Spectrum Scale Expert Talks

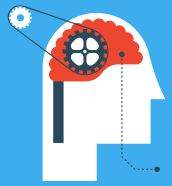
Episode 10:

Data Accelerator for AI and Analytics (DAAA)

IBM Spectrum Scale

Show notes:
www.spectrumscaleug.org/experttalks

Join our conversation:
www.spectrumscaleug.org/join



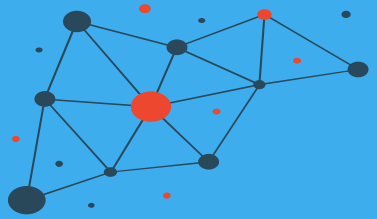
SSUG::Digital

Welcome to digital events!

IBM Spectrum Scale

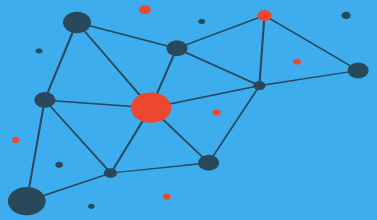
Show notes:
www.spectrumscaleug.org/experttalks

Join our conversation:
www.spectrumscaleug.org/join



About the user group

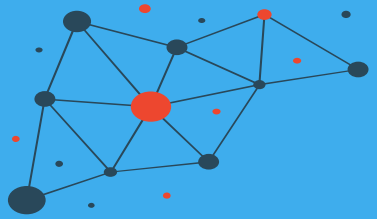
- Independent, work with IBM to develop events
- Not a replacement for PMR!
- Email and Slack community
- https://www.spectrumscaleug.org/join



We are …

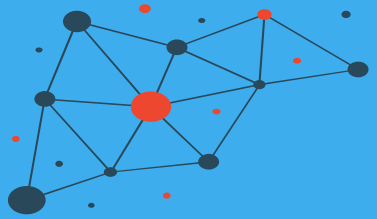
- Simon Thompson (UK)
- Kristy Kallback-Rose (USA)
- Bob Oesterlin (USA)
- Bill Anderson (USA)
- Chris Schipalius (Australia)

IBM CHAMPION



Check https://www.spectrumscaleug.org/experttalks for charts, show notes and upcoming talks

- Past talks:
  - 001: What is new in Spectrum Scale 5.0.5?
  - 002: Best practices for building a stretched cluster
  - 003: Strategy update
  - 004: Update on performance enhancements in Spectrum Scale (file create, MMAP, direct IO, ESS 5000)
  - 005: Update on functional enhancements in Spectrum Scale (inode management, vCPU scaling, NUMA considerations)
  - 006: Persistent Storage for Kubernetes and OpenShift environments
  - 007: Manage the lifecycle of your files using the policy engine
  - 008: Multi-node scaling of AI workloads using Nvidia DGX, OpenShift and Spectrum Scale
- Today:
  - Nov 16: Continental: Deep Thought – An AI Project for Autonomous Driving Development
  - Nov 16: Data Accelerator for AI and Analytics (DAAA)
- Next:
  - Nov 18: User Meeting at SC20 (Session 2) – What is new in Spectrum Scale 5.1? https://www.spectrumscaleug.org/event/sc20-meeting-session-2-what-is-new-in-spectrum-scale-5-1/



Speakers

- Simon Lorenz (IBM)
- Gero Schmidt Q&A (IBM)

Disclaimer

IBM's statements regarding its plans, directions, and intent are subject to change or withdrawal without notice at IBM's sole discretion. Information regarding potential future products is intended to outline our general product direction and it should not be relied on in making a purchasing decision. The information mentioned regarding potential future products is not a commitment, promise, or legal obligation to deliver any material, code, or functionality. The development, release, and timing of any future features or functionality described for our products remains at our sole discretion.

IBM reserves the right to change product specifications and offerings at any time without notice. This publication could include technical inaccuracies or typographical errors. References herein to IBM products and services do not imply that IBM intends to make them available in all countries.

IBM Spectrum Scale:
Data Accelerator for AI and Analytics (DAAA)

Simon Lorenz
Spectrum Scale BDA Architect

Gero Schmidt
Spectrum Scale BDA Software Engineer

Outline

- Customer Demands / Requirements
- Using IBM Spectrum Scale built-in Caching technology
- High Performance Tier (HPT)
- Data Accelerator for AI and Analytics (DAAA)
- Demo:
  AI Example Use Case: Autonomous Driving

Customer Demands / Requirements

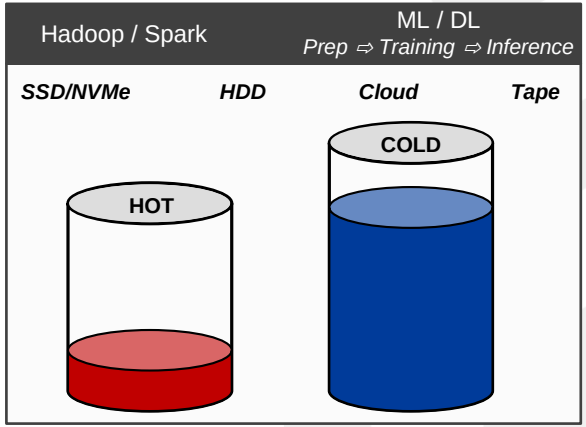
- Ease Data Scientist's and Storage Admins workflows:

  „Dean, a data scientist, is able to run his ML/DL jobs without any knowledge about the underlaying storage infrastructure and how to find and prefetch the data needed."

  - Lots of data! what do I need for my jobs, where is it placed?
  - How can I automate analytics on new ingested data?
  - How do I ensure my limited high-performance storage does not run out of space?
  - How can I have multiple Data Scientist's work in parallel on my system?
  - How do I ensure my data in the high-performance tier stays in sync with the capacity tier?
  - How can I easily / efficiently provision, test, deploy and scale my containerized workloads?

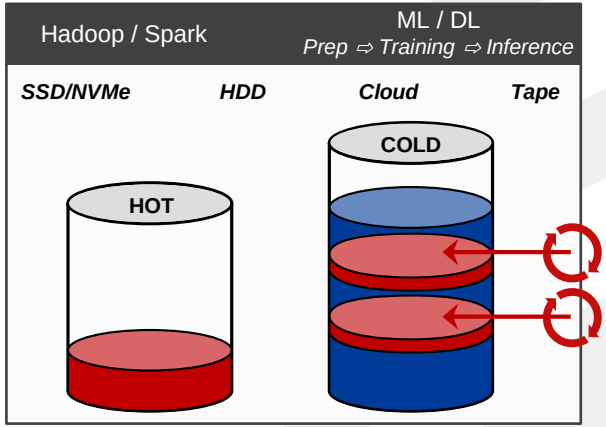
Storage tiering

- Customers across all Industries are creating large PB to EB data stores.
- Vast majority of data is relatively cold,



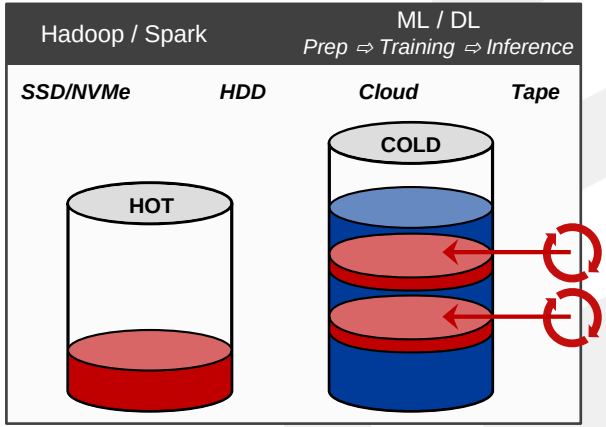
Storage tiering

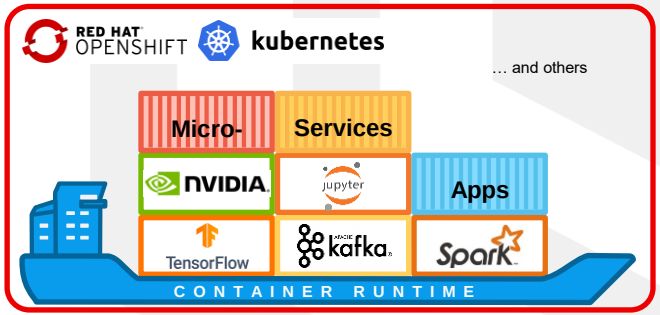
- Customers across all Industries are creating large PB to EB data stores.
- Vast majority of data is relatively cold, but still required for periodic trend analysis.



