X						
Entender y ajustar datos								
Estudiar documentación	X	X	X			X	X	
Evaluar y ajustar calidad de los datos		X	X	X	X	X		
Consultas								
Identificar consultas de interés factibles			X	X		X		
Priorizar y ajustar alcance			X	X		X		
Elaborar consultas priorizadas				X	X	X	X	X
BD de grafos								
Modelar								
Cargar en Neo4j			X			X	X	
Obtener recursos		X	X					
Preparar carga					X			
Ejecutar carga					X	X	X	
Identificar consultas de interés factibles						X		
Elaborar consultas priorizadas						X	X	X
Informe y defensa								
Definir estructura del informe		X	X					
Elaborar el informe			X			X	X	X
Revisar y ajustar								X

Cuadro 1: Actividades por semana terminada en un lunes

7.. Evaluación de los resultados proyecto

Esta sección presenta una evaluación de los resultados obtenidos en el marco del proyecto, así como proponen posibles trabajos futuros.

Se obtuvieron distintos tipos de resultados:

- a. Evaluación de la calidad de los datos publicados por ARCE, su consistencia con el estándar OCDS y el grado de cobertura del mismo.

- b. Un conjunto de consultas en MongoDB orientados a evaluar y caracterizar los datos publicados por ARCE.
- c. Responder a la pregunta de si valía la pena cargar los datos de las compras en una base de datos de grafos.

Con respecto a los datos publicados por ARCE, un primer comentario refiere a las condiciones de su publicación. En el momento de preparar la definición del proyecto, para el año 2020 los datos publicados eran unos y a la fecha de comienzo del proyecto los datos del 2020 habían cambiado sustancialmente. La cantidad de datos se había multiplicado por 24. No hemos encontrado ninguna advertencia de ese cambio. Sería bueno que junto con la publicación de los datos se suministrara algún tipo de dato de control, por ejemplo la cantidad de objetos contenidos en cada documento json. Esto facilitaría un tratamiento automático de la carga con evaluación posterior de la calidad de la misma. El contenido de los datos es esencialmente consistente con el estándar OCDS. Sin embargo corresponde destacar que no hay información publicada respecto a las ofertas recibidas. Solo hay información acerca de los llamados y de las adjudicaciones. Muchas de las consultas de interés para la detección de alertas de la integridad están relacionadas con la información de los llamados. Por ejemplo, identificar proveedores que se presentan siempre y nunca reciben una adjudicación o los proveedores que van rotando de manera sistemática quien recibe una adjudicación. En la sección 2.5 se detallan los principales problemas de calidad encontrados en los datos publicados.

Respecto a las consultas implementadas sobre MongoDB entendemos que pueden considerarse como un primer paso para desbrozar el camino hacia el análisis de los datos publicados por ARCE acerca de las compras del estado uruguayo. La experiencia con la utilización de un motor de bases de datos documental para este tipo de análisis arrojó resultados mixtos. Si bien la compatibilidad del modelo documental con el estándar de presentación de documentos facilitó varias acciones de carga y consulta, la falta de un esquema impuesto a la hora de almacenar los datos dificultó mucho la tarea de evaluar el cumplimiento del esquema proporcionado por el estándar (el cuál estuvo lejos de ser pleno). Por lo tanto, algunos problemas "de diseño" del conjunto de datos utilizados no se hicieron evidentes hasta etapas posteriores en la realización del informe, traduciéndose en un mayor costo de tiempo y esfuerzo.

En cuanto a la pregunta de si era conveniente pasar los datos a una base de grafos, la respuesta es un fuerte y claro sí. Las facilidades que brinda Neo4j para analizar las relaciones y visualizar grafos no tienen grado de comparación con lo que ofrece MongoDB.

