

Trabajo Nociones de Arquitectura de la Información

Versión: 2021.05.29 12:00

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. Juan David Ordoñez Noriega - 1017244729
2. Edwin Álvarez Pineda - 1040047474
3. Jenny Marcela Zapata Pulgarín - 1035832927
REPO EN GITHUB: https://github.com/Jezp27/Trabajo-Nociones-Arquitectura-Informacion

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.

Los datos se generan a partir de una recopilación de las posiciones laborales abiertas en Estados Unidos y específicamente asociadas con el desarrollo de software. La información es publicada en varios sitios web que permiten a los empleadores publicar ofertas laborales de acuerdo con sus necesidades, mientras los profesionales pueden explorar a cuáles de ellas aplicar e identificar las tendencias en el campo laboral. Dichas ofertas son recopiladas en un solo conjunto de datos donde se especifican las características de cada vacante (como el cargo al que se aplicaría y una descripción del mismo, el nombre de la compañía que oferta la vacante, su ubicación, la fecha de la publicación, entre otros) así como la página web en la cual fue publicada.

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.

Dada la gran cantidad de posiciones laborales que son puestas en línea diariamente, es difícil para un profesional analizar cuáles de ellas son las más convenientes de acuerdo con su ubicación, tipo de trabajo y características. De igual forma un empleador debe conocer cuáles son las empresas que en su misma zona están buscando personal en posiciones similares, pues estos se convierten en su competencia. Se pretende caracterizar las convocatorias relacionadas con el desarrollo de software que son publicadas en Estados Unidos para determinar estrategias de distribución de convocatorias e identificar las características que las hacen más atractivas, de esta forma se puede conectar los profesionales adecuados con las mejores ofertas del mercado. Los resultados obtenidos se pueden ver limitados si no se cuenta con una cantidad de ofertas lo suficientemente amplia que permita obtener resultados más cercanos a la realidad.

1.3 DETERMINACIÓN DE OBJETIVOS:

Describa 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).

- Identificar las categorías o roles con más demanda en el mercado laboral de software estadounidense.
- Presentar la información de manera legible y resumida de las convocatorias para que se puedan identificar las necesidades actuales del mercado laboral de software en Estados Unidos.
- Encontrar tendencias en el mercado laboral por zona de influencia y por tipo de empleado para entender cuales son los perfiles más solicitados por empresa.

1.4 EVALUACIÓN DE LA SITUACIÓN ACTUAL:

Describa 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.

En la actualidad cada una de las convocatorias por empresa es publicada de manera independiente y orientada a un público específico, lo cual puede facilitar un proceso de filtrado inicial pero que constituye un análisis superficial y puede estar limitando el alcance de las convocatorias al reducir el público objetivo. Adicionalmente, por lo general no se analizan las convocatorias respecto a otras similares publicadas, causando que se pierdan postulantes por errores en el enfoque de la convocatoria o por otros factores asociados al salario y la ubicación de la zona laboral; no siempre estos últimos se pueden cambiar fácilmente, pero se podría al menos mejorar las ofertas para hacerlas más atractivas y competitivas agregando valor con otros aspectos importantes que tal vez no fuesen incluidos sin un análisis riguroso del mercado.

2 COMPRENSIÓN DE LOS DATOS

2.1 RECOLECCIÓN DE DATOS

Describa 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

Los datos a utilizar han sido recopilados de algunas páginas web donde se han publicado vacantes para desarrolladores de software en Estados Unidos, dichas páginas son:

- <https://www.careerbuilder.com/job>
- <https://www.dice.com/jobs/>
- <https://www.indeed.com>
- <https://job-openings.monster.com/>

Desde cada una de las páginas mencionadas, las vacantes fueron inicialmente filtradas para obtener solo las correspondientes a Estados Unidos (Country = US/Usa) y el desarrollo de software; luego fueron agregadas dentro del conjunto de datos con los campos básicos de cada una de ellas. El periodo de dicha recolección está comprendido entre febrero y agosto de 2019. Entre los problemas encontrados se encuentra la falta de uniformidad entre las convocatorias publicadas, lo que causa que para ciertas columnas no hubiese valores disponibles o que no se usará una misma convención para los mismos, lo cual se debe resolver con un proceso de limpieza de la información antes de utilizarla para realizar el análisis correspondiente.

