Index-Free: The Architecture of Modern Log Management

For anyone responsible for log management, the concept of "index-free logging" might seem far-fetched. But this is the future architecture of log management solutions. Customers need more out of their logging solutions and using an index-free approach will enable them to get faster results at a fraction of the cost of traditional approaches continuing to use indexing.

It's important to recognize that the act of indexing itself is not all bad. There are many situations where indexing makes sense, but they are not ideal for log monitoring. Indexing is typically ideal in environments where there isn't a big load on ingest but a big load on searching large volumes of data. For log monitoring it's just the opposite. To handle logs, users need to be able to handle large loads of ingest and more occasional searches once incidents happen and that would typically be limited for a certain time period. Alternatives now exist that are faster, less expensive and more effective. Enterprises that need log management (and who doesn't nowadays) have to ask themselves, "Why aren't we implementing modern approaches to log management issues?"

Traditional Indexing Approach

Indexing has been an accepted approach to collecting and analyzing log data for well over a decade and many existing solutions are built on this traditional approach. But its time has passed. There are several things to understand about indexing. First, it is not free. It may appear to be since virtually every log-management vendor provides it, but there are tangible costs across multiple dimensions:

The act of indexing drives hardware costs, both on CPU and memory, to handle index processing at ingest and on storage footprint since indexes, in many cases, are larger than the log or event itself. In many cases, it will be too costly to monitor in real-time, which becomes critical if latency is too high.

Indexing is Less flexible when you need to explore or analyze your data since the search capabilities are dependent on the way you indexed it. This will limit what you can ask for or, worse even, force you to re-index the database. For large installations, re-indexing is a time-consuming and critical process that becomes necessary to do when applications are down or there is a security breach being investigated.

With the cost of the above issues, many companies spend more time on filtering data and deciding which logs to monitor rather than they actually spending resources on understanding the data they ingest to their log monitoring platform.

Perhaps the highest cost of all is the price paid when a critical piece of information in logs is overlooked, either because it was not found in time (or at all), or because a critical event was never logged in the first place (because the ideal goal of logging everything is perceived as too expensive, partially because indexing drives up costs).

Indexing was introduced for very good reasons and has solved a number of problems. But incremental improvements over the years in legacy log-management approaches pale in comparison to the results that can be achieved with index-free logging.

Indexing is better applied to applications other than log management, such as whole-text search, databases and Web searches. In legacy approaches to log management, indexing was used because it was “close enough” and worked sufficiently. But technology has moved on, modern, purpose-build technologies are available that are tailor-made specifically for log management, and thus address both concerns – cost and results.

A New Alternative to Indexing

Proponents of indexing will often claim that it’s essential; without indexing, they say, there would be no way to make sense out of the mountains of log data collected. That’s true. But as we have discussed, indexing is both costly and, in many instances, can compound the problem to be solved.

What, then, is the alternative? A better way of indexing? Emphatically, no. The answer is an entirely different approach that doesn’t involve indexing at all. This new wave of index-free log management incorporates three key approaches:

- reduce the amount of data you have to manage;
- reduce the amount of data you analyze; and
- trade off a slight decrease in analytics on historical data for much faster ingest, larger flexibility and better efficiency on hardware usage.

Let’s look at each of these in more detail.

The biggest barrier to effective log management is often the sheer amount of data to be managed. The more data, the longer it takes to process, and the more storage capacity required. So, if one wants to improve the log management process, a logical step is to reduce the amount of data to be handled. By removing the indexes, we drastically reduce the amount of data we need to handle. Adding on advanced compression technologies we can achieve reduction ratios as high as 30:1.

The second aspect of this modern approach is to analyze a smaller portion of the data. How is this done without increasing the likelihood of missing something? The answer is intelligent filtering. Filtering starts with the knowledge that large volumes of log data will not be of relevance to a specific query (for example, all logs outside a certain time parameter) while other sets of data have a higher chance of relevance (like starting in the "J" section of the phone book to find a person named Jones).

The key to filtering is that *all* logs are ingested up-front *without indexing*, which vastly speeds up the process. Then, at the point of extraction, filters are applied to both decrease the volume of logs analyzed and narrow the keys applied to the search. When a query is executed against historic information, filtering finds the hosts that have data in the desired time frame and scan that data looking for items that match the criteria each and every time. This is parallelized using a simple map/reduce approach.

Intelligent filtering of log data is a breakthrough concept even among modern log management solutions.

And by forgoing indexing, CPU cycles can be applied to another high-value step: real-time analysis of log data as it is being ingested. This can be extremely valuable in applications such as computer security, where an attack might be identified in progress rather than investigated after the fact.

The benefits of filtering are threefold. First, machine and human resources are not wasted on a process (indexing) that is both expensive and unnecessarily time-consuming. Second, results more closely fit the unique requirements of log management. And finally, the complete process – from ingest to insight – is completed in far less time and at a greatly reduced cost.

Summary

Index-free logging is leading the next wave of log management with solutions that eliminate the multi-dimensional limitations of indexing. As a result, it is now possible for enterprises to collect and analyze *all* log data while simultaneously reducing costs, delivering faster results and expanding the number of users who can produce new insights across a wide spectrum of security, business operations, DevOps and customer experience use cases.

The era of index-free logging is here. Join the movement #IndexFreeLogging.