COMP 120: Web Engineering
Leg 3: The Benefactor
Performance Plan

Andrew Burgos
Nolan Karpinsky
Ryan Havens
Jay DeStories

Overview

Issues

- We need to look at the larger tables in the database (potentially sharding the db)
- Database coupling can lead to issues in the future
- Lack of indexes. We need to begin to anticipate which tables need to be incorporated with clustered indexes vs non-clustered
- No memory caching for user data
- Potential difficulty monitoring the increased traffic for malicious activity
- Large websites are frequent targets of cyber attacks

Immediate Problems

- We cannot continue to use SQLite -- we plan to change to PostgreSQL.
- We need to employ a scalable method of hosting -- we plan to use Amazon EC2
- We need to separate media hosting from the rest of the site -- we plan to use Amazon S3

Non-Performance Problems

- Prior to benefactor advertising, we need a public domain name to market to our customers (Amazon Route 53)

COMP 120: Web Engineering
Leg 3: The Benefactor
Performance Plan

Optimization and Scaling

- Memory caching is important in reducing stress on servers and optimizing performance. Caching user data and frequent queries made to the database will allow us to handle more traffic. To do this we can use Memcached. http://memcached.org/
- UDP vs TCP connection. User Datagram Protocol is a faster but more unreliable connection. This system fires packets of data and does nothing to ensure they reach the location. Primarily used with audio and video files for fast uptime on stream. Transmission Control Protocol. This method is far more reliable but a little less efficient in performance. Data packets are fired in a specific order and are ensured to reach the destination.
- Web accelerator (f5.com)
- Since the tables are divided and distributed into multiple servers, the total number of rows in each table in each database is reduced. This reduces index size, which generally improves search performance. A database shard can be placed on separate hardware, and multiple shards can be placed on multiple machines. This enables a distribution of the database over a large number of machines, greatly improving performance
- Proper Indexing of data. Indexing is very important to the optimization of database queries.
    - Two types:
        - Clustered Index : Clustered indexes sort the tables in which they are included and allow easy access to database rows that are often queried. Tables that have primary keys are good options to have clustered indexes. Tables that are frequently updated or have wide keys are not good for clustered indexes. Because clustered indexes have an order in which they sort a table frequent changes to a table forces the database to continually move and rearrange data
        - Non-clustered Index: Non-clustered indexes do not sort the data in any specific way. Data is just thrown on the heap. These are best used for tables that are frequently getting updated and require no order. Because there is no order less queried tables favor non-clustered indexes
    - The storage location of indexes can improve query performance by increasing disk I/O performance. Storing a nonclustered index on a filegroup that is on a different disk than the table filegroup can improve performance because multiple disks can be read at the same time

COMP 120: Web Engineering
Leg 3: The Benefactor
Performance Plan

Security Challenges

- Encrypting user information allows us to establish trust between us and our clients. It would be incredibly harmful for our company image if an attack were to expose weak infrastructure.
- Obfuscation of important information in the database.
    - Passwords salted and hashed. A “salt” is random data that is used as an additional input to a one-way function that hashes a password or passphrase. Encrypting and obfuscating data
    - Avoid storing passwords and other important fields in plain-text at all
- Distributed Denial of Service attacks
    - Every year, malicious DDoS attacks break new records in disrupting web services worldwide. While attacks on Netflix and Twitter gain the most media spotlight, today nearly anyone could theoretically access open-source DDoS tools and carry out an attack on our website.
    - We can acquire DDoS protection through F5 or Cloudflare

COMP 120: Web Engineering
Leg 3: The Benefactor
Performance Plan

Testing Requirements

- Browser automation will be used to generate content for our system and ensure functionality when there is more content. We plan to use Selenium (docs.seleniumhq.org) to perform these tests.
- Stress tests will be performed on the live server before money is spent on advertising, both before and after content has been generated with browser automation. Many online services exist for this purpose, for example loadimpact.com. Depending on the performance of Selenium, it may be sufficient for stress-testing our expected loads.