Fuente de los datos: <https://www.kaggle.com/jobspikr/software-developer-job-listings-usa>

2.2 DESCRIPCIÓN DE DATOS (DICCIONARIO):

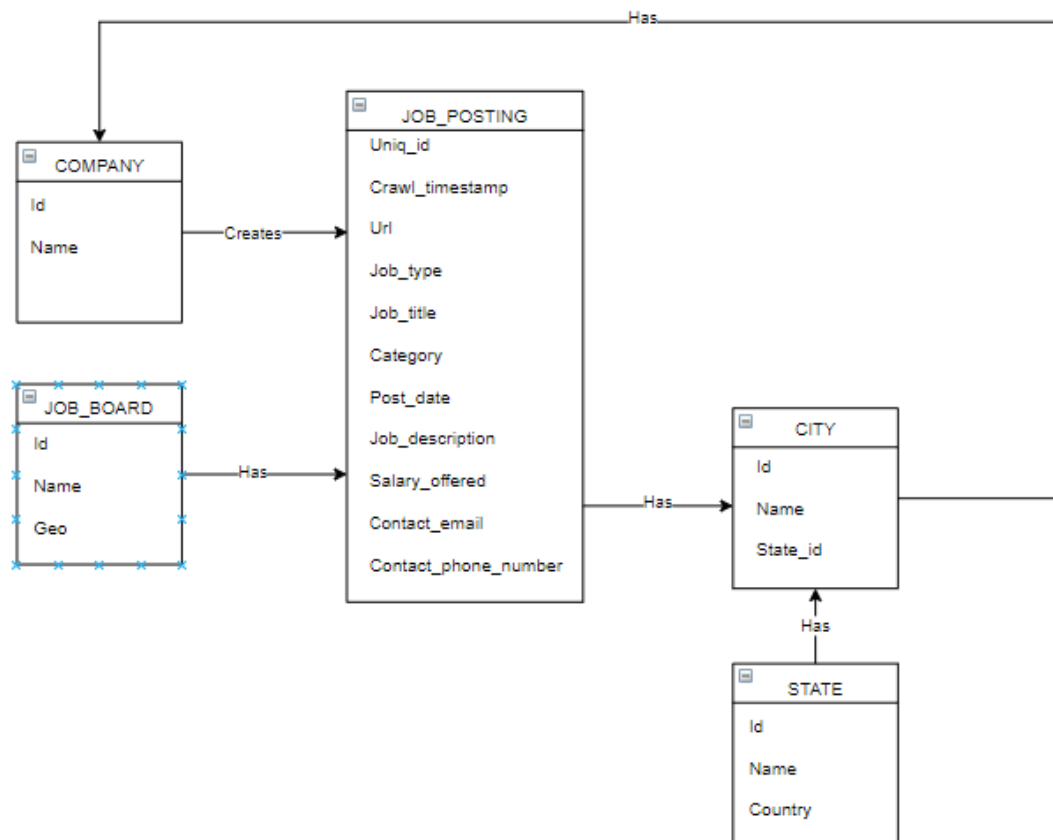
Diligencia la siguiente tabla, puede agregar otra columna si lo considera necesario.

Nombre del atributo / variable	Formato o Tipo de Dato	Descripción
crawl_timestamp	STRING	Fecha en la que la vacante fue agregada al conjunto de datos presentado.
url	STRING	Enlace específico de la convocatoria en cuestión.
job_title	STRING	Nombre del cargo al que se estaría aplicando.
category	STRING	Específica a qué área de negocio se relaciona la posición ofertada.
company_name	STRING	Nombre de la compañía a la que pertenece la vacante.
city	STRING	Ciudad donde se ubica la compañía oferente.
state	STRING	Siglas del estado al cual pertenece la compañía oferente.
country	STRING	País al cual pertenece la compañía oferente (US/Usa).
inferred_city	STRING	Ciudad en la cual debe desempeñar el cargo.
inferred_state	STRING	Estado al cual debe desempeñar el cargo.

inferred_country	STRING	País al cual debe desempeñar el cargo.
post_date	STRING	Fecha de publicación de la vacante.
job_description	STRING	Descripción de las funciones a desempeñar en dicho cargo.
job_type	STRING	Tipo de contrato al cual se aplicaría.
salary_offered	INT	Salario ofrecido por vacante (Null para todos los registros).
job_board	STRING	Plataforma/página web en la cual fue publicada la convocatoria.
geo	STRING	País en el cual es válida la convocatoria.
contact_email	STRING	Correo electrónico de contacto (Null para todos los registros)
contact_phone_number	STRING	Número de teléfono de contacto.
uniq_id	STRING	Número de identificación único por convocatoria (Algunos se repiten en el listado pero por variaciones en campos específicos de la convocatoria, no porque fuesen diferentes).