A partir de los resultados obtenidos surge un conjunto de posibles trabajos futuros, relacionados sobre todo con el uso de bases de datos de grafos para analizar las relaciones comprador-proveedor y posiblemente también las relaciones comprador-ítem-proveedor. En este proyecto el único factor de ponderación de la relación comprador-proveedor estuvo dado por el importe total de compras en un año. Un ponderador interesante a explorar es el porcentaje de participación de un proveedor en las compras totales de un comprador. De esta forma, para analizar las comunidades se podría partir

de las relaciones entre proveedores y compradores en las que el proveedor tiene un porcentaje de participación en el total de compras del año superior a cierto valor c , en lugar de considerar los montos totales vendidos por el proveedor en el año. Para ello en el modelo de datos de Neo4j bastaría con agregar un nodo "Año" y la relación "Compras totales" entre año y comprador con el dato del monto total comprado por el comprador en ese año. Otra posibilidad de ponderación que podría valer la pena explorar consiste en introducir los ítems en la relación, considerando la relación compuesta comprador-ítem-proveedor. Otra línea de trabajo interesante consiste en utilizar la biblioteca Neo4j Data Science Library para realizar análisis de comunidades. Si se trabaja con relaciones resumidas por año, la cantidad de relaciones totales permite analizar más años de los considerados en este proyecto.

Referencias

- [1] Alvarellos, A., Triñanes, J.: BDNR Definición del proyecto de fin de curso (2022)
- [2] ARCE: Datos históricos de compras, <https://catalogodatos.gub.uy/dataset/arce-datos-historicos-de-compras>, último acceso: 14 de mayo de 2022
- [3] mongoDBmanual: MongoDB database tools documentation (2021), <https://www.mongodb.com/docs/database-tools/>, último acceso: 4 de julio de 2022
- [4] Neo4j: The neo4j graph data science library manual v2.1 (2022), <https://neo4j.com/docs/graph-data-science/current/>, último acceso: 4 de julio de 2022
- [5] Open Contracting Partnership: About open contracting partnership, <https://www.open-contracting.org/about/>, último acceso: 14 de mayo de 2022
- [6] Open Contracting Partnership: Indicators to diagnose the performance of a procurement market, https://docs.google.com/document/d/1vSJk9-qWSTQEx9ZZc7BUhQZMHvTRcyDYVS2sl8HB__k/edit, último acceso: 14 de mayo de 2022
- [7] Open Contracting Partnership: OCDS Record Schema, https://standard.open-contracting.org/latest/en/schema/records_reference, último acceso: 3 de julio de 2022
- [8] Open Contracting Partnership: OCDS Release Merging, <https://standard.open-contracting.org/latest/en/schema/merging/>, último acceso: 3 de julio de 2022
- [9] Open Contracting Partnership: OCDS Release Schema, <https://standard.open-contracting.org/latest/en/schema/release/>, último acceso: 3 de julio de 2022

- [10] Open Contracting Partnership: OCDS Standard, <https://www.open-contracting.org/data-standard>, último acceso: 14 de mayo de 2022
- [11] Open Contracting Partnership: Red flags for integrity: Giving the green light to open data solutions, <https://www.open-contracting.org/wp-content/uploads/2016/11/OCP2016-Red-flags-for-integrityshared.pdf>, último acceso: 14 de mayo de 2022
- [12] Open Contracting Partnership: Red flags to ocds mapping, <https://docs.google.com/spreadsheets/d/12PFkUlQH09jQvcnORjcbh9-8d-NnIuk4mAQwdGiXeSM/edit#gid=656314485>, último acceso: 4 de julio de 2022
- [13] Open Contracting Partnership: Usability checks, <https://docs.google.com/spreadsheets/d/1nG7e52E1CX0XoUjz6pimW4Z7er9u3DJSs98QKdJJioE/edit#gid=1004816998>, último acceso: 4 de julio de 2022
- [14] Open Contracting Partnership: Use case guide: Indicators linked to ocds, https://docs.google.com/spreadsheets/d/1j-Y0ktZi0yhZzi-2GSabBCnzz6fF5lv8h1KYwi_Q9GM/edit#gid=1075539355, último acceso: 4 de julio de 2022
- [15] Open Contracting Partnership: Ocds item classification scheme (2017), <https://standard.open-contracting.org/1.1/en/schema/codelists/#item-classification-scheme>, último acceso: 4 de julio de 2022
- [16] Wachs, J., Fazekas, M., Kertész, J.: Corruption risk in contracting markets: a network science perspective. *International Journal of Data Science and Analytics* **12**, 45–60 (Jun 2021)

A. Anexo

A continuación se presenta la implementación (aggregation pipelines) para las consultas planteadas en MongoDB.

