Apply Filters to SQL Queries

Portfolio Activity Document | Cybersecurity Analyst

Project description

I put this activity together while working through a security review for my organization. Part of my role is keeping an eye on the system, tracking down anything that looks suspicious, and pushing updates out to employee machines when they're needed. Below are a few examples of how I used SQL filters to get through that work.

Retrieve after hours failed login attempts

We flagged a possible security incident that took place outside of normal business hours (after 18:00), so I needed to pull every failed login attempt that happened after that cutoff for review.

Here's the query I wrote to isolate those failed after-hours attempts:

```
MariaDB [organization]> SELECT * FROM log_in_attempts WHERE login_time > '18:00' AND success = FALSE;
+----------+----------+------------+------------+---------+-----------------+---------+
| event_id | username | login_date | login_time | country | ip_address      | success |
+----------+----------+------------+------------+---------+-----------------+---------+
|        2 | apatel   | 2022-05-10 | 20:27:27   | CAN     | 192.168.205.12  |       0 |
|       18 | pwashing | 2022-05-11 | 19:28:50   | US      | 192.168.66.142  |       0 |
|       20 | tshah    | 2022-05-12 | 18:56:36   | MEXICO  | 192.168.109.50  |       0 |
|       28 | aestrada | 2022-05-09 | 19:28:12   | MEXICO  | 192.168.27.57   |       0 |
|       34 | drosas   | 2022-05-11 | 21:02:04   | US      | 192.168.45.93   |       0 |
|       42 | cgriffin | 2022-05-09 | 23:04:05   | US      | 192.168.4.157   |       0 |
|       52 | cjackson | 2022-05-10 | 22:07:07   | CAN     | 192.168.58.57   |       0 |
|       69 | wjaffrey | 2022-05-11 | 19:55:15   | USA     | 192.168.100.17  |       0 |
|       82 | abernard | 2022-05-12 | 23:38:46   | MEX     | 192.168.234.49  |       0 |
|       87 | apatel   | 2022-05-08 | 22:38:31   | CANADA  | 192.168.132.153 |       0 |
|       96 | ivelasco | 2022-05-09 | 22:36:36   | CAN     | 192.168.84.194  |       0 |
|      104 | asundara | 2022-05-11 | 18:38:07   | US      | 192.168.96.200  |       0 |
|      107 | bisles   | 2022-05-12 | 20:25:57   | USA     | 192.168.116.187 |       0 |
|      111 | aestrada | 2022-05-10 | 22:00:26   | MEXICO  | 192.168.76.27   |       0 |
|      127 | abellmas | 2022-05-09 | 21:20:51   | CANADA  | 192.168.70.122  |       0 |
|      131 | bisles   | 2022-05-09 | 20:03:55   | US      | 192.168.113.171 |       0 |
|      155 | cgriffin | 2022-05-12 | 22:18:42   | USA     | 192.168.236.176 |       0 |
|      160 | jclark   | 2022-05-10 | 20:49:00   | CANADA  | 192.168.214.49  |       0 |
|      199 | yappiah  | 2022-05-11 | 19:34:48   | MEXICO  | 192.168.44.232  |       0 |
+----------+----------+------------+------------+---------+-----------------+---------+
19 rows in set (0.033 sec)
```

This one pulls failed login attempts that happened after 18:00. I started with a basic SELECT * against the log_in_attempts table, then narrowed it down with a WHERE clause using AND, since I needed both conditions to be true at once. The *login_time > '18:00'* piece grabs anything after 6pm, and *success = FALSE* keeps only the failed attempts.

Retrieve login attempts on specific dates

A suspicious event came up on 2022-05-09, so I needed to look at everything that happened that day plus the day before, in case there was any lead-up activity worth investigating.

The query below covers both dates:

```
MariaDB [organization]> SELECT * FROM log_in_attempts WHERE login_date = '2022-05-09' OR login_date = '2022-05-08';
+----------+----------+------------+------------+---------+----------------+---------+
| event_id | username | login_date | login_time | country | ip_address     | success |
+----------+----------+------------+------------+---------+----------------+---------+
|        1 | jrafael  | 2022-05-09 | 04:56:27   | CAN     | 192.168.243.140 |       1 |
|        3 | dkot     | 2022-05-09 | 06:47:41   | USA     | 192.168.151.162 |       1 |
|        4 | dkot     | 2022-05-08 | 02:00:39   | USA     | 192.168.178.71  |       0 |
|        8 | bisles   | 2022-05-08 | 01:30:17   | US      | 192.168.119.173 |       0 |
|       12 | dkot     | 2022-05-08 | 09:11:34   | USA     | 192.168.100.158 |       1 |
|      191 | cjackson | 2022-05-08 | 06:46:07   | CANADA  | 192.168.7.187   |       0 |
|      193 | lrodriqu | 2022-05-08 | 07:11:29   | US      | 192.168.125.240 |       0 |
|      197 | jsoto    | 2022-05-08 | 09:05:09   | US      | 192.168.36.21   |       0 |
+----------+----------+------------+------------+---------+-----------------+---------+
75 rows in set (0.001 sec)
```

This returned 75 login attempts across the two dates. Again I started from the full log_in_attempts table, but this time I used OR instead of AND, since I wanted rows matching either date, not both at once. One condition checks for *login_date = '2022-05-09'* and the other for *login_date = '2022-05-08'*.

