Using Spectrum Scale in the IoT world

Ray Coetzee - Pre-Sales Architect
GPFSUG presentation - May 2016

SEAGATE

# Internet of Things

From Wikipedia, the free encyclopedia

The **Internet of Things (IoT**) is the network of physical objects—devices, vehicles, buildings and other items embedded with electronics, software, sensors, and network connectivity that enables these objects to collect and exchange data. The IoT allows objects to be sensed and controlled remotely across existing network infrastructure, creating opportunities for more direct integration of the physical world into computer-based systems, and resulting in improved efficiency, accuracy and economic benefit when IoT is augmented with sensors and actuators, the technology becomes an instance of the more general class of cyber-physical systems, which also encompasses technologies such as smart grids, smart homes, intelligent transportation and smart cities.
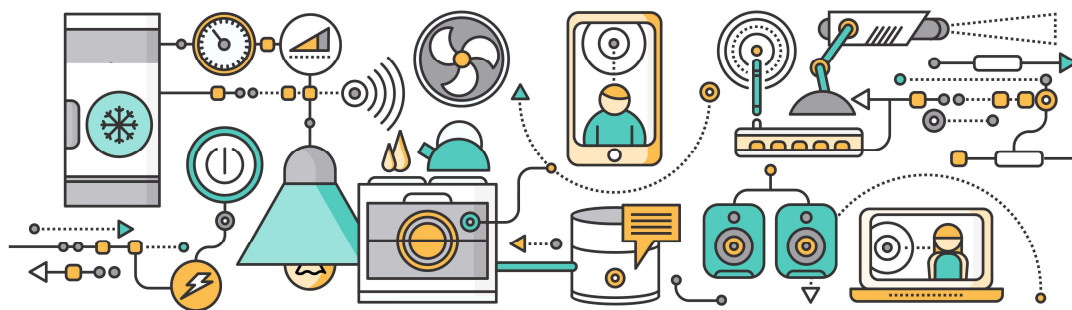
- Petroleum Exploration
- Digital Surveillance
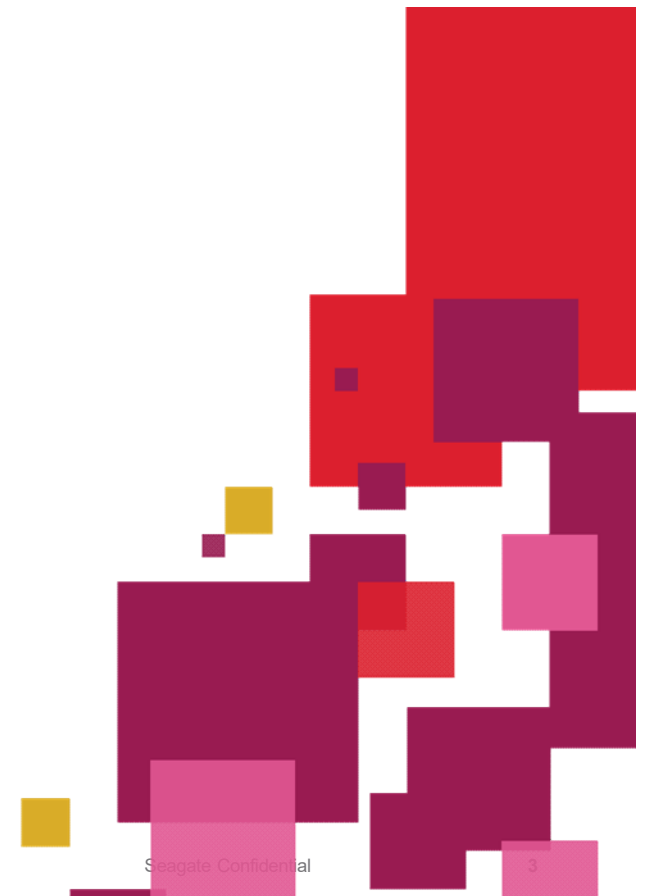
- Petroleum Exploration
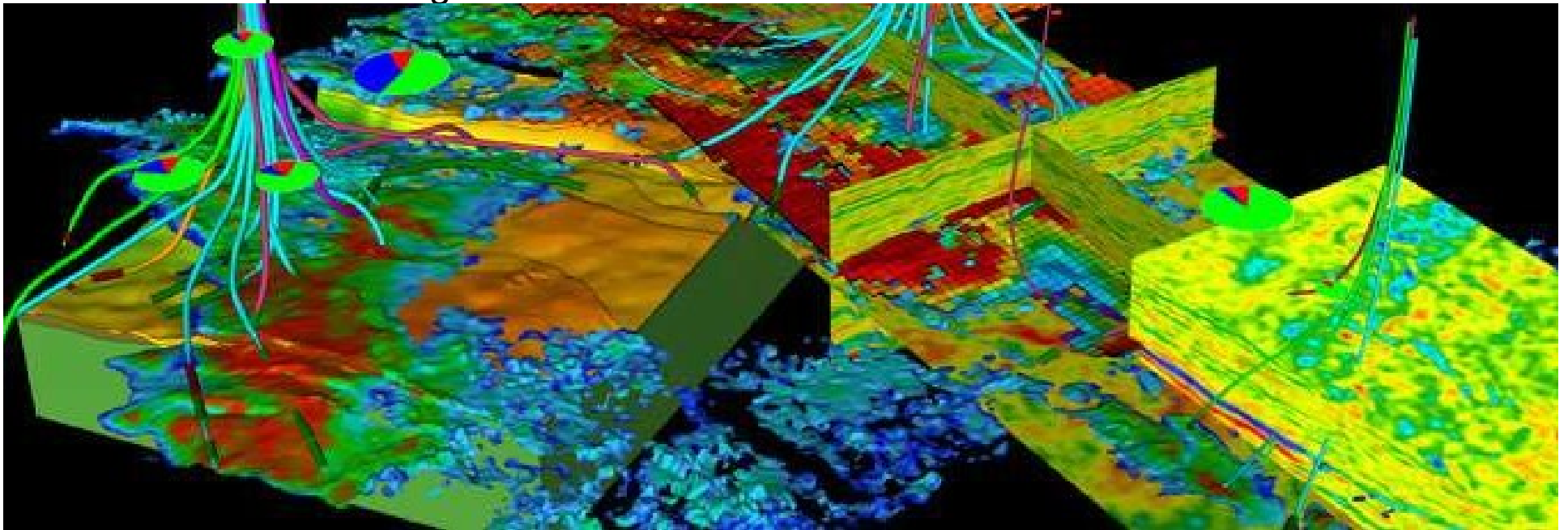- Digital Surveillance
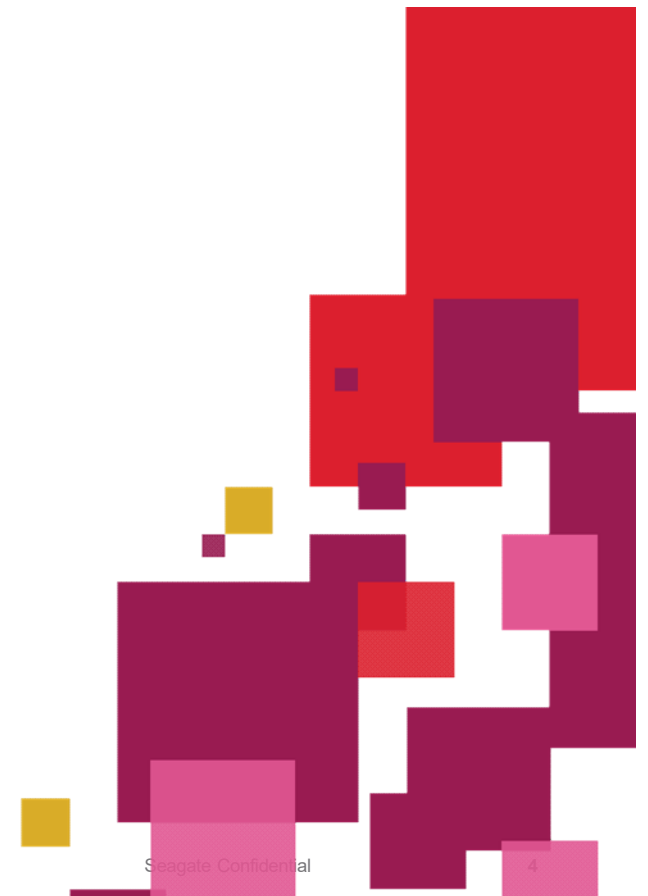
Petroleum exploration