1. Adjudicaciones

```
[{
  $unwind: '$awards'
}, {
  $match: {
    'awards.status': 'active'
  }
}, {
  $group: {
    _id: {
      $substr: [
        '$date',
        0,
        4
      ]
    },
    cantidad: {
      $sum: 1
    },
    monto: {
      $sum: '$awards.value.amount'
    }
  }
}]
```

Llamados. Los montos estimados no están presentes en el conjunto de datos.

```
[{
  $unwind: '$tender'
}, {
  $group: {
    _id: {
      $substr: [
        '$date',
        0,
        4
      ]
    }
  }
}]
```

```
    },  
    cantidad: {  
      $sum: 1  
    }  
  }  
}  
}]
```

```
2.  [{  
    $unwind: '$awards'  
  }, {  
    $match: {  
      'awards.status': {  
        $ne: 'active'  
      }  
    }  
  }, {  
    $group: {  
      _id: null,  
      cantidad: {  
        $sum: 1  
      },  
      monto: {  
        $sum: '$awards.value.amount'  
      }  
    }  
  }  
}]
```

```
3.  [{  
    $group: {  
      _id: null,  
      compradores: {  
        $addToSet: '$buyer.id'  
      }  
    }  
  }, {  
    $project: {  
      cant_compradores: {  
        $size: '$compradores'  
      }  
    }  
  }  
}]
```


4.

```
[{
  $unwind: '$awards'
}, {
  $match: {
    'awards.status': 'active'
  }
}, {
  $group: {
    _id: {
      $ifNull: [
        '$tender.procurementMethod',
        'Not informed'
      ]
    },
    cantidad: {
      $sum: 1
    },
    monto: {
      $sum: '$awards.value.amount'
    }
  }
}]
```
5.

```
[{
  $unwind: '$awards'
}, {
  $match: {
    'awards.status': 'active'
  }
}, {
  $unwind: '$awards.suppliers'
}, {
  $group: {
    _id: null,
    distintos_proveedores: {
      $addToSet: '$awards.suppliers.id'
    }
  }
}, {
  $project: {
    _id: 0,
```

```
    cantidad_proveedores: {
      $size: '$distintos_proveedores'
    }
  }
}]
```

6. **Considerando** como "Not informed" los casos en los que el proceso no tiene el campo correspondiente, se presentan proporciones para todos los tipos de llamados.

```
[{
  $match: {
    'awards.status': 'active',
    awards: {
      $elemMatch: {
        'value.amount': {
          $gt: 0
        }
      }
    }
  }
}, {
  $group: {
    _id: '$tender.procurementMethod',
    cantidad: {
      $sum: 1
    },
    monto: {
      $sum: {
        $reduce: {
          input: '$awards.value.amount',
          initialValue: 0,
          'in': {
            $add: [
              '$$this',
              '$$value'
            ]
          }
        }
      }
    }
  }
}, {
```

```
$group: {
  _id: null,
  procurementMethods: {
    $push: '$$ROOT'
  },
  cantidad_total: {
    $sum: '$cantidad'
  },
  monto_total: {
    $sum: '$monto'
  }
}
}, {
  $unwind: '$procurementMethods'
}, {
  $project: {
    _id: 0,
    procurementMethod: {
      $ifNull: [
        '$procurementMethods._id',
        'Not informed'
      ]
    },
    porcentaje_procesos: {
      $multiply: [
        100,
        {
          $divide: [
            '$procurementMethods.cantidad',
            '$cantidad_total'
          ]
        }
      ]
    },
    porcentaje_monto: {
      $multiply: [
        100,
        {
          $divide: [
            '$procurementMethods.monto',
            '$monto_total'
          ]
        }
      ]
    }
  }
}
```

```
]
}
}
}]
```

Sin considerar los casos en los que el proceso no tiene el campo correspondiente, se presentan proporciones para todos los tipos de llamados.

```
[{
  $match: {
    'awards.status': 'active',
    awards: {
      $elemMatch: {
        'value.amount': {
          $gt: 0
        }
      }
    },
    'tender.procurementMethod': {
      $exists: true
    }
  }
}, {
  $group: {
    _id: '$tender.procurementMethod',
    cantidad: {
      $sum: 1
    },
    monto: {
      $sum: {
        $reduce: {
          input: '$awards.value.amount',
          initialValue: 0,
          'in': {
            $add: [
              '$$this',
              '$$value'
            ]
          }
        }
      }
    }
  }
}
```

```
}, {
  $group: {
    _id: null,
    procurementMethods: {
      $push: '$$ROOT'
    },
    cantidad_total: {
      $sum: '$cantidad'
    },
    monto_total: {
      $sum: '$monto'
    }
  }
}, {
  $unwind: '$procurementMethods'
}, {
  $project: {
    _id: 0,
    procurementMethod: '$procurementMethods._id',
    porcentaje_procesos: {
      $multiply: [
        100,
        {
          $divide: [
            '$procurementMethods.cantidad',
            '$cantidad_total'
          ]
        }
      ]
    },
    porcentaje_monto: {
      $multiply: [
        100,
        {
          $divide: [
            '$procurementMethods.monto',
            '$monto_total'
          ]
        }
      ]
    }
  }
}
}]
```