2.3 MODELO DEL DOMINIO

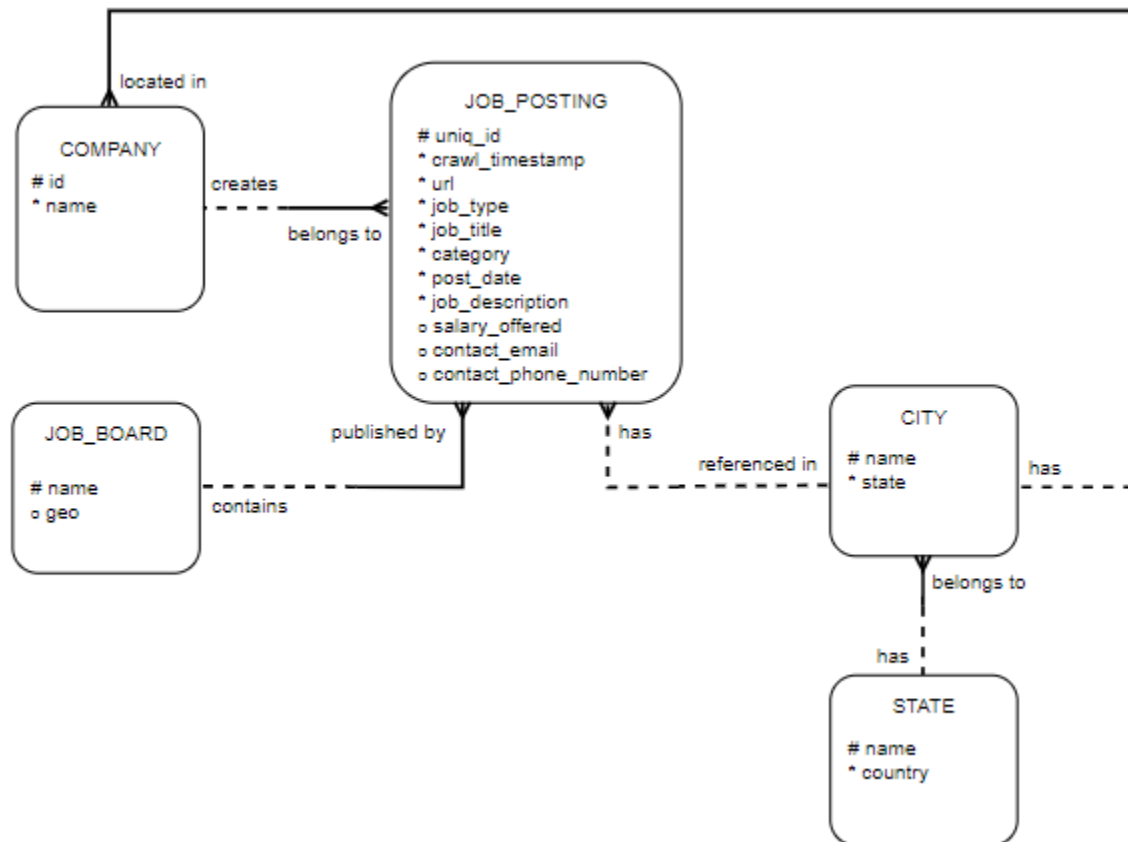
Observación: Incluya el gráfico del modelo del dominio que representa la estructura de datos de su problema.



3 MODELO ENTIDAD-RELACIÓN

3.1 TOMA DE PANTALLA DEL MODELO E-R

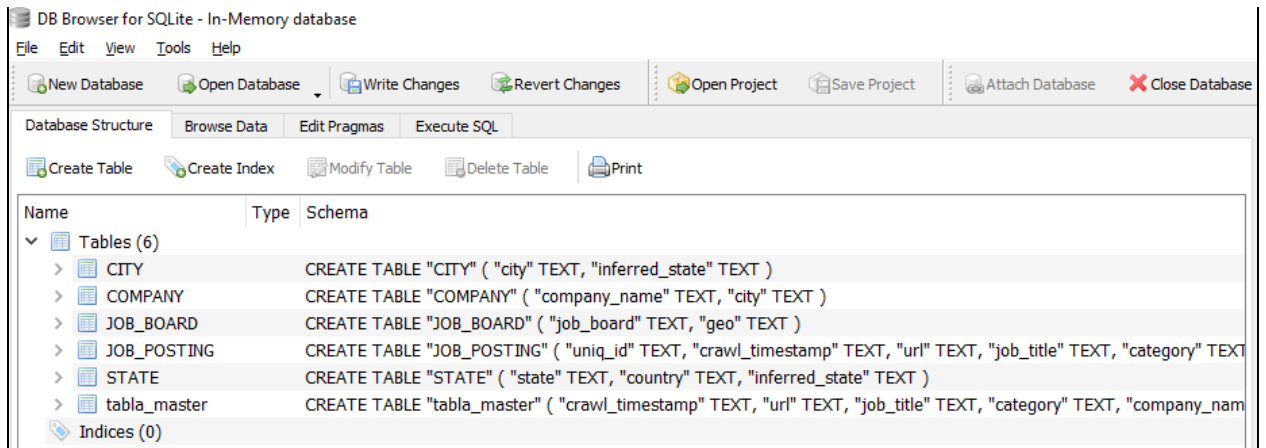
Observación: lo que se pide, puede usar <https://draw.io> o Microsoft Visio® y modele usando la notación de Barker.



