Health monitoring & predictive analytics
To lower the TCO in a datacenter

Christian B Madsen & Andrei Khurshudov
Seagate Technology
christian.b.madsen@seagate.com
Outline

1. The opportunity
2. Our vision & implementation
3. Use cases
4. Summary
The opportunity
What if…

Seagate offered you a technology that could help you

- Improve datacenter efficiency
- Optimize system management
- Reduce potential cost of operation
The problem

- **1 billion hard drives** will be used in cloud datacenters by 2020, highlighting the need to manage drive health at scale

- One total outage per datacenter is statistically expected every year

- 80% of those outages are not completely explained (or linked to root causes)

- $700,000 is the average cost per incident
  - $8,000 is the average cost per minute of an unplanned outage

- **Up to 10% of datacenter accidents are related to storage**

Source: Seagate Strategic Marketing and Research 2013
Better drive management will lower the TCO

Top 4 challenges in drive management

1. **Drive health monitoring**
   - Need reliable key performance indicators to track drive health status

2. **Drive failure prediction**
   - “Ultimately, we want to know when our drives will fail so we can take actions before that happens”

3. **Drive failure diagnostics and management automation**
   - Need to correctly identify and quickly resolve issues
   - Need to prevent false alerts to reduce cost of failure handling

4. **Drive lifespan extension**
   - Need to know how to optimize operating environment for better reliability
   - Need to reuse partially good drives (should be possible with in-drive diagnostic, IDD)
Our vision & implementation
Our vision

Monitoring, analytics, prediction and control – “The internet of things”

• Drive-centric health monitoring
• Analytics and predictive models
• Closed-loop automation
Functional diagram

- Monitoring, intelligent decisions and automation

Closed-loop automation

- Monitor
  - Compliance (threshold)
    - No
    - Exception (alert)
    - Recommended action
      - Choose action
    - Resolution

Example

- Monitor
  - Drive health
    - Not passing
    - Drive predicted to fail
      - Reset or turn off drive
        - Choose action
      - Turn off drive

Automation
Choosing action from recommended action can be automated by tying it to the specific application or saving choices.
Implementation

Architecture overview

Cloud Gazer™ Elements

- Storage Server
  - Drives
- Storage Server
  - Drives

Agents

Real-time metric aggregation

Data pool (10,000s of drives)

Server

- Clusters
- Servers
- Drives

Query data
Check thresholds
Manage drives

Cloud Gazer™ Dashboard

REST API Calls

Storage Software

Implementation Architecture overview
Use cases
Compliance (thresholds)

- Degradation and performance warnings
Compliance (thresholds)

- Degradation and performance warnings

**Overload detection**
Detecting and reporting when drive load exceeds design limits

---

**Projected Failure Rate Compared to Default (%)**

- Danger zone
- Overstressed region
- Datacenter failure rate

- **Overloaded drives**: 19 overloads
- Inspect the workload statistics page, heat map or list view to see if the trend is persistent and affects reliability

---

*Health Tiles (2 minutes ago)*

- **Failing Drives**: 0% (0 Drives / 402 Drives)
- **Drive Capacity Utilization**: 0% (0 Drives / 402 Drives)
- **Overheated Drives**: 72% (290 Drives / 402 Drives)
- **Overloaded Drives**: 5% (19 Drives / 402 Drives)
Compliance (thresholds)

- Degradation and performance warnings

**Recommended action**
How to increase drive reliability

**Overload detection**
Detecting and reporting when drive load exceeds design limits
Compliance (thresholds)

- Degradation and performance warnings

**Recommended action**
How to increase drive reliability

**Overload detection**
Detecting and reporting when drive load exceeds design limits

**HDD population failure rate**
Measuring stress and estimating failure acceleration of the disk drive population in real time. Relies on the proprietary failure prediction algorithms

---

**Projected Failure Rate Compared to Default (%)**

- **Overstressed region**: Datacenter failure rate
- **Danger zone**: Overload detection

**Health Tiles (2 minutes ago)**

- **19 overloaded drives**: Inspect the workload statistics page, heat map or list view to see if the trend is persistent and affects reliability.

