

TME3- WSM - Management Concerns with Large Websites

Tumblr.com

I have decided and started working on Tumblr.com from web100 website. It is one of the top social networking and microblogging platform for users. It is a platform where a user can express about themselves, their ideas via a blog which can be audio, video and photo blog but also includes a tab where you can find and ask questions. On this platform a user can follow other users and like and share and follow the blogs by other users. The current count of Tumblr users are 345 million registered users. It is ranked on 7th number among the other social networking sites. The home page lets you post a picture, video, blog, image or text. It also displays all the blogs by other users which you may like/comment or already follow. You may share, repost or like the blog by users. It gives you the accountability for taking an action on any post/blog. Tumblr only lets you see the content once you are a active member. You can take part in any activity or may explore the site if wanted. There are other tabs such as Home, Explore, Messaging, Latest activity in your profile, Notifications, Profile and Tab to start writing or post any activity on your account. At Home tab, you can view new posts by users where you can comment, like or share. Explore where you can explore the posts or blogs from around the world on a single page. The blogs that are shown to the users are based on their interests and likes. Tab to start new writing involves the interest of the user what they would like to post without wasting much time. As it is one of the top rated social networking platform, there may be many issues that can arise while maintaining the website for an administrator.

Configuration and Accountability

The problems that an administrators may face in context to configuration is that managing large traffic on their website as I mentioned earlier that Tumblr has over 345 million registered users and they are expecting 2.4 billion users by 2021. The configuration should be keeping in mind that the site has minimum or least number of bugs. Tumblr is cross platform which means it is open to run on operating systems as well as on phones and tablets. Another issue can be managing and controlling so many different users and their accounts on a single platform so administrator use AI technology and 200 different servers in the market which will help maintaining the traffic. This can also be done via collecting information on the regular users who log in everyday. Tumblr follows reliable repository architecture format where all the components which are same are grouped together. It can grouped using zip techniques to reduce the size but they do not compromise on their quality. As it is a social networking site, it is much better for the site to have as many users as possible. There are set of permissions that are given only to the admin to maintain the website and set of permissions given to users to use and explore the websites. There are various scripts that are written to filter the traffic and manage the traffic. Tumblr also uses SBT(Scale Based Tool) for filtering, managing the client activities and services. Example: when posting a blog or promoting a blog on Tumblr a person has to go through the filters which will decide whether the post can be posted or not. If it is not worth the effort, the post will not be posted giving user a warning which comes under scripting done in java. There are over 500 million views everyday and over 15 Billion in a month. Tumblr started using LAMB application in the beginning but now they are using Scala, HBase, Redis, Kafka, Finagle for building a powerful dashboard where there are million writes per second on tumblr and it will help them to solve various php problems in a short period of time. There are around 500 web servers and 200 database servers which are used at Tumblr. There are huge amount of updates that the administrator at tumblr manages everyday. Tumblr uses HAProxy on over 40 machine to keep up with the large scale users. Use of transparent shading takes place at Tumblr which is maintained by using dashboard notifications, URL shortener and a memcache proxy. The updates takes place every two weeks. They have about 200 servers due to availability reasons. They need the site to be so dependant and accountable including responsible availability at the time of the need so there is no chance of going down. It is available to users almost all the time which is a huge achievement in social networking sector. A FuncApp software is used to record the versions and o/s the app is running on.

Performance and Security

Tumblr uses various softwares for providing the best experience for the users. The operating system linux is being used to develop the website along with Apache servers for better performance .They also uses php, scala and ruby for development. Performance of a website can be measured by taking into account the traffic/audience and their impacts. Tumblr analysis the traffic sources along with bounce rate and average session time of the users. This is managed for best performance where tumblr optimized the images, files. It also minimizes HTTP requests for increasing the performance and the speed of the website. Another thing which is used in tumblr is they run a compression audio to reduce the server response time. Adding to this, Tumblr uses Func a secure and scriptable remote control framework and API by redhat for better performance. It also uses Capistrano to reliably deploy applications on the web servers. Tumblr also uses two components as a platform which are tumblelogs and dashboard. Tumblelog is where people deal with in terms of blog which is easy to maintain, cache as it is not dynamic. In Dashboard, it is like twitter real time entries and updates. Caching is not done here as all the request are real and active and if the media size increases per day which is along 50 gb would end up getting stored in S3 which is a cloud storage service provided by amazon.

