



Big Data com Hive e Impala

SQOOP

Sqoop



Ferramenta do Ecossistema
Hadoop para Importação de
Dados de RDBMS



Através de drive JDBC



Sqoop



Importa com 4 tarefas em paralelo (pode ser alterado usando `-m`)



Usa chave primária para dividir as tarefas

Espera-se que a chave primária esteja balanceada
Funciona com chaves primárias simples
Pode-se mudar a coluna usando-se `-split-by`



Sqoop

- ◇ Modo incremental
 - ◇ Append: para tabelas que tem inclusão de novos registros
 - ◇ Importa novos registros baseados em um ID
 - ◇ LastModified: quando os dados são atualizados
 - ◇ Utilizada um time stamp nos registros para verificar quais precisam atualização

Principais Funções

- ◆ Import: importa uma única tabela
- ◆ import-all-tables: importa todas as tabelas
- ◆ list-databases: lista bancos de dados
- ◆ list-tables: lista tabelas de um banco de dados

```
sqoop list-databases --connect  
jdbc:postgresql://127.0.0.1/ --  
username postgres --password 123456
```

Argumentos

- ◇ --connect : string de conexão, deve conter o servidor e o banco de dados
- ◇ --username: nome do usuário do banco de dados
- ◇ --password: senha na linha de comando
- ◇ --P: senha interativa (mais seguro)
- ◇ -- --schema: schema do banco de dados, se existir
- ◇ --table: importa tabela específica