SFD8 - Day 3 - Nimble

*Dan Leary, VP of products, solutions and alliances - Nimble storage adaptive flash platform overview
Overview
HQ in San Jose
>6000 Customers
>1000 partners

Operations in 30+ countries

1980s - DAS, 1995+ - Networked Storage
Today - Flash vs Disk, Cloud vs on-premises storage mgmt

Opportunity for a ground-up re-design of storage

Storage tiering - industry approach inherently complex

Adaptive flash - eliminating silos
- single cloud-based management system
- single storage architecture for all of your enterprise apps
- service levels
- comprehensive business continuity

CASL Architecture (pronounced *Castle*) - significantly better performance/$ and capacity/$
Scale to fit - non-disruptive flexible scaling to massive scale
Integrated protection - rapid backup and recovery
Ease of operations - proactive support and operational simplicity

NimbleOS - [refer to photo for notes]

SmartStack Converged Infrastructure Solutions - http://www.nimblestorage.com/solutions/smartstack/
An example of an Oracle solution can be found here: http://info.nimblestorage.com/rs/nimblestorage/images/nimblestorage-sb-smartstack-oracle.pdf

*Rod Bagg, VP Customer Support - InfoSight Overview
“Operational Intelligence”
In a connected world why can’t vendors proactively monitor customer deployed systems?
With modern data analytic tools can vendors predict snd prevent problems before they occur?

Design Philosophy
Be intuitive - present use cases and not just data
Be prescriptive - provide specific recommendations for immediate action
Be predictive - estimate future needs based on current and past learning

InfoSight Engine = Collect Sensors -> Data Analytics -> Visualise -> Learn and Automate
- Time to value
- Predictive analytics
- Full stack performance monitoring
- Proactive Support

94% of arrays are reporting back, 95% cases automatically opened by Nimble

Fun Facts
Deep Data
- 1000s of unique sensors recording operational data each second,
- 30-70M data points collected from every array every day,
- >20000000 heartbeats every week
Big Data
- 200B log events
- by-day view of every config element of every array
- lifetime data from day 1
Rich Analytics
- rich install base
- data from 1000s of arrays for 5+ years
- dedicated team of data scientists on support staff
- advanced analytics techniques, e.g.:
— differential equation models of IO flux
— correlation analysis of noisy neighbours
— bootstrapping and multiple regression analytics
— Monte-Carlo (https://en.wikipedia.org/wiki/Monte_Carlo_method) simulations for usage

Protecting your investment
Workload Sizing
- 1000s of system-years of real-world data
- workloads correlated to resource consumption
- exact configuration known before purchase
Predictive Capacity Recommendations
- continuous storage capacity prediction
- plan storage purchase in advance
Scale-to-fit Recommendations
- working-set analysis for cache and cpu optimisation
- optimal cache, scale-up, scale-out recommendations

Ensuring HA > 5 9s
100s of Health checks for proactive wellness
- adherence to best practices
- operational risk analysis
- automated sw fault analysis
Rigorous Deployment Process
- Nimble OS RC releases are always slowed-rolled to low risk environments first
- Pass / Fail GA metrics calculated / assessed daily
Risk Management

- if we know about a problem, it shouldn't happen to you
- Install base analytics for blacklisting and proactive fixes
- dynamic update path on every array ensures safety

"We believe Operational Intelligence will become the most important criteria for the CIO when making strategic vendor decisions"

*Devaki Kulkarni, Sr Product Manager - VMVision Demo
VMVision - http://uploads.nimblestorage.com/wp-content/uploads/2015/07/12132211/nimblestorage-vmvision.pdf
easily correlate VM performance data to physical host, neighbouring VMs and the storage it's sitting on

storage info from Nimble, VM data from vCenter

roadmap - looking at physical servers and Hyper-V

VMVsion portion of InfoSight is opt-in, not automatic

*Mark Cooke, Data Scientist - A peek under the hood of InfoSight
Data Quality
The big problems are the old problems
- how to combine information from different systems? Reporting may happen at different intervals, different levels of precision, have gaps, etc.
- problems can suddenly pop up in input that can have huge effects on output
- these issues are increasingly problematic due to volume, automation and interactions between different systems
- not specific to Nimble, or to storage in general. any data-centric environment has these problems
Solutions
- algorithms and analysis need to be built to handle input problems, with attention to the effects on results and the underlying reasons for the problems
- teach your systems what a number looks like
- focus continually on quality. Explore, and dig into anything that seems off.
- pay attention to the fundamentals

2 Examples - VMware and Faulty Disk Drives

Summary
- data quality is as much a cultural issue as a technical one
- new input problems can always creep in, these need to be monitored for and guarded against
- some problems will be unavoidable, these need to be dealt with in a comprehensive and robust manner
- a quick response to a critical situation requires a data warehouse that has been carefully collected and curated
- look for patterns that indicate normal behaviour
- when problems are found, identifying signatures that can be searched for in the install base

*Tomasz Barszczak, software development architect - building a reliable software foundation
[photo]
Guiding principles
Golden rules
1. Integrity - don't return incorrect data
2. Durability - don't lose data
3. Availability - don't lose access to data

1. Integrity
Ensure what we read matches what we've written - use checksum and self-ID

2. Durability
Drives fail - larger drive sizes - data loss during RAID reconstruction

NimbleOS 1.x - DP RAID
NimbleOS 2.x - Triple parity RAID without performance impact (sequential layout - space efficient parity RAID without performance penalty)
Additional parity is stored within RAID chunk on the same disk (intra drive parity)

3. Availability
5 9s across 6000 production customers

Factors:
Hardware - no SPoF
CASL - HA using active - standby controllers and failover (managed and unmanaged)
InfoSight - monitoring, data analysis, failure prediction and intervention
Processes - tight support-engineering loop

61% customers upgrading via NDU during their business hours

Protect from External Risks
- user error or app error - snapshots (RoW)
- disaster - replication
- data theft - encryption

snapshots
- embedded into data layout and index, without performance penalty, preserves write-time locality
- space efficient: application specific block size, compressed, on disk
- protection for little cost (primary space, controlled by retention)

replication (snapshot replication - async - 2 minutes? a lot of customers doing 5)
- compressed blocks - no processing for compressed replication
- non-chatty protocol tolerates large network latency

encryption
- global or per-volume

- encrypted replication

*Suresh Vasudevan, CEO - Fireside chat - https://www.linkedin.com/pub/suresh-vasudevan/10/381/a50
[photo]

How have things changed?
- Hire well
- Manage your talent retention well