workable

Assignment for the Operations Engineer position

Task 1

The file 2014-09-03.log is a log file from the demo environment of Workable, including all entries for a particular day. You can access this demo environment here: http://workabledemo.com/backdoor

Using this log as input, please answer the following questions.

1. List of URLs that were not found (404 error), including number of times each URL was requested
2. Average time to serve a page
3. Which database table is most frequently loaded?
4. Is any URL redirection taking place?
5. Are there any server errors? Ideas about possible causes?

Deliverables

For each question, please provide the following:

- Answer
- Explanation of how you arrived at the result, including any scripts or other commands that you used.

Task 2

You are working for a business that owns two DVD rental stores. Both stores use software to manage the DVD rentals, but the software has no reporting capabilities.
The owner wants to better understand various aspects of his business, and has asked you a few questions.
In order to answer his questions, you will design and execute a few queries on the database of the DVD rental software that is used by the stores (there is one database for both stores).
You can view the database schema at the following link:
http://www.postgresqltutorial.com/postgresql-sample-database/
This page also contains instructions to download the full database and restore it in a PostgreSQL instance that you will set up.

Please provide the SQL queries that you will use to answer the owner's questions, as well as the actual answers.

Tip: You may receive a few errors on alter table statements when restoring the database, which can be ignored.
Tip: You may find it useful to use the WITH statement.

1. Which customer has made the most rentals at store 2?
2. What percentage of movies were out of stock at each store on 29/7/2005 at midnight?

3. Is the employee that performs the rental of a DVD usually the one who also takes the payment?
4. How many rentals do we do per month?
5. What percentage of our customers are active at any given month? We define active as performing at least one rental during that month.
6. Are there some films that are particularly popular and are rented all the time, or do people tend to spread their choice evenly among the available films?
7. Which film category is the most popular among our customers?
8. Are there any other insights that you can gather from the data that would be helpful to me, as the owner of this business?

Deliverables

For each question, please provide the following:

- Answer
- SQL query
- For question 8, also explain the way in which the insights that you provide will be useful to the business

Task 3

Your objective is to build a small project named MovieData. The application will use the API from Rotten Tomatoes (http://developer.rottentomatoes.com/). Use the following API key: qtqep7qydngcc7grk4r4hyd9 or create a new one on the site.

The goal of this assignment is to implement an application or script that will retrieve and store movie information.

The application must retrieve a list of the movies currently in theaters, along with at least the following attributes

- **name**
- **description**
- **year of production**
- **List of starring actors**

For each actor, it must try to find that actor's entry on Wikipedia and pull the first paragraph of the article. It should not use Wikipedia's API to extract the text, but rather any form of page scraping.

All the above information will be stored in a relational database. The data in the database will be updated each time the process runs.

You are free to include any additional features / optimizations that you may find relevant or could showcase your skills but please bear in mind that you should cover the core requirements first before attempting any improvements.

Deliverables

The final deliverable should contain:

- Source code
- DDL for Database schema
- A simple Readme.txt that will describe the way to build and use the application along with any other external systems that you may have used.

Notes

You will need to provide a working prototype of the application. You are free to implement your project in any technology stack you prefer between Python/Perl/PHP/Ruby/Any scripting language/Java, and also to use any tools that you see fit.

You will be assessed on the following:
a) Fitness for purpose (i.e. your application does what its specifications require)
b) Simplicity (the simpler / smaller the solution, the better)
c) Code quality