SFD6

Coho Data

@andywarfield - co-founder and CTO

Coho Datastream Overview

When they started 3 years ago, they thought they were going to build something else, but then they looked at flash

sw problem - taking advantage of the perf capabilities of things like Fusion-IO cards

sw stack that qualifies and ships on commodity hw
nfs v3 server for VMware

bottom half effectively virtualises flash [photo]

Data profiles - effectively a roadmap. Right now they're doing NFS and a few other things.

what / who / why
Flash is dropping in price, doubling in density every 12 - 18 months

PCIe Flash is unreliable, so let's aggregate. And then create a volume mgr, and snapshots. And there's a lot of sw in play.

What Coho does though, is only dictate multi tenants and control the Flash Translation Layer

About to ship 2.0
Adds remote asynchronous site to site replication -
http://www.cohodata.com/blog/2014/11/05/scale-out-storage-site-site-replication/
Some API additions

Product UI demo

API browser is very cool - fast way to learn about how to use the API

*Your storage data path is a network datapath

Media is stratifying (more and faster than ever before)
- 1 Mx difference in slowest to fastest latency
- 400x difference in slowest to fastest cost
- <sarcasm> NVDIMM is a great deal

Network pain becomes storage pain a few years later

*The inevitability of hierarchical storage

Tiering and caching need a lot of metadata to do the job properly
This can be expensive though

http://goo.gl/eV9Ffx

Two applications of counter stacks (UI, OnStream, and noisy neighbours)

Used to show age of data in the array
"Simulate purchasing more storage a week ago"

http://vsamurai.com/coho-data-intelligent-analytics/

*Careful, calculating convergence

"Lazy isn't bad, lazy is just … lazy"

Hyper-converged makes sense at small scale

You buy the storage system for a peak demand, VMs add to the peaks.
There's still an opportunity to use the system with good resource mgmt

2 PoCs
"Sprawl and withdrawal"
Putting Cloudera CDH5 on the storage and integrating with it
Deployed as Dockerized tenants on VXLAN
"preserving topology awareness in a virtual environment is important for job scheduling (e.g. rack locality)"

Ingest -> Filter / Transform -> Analyze -> Act (many customers have this pattern)