Exploration starts with collecting seismic field surveys
By the miracle of HPC, millions of data points are combined to form a
Seismic cross section
Further processing leads to a 3D/4D model of the reservoir

Petroleum exploration

Challenges of Petroleum Engineering

- Petroleum exploration has a significantly complex processing pipeline which consists of multiple stages of signal processing, 3D modeling, yield analysis & projections of each potential and existing well.
- Each of the stages has a different IO pattern for example N-N, N-1, large sequential reads, small strided reads and each feeds into the next stage
- Various IO & file sizes
- Large amount of results need to be retained for historical comparisons & projections

Where is the IoT?

Petroleum exploration

The Horizontal drill bit
- Contains pressure, temperature, positional sensors & directional actuators to control descent
- Real-time telemetry is streamed to the data center
- Obstacle avoidance algorithms are used in conjunction with the 3D reservoir model to steer the drill around hard rock formation, gas pockets or to avoid contaminating ground water
- Smaller reservoirs can be linked by a single well
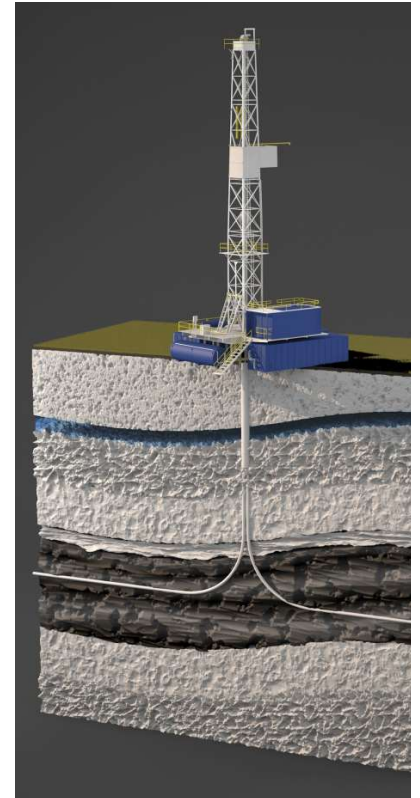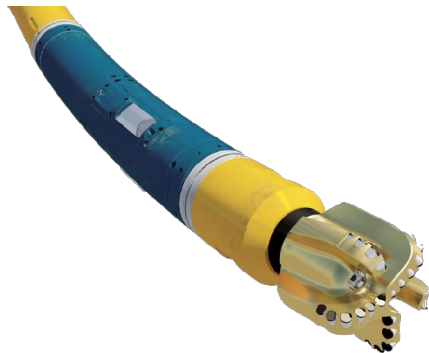
Petroleum exploration

