

CRITIQUE - 10 [Lecture - 11 Storage Performance]

The main topic of today's lecture is about storage performance. We discussed automation performance regression testing - a way of measuring performance from one build to the next build and analyzing it. Some of the standard benchmarks in storage performance measures like SPC1 and SPC2 were discussed briefly. We also talked about how to find performance bottlenecks, some techniques to fix performance and ZFS prefetch to enhance performance in streaming workloads.

The lecture started with giving us some information about the what the storage performance team at oracle does and quick peek into the topic that was going to be discussed. I appreciate that Natalie took an example of customer use-case to illustrate some of the typical uses of storage system like databases, streaming workloads (Netflix) etc and the possible performance bottlenecks that may arise. To get a rough estimate of the storage performance, the customer workloads are simulated using Vdbench. I got to know that you can set the workload parameters with Vdbench like IOPS, cache hits etc. we discussed about the bottlenecks in customer use case, like slow network (limitation on the network bandwidth), slow rotating disks, software issues(client and server bottlenecks), hardware issues (PCI limitations) and non-optimal configuration parameters. All of these issues will diminish the performance of storage systems.

The first step to fixing the performance bottlenecks is to define the problem statement and success criteria. There are a lot of utility tools available on solaris and linux like system for measuring performance. For instance to measuring the IOPS we can use tools like iostat, vmstat for statistics on virtual memory, dtrace is a dynamic instrumentation tool for probing the kernel code, kstat utility examines the available kernel statistics. CPU profiling can be done using dtrace and flame graph. Flame graphs are a visualization of profiled software, allowing the most frequent code-paths to be identified quickly and accurately. For monitoring the kernel locking and profiling statistics we can use lockstat. This part was interesting and I learned a bunch of profiling tools for detecting the performance issues.

I learnt about the different types of standard performance benchmarks like SPC1 and SPC2. SPC1 is used for simulating database type workloads and generates various levels of IOPS. SPC2 is focused on streaming workloads like video on demand, large file processing, large query data etc. Also from this <http://www.storageperformance.org> link, I got to know that SPC1 is characterized predominantly by random IO operation whereas SPC2 is characterized by large IO operations organized into one or more

concurrent sequential patterns. SPECsfs 2014 is the latest version of standard performance evaluation benchmark suite for measuring the file server throughput and response time. It provides a standardized way of measuring and comparing performance among different storage vendors.

In the second half of the lecture, we discussed ZFS performance projects on enhancing the ZFS prefetch. Prefetch is ZFS read-ahead feature, to predict and pre-cache the blocks for improved performance and reduced latency. Discussion on some of the considerations for prefetch gave me an idea on potential conflicts between ARC cached blocks and prefetched blocks, readahead thrashing and double IO's. Prefetch algorithms generate triggers to load the blocks from disk to the cache. It loads the indirect blocks asynchronously. Prefetch algorithm doubles the window size and generates prefetch trigger every time a sequential stream is read. The side by side comparison of the performance metrics of old prefetch to the new trigger based prefetch was helpful and gave me an idea of the impact of new algorithm on performance.

Overall, the lecture was good. I like customer use-case that was taken as a running example to illustrate all the performance issues, ways to measure the performance using utilities like iostat, PDM, flame graph, DTrace etc, and the possible techniques that can be employed to improve the performance and reduce latency. I really appreciate the fact that all the examples were visually illustrated especially use-case example and ZFS prefetching.

Reference:

1. Storage Performance - Lecture slides by Natalie Ross.
2. <http://www.storageperformance.org>
3. <https://www.spec.org/sfs2014/>