

Taller Práctico

Carga de datos

1. Iniciar Ambari
2. Ir a Files View
3. Cargar navegar hasta la carpeta de maria_dev
4. Crear carpeta "taller"
5. Subir el archivo a hadoop
6. Ir a Hive View
7. Cargar el archivo desde hadoop /user/maria_dev/taller/u.data, desde csv y con delimitador TAB
8. Nombrar la tabla "ratings"
 - Columnas:
 - user_id
 - movie_id
 - rating
 - rating_time
9. Click en Upload Table
10. Realizar el mismo proceso, con u.item
 - Llamar a la tabla "movie_names"
 - Columnas:
 - movie_id
 - movie_name
 - Las demas columnas no serán relevantes por el momento

Análisis de Datos

1. Ir a Hive view
2. Ver las tablas cargadas
3. Ejecutar

```
SELECT movie_id, count(movie_id) as ratingCount
FROM ratings
GROUP BY movie_id
ORDER BY ratingCount
DESC
```

4. Chequear resultados
5. Chequear graph
 - x: movie_id
 - y: ratingCount