

---

# DATA MANAGEMENT FOR UNSTRUCTURED DATA

Olivier Rivard  
Software developer  
[olivier@ordigraphe.com](mailto:olivier@ordigraphe.com)



DataIntell

# ORDIGRAPHE

- 35 years of experience
- Integration in workflows
- External view of challenges
- Post-production and VFX in Montreal



ORDIGRAPHE

# DATA MANAGEMENT SOLUTION



GLOBAL VIEW OF  
THE DATA



HOW MUCH  
DATA HAS NOT  
BEEN ACCESSED  
IN HOW LONG



HOW MANY FILES  
ARE DUPLICATE



HOW MUCH  
DATA WAS ADDED  
IN THE LAST  
WEEK



WHERE ARE THE  
FILES LOCATED IN  
THE  
INFRASTRUCTURE

# VENDORS SOLUTION

- CPU Load
- Disk health check
- File system health check
- IOPS

## Host Status Summary

Up	Down	Unreachable	Pending
53	11	3	0
Unhandled		Problems	All
64		64	117

Last Updated: 2017-10-05 16:06:57

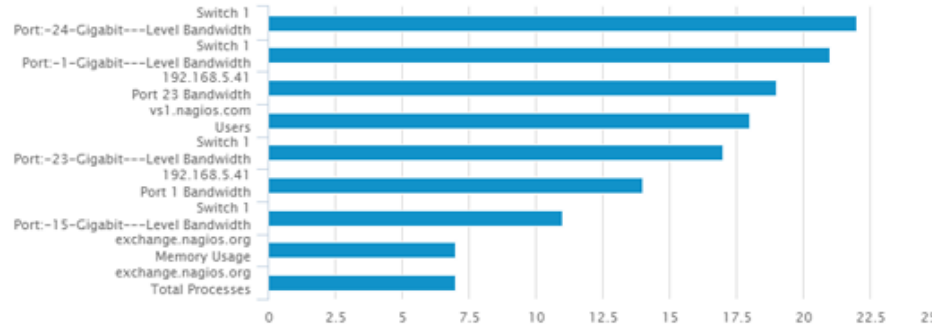
## Service Status Summary

Ok	Warning	Unknown	Critical	Pending
226	12	84	271	2
Unhandled		Problems	All	
366		367	595	

Last Updated: 2017-10-05 16:06:57

## My Graph

Top Alert Producers Last 24 Hours



## Metrics Overview

### Disk Usage

Host	Service	% Utilization	Details
localhost	Root Partition	78.67%	DISK WARNING - free space: / 1207 MB (17% inode=68%):
vs1.nagios.com	/ Disk Usage	37.30%	DISK OK - free space: / 117214 MB (61% inode=99%):
exchange.nagios.org	/ Disk Usage	13.22%	DISK OK - free space: / 68067 MB (86% inode=97%):

Last Updated: 2017-10-05 16:06:58

## Hostgroup Status Summary

### Status Summary For All Host Groups

Host Group	Hosts	Services
All EMC SAN Hosts (all_emc_hosts)	1 Up	4 Ok 1 Critical
Firewalls (firewalls)	1 Up	1 Ok
Host Deadpool (host-deadpool)	3 Up 1 Down 1 Unreachable	8 Ok 7 Critical
Linux Servers (linux-servers)	5 Up	52 Ok 3 Warning 9 Unknown 6 Critical
new group (new group)	8 Up 1 Down 2 Unreachable	58 Ok 3 Warning 9 Unknown 11 Critical
Printers (printers)	1 Up 2 Unreachable	2 Ok 3 Critical
Websites (websites)	3 Up	20 Ok 2 Warning 2 Critical
Windows Servers (windows-servers)	1 Down	6 Critical