7. **No se puede calcular** ya que no hay información disponible sobre la cantidad de ofertas, además de la ganadora.
8. **No se puede calcular** ya que no hay información disponible sobre la cantidad de ofertas, además de la ganadora.
9. Se presenta el pipeline para el cálculo incluyendo además el monto total adjudicado en el mercado de ese artículo y la cantidad de adjudicaciones. Estos campos ayudan a analizar la situación de los mercados con alta concentración. A modo de ejemplo se agrega al final una etapa de `$match` adicional que realiza un corte en particular sobre los datos. Además, en esta consulta, se intenta inferir el valor de cada artículo para los casos en el que el único valor presente es el total de la adjudicación.

```
[{
  $unwind: '$awards'
}, {
  $match: {
    'awards.status': 'active'
  }
}, {
  $unwind: {
    path: '$awards.items',
    preserveNullAndEmptyArrays: true
  }
}, {
  $unwind: {
    path: '$awards.suppliers',
    preserveNullAndEmptyArrays: true
  }
}, {
  $addFields: {
    'awards.items.classification.id': {
      $ifNull: [
        '$awards.items.classification.id',
        'Not informed'
      ]
    }
  },
  'awards.items.quantity': {
    $cond: {
      'if': {
        $and: [
```

```
{
  $eq: [
    '$awards.items.unit.value.amount',
    null
  ]
},
{
  $ne: [
    '$awards.value',
    null
  ]
}
],
then: 1,
'else': {
  $ifNull: [
    '$awards.items.quantity',
    1
  ]
}
},
'awards.items.unit.value.amount': {
  $cond: {
    'if': {
      $eq: [
        '$awards.items.quantity',
        null
      ]
    },
    then: {
      $ifNull: [
        '$awards.value',
        '$awards.items.unit.value.amount',
        0
      ]
    },
    'else': {
      $ifNull: [
        '$awards.items.unit.value.amount',
        '$awards.value.amount',
        0
      ]
    }
  }
}
```

```
    ]
  }
}
},
'awards.suppliers.id': {
  $ifNull: [
    '$awards.suppliers.id',
    'Not informed'
  ]
}
}
}, {
$group: {
  _id: {
    item: '$awards.items.classification.id',
    proveedor: '$awards.suppliers.id'
  },
  item_desc: {
    $first: '$awards.items.classification.description'
  },
  proveedor_nombre: {
    $first: '$awards.suppliers.name'
  },
  monto_item_proveedor: {
    $sum: {
      $multiply: [
        '$awards.items.quantity',
        '$awards.items.unit.value.amount'
      ]
    }
  },
  cantidad_items_proveedor: {
    $sum: '$awards.items.quantity'
  },
  cantidad_adjudicaciones: {
    $sum: 1
  }
}
}, {
$group: {
  _id: '$_id.item',
  desc: {
    $first: '$item_desc'
  }
}
```



```
    },
    monto_item_por_proveedor: {
      $push: '$monto_item_proveedor'
    },
    monto_total_item: {
      $sum: '$monto_item_proveedor'
    },
    cantidad_items_adjudicados: {
      $sum: '$cantidad_items_proveedor'
    },
    cantidad_adjudicaciones: {
      $sum: '$cantidad_adjudicaciones'
    }
  }
}, {
  $project: {
    desc: 1,
    monto_total_item: 1,
    cantidad_items_adjudicados: 1,
    cantidad_adjudicaciones: 1,
    hhi: {
      $reduce: {
        input: '$monto_item_por_proveedor',
        initialValue: 0,
        'in': {
          $add: [
            '$$value',
            {
              $pow: [
                {
                  $multiply: [
                    {
                      $divide: [
                        '$$this',
                        '$monto_total_item'
                      ]
                    },
                    100
                  ]
                },
                2
              ]
            }
          ]
        }
      }
    }
  }
}
```

```
    ]
  }
}
}
}, {
  $match: {
    hhi: {
      $gt: 4000
    },
    monto_total_item: {
      $gt: 1000000
    },
    cantidad_adjudicaciones: {
      $gt: 30
    }
  }
}]
```

```
10. [{
  $unwind: '$awards'
}, {
  $unwind: '$awards.items'
}, {
  $unwind: '$awards.suppliers'
}, {
  $group: {
    _id: '$awards.items.classification.id',
    descripcion: {
      $first: '$awards.items.classification.description'
    },
    proveedores: {
      $addToSet: '$awards.suppliers.id'
    },
    adjudicaciones: {
      $sum: 1
    }
  }
}, {
  $addFields: {
    ratio: {
      $divide: [
        '$adjudicaciones',
```

```
    {
      $size: '$proveedores'
    }
  ]
}
}
}, {
  $sort: {
    ratio: -1
  }
}]
```

```
11.  [{
      $unwind: '$awards'
    }, {
      $match: {
        'awards.status': 'active'
      }
    }, {
      $addFields: {
        buyer: {
          $ifNull: [
            '$buyer',
            {
              id: '-1',
              name: 'Not informed'
            }
          ]
        }
      }
    }, {
      $group: {
        _id: '$buyer.id',
        nombre: {
          $first: '$buyer.name'
        },
        cantidad_adjudicaciones: {
          $sum: 1
        },
        monto_adjudicado: {
          $sum: '$awards.value.amount'
        },
        cant_proveedores: {
```

```
    $addToSet: '$awards.suppliers'
  }
}
}, {
  $addFields: {
    cant_proveedores: {
      $size: {
        $reduce: {
          input: '$cant_proveedores',
          initialValue: [],
          'in': {
            $setUnion: [
              '$$value',
              '$$this.id'
            ]
          }
        }
      }
    }
  }
}, {
  $sort: {
    monto_adjudicado: -1
  }
}, {
  $limit: 10
}]
```

```
12. [{
  $unwind: '$awards'
}, {
  $match: {
    'awards.status': 'active'
  }
}, {
  $addFields: {
    'awards.suppliers': {
      $ifNull: [
        '$awards.suppliers',
        [
          {
            id: '-1',
            name: 'Not informed'
          }
        ]
      ]
    }
  }
}
```

```
    }
  ]
]
}
}
}, {
  $unwind: '$awards.suppliers'
}, {
  $group: {
    _id: '$awards.suppliers.id',
    nombre: {
      $first: '$awards.suppliers.name'
    },
    cantidad_adjudicaciones: {
      $sum: 1
    },
    monto_adjudicado: {
      $sum: '$awards.value.amount'
    }
  }
}, {
  $sort: {
    monto_adjudicado: -1
  }
}, {
  $limit: 10
}]
```