- **Failing Drives**: 0% 0 Drives / 402 Drives
- **Drive Capacity Utilization**: 0% 0 Drives / 402 Drives
- **Overheated Drives**: 72% 290 Drives / 402 Drives
- **Overloaded Drives**: 5% 19 Drives / 402 Drives
Compliance (thresholds)

Degradation and performance warnings

**Failure detection**
Warning about expected drive failures. Relies on the proprietary failure prediction algorithms that use unsupervised machine learning techniques. Expected average failure prediction time window is from 9 days to 12 days.

**Overload detection**
Detecting and reporting when drive load exceeds design limits.

**Recommended action**
How to increase drive reliability.

**HDD population failure rate**
Measuring stress and estimating failure acceleration of the disk drive population in real time. Relies on the proprietary failure prediction algorithms.

**Projected Failure Rate Compared to Default (%)**
- Projected
- Default
- Overstressed region

**Datacenter failure rate**

Failing Drives: 0%
Drive Capacity Utilization: 0%
Overheated Drives: 72%
Overloaded Drives: 5%

19 overloaded drives: Inspect the workload statistics page, heat map or list view to see if the trend is persistent and affects reliability.

Health Tiles (2 minutes ago)
- Watching: 1 clusters, 1 servers, 3 drives
- Failures: 0 Drives / 402 Drives
- Overheats: 290 Drives / 402 Drives
- Overloads: 19 Drives / 402 Drives

2015 Data Storage Innovation Conference. © Seagate Technology. All Rights Reserved.
Workload optimization

Drive visibility tools to improve workload balancing

**Before** Load balancing issues

Workload predominantly hitting one server

**After** Workload distributed over servers and time

Workload peaked on Sunday
Unsupervised machine learning and failure prediction

- No interaction between drive set, no prior knowledge

For now, an average failure prediction window is on the order of 9 to 12 days
Failure prediction accuracy ranges from 55% to 90%
Prediction and follow up actions

- Heat map indicates drives at risk and you can issue drive tests (DST, IDD, …) to resolve or corroborate.

Systematic failure predicted: 3 out of 5 drives predicted to fail sit in end location of servers.
Find failure triggers

Root cause tools including a drive temperature heat map can help you triage the cause of your drive issues.

- Systematic failure predicted: 3 out of 5 drives predicted to fail sit in end location of servers.

- Common factors for drives in the end position is a cooler temperature. Therefore increasing the server temperature may reduce the (dominant) failure mechanism and increase drive reliability.
Failure prediction lead time

We can predict drives will fail on average 9 -10 days before the failure

Case study 1, we predicted most drives (118 drives) to fail 12 days prior to failure

Case study 2, we predicted 5 drives to failed 23 days prior to failure, 2 drives 22 days prior to failure,… 2 drives just one day in advance

Currently catch 55-90% of failures ahead of time
Summary
Why Cloud Gazer™?

- Truly drive-centric management tool for the cloud
- Most efficient tool for extracting drive health information using Seagate IP
  - Nobody knows drives better than us
  - Freeware utilities are frequently wrong
- Runs on any Linux system with little overhead (<1%)
  - Windows is next
- Data can be collected, monitored and analyzed locally or in the Cloud
- ReSTful API to interact with other software
- New analytics, prediction, AI, and control capabilities are added continually
- Drive repair will be possible with in-drive diagnostic
- Enclosure control will be possible by summer 2015

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Seagate's CloudGazer™</th>
<th>Competition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw Error Rate</td>
<td>Yes</td>
<td>Partial</td>
</tr>
<tr>
<td>Spin Up Time</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Start/Stop Count</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Retried Sectors Count</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Seek Error Rate</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Power On Hours</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Power Cycle Count</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Reported Uncorrectable</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Command Timeout</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>High Fly Wristas</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Emergency Retact Count</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Load/Unload Count</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Temperature</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>ECC Count</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Load Cycle Count</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Pending Sparing Count</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Head Flight Hours</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Lifetime Wristas</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Lifetime Reads</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Flags 1</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>RV Abs Mian</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Motor Power</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Critical Event Errors</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

*Seagate drives*
Questions?