Proyecto No. 3
HBase - *PAREJAS o TRÍOS*

I. Modalidad y fecha de entrega

a) El proyecto debe realizarse en parejas o grupos de tres personas. Deberán asignarse a los grupos en canvas.
b) Debe ser enviado antes de la fecha límite de entrega: Martes 28 de mayo a las 23:59 horas a más tardar.
c) El orden de los grupos a presentar será sorteado aleatoriamente previo a la semana de entrega. Las presentaciones de proyecto se basarán en la dinámica de *show and tell* al catedrático durante el período de clase, cada grupo deberá presentar las funcionalidades (requeridas y extras) de su proyecto.
d) Algunos grupos podrán escoger entregar hasta el jueves 30 de mayo antes de las 17:20 horas, pero deberán cambiar de lugar con otro grupo que presentaba ese día (en caso que originalmente presentaran el miércoles). Ambos grupos deberán acordar este movimiento. Si decide entregar hasta el jueves, no podrá obtener más del 75% de su nota.
e) Consultar la rúbrica de evaluación.

II. Objetivo y descripción de la actividad

El objetivo de este proyecto es desarrollar una aplicación de línea de comando que facilite a los usuarios la gestión y almacenamiento de datos estructurados como archivos columnares. Esta aplicación simulará el funcionamiento de HBase, implementando las funciones DDL y DML comúnmente utilizadas en HBase. La aplicación abordará aspectos clave, como la definición y estructura de almacenamiento de los archivos, la capacidad de manipular múltiples tablas y familias de columnas, y la manipulación de los diferentes tipos de datos en las celdas (texto, numérico, listas, fechas, booleanos, etc).

Instrucciones generales y observaciones

Deberá implementar, en su lenguaje de programación de preferencia, el funcionamiento de HBase. Hoy en día existen muchas librerías que se conectan directamente a HBase, y la idea es que puedan simular el funcionamiento de este sin utilizar ninguna librería adicional.

No será necesario manejar múltiples bases de datos, pero sí deberá tener la capacidad de poder crear y manejar múltiples tablas.

Deberá cargar un set de datos inicial. El volumen de estos datos es indiferente, pero la idea es que pueda manipular información en distintas tablas.

Cada función que implemente la puede llamar como gusten bajo la sintaxis que el grupo considere, pero sí tiene que dar el mismo resultado que HBase tomando los mismos argumentos. Los metadatos que maneje su programa deberán de simular el comportamiento del HFile: row_key ordenados, column families correspondientes, timestamp de la columna específica y el valor. El formato y la forma en la que estos HFiles se almacenen quedarán a discreción del grupo. Recuerde que cada consulta que se haga hacia HBase tiene que ir a consultar de la misma manera estos HFiles. Tome como referencia la imagen a continuación:

Row Key	CF1:Col	Timestamp	Value
1	CF1:A	1591649830	Val_1
1	CF1:B	1591649830	Val_2
2	CF1:A	1591649830	Val_3
2	CF1:B	1591649830	Val_4

Row Key	CF2:Col	Timestamp	Value
1	CF2:C	1591649830	Val_5
1	CF2:D	1591649830	Val_6
2	CF2:C	1591649830	Val_7
2	CF2:D	1591649830	Val_8

Ejemplo de HFiles

Deberán de existir tantos HFiles según la cantidad de Regions utilizados, y deberá asumir que cada tabla usa únicamente un Region y no será necesario manejar temas de alta disponibilidad.

Su programa deberá tener una interfaz (GUI) que permita escribir estas líneas de código para consultar HBase y que sobre esta misma interfaz devuelva los resultados.

Se recomienda consultar la rúbrica de evaluación para mayor detalle. Lo más importante es la interacción que hay con los datos que maneja.

Funciones a implementar

DDL (Lenguade de definición de datos)

- Create
- List
- Disable
- Is_enabled
- Alter
- Drop
- Drop All
- Describe

DML (Lenguaje de manipulación de datos)

- Put (como función para insertar y actualizar. Si actualiza el timestamp deberá ser actualizado).
- Get
- Scan
- Delete
- Deleteall
- Count
- Truncate (deberá replicar el disable, drop y recreate de la tabla)

ETAPA 01 - Propuesta de Proyecto

Una vez definidos los grupos y asignados en canvas, deberán pensar en la estructura de los archivos que generarán para simular el funcionamiento de HBase.

ETAPA 02 - Elaboración del proyecto

Con la propuesta aprobada, se encargará de implementar la solución en el lenguaje de programación a su preferencia. Lo más importante será cómo implementen las funciones DDL y DML con los aspectos a evaluar de la rúbrica.

ETAPA 03 - Presentación

Presentará los resultados de su proyecto y el funcionamiento del mismo durante los períodos de clase. La forma de evaluación será empezar explicando su proyecto y la estructura de almacenamiento de sus datos, e ir mostrando cómo su solución cumple con cada uno de los aspectos a evaluar.

III. Temas a reforzar

- Funcionamiento general de HBase
- Estructura y almacenamiento de datos
- Funciones DDL y DML
- Uso de múltiples tipos de datos
- Usabilidad de una base de datos orientada a columnas

IV. Entregables

1. Código fuente desarrollado (url de repositorio, con historial de cambios).
2. Video general del funcionamiento de su aplicación, que no dure más de **5 minutos**. En caso de no subir el video de funcionamiento en su entrega, perderá hasta el 10% de la nota.
3. Documento escrito que incluya un diagrama del modelo de datos desarrollado y la explicación de este. Deberá incluir la definición de las diferentes etiquetas de nodos y tipos de relaciones, con sus respectivas propiedades y tipos de datos.

V. Rúbrica de Evaluación

Categoría	Criterio	Puntaje
General	Dataset inicial cargado y funcional Manejo de distintas tablas, distintos column families por tabla, manejo de timestamps, valores asignados. Las tablas deberán tener sentido lógico, el dataset deberá ser representativo. Puede apoyarse de cualquier material gráfico que desee	10
	Simulación de HFiles RowKeys ordenados, formatos adecuados, funcionamiento coherente, asociado a un único Region. Todos los cambios realizados a las tablas se deberán ver reflejados en estos HFiles	10
Funciones DDL	Create	5
	List	5

	Disable	3
	Is enable	3
	Alter	3
	Drop	3
	Drop all	3
	Describe	5
Funciones DML	Put Insertar y actualizar	10
	Get Output de manera "limpia y ordenada" (output in a readable, pretty way)	8
	Scan Output en formato "HTable". Deberá incluir metadatos como el timestamp. Este último debe cambiar en caso se realice un update	10
	Delete	5
	Delete all	5
	Count	5
	Truncate Deberá mostrar un output que indique el paso a paso de lo que se está ejecutando por detrás y seguir el mismo lineamiento del truncate (Disable, truncate, etc.)	7
Extras	Insert many	5
	Update many	5
	Indexado	5
	UI	5

Nota: Lo marcado en amarillo hace referencia a los puntos extra. El proyecto tiene un valor total de 15 puntos netos pero será calificado sobre 100. Podrá llegar a obtener hasta 120 puntos en total.