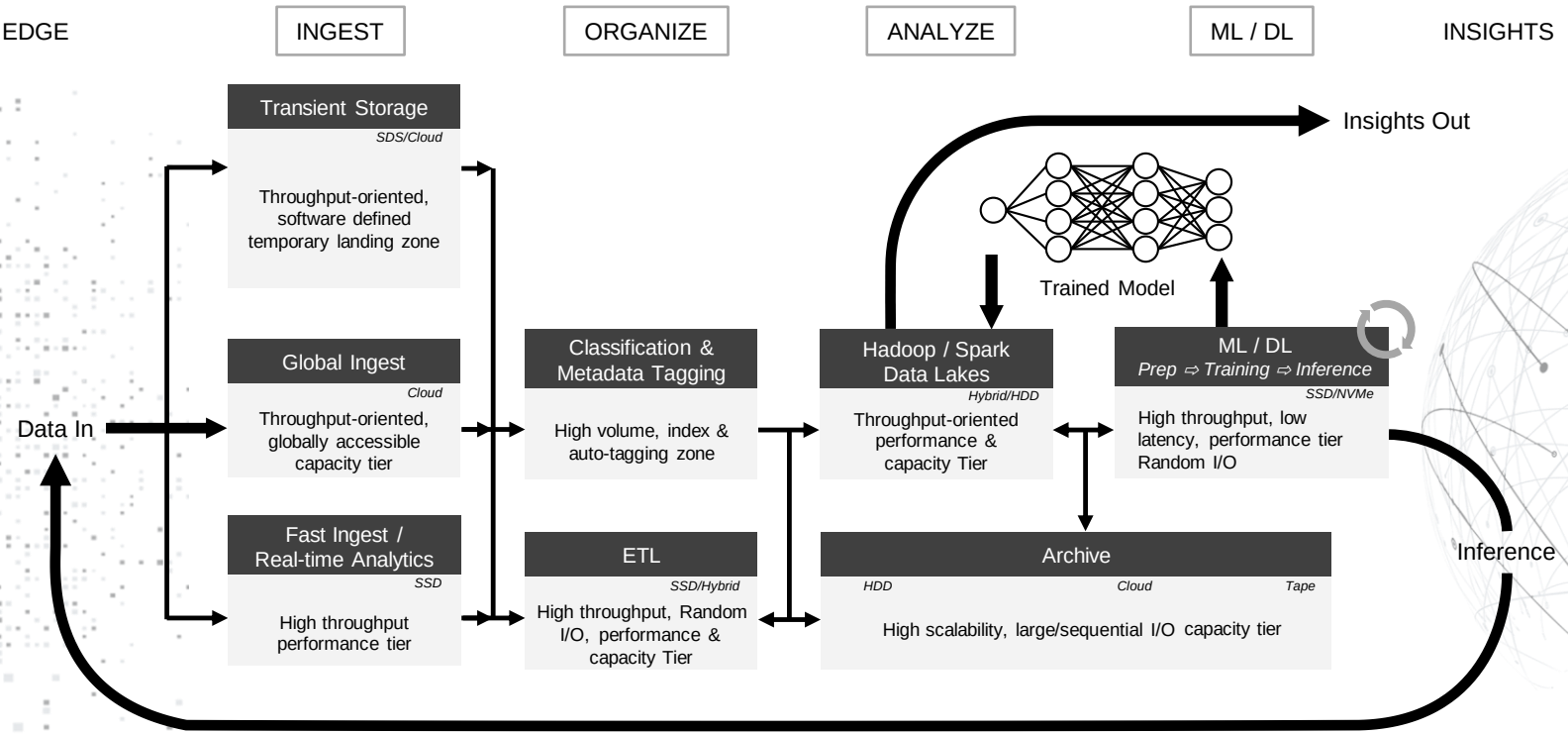
Storage tiering

- Customers across all Industries are creating large PB to EB data stores.
- Vast majority of data is relatively cold, but still required for periodic trend analysis.
- Workloads more and more move into containers
- AI / Analytics requires high performance, low latency storage to keep expensive CPU / GPU / TPU / FPGA busy.