B. Anexo: Resultados de consultas en base documental

Aquí se presenta una porción de los resultados obtenidos de cada una de las consultas señaladas en esta la sección 4., para la totalidad de los datos disponibles, de acuerdo al proceso descrito en las secciones 2. y 3.. Para algunas consultas se filtraron los datos resultantes para la elaboración del documento

Un análisis preliminar de los datos obtenidos puede encontrarse en la sección 4.1

1. Cantidad e importes totales de llamados y adjudicaciones por año.
Adjudicaciones.

```
[{
  "_id": "2017",
  "cantidad": 80286,
```

```
"monto": 117863818532.80
},{
  "_id": "2004",
  "cantidad": 38305,
  "monto": 13869110286414.65
},{
  "_id": "2007",
  "cantidad": 60935,
  "monto": 111067182978.97
},{
  "_id": "2009",
  "cantidad": 39205,
  "monto": 15758577095.3005
},{
  "_id": "2014",
  "cantidad": 54496,
  "monto": 55604875237.4288
},{
  "_id": "2011",
  "cantidad": 50117,
  "monto": 212869014416.4111
},{
  "_id": "2015",
  "cantidad": 62819,
  "monto": 71085304643.70805
},{
  "_id": "2021",
  "cantidad": 96825,
  "monto": 410031778743.2655
},{
  "_id": "2008",
  "cantidad": 73191,
  "monto": 39536928133.944954
},{
  "_id": "2005",
  "cantidad": 39769,
  "monto": 4332388570380.2217
},{
  "_id": "2018",
  "cantidad": 92098,
  "monto": 245431647903.70322
},{
  "_id": "2020",
```

```
    "cantidad": 86861,
    "monto": 347663093809.8209
  },{
    "_id": "2019",
    "cantidad": 100234,
    "monto": 267645487494.48904
  },{
    "_id": "2016",
    "cantidad": 72718,
    "monto": 83858956672.57964
  },{
    "_id": "2012",
    "cantidad": 51076,
    "monto": 42245852444.65257
  },{
    "_id": "2002",
    "cantidad": 15836,
    "monto": 39588507435.58215
  },{
    "_id": "2010",
    "cantidad": 44427,
    "monto": 25550079340.069824
  },{
    "_id": "2013",
    "cantidad": 50230,
    "monto": 43543393086.053604
  },{
    "_id": "2003",
    "cantidad": 33626,
    "monto": 827881431836.3223
  },{
    "_id": "2006",
    "cantidad": 40012,
    "monto": 73054227792.44075
  }
]
```