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

Se recomienda repasar SQL en <https://www.w3schools.com/sql/default.asp>



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**

		url	job_title	
		Filter	Filter	Filter
1	2019-02-06 05:31:44 +0000	https://www.careerbuilder.com/job/...	Sr. Software Developer	architecture
2	2019-02-06 05:31:44 +0000	https://www.dice.com/jobs/detail/C%2523-Lead...	C# Lead Software Developer	NULL
3	2019-02-06 05:31:44 +0000	https://www.dice.com/jobs/detail/Senior-...	Senior Software Developer	NULL
4	2019-02-06 05:27:30 +0000	https://www.dice.com/jobs/detail/Senior-...	Senior Software Developer	NULL
5	2019-02-06 05:37:30 +0000	https://www.dice.com/jobs/detail/C%2523-Lead...	C# Lead Software Developer	NULL
6	2019-02-06 05:47:13 +0000	https://www.dice.com/jobs/detail/Software-...	Software Developer 3	NULL
7	2019-02-06 05:53:38 +0000	https://www.dice.com/jobs/detail/C%2528%252...	C++ Software Developer	NULL
8	2019-02-06 05:55:55 +0000	https://www.dice.com/jobs/detail/C%2528%252...	C++ Software Developer - Local W2 only	NULL
9	2019-02-06 05:58:22 +0000	https://www.careerbuilder.com/job/...	Software Developer - .Net	architecture
10	2019-02-06 05:45:09 +0000	https://www.careerbuilder.com/job/...	.Net Software Developer	architecture
11	2019-02-06 06:00:32 +0000	https://www.careerbuilder.com/job/...	Senior Java Software Developer	architecture
12	2019-02-06 05:55:32 +0000	https://www.careerbuilder.com/job/...	Software Developer/Engineer – Self-Checkout	architecture
13	2019-02-06 06:02:30 +0000	https://www.dice.com/jobs/detail/Software-...	Software Developer	NULL
14	2019-02-06 06:02:22 +0000	https://www.dice.com/jobs/detail/C%2528%252...	C++ Software Developer	NULL
15	2019-02-06 06:03:00 +0000	https://www.careerbuilder.com/job/...	Software Developer	architecture
16	2019-02-06 06:13:57 +0000	https://www.dice.com/jobs/detail/Senior-...	Senior Software Developer	NULL
17	2019-02-06 06:19:13 +0000	https://www.indeed.com/viewjob?...	IT Software Developer	Computer/In

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*

1)

SQL 1

```
1 SELECT category, COUNT(uniq_id) AS AMOUNT
2 FROM JOB_POSTING
3 GROUP BY (category)
4 ORDER BY AMOUNT DESC;
```

	category	AMOUNT
1	Computer-or-internet	3115
2	Computer/Internet	1798
3	NULL	839
4	arts design and media	623
5	technology	622
6	computer jobs	608
7	architecture and engineering	526
8	Engineering-or-architecture	225
9	Construction-or-facilities	139
10	life physical and social science	133

Execution finished without errors.
Result: 111 rows returned in 141ms
At line 1:
SELECT category, COUNT(uniq_id) AS AMOUNT
FROM JOB_POSTING
GROUP BY (category)
ORDER BY AMOUNT DESC;

2)

SQL 1

```
1 SELECT job_type, COUNT(uniq_id) AS AMOUNT
2 FROM JOB_POSTING
3 GROUP BY (job_type)
4 ORDER BY AMOUNT DESC;
```

	job_type	AMOUNT
1	Undefined	6678
2	Full Time	2685
3	Contract	628
4	Part Time	6
5	Internship	3

Execution finished without errors.
Result: 5 rows returned in 123ms
At line 1:
SELECT job_type, COUNT(uniq_id) AS AMOUNT
FROM JOB_POSTING
GROUP BY (job_type)
ORDER BY AMOUNT DESC;

3)

SQL 1

```
1 SELECT JB.job_board,JB.geo,COUNT(uniq_id) AS AMOUNT
2 FROM JOB_POSTING JP
3 LEFT JOIN JOB_BOARD JB ON JP.job_board=JB.job_board
4 GROUP BY (JB.job_board)
5 ORDER BY AMOUNT DESC;
```

	job_board	geo	AMOUNT
1	indeed	usa	5926
2	careerbuilder	usa	2325
3	monster	usa	910
4	dice	usa	839

Execution finished without errors.