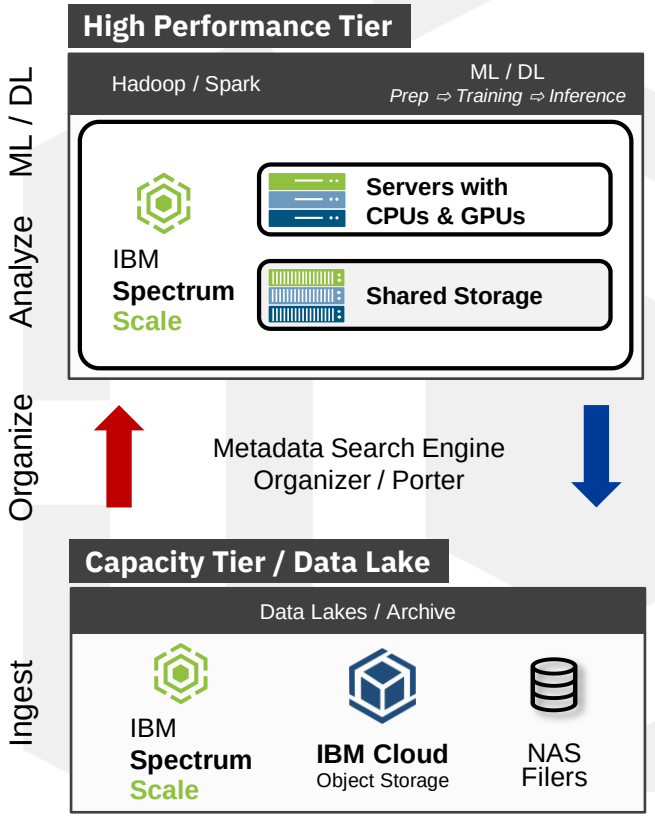


Enterprise Data Pipeline with IBM Spectrum Storage

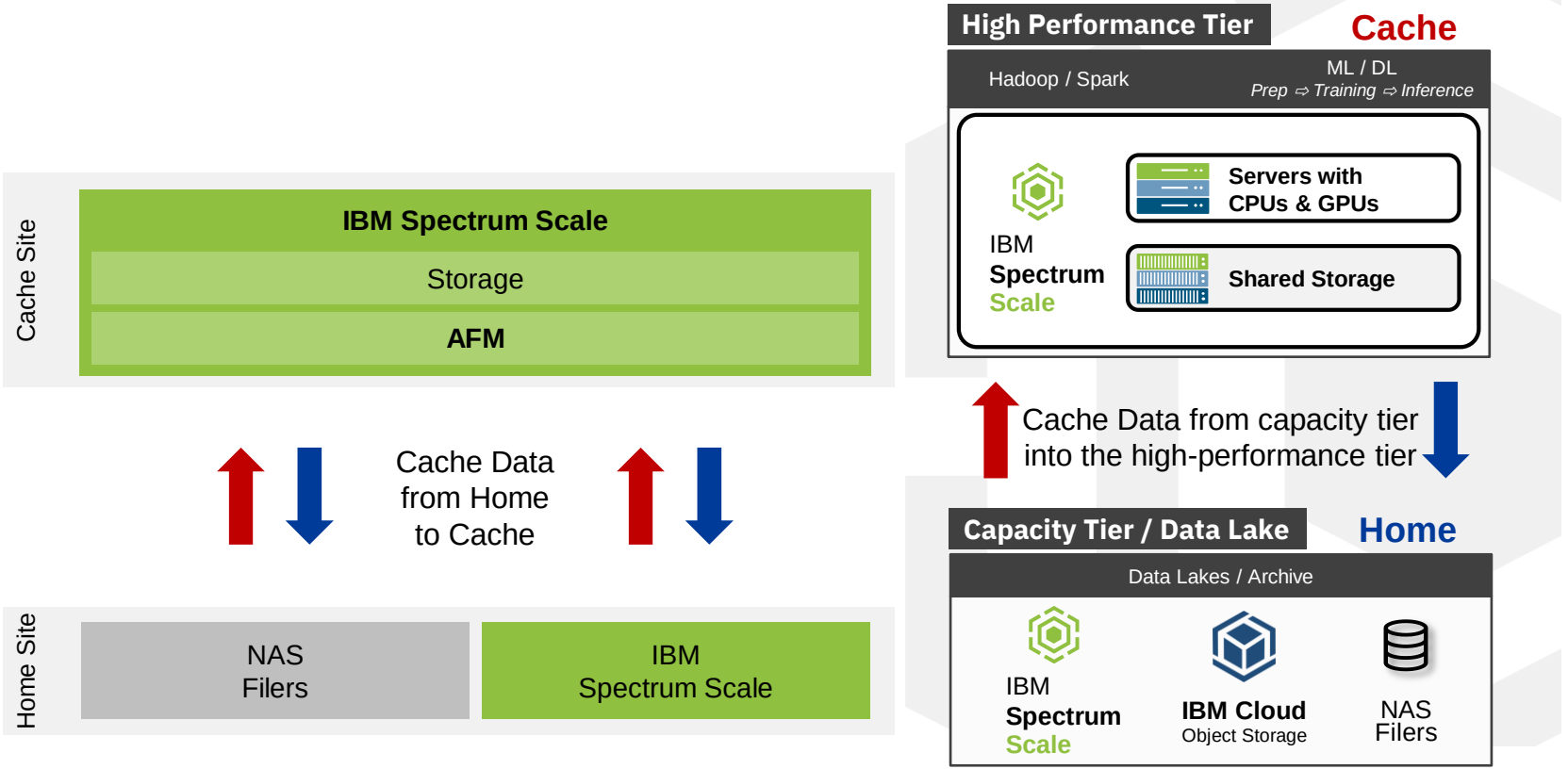


Performance vs. Capacity Storage

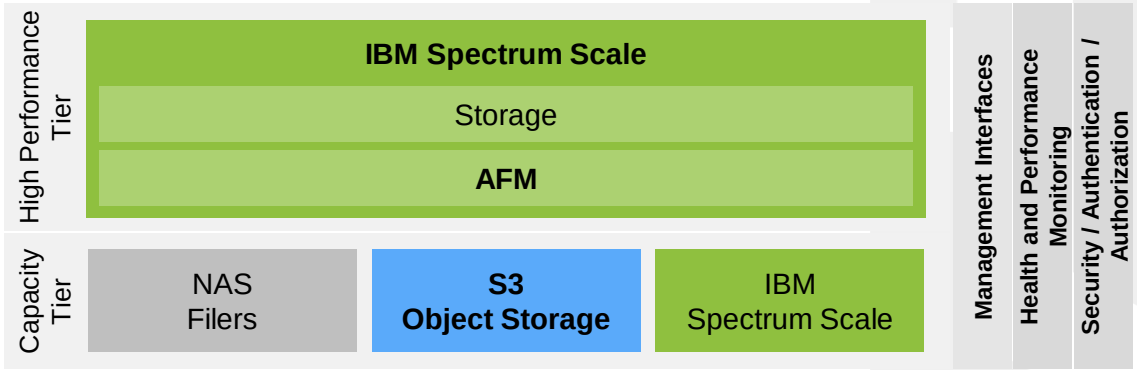
- **Performance Tier**
  - **AI and Analytics with high performing Accelerators**
  - **Maximize performance** of storage: **$/IOps & $/GBps are key**
    - Low latency random I/O & High bandwidth sequential
    - Relatively small compared to Capacity tier (say 5-25%)
    - Can be Lower Durability, Lower Availability, Lower Reliability, if Architected properly
    - No Geo-distribution
- **Capacity Tier** (aka “Data Lake”)
  - Data usually gets ingested here, large volumes
  - **Minimize** the cost of storage: **$/TB is key**
    - High Durability, Availability, Reliability, Geo-distribution



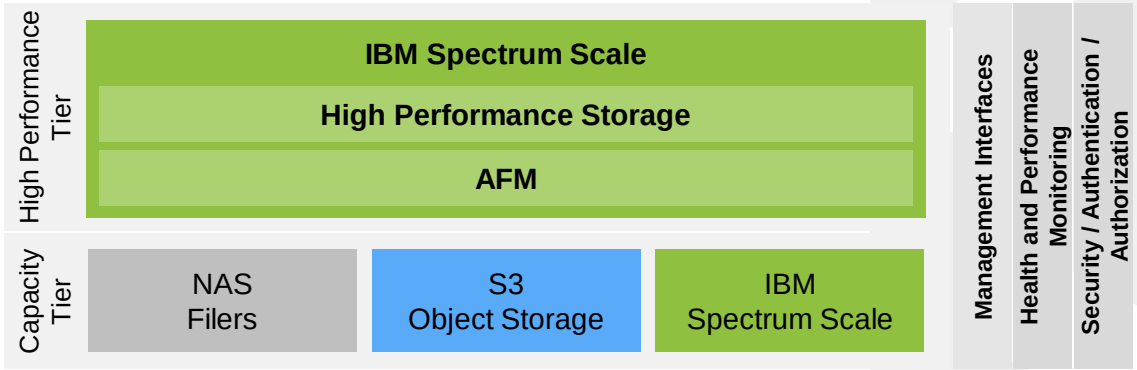
Using IBM Spectrum Scale built-in Caching technology



