Introducing YugabyteDB Performance Advisor

Sushant Mishra
July/22/2022

YFTT
YUGABYTEDB
FRIDAY
TECH TALKS

yugabyteDB

Problem of finding performance issues in distributed systems

- It is hard to operationally manage the distributed database cluster.

- In single node database to debug performance issues.
  - Monitor metrics such as CPU, memory, network and IO utilization of specific system.
  - Monitor database metrics and events.
  - Correlated all the information collected to find the issues.

- For distributed system, the same analysis needs to be performed on multiple nodes in the cluster and subsequently required to correlated those metrics across multiple different nodes.
  - A query processed on initiated on node A may involved multiple nodes in the cluster, necessitating the need to analyse all the nodes which processed this query.

What is the objective of Performance Advisor?

- Auto monitor the YugabyteDB cluster and various metrics to find the bottlenecks and performance issues.
- Reduce the need to manually monitor the cluster for most of the common problems faced.
- Provide end user with recommendations based on which options they can take concrete action to fix the underlying problem.

Performance Advisor: Delete unused indexes

- An index allows the database to retrieve specific rows much faster than it could do without an index.
  - Without using an index it results in querying the whole table to find the relevant rows.
  - Adding an index is a common recommended way to enhance database performance.
- Index can also possibly add to the latency of the writes to the database. When a row is updated, the corresponding index is updated if the updated columns contains index columns.
- It is preferable to delete the unused indexes as they may help improve the write performance.

Performance Advisor: Hash vs. Range Sharding

- YugabyteDB supports hash and range sharding for the data
  - https://docs.yugabyte.com/preview/architecture/docdb-sharding/sharding/
- Default sharding in YugabyteDB is hash
  - CREATE TABLE **Employee** (emp_id int, name text, department text, joining_date timestamp, PRIMARY KEY (emp_id));
    - **Employee** table is hash_shareded on emp_id
  - CREATE INDEX *joining_date_index* on employee (joining_date);
    - joining_date_index is hash sharded

.

- Common query to find employees joined in January 2020
  - **SELECT** *name*, *department* from employee where joining_date >= '2020-01-01' AND joining_date < '2020-01-31' ;
  - Due to ***joining_date_index*** being hash shared, the query results in seq_scan despite having index
  - It is preferable to have range sharding for ***joining_date*** columns
    - Create index joining_date_index on employee (joining_date ASC);
    - It allows index scan to be used for range based queries

Performance Advisor: Nodes utilization imbalance

- There are various ways to connect to nodes in the YugabyteDB cluster
  - Using a load balancer
  - Smart driver
  - Application managing the connection to each individual node in the cluster
- Due to various issues based on how the application connects to the nodes in the database cluster, there could be imbalance based on the connections, queries processed and CPU utilization

Performance Advisor: Connection skew

- Measures number of connections created against each node and detects if any node is handling a higher number of connections then other nodes in the cluster.
- If a node handles the majority of connections, it may experience performance degradation.
- **Recommendation:** Balance connections evenly across nodes in the cluster.

Performance Advisor: Query askew

- Measures number of Queries executed against each node and detects if any node is handling a higher number of Queries then other nodes in the cluster.
- If a node handles the higher number of queries, then it may experience performance degradation.
- **Recommendation:** Review connections which are executing the queries and rebalance them across nodes in the cluster.

Performance Advisor: CPU skew and High CPU utilization

- CPU Skew measures whether CPU usage is higher on a specific node compared to other nodes in the cluster.
- Monitor the node with highest CPU utilization in the cluster.
- High CPU results in degraded performance for the queries executing on the node with higher CPU utilization.
- **Recommendation:** To identify why CPU usage is higher on a given node vs other nodes on cluster.

Thank You

Join us on Slack: yugabyte.com/slack (#yftt channel)

Star us on Github: github.com/yugabyte/yugabyte-db

Yugabyte YSQL Performance Advisor: https://docs.yugabyte.com/preview/yugabyte-cloud/cloud-monitor/cloud-advisor/

YFTT YUGABYTEDB FRIDAY TECH TALKS

yugabyteDB