Data Platform for ADAS and Autonomous Driving Development

Nils Haustein, STSM, ESCC
Thanks to Frank Kraemer for the contributions

EMEA Storage Competence Center

IBM Spectrum Scale Strategy Days 2019

EMEA Storage Competence Center

The EMEA Storage Competence Center (ESCC) provides **Pre- to Post-Sales Storage services to assist you across IBM's complete Storage portfolio (Hardware / Software / Solutions)** via:

Advanced Technical Skills (ATS)
Skill Transfer, New Product Intro., Solution Enablement, Architectural Guidance

Lab Services (LS)
Skill Enablement, Hands-On-Training, Implementation Assistance, Data Migration & Performance Tuning

Development & Product Test Systems Lab
Proof of Concept (PoC), Hands-on Product Training …

Product Field Engineering (PFE)
Last level defect support

Client Care
Pro-active Customer Care Management

ESCC@IBM Intranet

http://escc.mainz.de.ibm.com/e_dir/e4uweb.nsf/site.xsp

ESCC@IBM Partnerworld

https://www.ibm.com/partnerworld/page/pw_com_bp_ecampus

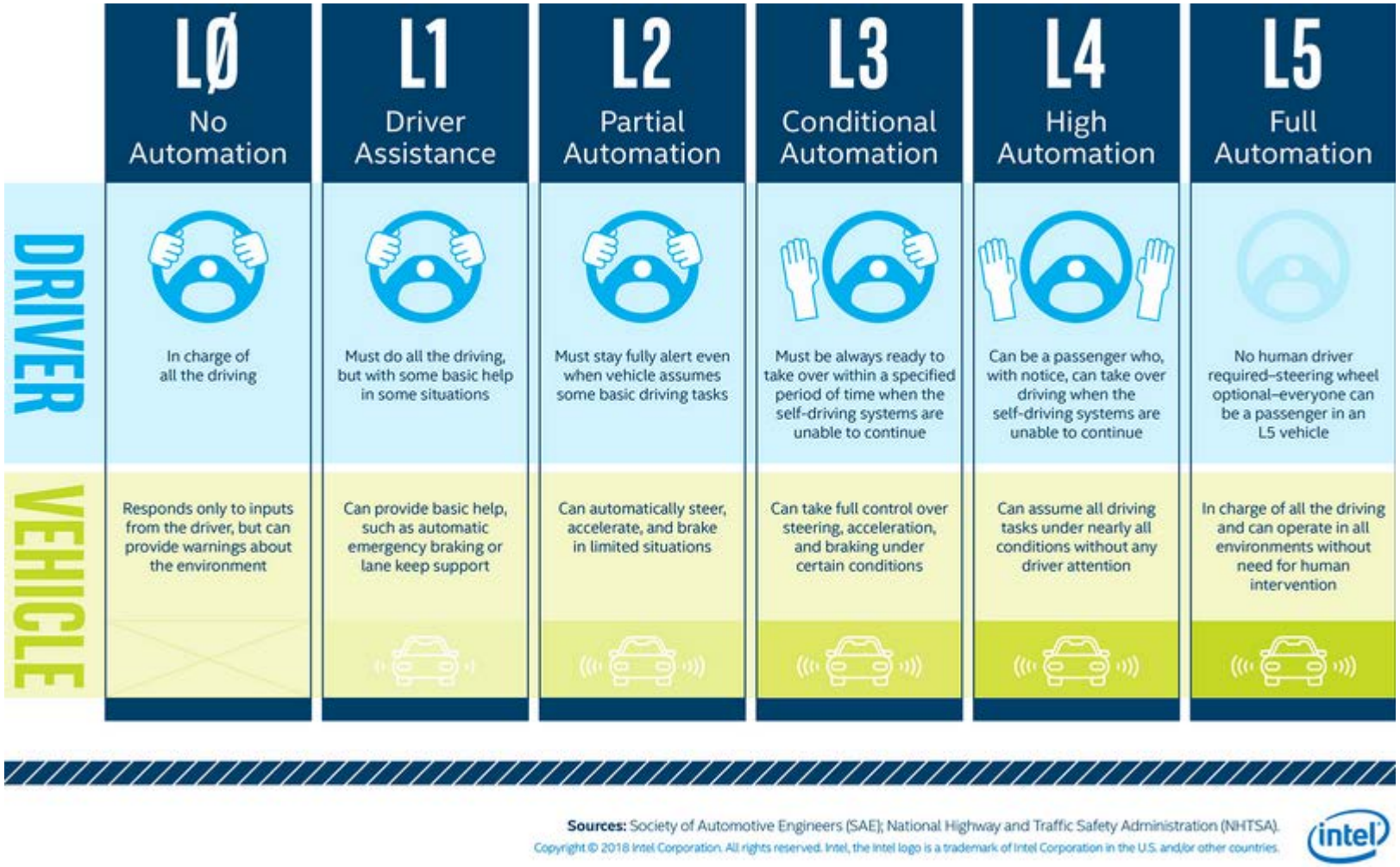
Agenda

Autonomous Driving / ADAS overview and data flows

Data Platform for ADAS / autonomous driving development

Autonomous driving overview

Autonomous driving (AD) is based on a self-driving car that is capable of sensing its environment and moving with little or no human input, There are 6 levels of autonomous driving