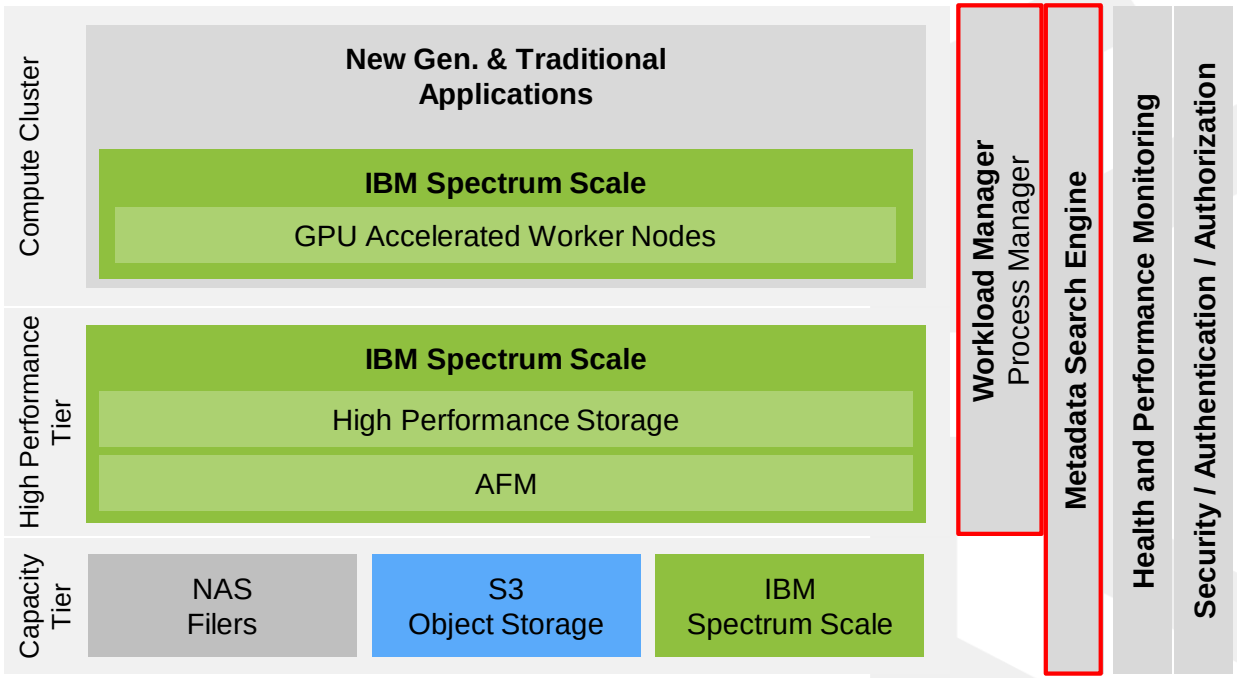
Using IBM Spectrum Scale built-in Caching technology



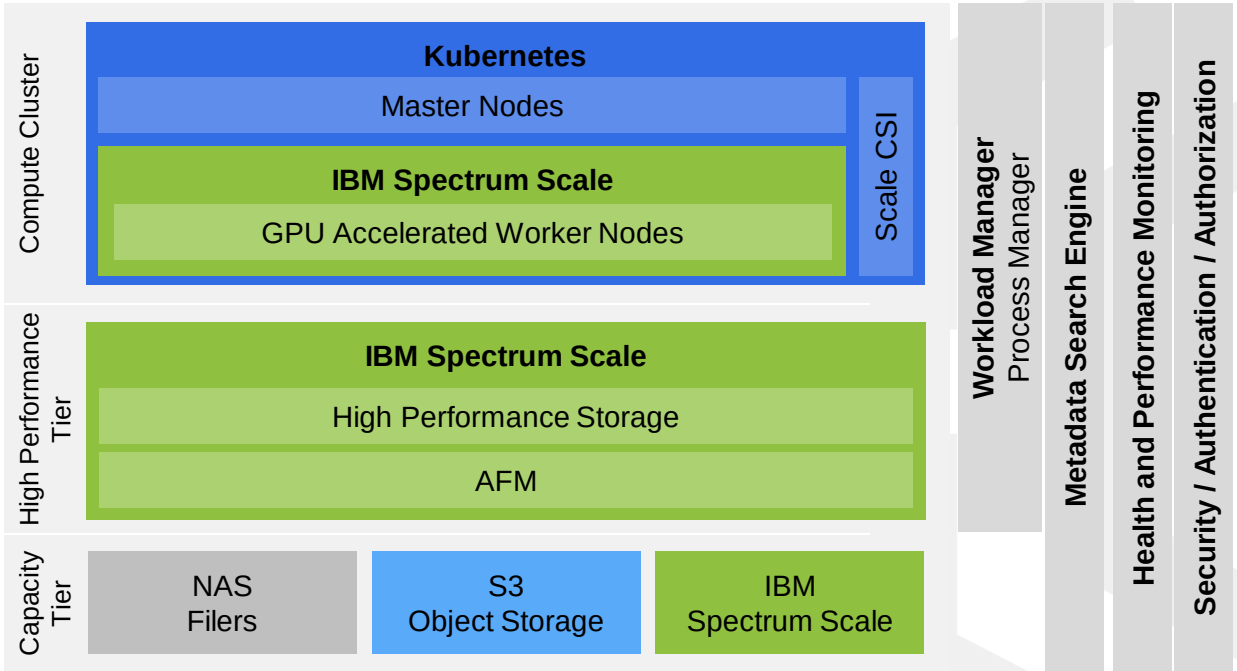
High Performance Tier (HPT) Technology



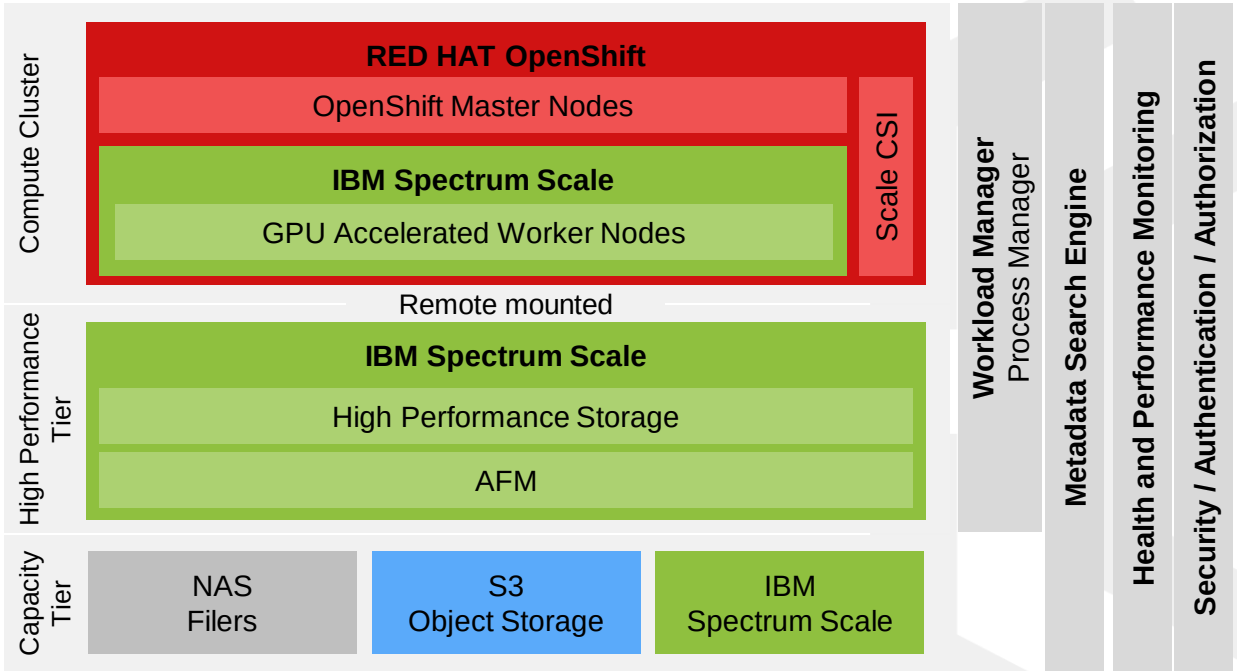
Data Accelerator for AI and Analytics (DAAA) Solution



