Working with Big Data in SQL

Bulk Inserts

Embedding SQL in Python

Importing Large Amounts of Data

◆Assuming data is in a text format (e.g., a comma separated file)

◆Two options

1. INSERT INTO
2. From command line using *mysqlimport*

◆Inherently the same

Using INSERT INTO

◆ LOAD DATA LOCAL INFILE <filename.csv> INTO TABLE <tablename> FIELDS TERMINATED BY ',' ESCAPED BY "" IGNORE 1 LINES;

◆ IGNORE header if exists

From Command Line

◆ mysqlimport --local --fields-terminated-by=, --fields-enclosed-by='\"' --ignore-lines=1 –u <username> -p <dbname> <filename>

◆ Expects the filename to have the same prefix as the table name

◆ Calls LOAD DATA internally

Connecting to Database from Python

◆ Need a connector

  ◗ This is true for any programming language

◆ MySQLdb or PyMySQL

  ◗ MySQLdb is a thin wrapper around the _mysql API (originally developed for C)

  ◗ PyMySQL is a pure Python library

◆ We will use PyMySQL