Lessons learned for Spectrum Scale:
- IO profiling of applications can give you better performance improvements than brute forcing a bottle neck with larger pagepools or cache.
- Segregate access patterns with multiple filesystem & multi-clustering.
- Don't assume customer workloads have ever been optimized i.e. constant stat queries on files flooding metadata or massively large flat folder structures.
- Few network engineers can troubleshoot a network like Spectrum Scale
- **Educate users/customers new to Spectrum Scale before starting the installation.**
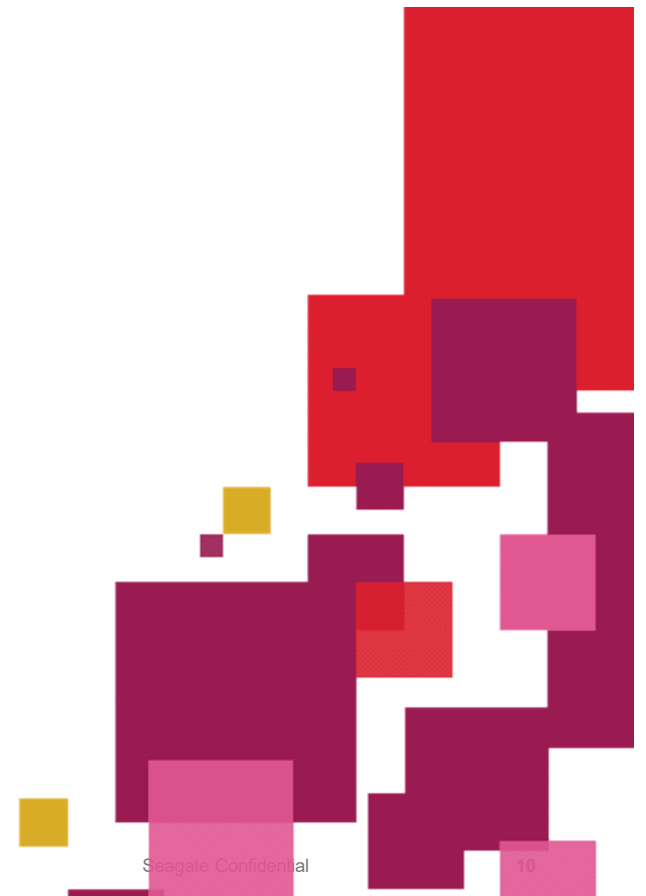
- **Petroleum Exploration**
- **Digital Surveillance**

Digital surveillance

Public surveillance
License plate recognition
Commercial surveillance
WAN
Lawful Interception
Social Media feeds
WAN
Surveillance Drones

Virtualized recording infrastructure natively mounting Spectrum Scale filesystems over Data Network

Virtual Recording Servers
VMware vSphere

Virtual Desktops
VMware vSphere

Analytic Servers

10/40GbE Data Network

Governmental Departments

Police Departments

Ministry of Interior

Large scale projects contain:
- 1000's of CCTV streams
- Variable IO sizes
- Audio sensors
- Sonar
- Chemical detectors
- Drones
- Social media feeds
- Multiple forms of analytics

Digital surveillance

Challenges of Digital Surveillance
- Packet loss needs to be reduced to a absolute minimum due to predominantly using UDP traffic.
- QoS must be maintained during the most demanding failure scenarios
- Low latency access is crucial
- 30-100 threads accessing the filesystem per recording server
- CCTV suffers from severe fragmentation due to write-delete-write cycles
- Small application IO sizes ranging from 4byte to 64Kbyte
- Solution must maintain 24x7x365 availability
- Write/Read ratio is normally 90/10 but can become 90/10 Read/Write when doing analytics while still maintaining the same Write MB/s.
- 100's of Millions of files
- **Applications are dependent on Microsoft Windows which has TCP/IP limitations**

Digital surveillance

Why to use Spectrum Scale:
- The pagepool allows buffering of IO during failover events
- Support for thousands of open threads
- Native Microsoft Windows support
- Online background defragmentation
- Efficient capacity utilization with sub-blocks
- Efficient network load balancing
- Fast playback of multiple streams
- QoS abilities
- Provides CIFS/NFS exports (Protocol Servers)

Digital surveillance

Tips for a successful project at scale:

- A end-to-end POC is a must
- A MOS (Mean Opinion Score) network test must be done
- Do not assume Ethernet hardware are compatible by default (Cisco vs. Mellanox vs. "Force10")
- Create a standardized configuration for the entire solution
- Different brands of NIC/CNA (Broadcom, Intel, Brocade) failover and flow control differently

Digital surveillance

Lessons learned for Spectrum Scale:
- LACP Bonded network interfaces are a must for fast link failover
- Rx/Tx Flow control must be enabled through the entire Data Network
- The pagepool can be sized sufficiently to mask the recovery time of a hardware failure i.e. switch, network link, NSD server, controller
- Minimize the pagepool size on the NSD servers
- Be careful with using Tiebreaker disks when trying to expedite failure detection & failover times as Deadman timers can be tripped very easily.
- Default tuning parameters are "good enough" in some situations
- Take the time to automate and standardize the Spectrum Scale for Windows client deployment.
- vMotion of a Windows Spectrum Scale client works flawlessly while streaming IO.

THANK YOU

SEAGATE