Data Accelerator for AI and Analytics (DAAA) Solution

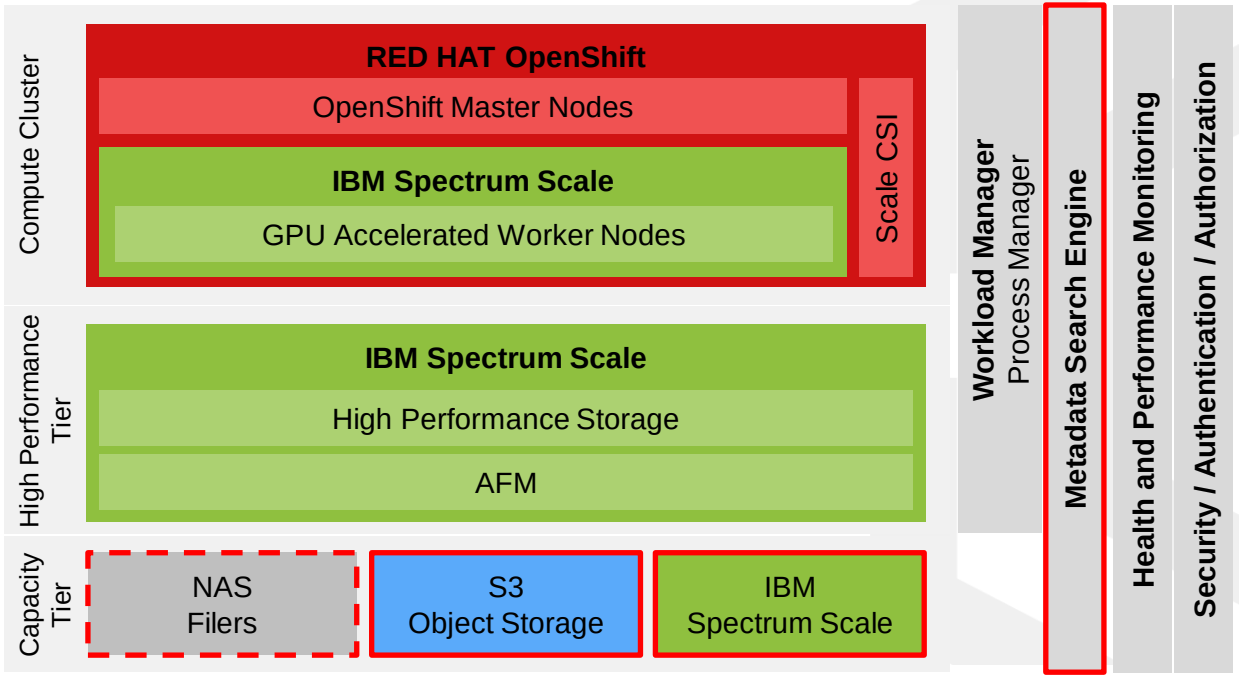


Data Accelerator for AI and Analytics (DAAA) Solution



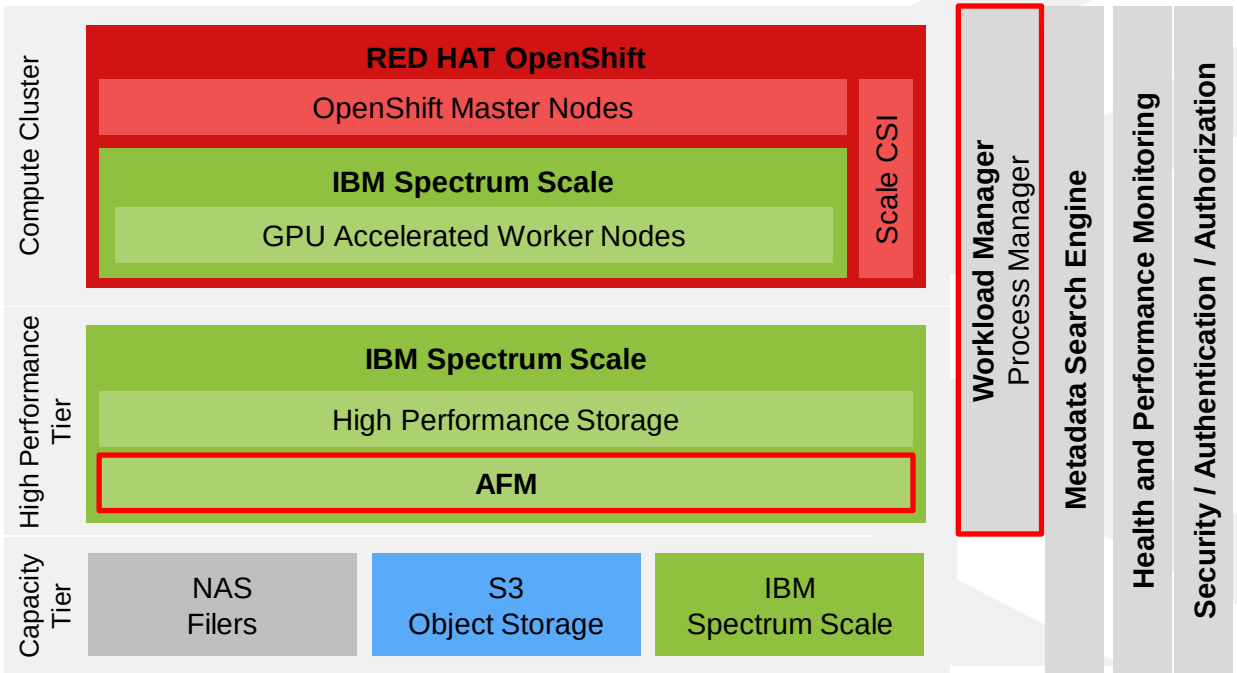
Data Accelerator for AI and Analytics (DAAA)

Lots of data! what do I need for my jobs, where is it placed?
How can I automate analytics on new ingested data?



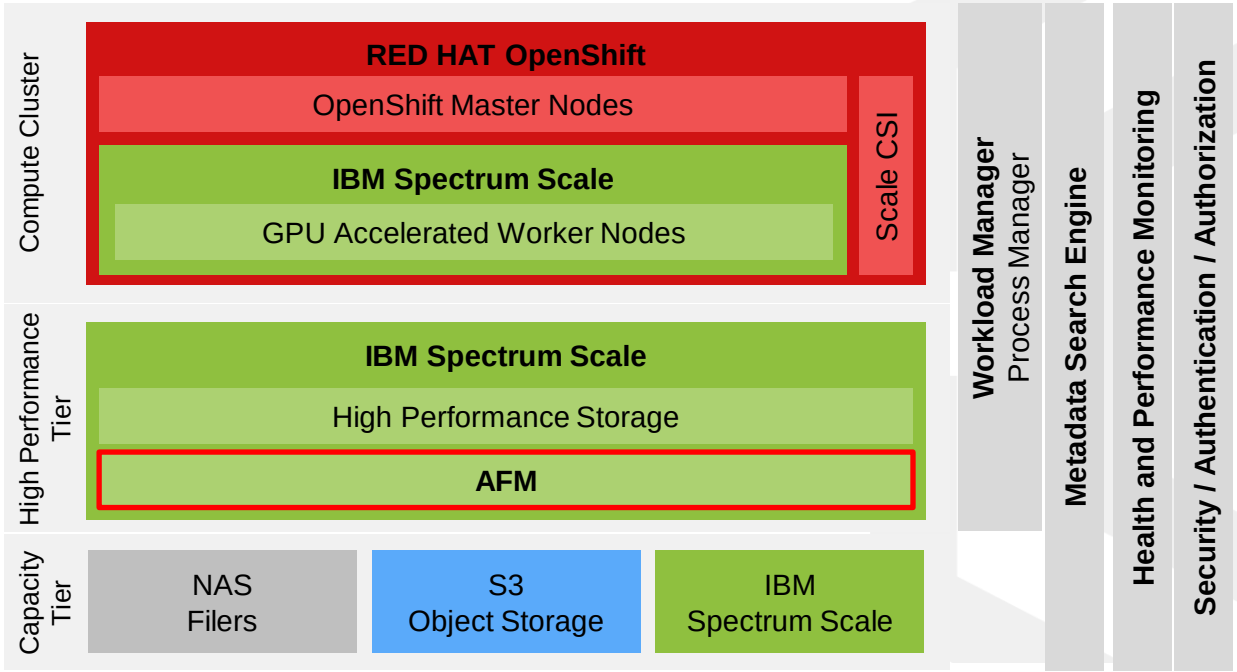
Data Accelerator for AI and Analytics (DAAA)

How do I ensure my limited high performance storage does not run out of space?
How can I have multiple Data Scientist's work in parallel on my system?



