SPONSORED CONTENT

AWS Insider.net

5 TIPS FOR CREATING A MASSIVELY SCALABLE LOG ANALYTICS SOLUTION FROM YOUR AWS S3 STORAGE

Expert tips, best practices, and insights from the recent webcast, "Data Lakes with a Purpose: How To Turn Your S3 into a Log Analytics Data Lake."

Your experts:

- Thomas Hazel, CTO and Founder of ChaosSearch
- Corey Quinn, Chief Cloud Economist at The Duckbill Group

CoreDESIGN / Shutterstock.com

Tip #1: Storing Your Data in Amazon S3 is More Cost-Effective

"Even after the recent price cuts, per gigabyte of disk volume is eight cents per gigabyte per month, and it's not particularly durable. S3 is 2.3 cents per gigabyte now, and that is a ridiculous number of nines of durability. There's no cross-AZ [availability zone] charges…. it's where everyone's data already lives. So why pull all that out in order to wind up running analysis on it? It effectively cuts against the grain of the way data exists in its native state in a cloud environment." —*C.Q.*

Tip #2: Overprovision! You Won't Be Storing Data Only Once

"We absolutely need to keep those ancient logs from 2012 full of monitoring checks. We don't want to have everything living in one place. So we need to store it both online and in multiple locations and have multiple copies. So for every gigabyte you wind up wanting to store, you end up inherently having [to] overprovision because that's how disks work, and store it multiple times." —*C.Q.*

Tip #3: Maintaining a DIY Solution is Time-Consuming and Unreliable

"When the ELK stack really got popular about eight years ago, people started rolling their own, where they take their data, bring in Logstash to send this data to the Elasticsearch database, and then leverage Kibana visualization to do any type of analysis. Now, it did reduce the cost per se, but they added all the pain and expense of managing these clusters. And you can imagine the horror stories of clusters falling over and systems being down—it's all well-documented."

"When you're building out your log system, particularly with ELK, there's a process of sizing, the infrastructure, the design, the cost, the configuration

of these workloads.... All this takes time, typically weeks, if not months, to really build out. Then imagine: now, all of a sudden, you have a new requirement, a new log, and you have to rebuild the pipeline to support that scale. That complexity is going to have something fail. So our viewpoint is there's got to be a better way...a better way to reduce complexity, time and cost. Imagine if you didn't have to move this data around at all!" —*T.H.*

> We created new index technology [at ChaosSearch] around a new architecture, a distributed fabric that allows us to get high performance on the first read–and every read after that." *–Thomas Hazel, CTO and Founder of ChaosSearch*

Tip #4: Find Solutions that Keep Your Data in S3 and Give You Top Performance

"We created new index technology [at ChaosSearch] around a new architecture, a distributed fabric that allows us to get high performance on the first read—and every read after that."

"But we didn't stop there.... We also built in things that you'd want in a data lake, like data discovery. What's in your bucket? Data cataloging. So you can search and analyze what's in your bucket, and have the ability to automatically index data as it's coming in [and] it will manage the schema for you over time."

"Other aspects of our product, like the data refinery, create different lenses into the data, based on what you want to do with it. Maybe you want to have this ELB log to be given role-based access to the IT department over here, or maybe the flow logs over there. The idea is that this refinery allows you to easily—through a wizard—create different lenses, different index patterns." —*T.H.*

CoreDESIGN / Shutterstock.com

Tip #5: Leverage Technology You Can Spin Up Quickly with Your Data

"In our free trial, you log in, set up a role ARN policy to access your S3 read-only, and a location for us to index the data and store those indices into your account. It's literally a click, click, click with a cloud formation-type template. It's really five minutes and you're up and running—the great thing about cloud storage is that your data is already there. Imagine if you had a petabyte of data and you had to move it from out of your cloud storage into another service. No, that's time. That's cost. With us, you won't have any of that."

"The key thing to remember is that you own this data. We're just a distributed compute fabric that does the intelligent work of indexing, and we put those indices into your account. And then we have a query execution on that data. It's pretty exciting!" —*T.H.*

SPONSORED BY:
CHAOSSEARCH
Find out more:
https://www.chaossearch.io/