ADAS = Advanced Driving Assistance System → help the driver

Systems and functions for ADAS are deployed for autonomous driving which focuses on **driver-less driving**

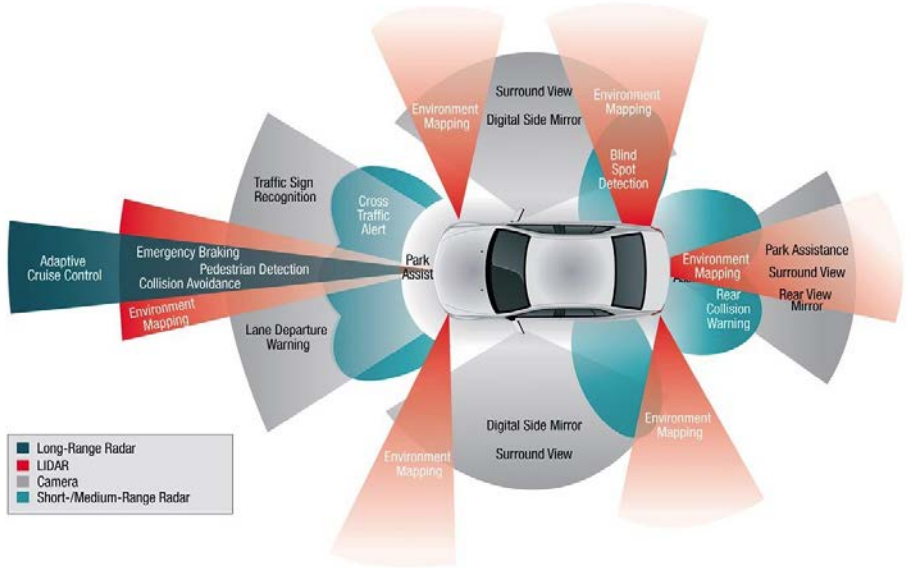
Autonomous driving challenges

— To get to level 3 and beyond heavy simulation has to be done
  - Achieving level 5 would take 100+ years with ordinary driving

— Simulation is based on sensor data captured by test cars

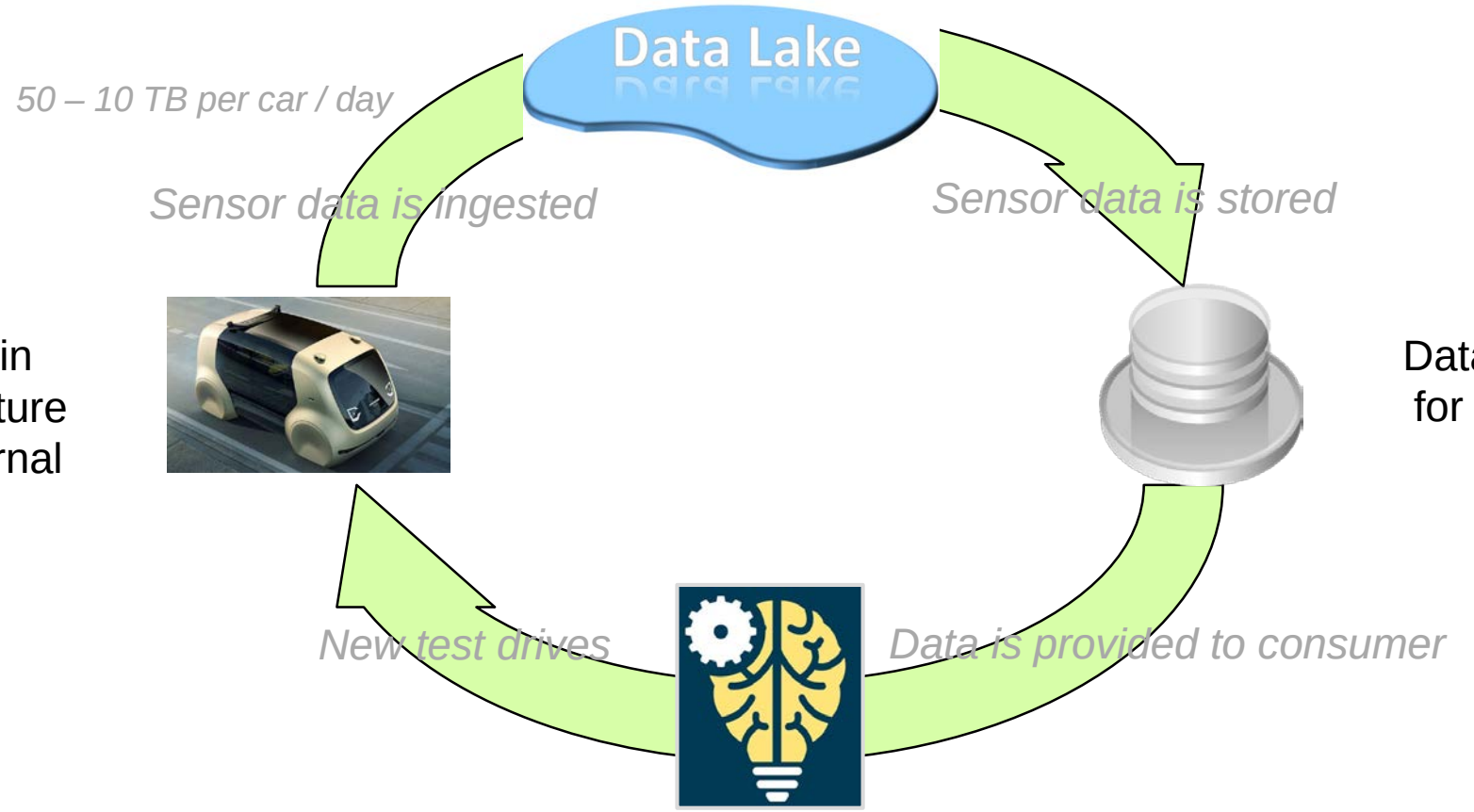
— Test cars are equipped with sensors and record different driving scenarios
  - Cameras, RADAR, LIDAR, etc.
  - Sensor data is stored in logging device within the test car

