Ejercicios HDFS: Block Size & Balanceamiento

```
# Para iniciar el Cluster de HDFS
start -dfs.sh

# Para parar el Cluster de HDFS
stop -dfs.sh

# Para insertar un fichero
hdfs dfs -D dfs. blocksize =Xm -D dfs. replication =Y -
copyFromLocal ficheroOrigen destino

# Para leer un fichero
hdfs dfs -cat fichero

# Para leer un fichero sin mostrar contenido y ademas contar tiempo
time hdfs dfs -cat fichero >/dev/ null

# Para borrar un fichero
hdfs dfs -rm fichero

# Para conocer numero de bloques de un fichero en un determinado slave
hdfs fsck / ruta / fichero -files -blocks -locations | grep
IPslave | wc -l

# Para conocer estadisticas sobre un fichero
hdfs fsck fichero
```

1. Block size

1.1. Ejercicio 1

Con algún fichero .txt (>1024MB) haga diversas pruebas modificando el block_size y contestando a las cuestiones.

Nota: Tenga en cuenta que el fichero tiene un factor de replicación por defecto generar a dos copias al tener dos slaves, lo que duplica el espacio, considere suprimir los ficheros de HDFS una vez acabe el ejercicio.

1. Complete la tabla

block size	# blocks	input time	read time 1	read time 2	read time 3	Avg read time
2m						
32m						
128m						
2GB						

2. ¿Tiene sentido el número de bloques creado? ¿es el mínimo número de bloques posible?

3. Sobre el tiempo de inserción, ¿cómo afecta el tamaño de los bloques?

4. Sobre el tiempo de lectura, ¿cómo afecta el tamaño de los bloques?

5. Suponga que el fichero es del orden de Terabytes, ¿consideraría bloques de tamaño por defecto (128 MB) ?, ¿por qué?

2. Balanceamiento

2.1. Ejercicio 2

Suprima los ficheros que haya insertado en el ejercicio 1 para liberar espacio en HDFS, vuelve a insertar el mismo fichero con diferentes tamaños de bloque pero fijando siempre a 1 el número de replicaciones.

1. Complete la tabla

block size	# blocks slave1		# blocks slave2	
2m				
32m				
128m				

2. ¿Cómo cree que HDFS distribuye los bloques?

3. ¿Cómo afecta el tamaño de bloques a la distribución?, ¿cuándo se alcanza un mejor balanceamiento?

4. ¿Por qué fijar número réplicas a 1, que sucedería si fuesen 2 o 3?