Result: 4 rows returned in 197ms

At line 1:

```
SELECT JB.job_board,JB.geo,COUNT(uniq_id) AS AMOUNT
FROM JOB_POSTING JP
LEFT JOIN JOB_BOARD JB ON JP.job_board=JB.job_board
GROUP BY (JB.job_board)
ORDER BY AMOUNT DESC;
```

4)

SQL 1

```
1 SELECT job_title,COUNT(uniq_id) AS AMOUNT
2 FROM JOB_POSTING
3 GROUP BY (job_title)
4 ORDER BY (AMOUNT) DESC
5 LIMIT 10;
```

	job_title	AMOUNT
1	Software Developer	2217
2	Senior Software Developer	714
3	Junior Software Developer	260
4	Software Developer (JAVA SWING)	182
5	Software Developer II	176
6	Associate Software Developer	149
7	Sr. Software Developer	147
8	Entry Level / Junior Software Developer	117
9	Software Developer- App&Emerging Tech- ...	95
10	Software Developer (Java)	80

Execution finished without errors.
Result: 10 rows returned in 129ms
At line 1:
SELECT job_title,COUNT(uniq_id) AS AMOUNT
FROM JOB_POSTING
GROUP BY (job_title)
ORDER BY (AMOUNT) DESC
LIMIT 10;

5)

SQL 1

```

1  SELECT CP.company_name,ifnull(CT.city,'Remote') AS COMPANY_CITY,
2  COUNT(DISTINCT(uniq_id)) AS AMOUNT
3  FROM JOB_POSTING JP
4  LEFT JOIN COMPANY CP ON JP.company_name=CP.company_name
5  LEFT JOIN CITY CT ON CP.city=CT.city
6  GROUP BY (CP.company_name)
7  ORDER BY AMOUNT DESC;

```

	company_name	COMPANY_CITY	AMOUNT
1	Robert Half Technology	Remote	338
2	CyberCoders	Remote	263
3	AutomationDirect	Computer-or-internet	162
4	Kforce Technology	Ames	140
5	Lionbridge Technologies	Chicago	122
6	Leidos	Aberdeen	122
7	General Dynamics Information Technology	Alexandria	117
8	Smoothstack Inc.	Computer-or-internet	104
9	TEKsystems, Inc	Aberdeen	101
10	Kforce Technology Staffing	Arlington	101

Execution finished without errors.

Result: 3044 rows returned in 1958ms

At line 1:

```

SELECT CP.company_name,ifnull(CT.city,'Remote') AS COMPANY_CITY,COUNT(DISTINCT(uniq_id)) AS
AMOUNT
FROM JOB_POSTING JP
LEFT JOIN COMPANY CP ON JP.company_name=CP.company_name
LEFT JOIN CITY CT ON CP.city=CT.city
GROUP BY (CP.company_name)
ORDER BY AMOUNT DESC;

```

6)

```
SQL 1 X
7
8 SELECT MIN(strftime('%Y %m %d',post_date)) AS DATE_FIRST_OFFERT FROM JOB_POSTING;
9
<
DATE_FIRST_OFFERT
1 2018 09 03

Execution finished without errors.
Result: 1 rows returned in 4989ms
At line 8:
SELECT MIN(strftime('%Y %m %d',post_date)) AS DATE_FIRST_OFFERT FROM JOB_POSTING;
```

7)

```
SQL 1 X
7
8 SELECT MAX(strftime('%Y %m %d',post_date)) AS DATE_LAST_OFFERT FROM JOB_POSTING;
9
<
DATE_LAST_OFFERT
1 2019 08 02

Execution finished without errors.
Result: 1 rows returned in 242ms
At line 8:
SELECT MAX(strftime('%Y %m %d',post_date)) AS DATE_LAST_OFFERT FROM JOB_POSTING;
```

8)

SQL 1

```
1 SELECT strftime('%Y',post_date) AS YEAR,
2    strftime('%m',post_date) AS MONTH, COUNT(uniq_id) AS AMOUNT
3 FROM JOB_POSTING
4 GROUP BY YEAR,MONTH
5 ORDER BY AMOUNT DESC;
```

	YEAR	MONTH	AMOUNT
1	2019	06	2827
2	2019	07	2345
3	2019	02	1454
4	2019	05	1370
5	2019	04	999
6	2019	03	949
7	2019	08	38
8	2019	01	13
9	2018	09	3
10	2018	10	2

Execution finished without errors.

Result: 10 rows returned in 250ms

At line 1:

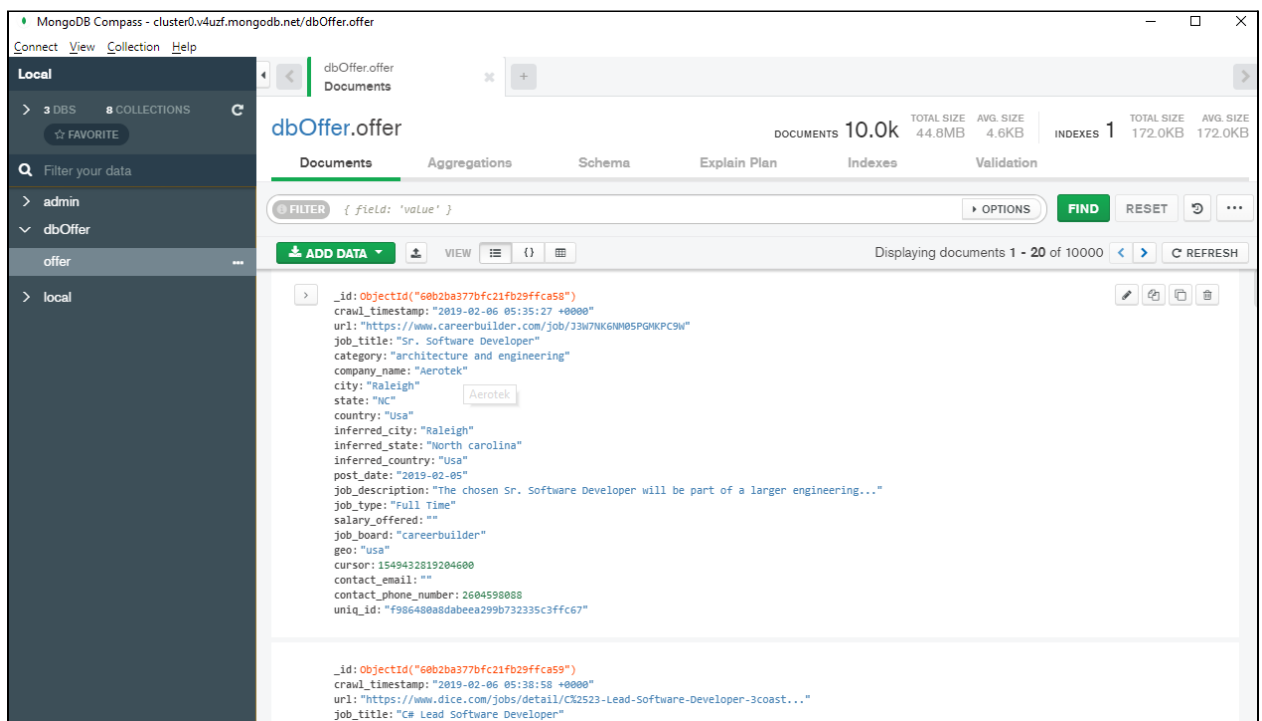
```
SELECT strftime('%Y',post_date) AS YEAR, strftime('%m',post_date) AS MONTH, COUNT(uniq_id) AS AMOUNT
FROM JOB_POSTING
GROUP BY YEAR,MONTH
ORDER BY AMOUNT DESC;
```

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

```
jdordonezn@DESKTOP-839T26K MINGW64 ~/Downloads/archive
$ mongoimport --uri "mongodb+srv://jdordonezn:cluster0.v4ufz.mongodb.net/" --type csv -d dboffer -c offer --headerline --drop --file software_develope
connected to: mongodb+srv://[REDACTED]@cluster0.v4ufz.mongodb.net/
dropping: dboffer.offer
[#####] .....] dboffer.offer 3.13MB/41.5MB (7.5%)
[#####] .....] dboffer.offer 6.48MB/41.5MB (15.6%)
[#####] .....] dboffer.offer 6.54MB/41.5MB (15.8%)
[#####] .....] dboffer.offer 10.6MB/41.5MB (25.5%)
[#####] .....] dboffer.offer 10.6MB/41.5MB (25.5%)
[#####] .....] dboffer.offer 10.6MB/41.5MB (25.5%)
[#####] .....] dboffer.offer 15.6MB/41.5MB (37.4%)
[#####] .....] dboffer.offer 15.6MB/41.5MB (37.4%)
[#####] .....] dboffer.offer 20.5MB/41.5MB (49.3%)
[#####] .....] dboffer.offer 20.5MB/41.5MB (49.3%)
[#####] .....] dboffer.offer 20.5MB/41.5MB (49.3%)
[#####] .....] dboffer.offer 26.0MB/41.5MB (62.7%)
[#####] .....] dboffer.offer 26.0MB/41.5MB (62.7%)
[#####] .....] dboffer.offer 26.0MB/41.5MB (62.7%)
[#####] .....] dboffer.offer 30.9MB/41.5MB (74.3%)
[#####] .....] dboffer.offer 30.9MB/41.5MB (74.3%)
[#####] .....] dboffer.offer 34.4MB/41.5MB (82.9%)
[#####] .....] dboffer.offer 34.4MB/41.5MB (82.9%)
[#####] .....] dboffer.offer 38.0MB/41.5MB (91.5%)
[#####] .....] dboffer.offer 38.0MB/41.5MB (91.5%)
[#####] .....] dboffer.offer 41.5MB/41.5MB (100.0%)
[#####] .....] dboffer.offer 41.5MB/41.5MB (100.0%)
[#####] .....] dboffer.offer 41.5MB/41.5MB (100.0%)
10000 document(s) imported successfully. 0 document(s) failed to import.
```