Llamados

```
[{
  "_id": "2010",
  "cantidad": 23893
},{
  "_id": "2004",
```

```
"cantidad": 14906
},{
  "_id": "1970",
  "cantidad": 5
},{
  "_id": "2002",
  "cantidad": 876
},{
  "_id": "2006",
  "cantidad": 15057
},{
  "_id": "2013",
  "cantidad": 25079
},{
  "_id": "2003",
  "cantidad": 8315
},{
  "_id": "2018",
  "cantidad": 33518
},{
  "_id": "2017",
  "cantidad": 28425
},{
  "_id": "2008",
  "cantidad": 22744
},{
  "_id": "2005",
  "cantidad": 13412
},{
  "_id": "2012",
  "cantidad": 25486
},{
  "_id": "2020",
  "cantidad": 30754
},{
  "_id": "2019",
  "cantidad": 34924
},{
  "_id": "2016",
  "cantidad": 25025
},{
  "_id": "2009",
  "cantidad": 21521
```



```
},{
  "_id": "2014",
  "cantidad": 26297
},{
  "_id": "2011",
  "cantidad": 27519
},{
  "_id": "2007",
  "cantidad": 19638
},{
  "_id": "2015",
  "cantidad": 25585
},{
  "_id": "2021",
  "cantidad": 42497
}]
```

2. Cantidad de anulaciones / cancelaciones.

```
[{
  "_id": null,
  "cantidad": 27227,
  "monto": 732803748.803975
}]
```

3. Cantidad de entidades compradoras.

```
[{
  "_id": null,
  "cant_compradores": 395
}]
```

4. Cantidad e importes por tipo de compra.

```
[{
  "_id": "open",
  "cantidad": 453372,
  "monto": 19646476518576.65
},{
  "_id": "Not informed",
  "cantidad": 715623,
```

```
"monto": 887830909913.7101
},{
  "_id": "limited",
  "cantidad": 1764,
  "monto": 355798822174.459
},{
  "_id": "direct",
  "cantidad": 11960,
  "monto": 336929230567.1251
},{
  "_id": "selective",
  "cantidad": 347,
  "monto": 4743533160.4674
}]
```

5. Cantidad de proveedores.

```
{
  "cantidad_proveedores": 29384
}
```

6. Proporción de llamados abiertos (competitivos) respecto al total (en cantidad y en importes).

Considerando los no informados

```
[{
  "procurementMethod": "Not informed",
  "porcentaje_procesos": 66.72940523834053,
  "porcentaje_monto": 4.180118143667077
},{
  "procurementMethod": "open",
  "porcentaje_procesos": 32.3532908031858,
  "porcentaje_monto": 92.48007659418842
},{
  "procurementMethod": "selective",
  "porcentaje_procesos": 0.024504539074833732,
  "porcentaje_monto": 0.022328774000922667
},{
  "procurementMethod": "direct",
  "porcentaje_procesos": 0.8774710481477697,
  "porcentaje_monto": 1.586033940907821
},{
```

```
"procurementMethod": "limited",
"porcentaje_procesos": 0.015328371251066208,
"porcentaje_monto": 1.7314425472357446
}]
```

