

Universidad Nacional Autónoma de Honduras - Valle de Sula

TEMA:

Análisis Estadístico de Historial Crediticio

INTEGRANTES:

Luis Felipe Banegas Ayestas

<https://github.com/FELIPEBANEGAS1>

Jorge Rivera

<https://github.com/JRivera16>



Índice

Portada

1. Índice
2. Objetivo
3. Adquisición de Datos
4. Almacenamiento de la Información
5. Software y CPU para Análisis de la Información
6. Internet para el acceso a Resultados
7. Distribución de Costos

1. INTRODUCCION

- ▶ El presente documento propone el montaje de una plataforma para el análisis estadístico del historial crediticio de los usuarios bancarios ya sea de uno o varios bancos.
- ▶ El proyecto consiste en recolectar la información de una gran cantidad de usuarios. Dicha información incluirá datos tales como fecha de nacimiento, lugar de nacimiento, lugar de domicilio, lugar de trabajo, historial de pagos, etc.

2. Objetivo

- El objetivo principal será de procesar la información brindada por los bancos para dar como valor agregado datos importantes sobre el riesgo que podría representar brindar un crédito a una persona en específico.



DATA
SCIENCE

3. ADQUISICION DE DATOS

- La información será brindada por la institución ya que es un proceso común en las instituciones financieras manejar una base de datos de los clientes, dicha base de datos incluye información tal como información demográfica, estatus familiar, estatus de empleo, historial crediticio actual, datos personales generales.



3. ADQUISICION DE DATOS

- La institución financiera deberá proveer una conexión directa a su base de datos. Por lo general estos datos están almacenados en bases de datos relacionales y se consumen haciendo uso de lenguaje SQL.



4. ALMACENAMIENTO DE LA INFORMACION


- ▶ Una vez obtenida la conexión a la base de datos de la institución financiera, los datos serán cargados a una base de datos totalmente aparte de la institución.
- ▶ Se estima que la cantidad de datos a almacenar estará en el orden de Terabytes debido a la posible cantidad de instituciones bancarias que hagan uso del servicio.





4. ALMACENAMIENTO DE LA INFORMACION

Se usará un servicio de base de datos relacionales en la nube tal como Google Cloud SQL, esta decisión se toma para reducir los costos de mantenimiento y ahorrar costos de personal administrador de redes.



4. ALMACENAMIENTO DE LA INFORMACION

- Para el almacenamiento de la información se hizo la cotización de un servicio base de datos relacional en la nube, con las siguientes características 4 vCPUs, 16 GB de RAM, 1 TB de almacenamiento en SSD por \$ 453.25/mes, comparado al salario mensual de un administrador de sistemas que puede rondar \$ 1300.00.

Estimate

Cloud SQL for PostgreSQL

DB-STANDARD-4

of instances: 1

Instance type: db-standard-4

Location: Iowa

730.0 total hours per month

SSD Storage: 1,024.0 GiB

Backup: 1,024.0 GiB

USD 453.25

Total Estimated Cost: USD 453.25 per 1 month

Estimate Currency

USD - US Dollar

EMAIL ESTIMATE

SAVE ESTIMATE

<https://cloud.google.com/products/calculator/#id=fcb5d424-0487-4ee3-8255-60b0ee5d0680>

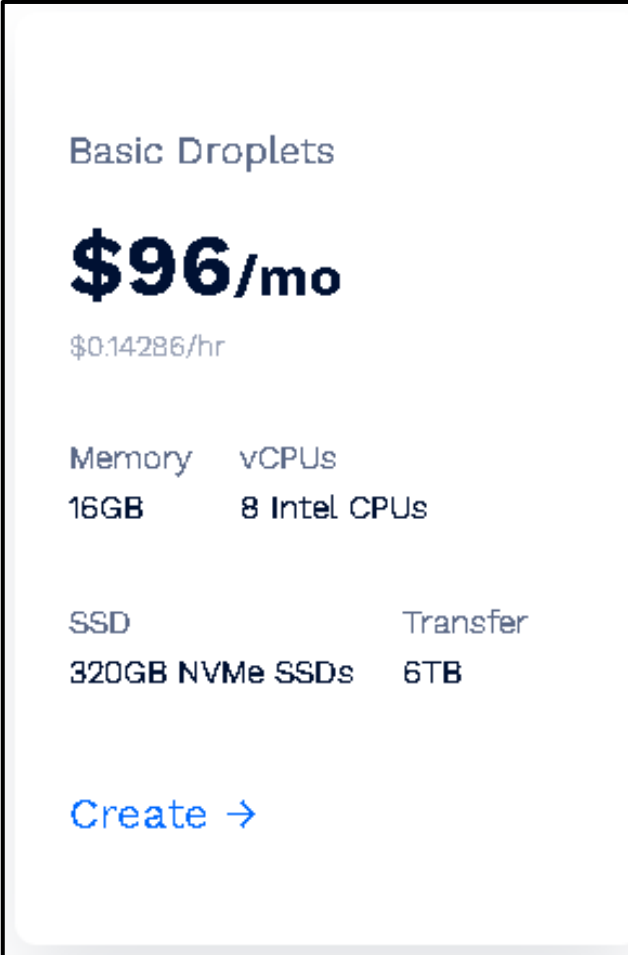


5. SOFTWARE Y CPU PARA ANALISIS DE LA INFORMACION

- Para el análisis de la información se utilizará un lenguaje libre de código abierto tal como Python y también se utilizaran librerías de machine learning tales como PyTorch.

5. SOFTWARE Y CPU PARA ANALISIS DE LA INFORMACION

- Los modelos desarrollados serán entrenados en una máquina virtual Linux en la nube con las siguientes características 8 vCPUs, 16 GB de RAM, 320 GB de almacenamiento SSD, tal como la se muestra en la imagen a la derecha.



Basic Droplets

\$96/mo

\$0.14286/hr

Memory	vCPUs
16GB	8 Intel CPUs
SSD	Transfer
320GB NVMe SSDs	6TB

Create →

<https://www.digitalocean.com/pricing>

6. INTERNET PARA EL ACCESO A RESULTADOS

50 MB

\$ 70

Televisión Analógica
Caja HD+ 100 canales digitales+ 100 canales HD
Soporte Técnico Personalizado

Contratar

- Para mantener el acceso a los resultados se contratara un proveedor de internet local con red de fibra óptica con ancho de banda de 50 MB.

http://www.multicable.hn/index.php?option=com_sppagebuilder&view=page&id=39&Itemid=169

7. DISTRIBUCION DE COSTOS

Distribución de costos

Descrpción	Porcentaje
Software	0%
Hard Disks	4%
CPUs	89%
Network (internet or others)	7%

