|  |
| --- |
| **FACULTAD DE INGENIERÍA Y ARQUITECTURA**  **CARRERA DE INGENIERÍA DE SISTEMAS**  **CURSO: ANALÍTICA CON BIG DATA** |

**Laboratorio 2 (parte 1)**

**Apache Cassandra**

En esta parte de la asignación se utilizará el servidor online de Cassandra llamado ASTRA y que se encuentra en <https://astra.datastax.com>. Si no se tiene una cuenta, se tendrá que crear una cuenta, de preferencia con el correo de la universidad. Para la parte donde se solicita el uso de CQL se puede utilizar a través de la misma interfaz online. Para la parte del driver de Python, se sugiere utilizar Google Colab en la nube o algún sistema local como Jupyter Notebook (sin embargo, igualmente se podría utilizar cualquier Python desde el terminal o desde algún otro IDE).

1. Suponer que se tiene una cadena de farmacias con distintas sucursales y que los clientes no son exclusivos de una sucursal específica (pueden comprar en diferentes farmacias de la cadena). Para cada cliente, se desea almacenar información relacionada con el producto que se compra, como el ID del producto, la cantidad, cuándo se compra, el precio que se paga, etc. Usando Apache Cassandra, proponer una solución a este problema, creando al menos 8 registros. Se debe usar directamente comandos de CQL (NO Python) Indicar lo creado, los comandos usados, y una captura de pantalla de la tabla final. **(3 puntos)**
2. El promedio industrial Dow Jones es un índice bursátil que mide el desempeño de las 30 mayores sociedades anónimas que cotizan en el mercado bursátil de Estados Unidos. El archivo “DowJones20.csv” muestra información de un periodo de 5 años sobre este índice (y ha sido obtenido de https://finance.yahoo.com/quote/%5EDJI/history). Usando Python, cargar los datos de este archivo en una tabla de Cassandra, escogiendo tipos de datos adecuados para cada una de las columnas. Mostrar las instrucciones utilizadas y una verificación del resultado obtenido. **(3 puntos)**
3. Usando el driver de Python, y posiblemente alguna librería de Python, obtener únicamente los registros correspondientes al año 2020 de la tabla de Cassandra creada en el punto anterior. Luego, para estos registros, indicar cuál fue el volumen promedio, el volumen total, en qué fecha se dio el más alto y el más bajo valor al cierre (columna Close). Brindar alguna otra información que se pueda considerar relevante, y mostrar al menos dos gráficas relevantes a partir de los datos, usando matplotlib. **(4 puntos)**