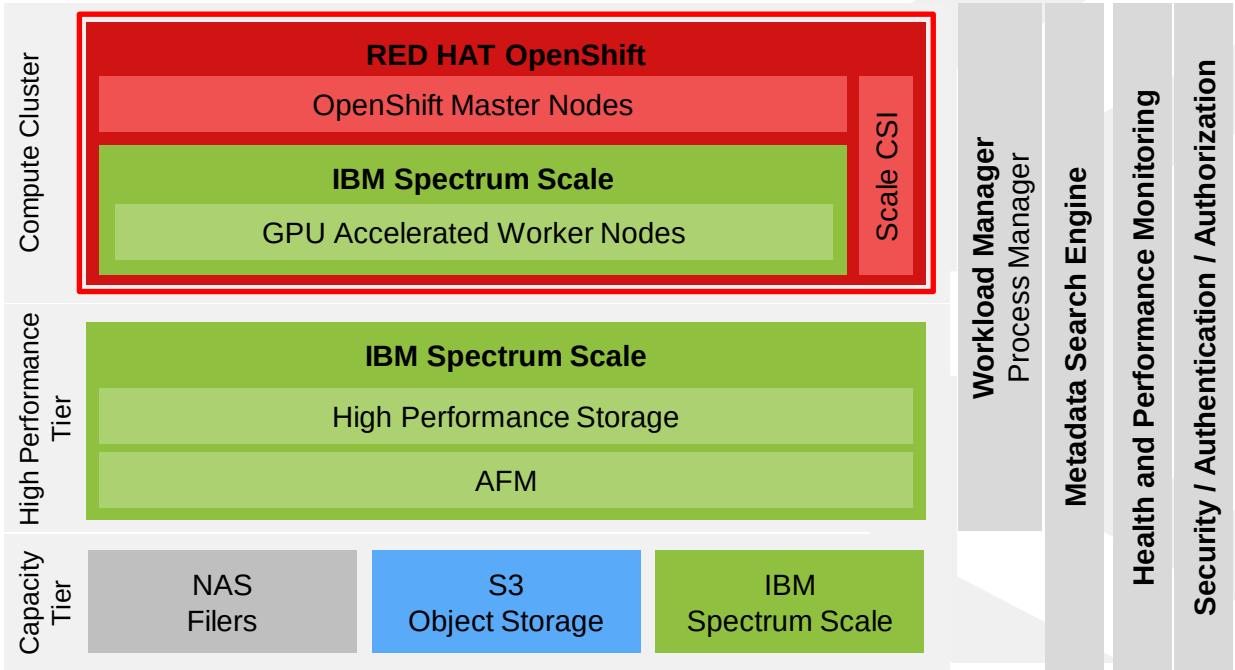
Data Accelerator for AI and Analytics (DAAA)

How do I ensure my data in the high performance tier stays in sync with the capacity tier?



Data Accelerator for AI and Analytics (DAAA)

How can I easily / efficiently provision, test, deploy and scale my containerized workloads?



Summary: Data Accelerator for AI and Analytics (DAAA)

- Brings the right data closer to AI & analytics applications at the right time
  - Faster model training, inference or real-time analytics
  - Without another “persistent” copy of the data, keeping a single source of truth
- Integrates with
  - Workload Managers (e.g. IBM Spectrum LSF)

    Handle job scheduling and balancing load on compute nodes
  - Metadata Search Engines (e.g. IBM Spectrum Discover)

    Enable meta-data management functions such as tagging, classification to automate selection of right datasets for prefetching into the cache
- End to end security and monitoring
- Eases Data Scientist’s and System Admins workflows

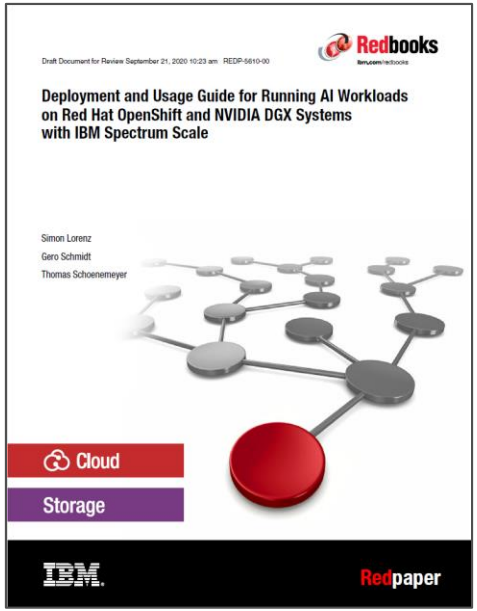
AI Example Use Case: Autonomous Driving

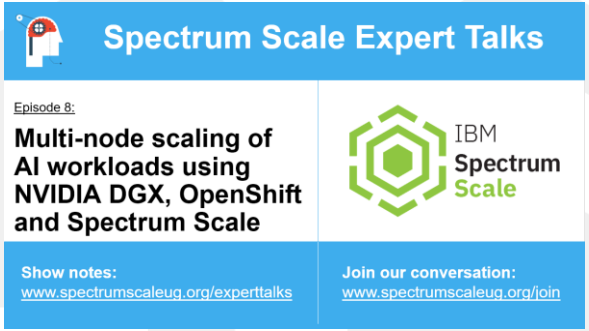
Worked on an IBM Redpaper:

Deployment and Usage Guide for Running AI Workloads on Red Hat OpenShift and NVIDIA DGX Systems with IBM Spectrum Scale

Visit:

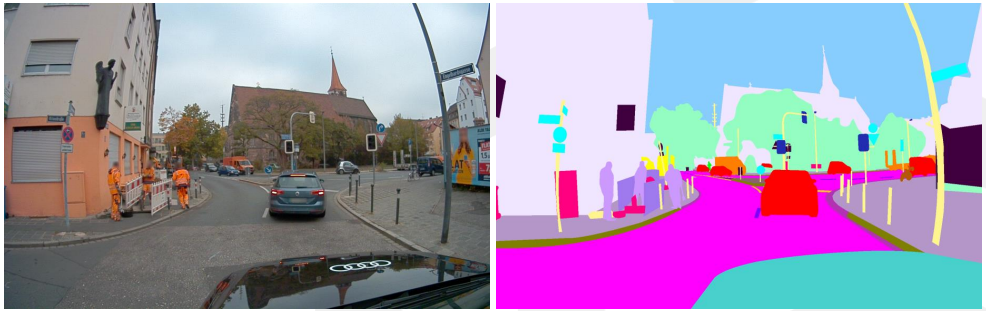
http://www.redbooks.ibm.com/redpieces/abstracts/redp5610.html





Autonomous Driving Dataset used

- Audi Autonomous Driving Dataset (A2D2) published by Audi https://www.a2d2.audi
- Six cameras and five Li-DAR units, providing full 360° coverage
- Data is time synchronized and mutually registered
- 41,277 frames with semantic segmentation image and point cloud labels
- Semantic segmentation dataset features 38 categories



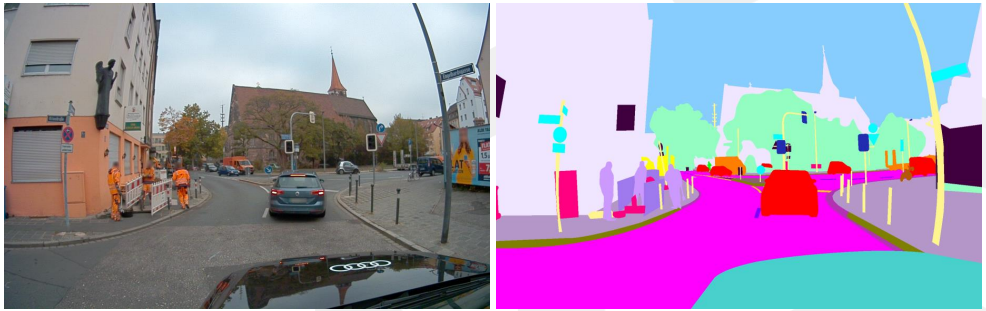
(A2D2- Jacob Geyer et al, 2020, https://arxiv.org/abs/2004.06320)