Retrieve login attempts outside of Mexico

Digging further into the login data, I noticed a pattern worth flagging in the attempts that came from outside of Mexico, so those needed a closer look too.

Here's how I filtered for everything outside Mexico:

```
MariaDB [organization]> SELECT * FROM log_in_attempts WHERE NOT country LIKE 'MEX%';
+----------+----------+------------+------------+---------+-----------------+---------+
| event_id | username | login_date | login_time | country | ip_address      | success |
+----------+----------+------------+------------+---------+-----------------+---------+
|        1 | jrafael  | 2022-05-09 | 04:56:27   | CAN     | 192.168.243.140 |       1 |
|        2 | apatel   | 2022-05-10 | 20:27:27   | CAN     | 192.168.205.12  |       0 |
|      197 | jsoto    | 2022-05-08 | 09:05:09   | US      | 192.168.36.21   |       0 |
|      200 | jclark   | 2022-05-12 | 01:11:45   | CANADA  | 192.168.91.103  |       1 |
+----------+----------+------------+------------+---------+-----------------+---------+
144 rows in set (0.018 sec)
```

This gives every login attempt from a country other than Mexico. I used NOT combined with LIKE and the *MEX%* pattern, since the dataset lists Mexico two different ways (MEX and MEXICO), and the % wildcard covers both by matching any characters that follow.

Retrieve employees in Marketing

Next up, my team needed to push updates to specific Marketing department machines, so I had to pull a list of which employees and devices that covered.

This query gets me the Marketing employees in the East building:

```
MariaDB [organization]> SELECT * FROM employees WHERE department = 'Marketing' AND office LIKE 'East%';
+-------------+--------------+----------+------------+----------+
| employee_id | device_id    | username | department | office   |
+-------------+--------------+----------+------------+----------+
|        1000 | a320b137c219 | elarson  | Marketing  | East-170 |
|        1052 | a192b174c940 | jdarosa  | Marketing  | East-195 |
|        1075 | x573y883z772 | fbautist | Marketing  | East-267 |
|        1088 | k8651965m233 | rgosh    | Marketing  | East-157 |
|        1103 | NULL         | randerss | Marketing  | East-460 |
|        1156 | a184b775c707 | dellery  | Marketing  | East-417 |
|        1163 | h679i515j339 | cwilliam | Marketing  | East-216 |
+-------------+--------------+----------+------------+----------+
7 rows in set (0.001 sec)
```

This returns everyone in Marketing who sits in the East building. I combined two conditions with AND since both needed to be true: *department = 'Marketing'* and *office LIKE 'East%'*. I used LIKE with the wildcard here because the office column stores a specific office number after the building name, so East% catches all of them.

Retrieve employees in Finance or Sales

Finance and Sales were next in line for a different update, so this time I only needed employees from those two departments.

Here's the query for pulling both departments at once:

```
MariaDB [organization]> SELECT * FROM employees WHERE department = 'Finance' OR department = 'Sales';
+-------------+--------------+----------+------------+-------------+
| employee_id | device_id    | username | department | office      |
+-------------+--------------+----------+------------+-------------+
|        1003 | d394e816f943 | sgilmore | Finance    | South-153   |
|        1007 | h174i497j413 | wjaffrey | Finance    | North-406   |
|        1008 | i858j583k571 | abernard | Finance    | South-170   |
|        1009 | NULL         | lrodriqu | Sales      | South-134   |
```

```
|        1187 | f963g637h851 | bbode    | Finance    | East-351    |
|        1188 | g164h566i795 | noshiro  | Finance    | West-252    |
|        1195 | n516o853p957 | orainier | Finance    | East-346    |
+-------------+--------------+----------+------------+-------------+
71 rows in set (0.001 sec)
```

This pulls everyone from Finance and Sales. I went with OR this time instead of AND, since I wanted anyone matching either department rather than both. The two conditions are *department = 'Finance'* and *department = 'Sales'*.

Retrieve all employees not in IT

The last update needed to go out to everyone outside of the IT department, so I pulled that list before rolling anything out.

This is the query I used to exclude IT:

```
MariaDB [organization]> SELECT * FROM employees WHERE NOT department = 'Information Technology';
+-------------+--------------+----------+-----------------+-------------+
| employee_id | device_id    | username | department      | office      |
+-------------+--------------+----------+-----------------+-------------+
|        1000 | a320b137c219 | elarson  | Marketing       | East-170    |
|        1001 | b239c825d303 | bmoreno  | Marketing       | Central-276 |
|        1002 | c116d593e558 | tshah    | Human Resources | North-434   |
|        1003 | d394e816f943 | sgilmore | Finance         | South-153   |
|        1195 | n516o853p957 | orainier | Finance         | East-346    |
|        1198 | q308r573s459 | jmartine | Marketing       | South-117   |
|        1199 | r520s571t459 | areyes   | Human Resources | East-100    |
+-------------+--------------+----------+-----------------+-------------+
161 rows in set (0.001 sec)
```

This returns everyone who isn't in the Information Technology department. I used NOT against *department = 'Information Technology'* to exclude that group and keep everyone else.

Summary

Across these tasks I used SQL filters to pull the exact login and employee data I needed, working with the log_in_attempts and employees tables. I relied on AND, OR, and NOT depending on whether I needed conditions to overlap or exclude each other, and used LIKE with the % wildcard whenever I needed to match a pattern rather than an exact value.