Sin considerar los no informados

```
[{
  "procurementMethod": "direct",
  "porcentaje_procesos": 2.637377102614811,
  "porcentaje_monto": 1.6552242709773255
},{
  "procurementMethod": "limited",
  "porcentaje_procesos": 0.04607182817401988,
  "porcentaje_monto": 1.806976291028802
},{
  "procurementMethod": "selective",
  "porcentaje_procesos": 0.07365224231901137,
  "porcentaje_monto": 0.02330286112688096
},{
  "procurementMethod": "open",
  "porcentaje_procesos": 97.24289882689216,
  "porcentaje_monto": 96.51449657686699
}]
```

7. Proporción de llamados con un único oferente en los llamados abiertos respecto al total (en cantidad y en importes). - **No se puede realizar**
8. Promedio y mediana de cantidad de oferentes en llamados competitivos. - **No se puede realizar**
9. Índice de Herfindahl-Hirschman.

```
[{
  "_id": "26518",
  "desc": "MANTENIMIENTO PREVENTIVO PARA EQUIPOS DE TRANSPORTE (SERVICE)",
  "monto_total_item": 4203719005.6766996,
  "cantidad_items_adjudicados": 5034166.35,
  "cantidad_adjudicaciones": 7951,
  "hhi": 9261.95404724973
},{
  "_id": "1070",
```

```
"desc": "FORMULARIO",
"monto_total_item": 435298219781.9067,
"cantidad_items_adjudicados": 109576544.22,
"cantidad_adjudicaciones": 3970,
"hhi": 7313.631720242245
},{
  "_id": "6679",
  "desc": "CARPETA CON TAPA TRANSPARENTE",
  "monto_total_item": 71907333.2989,
  "cantidad_items_adjudicados": 1373145.49,
  "cantidad_adjudicaciones": 2698,
  "hhi": 7996.152813184404
},{
  "_id": "747",
  "desc": "ENTREGA DE ENCOMIENDAS Y PAQUETES DENTRO DEL PAIS",
  "monto_total_item": 4572688314.4377,
  "cantidad_items_adjudicados": 31353043,
  "cantidad_adjudicaciones": 4206,
  "hhi": 6616.643619773819
},{
  "_id": "75247",
  "desc": "PEMBROLIZUMAB ( INYECTABLE)",
  "monto_total_item": 526399514.125,
  "cantidad_items_adjudicados": 3709,
  "cantidad_adjudicaciones": 1789,
  "hhi": 9420.075595302202
},{
  "_id": "28031",
  "desc": "CONTRATACION DE MANO DE OBRA",
  "monto_total_item": 2571059742.039825,
  "cantidad_items_adjudicados": 2732893.87,
  "cantidad_adjudicaciones": 2662,
  "hhi": 5420.554965671543
},{
  "_id": "3714",
  "desc": "ETIQUETA AUTOADHESIVA",
  "monto_total_item": 103621385.70765,
  "cantidad_items_adjudicados": 9318904,
  "cantidad_adjudicaciones": 1893,
  "hhi": 7639.974431373058
},{
  "_id": "Not informed",
  "desc": null,
```

```
"monto_total_item": 270524841530.515,
"cantidad_items_adjudicados": 43851,
"cantidad_adjudicaciones": 43851,
"hhi": 10000
},{
  "_id": "8194",
  "desc": "TEXTOS Y PUBLICACIONES OFICIALES",
  "monto_total_item": 310342879.932,
  "cantidad_items_adjudicados": 161694.32,
  "cantidad_adjudicaciones": 6699,
  "hhi": 4512.300570661517
},{
  "_id": "284",
  "desc": "TUBO FLUORESCENTE",
  "monto_total_item": 1097378040.498,
  "cantidad_items_adjudicados": 202662,
  "cantidad_adjudicaciones": 1791,
  "hhi": 6634.0569753995405
},{
  "_id": "",
  "desc": "Fibrobroncoscopia y Fibrobronco aspiracion en ni?os",
  "monto_total_item": 138367436123.2514,
  "cantidad_items_adjudicados": 14167743.77,
  "cantidad_adjudicaciones": 6257,
  "hhi": 5666.226955966111
},{
  "_id": "4418",
  "desc": "DIARIO",
  "monto_total_item": 292883932.9665,
  "cantidad_items_adjudicados": 234370,
  "cantidad_adjudicaciones": 4072,
  "hhi": 6270.735194319976
},{
  "_id": "15595",
  "desc": "SUPERGAS",
  "monto_total_item": 117016265.938915,
  "cantidad_items_adjudicados": 797064.42,
  "cantidad_adjudicaciones": 2596,
  "hhi": 6998.267502489828
}]
```