Autonomous Driving Dataset used

- Audi Autonomous Driving Dataset (A2D2)
- Semantic segmentation dataset features 38 categories

Autonomous Driving Dataset used

- Audi Autonomous Driving Dataset (A2D2) published by Audi https://www.a2d2.audi
- Six cameras and five Li-DAR units, providing full 360° coverage
- Data is time synchronized and mutually registered
- 41,277 frames with semantic segmentation image and point cloud labels
- Semantic segmentation dataset features 38 categories



(A2D2- Jacob Geyer et al, 2020, https://arxiv.org/abs/2004.06320)

Autonomous Driving Dataset used

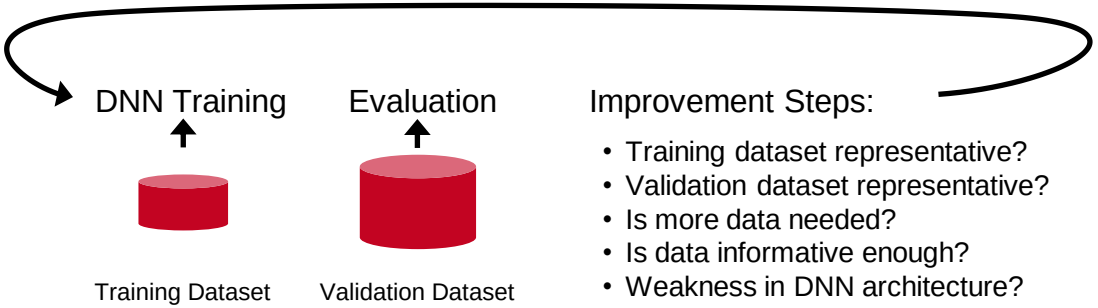
- Audi Autonomous Driving Dataset (A2D2) published by Audi https://www.a2d2.audi
- Six cameras and five Li-DAR units, providing full 360° coverage
- Data is time synchronized and mutually registered
- 41,277 frames with semantic segmentation image and point cloud labels
- Semantic segmentation dataset features 38 categories

(A2D2- Jacob Geyer et al, 2020, https://arxiv.org/abs/2004.06320)

```
26    "#ffffc8": "Utility vehicle 2",
27    "#e96400": "Sidebars",
28    "#6e6e00": "Speed bumper",
29    "#808000": "Curbstone",
30    "#ff[illegible]5": "Solid line",
31    "#4000[illegible]0": "Irrelevant signs",
32    "#b97a[illegible]7": "Road blocks",
33    "#000064": "Tractor",
34    "#3b636c": "Non-drivable street",
35    "#d23273": "Zebra crossing",
36    "#ff0080": "Obstacles / trash",
37    "#fff68f": "Poles",
38    "#960096": "RD restricted area",
39    "#ccff99": "Animals",
```

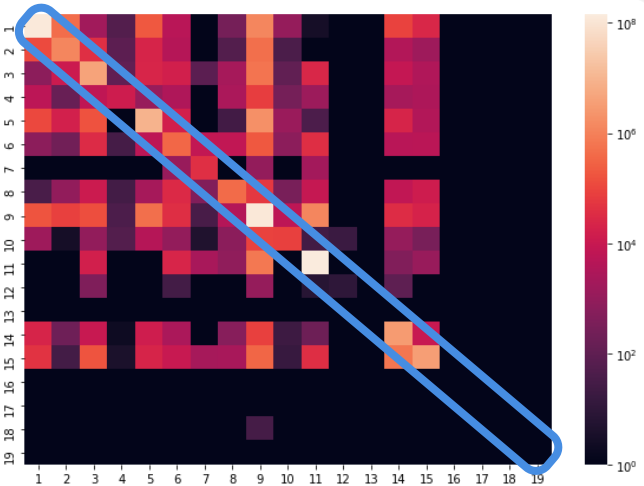
	A	H	I	J	K	L	M	N	O	P
1	Filename	Car	Bicycle	Pedestrian	Truck	Small vehicles	Traffic signal	Traffic sign	Utility vehicle	Sidebars
50	62316.png	39564	315	2302	0	0	0	7626	0	0
51	78793.png	48538	0	0	7671	0	113	465	0	0
52	12593.png	238144	1511	0	0	1342	0	8956	0	0
53	20783.png	31558	1525	4213	0	0	3337	20014	3917	0
54	27813.png	10728	14712	585	3514	0	5184	3814	1762	0
55	70939.png	31037	1889	187	0	0	0	18397	0	0
56	00488.png	11482	0	0	0	0	130	180	0	0
57	42035.png	299263	4371	0	0	0	4149	3565	0	0
58	73496.png	15546	0	0	0	0	0	7331	0	1955
59	07232.png	25964	4476	5941	0	0	0	17881	0	0
60	74320.png	80	0	0	0	0	0	233	0	4041
61	75272.png	2796	0	0	0	0	0	835	0	137
62	21604.png	172247	3282	0	0	0	180	1413	0	0
63	41344.png	32295	14794	243	13485	0	2348	2346	0	0
64	73456.png	43828	0	0	0	0	0	6814	0	5598
65	68156.png	108909	2071	0	498	0	0	3995	0	0
66	46816.png	213287	0	103	197	0	332	5287	0	0
67	56672.png	102111	0	0	1794	0	0	25615	34367	0
68	22033.png	3783	169	16855	0	585	688	1846	0	0
69	06963.png	28567	0	210	9933	0	0	3346	0	0
70	50297.png	232642	0	8496	0	0	0	7660	0	0
71	62375.png	39551	3061	3878	0	0	0	7880	0	0
72	71954.png	156984	0	26788	41861	0	0	80	0	0

Building the right training and validation dataset



Improvement Steps:

- Training dataset representative?
- Validation dataset representative?
- Is more data needed?
- Is data informative enough?
- Weakness in DNN architecture?