Last Updated: 2017-10-05 16:06:57

# MEDIA ASSET MANAGEMENT

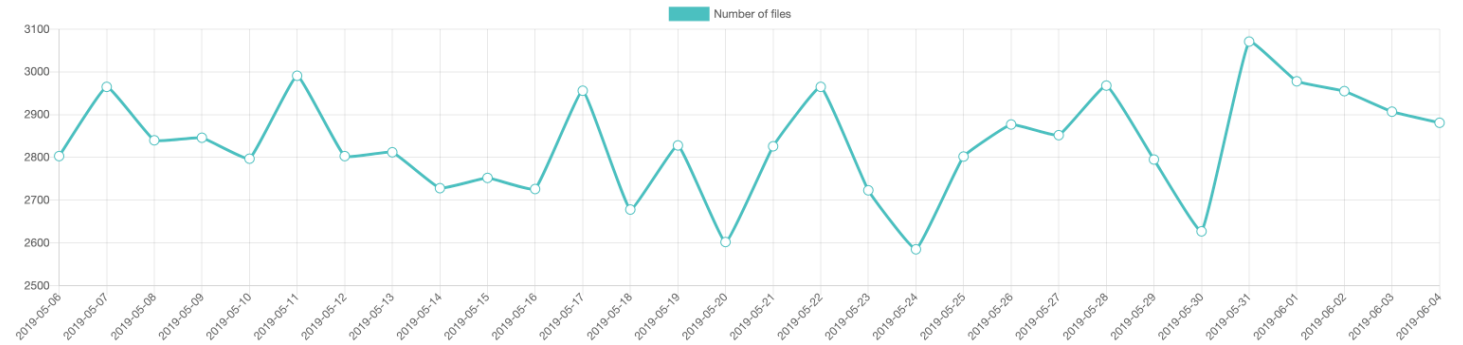
- Automation on asset
- Transcoding
- Proxy generation
- Structured data
- Good file management from ingest to archive
- It's not managing all the infrastructure



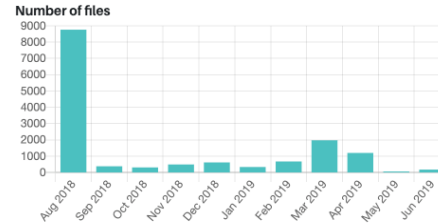
# DATA MANAGEMENT SOLUTION

- Agnostic of the infrastructure
- Global view of all data
- Unstructured data
- Analytics on data
- Trending on data

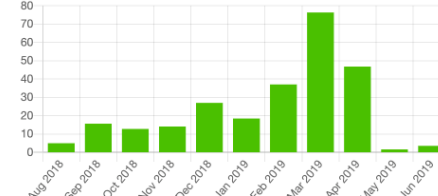
Number of duplicate files



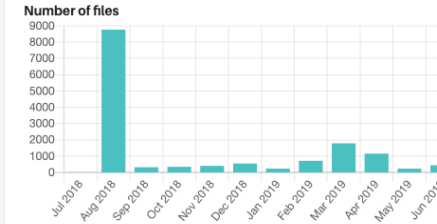
Creation date by month



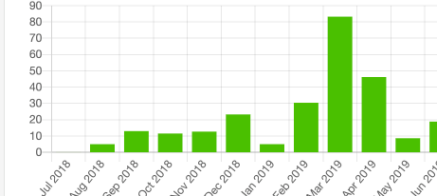
Total size of files



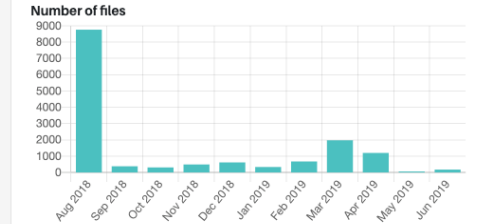
Last access date by month



Total size of files



Last modified date by month



Total size of files



---

DEMO



DataIntell

# PROJECTS

- Global view across multiple storages
- Data federation
- Analytics of the data
- Space allocation
- Actual needs of a project



## SPEED OF THE SCAN

- Scanning millions or billions of files
- Multiple servers for scan
- Indexing delta changes
- Integration with Vendors API



## USEFUL INFORMATION AT THE RIGHT TIME

- Different information depending on the user type
- Show different type of data at different moment in time
- Show suggestion with alerts