Previously Tumblr was using LAMP application build in PHP. It had one server, one web server and a php application from where it started growing. In the begging they started doing caching followed by HAProxy along with SQL sharding. SQL sharding was a huge help and a success for tumblr in the begging for performance issues. Currently Tumblr has JVM centric approach for more speed and better development. The main goal for now is to close the PHP application and start making the app where it has a thin layer of services requested like administration, presentation etc. Finagle is used which is a library by twitter. It helps in better performance, distributed tracing, service discovery, and service registration. HBase backs up their data which is historical. Data is handled better for better performance such as the active user who has million of users has a different mode and then there's one for the normal users. It helps keeping data together and best performance.

Security is one of the top most concerns for any or every site on the web. Nobody wants their site to go down by any attacker. So to prevent from those attacks and viruses, Tumblr takes it's security very seriously. There was an attack on tumblr in recently in 2018, june where the email and passwords of users were compromised. It was reported by OATH bug bounty program team which works for tumblr and is owned by Yahoo. To prevent such silly attacks, Tumblr has made a safe mode where the user can navigate for hours and will not be monitored or exposed under any circumstances. Also to add the security they have introduced SSL(Secure Socket Layer) with cache where the data is transferred from your tumblr server to the computer in an encrypted manner. It decreases the chances of getting seing. This was done keeping in mind the security of the Tumblr as well as their users. They have also put together the firewall configurations to stay away from attacks.

Privacy, Ethical and Legal Concerns

Tumblr is build to celebrate the freedom given to everyone who would like to share something a blog, idea, question or post a video or audio. During all this tumblr never lets anyone to hurt someone's sentiment or feelings. It is totally against terrorism, hate speech, harm to minors, promotions or glorification of self harm, threats, deceptive and fraudulent links, username abuse, spam, mass registration of an account, harassment to any user, privacy violation, human trafficking and prostitution and themes distributed by other parties. Any such acts against any figure or anyone would be reported and a action will be taken. There was a huge concern by June, 2018 where attackers were getting information on username and password which comes under privacy of a user. This was reported and the attack was saved by the developers. They have developed security protocols where the privacy of the user will not ever be compromised. Any post where the above things are involved, there will be a legal and strict action against the one who did it. There will be a notice given to the user by email who will be pushing these guidelines. Tumblr has strong security protocols for their site where privacy is the far most priority.

Conclusion

Tumblr has become one of the most used social networking platform which is hard to handle and more complicated than twitter. It is due to the large number of users adopting Tumblr and enjoy what the platform provides them. As discussed earlier the configuration, accounting, performance, security issues are well addressed and handled by Tumblr, it does make it one of the best platforms for the users to use. They have a strong backend in MYSQL and they use good programming methods. Tumblr holds well maintained and easy configurations along with best performance. The security is one of the main things they would like to increase their attention on. Having currently over 345 million users, it is their responsibility to maintain the platform for the security of these people as well as the privacy. Tumblr has overcome all these concerns and issues and is delivering one of the best social networking experience. Tumblr is number 9 among other social networking platforms which is a huge achievement and they would grow more as they are pursuing. They have made huge accomplishments in almost every sector and they have good availability of the software on every platform that can ever be imagined. They have various hard working teams in infrastructure, platform, SRE, web operations and services.

List of Reference

<http://highscalability.com/blog/2012/2/13/tumblr-architecture-15-billion-page-views-a-month-and-harder.html>

<https://www.tumblr.com/explore/asks>

<http://www.web100.com/>

<https://security.tumblr.com/>

<https://support.tumblr.com/>

<https://www.alexa.com/siteinfo/tumblr.com>

<https://www.cnet.com/news/tumblr-activates-ssl-but-with-a-catch/>

<https://www.tumblr.com/policy/en/community>