Confusion matrix presenting the training evaluation

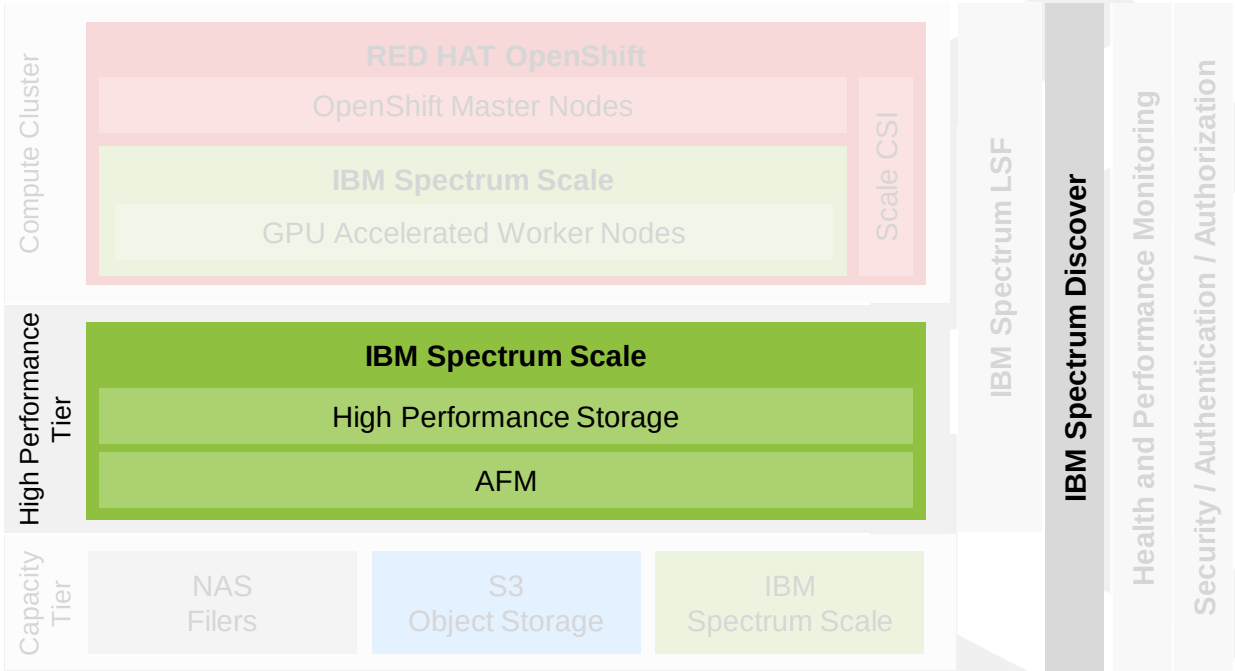
- Building a representative validation dataset is challenging
- Validation dataset is significantly larger than the training dataset
- Validating the trained DNN against a large validation dataset is critical to understand its weaknesses

For a perfect predictor:

Diagram would only have a diagonal line from top left to bottom right.

That would read as the network would have classified all pixels of a certain class right.

IBM Spectrum Scale AFM and Spectrum Discover



DAAA Insides / Demo

- AFM Background
- Data Accelerator for AI and Analytics Use Cases
  - Data Scientist Dean starts an AI Job
  - Data Ingest starts an AI Job
  - DAAA Workflow Steps with Jupyter Notebook

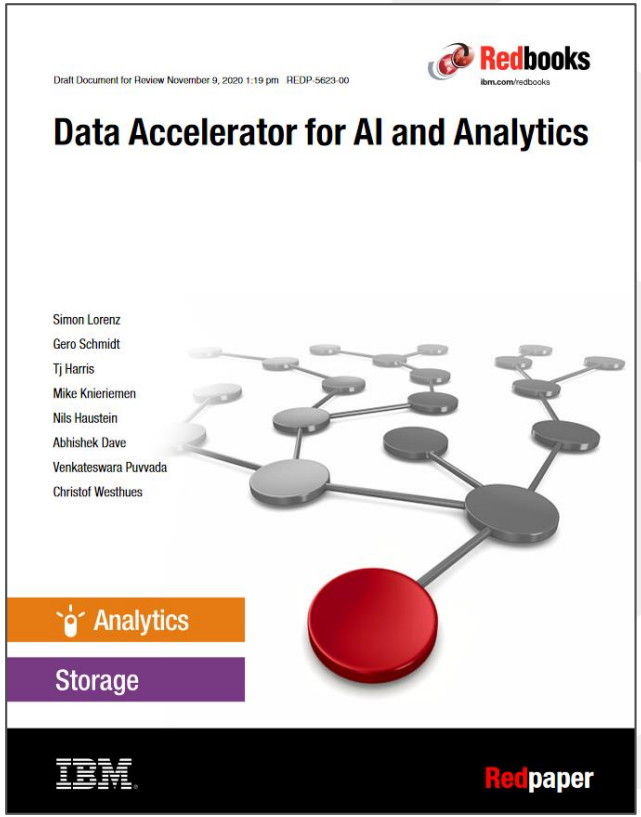
25min. Demo and additional Insights: Not available online

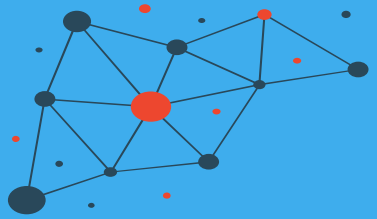
Data Accelerator for AI and Analytics RedPaper

Data Accelerator for AI and Analytics

Visit:

http://www.redbooks.ibm.com/redpieces/abstracts/redp5623.html
(published November 09, 2020)





Check https://www.spectrumscaleug.org/experttalks for charts, show notes and upcoming talks

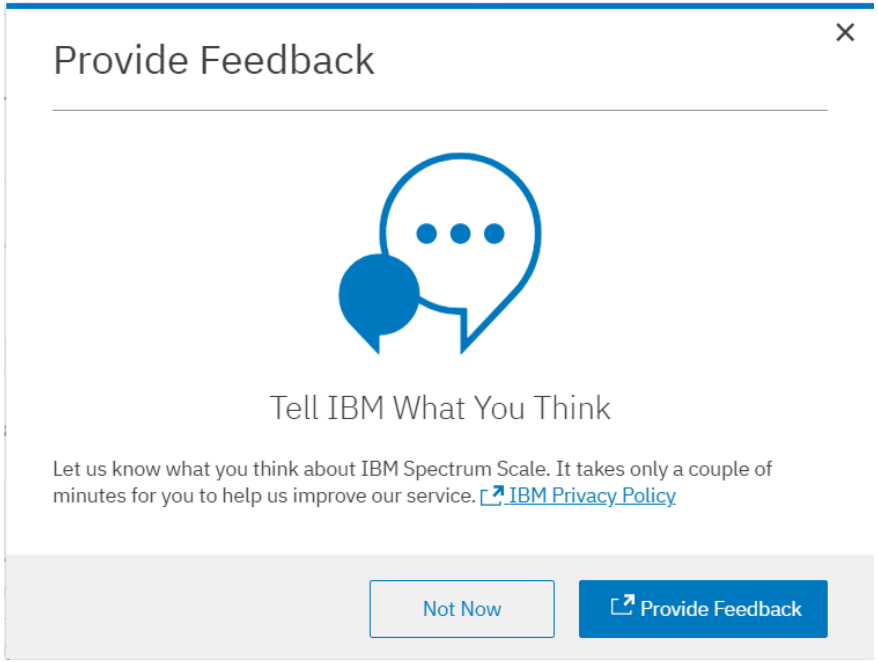
- Past talks:
  - 001: What is new in Spectrum Scale 5.0.5?
  - 002: Best practices for building a stretched cluster
  - 003: Strategy update
  - 004: Update on performance enhancements in Spectrum Scale (file create, MMAP, direct IO, ESS 5000)
  - 005: Update on functional enhancements in Spectrum Scale (inode management, vCPU scaling, NUMA considerations)
  - 006: Persistent Storage for Kubernetes and OpenShift environments
  - 007: Manage the lifecycle of your files using the policy engine
  - 008: Multi-node scaling of AI workloads using Nvidia DGX, OpenShift and Spectrum Scale
- Today:
  - Nov 16: Continental: Deep Thought – An AI Project for Autonomous Driving Development
  - Nov 16: Data Accelerator for AI and Analytics (DAAA)
- Next:
  - Nov 18: User Meeting at SC20 (Session 2) – What is new in Spectrum Scale 5.1? https://www.spectrumscaleug.org/event/sc20-meeting-session-2-what-is-new-in-spectrum-scale-5-1/

Thank you!

IBM Spectrum Scale

Please help us to improve Spectrum Scale with your feedback

- If you get a survey in email or a popup from the GUI, please respond
- We read every single reply



Spectrum Scale User Group

The Spectrum Scale (GPFS) User Group is free to join and open to all using, interested in using or integrating IBM Spectrum Scale.

The format of the group is as a web community with events held during the year, hosted by our members or by IBM.

See our web page for upcoming events and presentations of past events. Join our conversation via mail and Slack.

www.spectrumscaleug.org