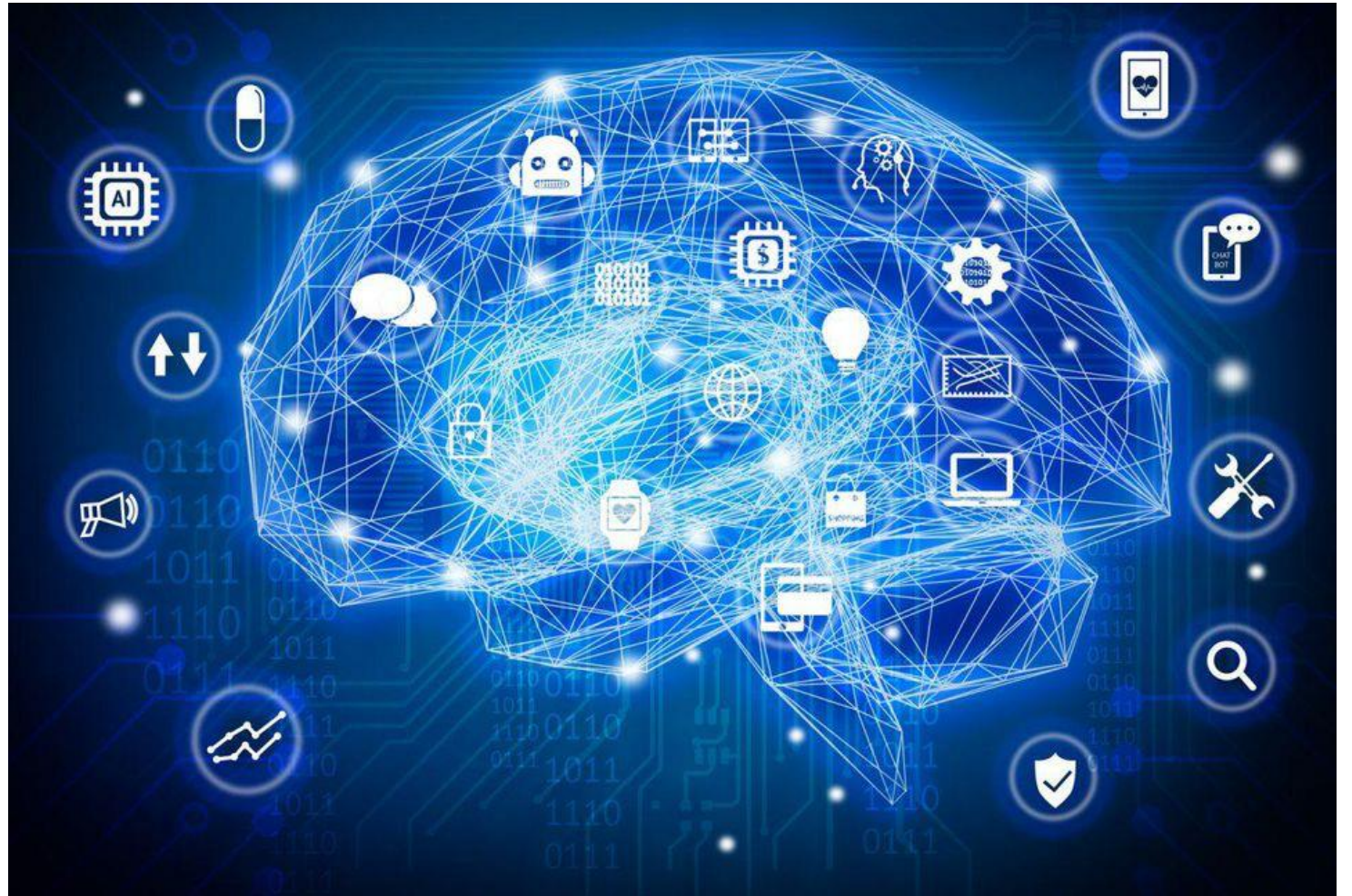
# INTEGRATIONS

- Storage vendors
- Backup and archive solutions
- Cloud providers
- Media asset management



# MACHINE LEARNING

- Predict if a file will be deleted in the next month
- Moving data based on usage
- Evaluate the needs for new storage
- Scan most used folders more frequently
- Best data to show to users



# DATA MANAGEMENT SOLUTION

- More proactive than reactive
- Plan infrastructure needs
- Information tool for emergency
- From management to IT
- Quickly find files

---

QUESTIONS?



DataIntell