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**

1)

```
MongoDB Enterprise atlas-pql2so-shard-0:PRIMARY> db.offer.aggregate(
... [
...   {
...     '$group': {
...       '_id': '$category',
...       'amount': {
...         '$sum': 1
...       }
...     }, {
...       '$sort': {
...         'amount': -1
...       }
...     }, {
...       '$limit': 10
...     }
...   ]
... );
{ "_id" : "Computer-or-internet", "amount" : 3115 }
{ "_id" : "Computer/Internet", "amount" : 1798 }
{ "_id" : "", "amount" : 839 }
{ "_id" : "arts design and media", "amount" : 623 }
{ "_id" : "technology", "amount" : 622 }
{ "_id" : "computer jobs", "amount" : 608 }
{ "_id" : "architecture and engineering", "amount" : 526 }
{ "_id" : "Engineering-or-architecture", "amount" : 225 }
{ "_id" : "Construction-or-facilities", "amount" : 139 }
{ "_id" : "life physical and social science", "amount" : 133 }
MongoDB Enterprise atlas-pql2so-shard-0:PRIMARY> _
```

2)

```
MongoDB Enterprise atlas-pql2so-shard-0:PRIMARY> db.offer.aggregate(
... [
...   {
...     '$group': {
...       '_id': '$job_type',
...       'amount': {
...         '$sum': 1
...       }
...     }, {
...       '$sort': {
...         'amount': -1
...       }
...     }
...   ]
... );
{ "_id" : "Undefined", "amount" : 6678 }
{ "_id" : "Full Time", "amount" : 2685 }
{ "_id" : "Contract", "amount" : 628 }
{ "_id" : "Part Time", "amount" : 6 }
{ "_id" : "Internship", "amount" : 3 }
MongoDB Enterprise atlas-pql2so-shard-0:PRIMARY>
```

3)

```
MongoDB Enterprise atlas-pql2so-shard-0:PRIMARY> db.offer.aggregate(
... [{$group: {
...   _id: "$job_board",
...   amount: {
...     $sum: 1
...   },
...   geo: {$first:"$geo"}
... }}, {$sort: {
...   amount: -1
... }}]
... );
{ "_id" : "indeed", "amount" : 5926, "geo" : "usa" }
{ "_id" : "careerbuilder", "amount" : 2325, "geo" : "usa" }
{ "_id" : "monster", "amount" : 910, "geo" : "usa" }
{ "_id" : "dice", "amount" : 839, "geo" : "usa" }
MongoDB Enterprise atlas-pql2so-shard-0:PRIMARY>
```

4)

```
MongoDB Enterprise atlas-pql2so-shard-0:PRIMARY> db.offer.aggregate(
... [
...   {
...     '$group': {
...       '_id': '$job_title',
...       'amount': {
...         '$sum': 1
...       }
...     }, {
...       '$sort': {
...         'amount': -1
...       }
...     }, {
...       '$limit': 10
...     }
...   ]
... );
{ "_id" : "Software Developer", "amount" : 2218 }
{ "_id" : "Senior Software Developer", "amount" : 714 }
{ "_id" : "Junior Software Developer", "amount" : 260 }
{ "_id" : "Software Developer (JAVA SWING)", "amount" : 182 }
{ "_id" : "Software Developer II", "amount" : 176 }
{ "_id" : "Associate Software Developer", "amount" : 149 }
{ "_id" : "Sr. Software Developer", "amount" : 147 }
{ "_id" : "Entry Level / Junior Software Developer", "amount" : 117 }
{ "_id" : "Software Developer- App&Emerging Tech- Experienced Associate", "amount" : 95 }
{ "_id" : "Software Developer (Java)", "amount" : 80 }
MongoDB Enterprise atlas-pql2so-shard-0:PRIMARY>
```

5)

```
MongoDB Enterprise atlas-pql2so-shard-0:PRIMARY> db.offer.aggregate(
... [
...   {
...     '$group': {
...       '_id': '$company_name',
...       'company_city': {
...         '$first': '$city'
...       },
...       'amount': {
...         '$sum': 1
...       }
...     }, {
...       '$sort': {
...         'amount': -1
...       }
...     }, {
...       '$limit': 10
...     }
...   ]
... );
{ "_id" : "Robert Half Technology", "company_city" : "Madison", "amount" : 351 }
{ "_id" : "CyberCoders", "company_city" : "Chicago", "amount" : 263 }
{ "_id" : "AutomationDirect", "company_city" : "Computer-or-internet", "amount" : 163 }
{ "_id" : "Kforce Technology", "company_city" : "Windsor", "amount" : 140 }
{ "_id" : "Lionbridge Technologies", "company_city" : "Chicago", "amount" : 129 }
{ "_id" : "Leidos", "company_city" : "King Of Prussia", "amount" : 122 }
{ "_id" : "General Dynamics Information Technology", "company_city" : "Alexandria", "amount" : 118 }
{ "_id" : "Kforce Technology Staffing", "company_city" : "Mountain View", "amount" : 110 }
{ "_id" : "Apex Systems", "company_city" : "Columbia", "amount" : 104 }
{ "_id" : "Smoothstack Inc.", "company_city" : "Computer-or-internet", "amount" : 104 }
MongoDB Enterprise atlas-pql2so-shard-0:PRIMARY> _
```