10. Ratio número de adjudicaciones/número de proveedores para un ítem.

```
[{
  "descripcion": "RIBOCICLIB ( COMPRIMIDOS)",
  "adjudicaciones": 759,
  "ratio": 759
},{
  "descripcion": "NIVOLUMAB ( INYECTABLE)",
  "adjudicaciones": 1244,
  "ratio": 622
},{
  "descripcion": "CETUXIMAB (INYECTABLE)",
  "adjudicaciones": 1153,
  "ratio": 576.5
},{
  "descripcion": "PEMBROLIZUMAB ( INYECTABLE)",
  "adjudicaciones": 1789,
  "ratio": 447.25
},{
  "descripcion": "OMALIZUMAB ( INYECTABLE)",
  "adjudicaciones": 749,
  "ratio": 374.5
},{
  "descripcion": "RUXOLITINIB (COMPRIMIDOS)",
  "adjudicaciones": 668,
  "ratio": 334
},{
  "descripcion": "BEVACIZUMAB (INYECTABLE)",
  "adjudicaciones": 626,
  "ratio": 313
},{
  "descripcion": "TREMETINIB ( COMPRIMIDOS)",
  "adjudicaciones": 300,
  "ratio": 300
},{
  "descripcion": "IBRUTINIB ( COMPRIMIDOS)",
  "adjudicaciones": 893,
  "ratio": 297.6666666666667
},{
  "descripcion": "DABRAFENIB (COMPRIMIDOS)",
  "adjudicaciones": 271,
  "ratio": 271
}...]
```

11. Información por entidad compradora: Cantidad de proveedores, Cantidad de

adjudicaciones, Importe adjudicado, Top 10 por importe adjudicado ["id" : "3 – 4", "nombre" : "ComandoGeneraldelEjército", "cantidad_adjudicaciones" : 9741, "monto_adjudicado" : 16006623100421,55, "cant_proveedores" : 1436, "id" : "67 – 1", "nombre" : "Administración Nacional de Aduanas"]

12. [{
 "_id": "R/210516290010",
 "nombre": "TURISMAR S.R.L.",
 "cantidad_adjudicaciones": 1132,
 "monto_adjudicado": 1309403895551.3987
}, {
 "_id": "R/210237200015",
 "nombre": "AGENCIA CENTRAL S.A.",
 "cantidad_adjudicaciones": 2471,
 "monto_adjudicado": 1309371589399.8118
}, {
 "_id": "R/140026100016",
 "nombre": "TURIL S.A.",
 "cantidad_adjudicaciones": 2346,
 "monto_adjudicado": 1309342364329.893
}, {
 "_id": "R/210193770016",
 "nombre": "COMPAÑIA ORIENTAL DE TRANSPORTE S.A.",
 "cantidad_adjudicaciones": 1681,
 "monto_adjudicado": 1309289408904.2065
}, {
 "_id": "R/030083940016",
 "nombre": "DIOMAR G. NUÑEZ PEREIRA SOCIEDAD ANONIMA",
 "cantidad_adjudicaciones": 2084,
 "monto_adjudicado": 1309286968402.6404
}, {
 "_id": "R/160000820016",
 "nombre": "CHADRE S.A.",
 "cantidad_adjudicaciones": 255,
 "monto_adjudicado": 1309206016732.0425
}, {
 "_id": "R/210148540015",
 "nombre": "VIAJES CYNOSA S.A.",
 "cantidad_adjudicaciones": 905,
 "monto_adjudicado": 1309178610732.6038
}, {
 "_id": "R/210354700014",
 "nombre": "COMPAÑIA URUGUAYA TRANSPORTES LTDA. C.U.T.",
 "cantidad_adjudicaciones": 1248,
 "monto_adjudicado": 1309170863443.177
}

```
},{
  "_id": "R/210233220011",
  "nombre": "COOP. DE OMNIBUS RUTAS DEL PLATA (CORPLA)",
  "cantidad_adjudicaciones": 701,
  "monto_adjudicado": 1309163778579.8684
},{
  "_id": "R/090008860012",
  "nombre": "CORPORACION DE OMNIBUS S.A.",
  "cantidad_adjudicaciones": 865,
  "monto_adjudicado": 1309162041531.2952
}]
```