6)

```
MongoDB Enterprise atlas-pql2so-shard-0:PRIMARY> db.offer.aggregate(
... [
...   {
...     '$project': {
...       'date_first_offer': {
...         '$dateFromString': {
...           'dateString': '$post_date'
...         }
...       }
...     }
...   }, {
...     '$sort': {
...       'date_first_offer': 1
...     }
...   }, {
...     '$limit': 1
...   }
... ]
... );
{ "_id" : ObjectId("60b2ba627bfc21fb29ffe1f1"), "date_first_offer" : ISODate("2018-09-03T00:00:00Z") }
MongoDB Enterprise atlas-pql2so-shard-0:PRIMARY> _
```

7)

```
MongoDB Enterprise atlas-pql2so-shard-0:PRIMARY> db.offer.aggregate(  
... [  
...   {  
...     '$project': {  
...       'date_last_offer': {  
...         '$dateFromString': {  
...           'dateString': '$post_date'  
...         }  
...       }  
...     }  
...   }, {  
...     '$sort': {  
...       'date_last_offer': -1  
...     }  
...   }, {  
...     '$limit': 1  
...   }  
... ]  
... );  
{ "_id" : ObjectId("60b2ba757bfc21fb29fff159"), "date_last_offer" : ISODate("2019-08-02T00:00:00Z") }  
MongoDB Enterprise atlas-pql2so-shard-0:PRIMARY> _
```


8)

```
MongoDB Enterprise atlas-pql2so-shard-0:PRIMARY> db.offer.aggregate(
... [
...   {
...     '$project': {
...       'date': {
...         '$dateFromString': {
...           'dateString': '$post_date'
...         }
...       }
...     }
...   }, {
...     '$project': {
...       'year': {
...         '$year': '$date'
...       },
...       'month': {
...         '$month': '$date'
...       }
...     }
...   }, {
...     '$group': {
...       '_id': {
...         'year': '$year',
...         'month': '$month'
...       },
...       'year': {
...         '$first': '$year'
...       },
...       'mont': {
...         '$first': '$month'
...       },
...       'amount': {
...         '$sum': 1
...       }
...     }
...   }, {
...     '$sort': {
...       'amount': -1
...     }
...   }
... ]
... );
{ "_id" : { "year" : 2019, "month" : 6 }, "year" : 2019, "mont" : 6, "amount" : 2827 }
{ "_id" : { "year" : 2019, "month" : 7 }, "year" : 2019, "mont" : 7, "amount" : 2345 }
{ "_id" : { "year" : 2019, "month" : 2 }, "year" : 2019, "mont" : 2, "amount" : 1454 }
{ "_id" : { "year" : 2019, "month" : 5 }, "year" : 2019, "mont" : 5, "amount" : 1370 }
{ "_id" : { "year" : 2019, "month" : 4 }, "year" : 2019, "mont" : 4, "amount" : 999 }
{ "_id" : { "year" : 2019, "month" : 3 }, "year" : 2019, "mont" : 3, "amount" : 949 }
{ "_id" : { "year" : 2019, "month" : 8 }, "year" : 2019, "mont" : 8, "amount" : 38 }
{ "_id" : { "year" : 2019, "month" : 1 }, "year" : 2019, "mont" : 1, "amount" : 13 }
{ "_id" : { "year" : 2018, "month" : 9 }, "year" : 2018, "mont" : 9, "amount" : 3 }
```

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?
 - Se requiere conocimiento de SQL y NoSQL.
 - Tiempos de migración.
 - Mecanismos de sincronización.
 - Requiere conocimientos experto en el estado actual del modelo relacional.
 - Denormalizaciones en el modelo relacional pueden dificultar el proceso de migración.
2. ¿Cuáles limitaciones adicionales se deben considerar, a parte de las mencionadas en el artículo?
 - Se debe tener en cuenta los mecanismos de sincronización para no perder información mientras se los usuarios finales siguen usando la aplicación.
 - Se debe evaluar los mejores tiempos y horarios de migración para evitar bloqueos de la DB en producción.
 - Se debe contar con tiempo en el equipo de desarrollo para ejecutar las migraciones (muchas veces estos procesos sacrifican nuevas funcionalidades planeadas por negocio).
3. ¿Cuáles son las razones (criterios) que se deben considerar para migrar un conjunto de datos relacionados a NoSQL?
 - Ejecutar consultas complejas en múltiples tablas y en tiempo reducido es uno de los criterios para una migración, ya que SQL no satisface este tipo de necesidades de negocio.
 - Realizar cambios drásticos en el modelo de negocio en cortos periodos de tiempo.