— Sensor data needs to be captured, cleaned, tagged and provided to other workflows supporting AD development
  - A test car produces ~ 50 – 70 TB per day



High level Data flow for AD and ADAS development

Preparation and Calibration
Sensor data is processed, calibrated and prepared



Test drives:
Drive under certain conditions and capture sensor data in internal logger devices

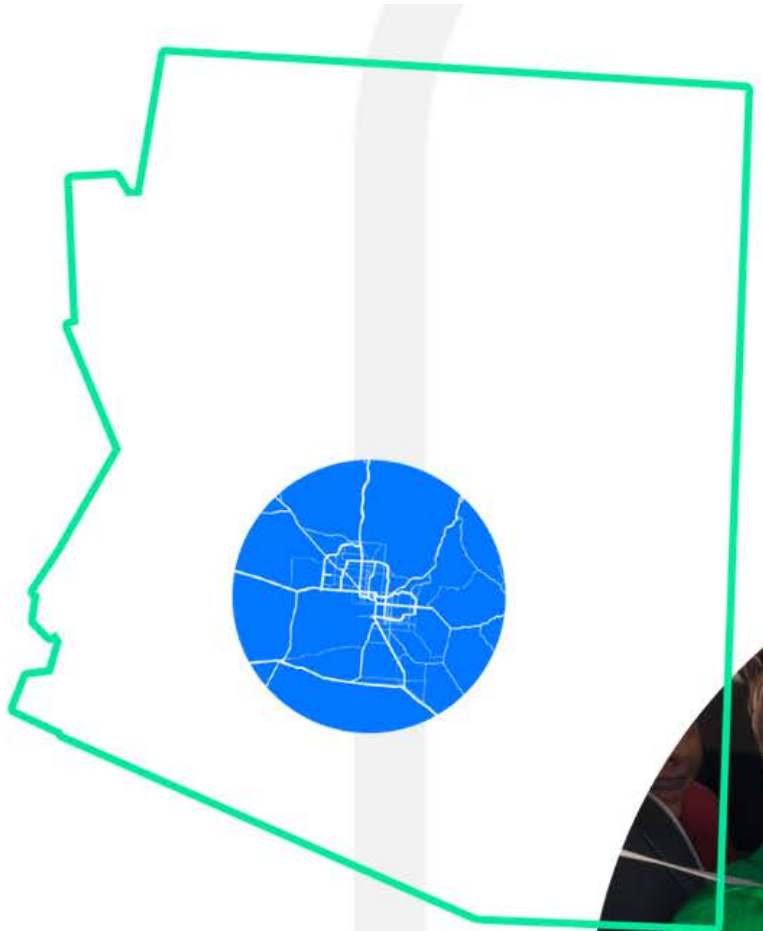
Mass Storage:
Data is stored and provided for subsequent processing

100s of Petabytes

Development, training, simulation and testing
Machine learning, simulation, etc.

Autonomous driving is reality (in test mode)

Waymo (former Google X project) offers driverless (test) rides in certain cities in the US



Waymo's early rider program begins

We invited residents in Phoenix, AZ to join us in the first public trial of self-driving vehicles. Their feedback has been integral in helping shape our technology, service, and customer experience.

Be an early rider

Live in Metro Phoenix? Apply for Waymo's early rider program and you can use our self-driving cars to go places every day. As one of our first riders, you'll be able to share your thoughts and experiences directly with our team to help shape the future of self driving.

To get started, tell us a bit about yourself and your typical commute.

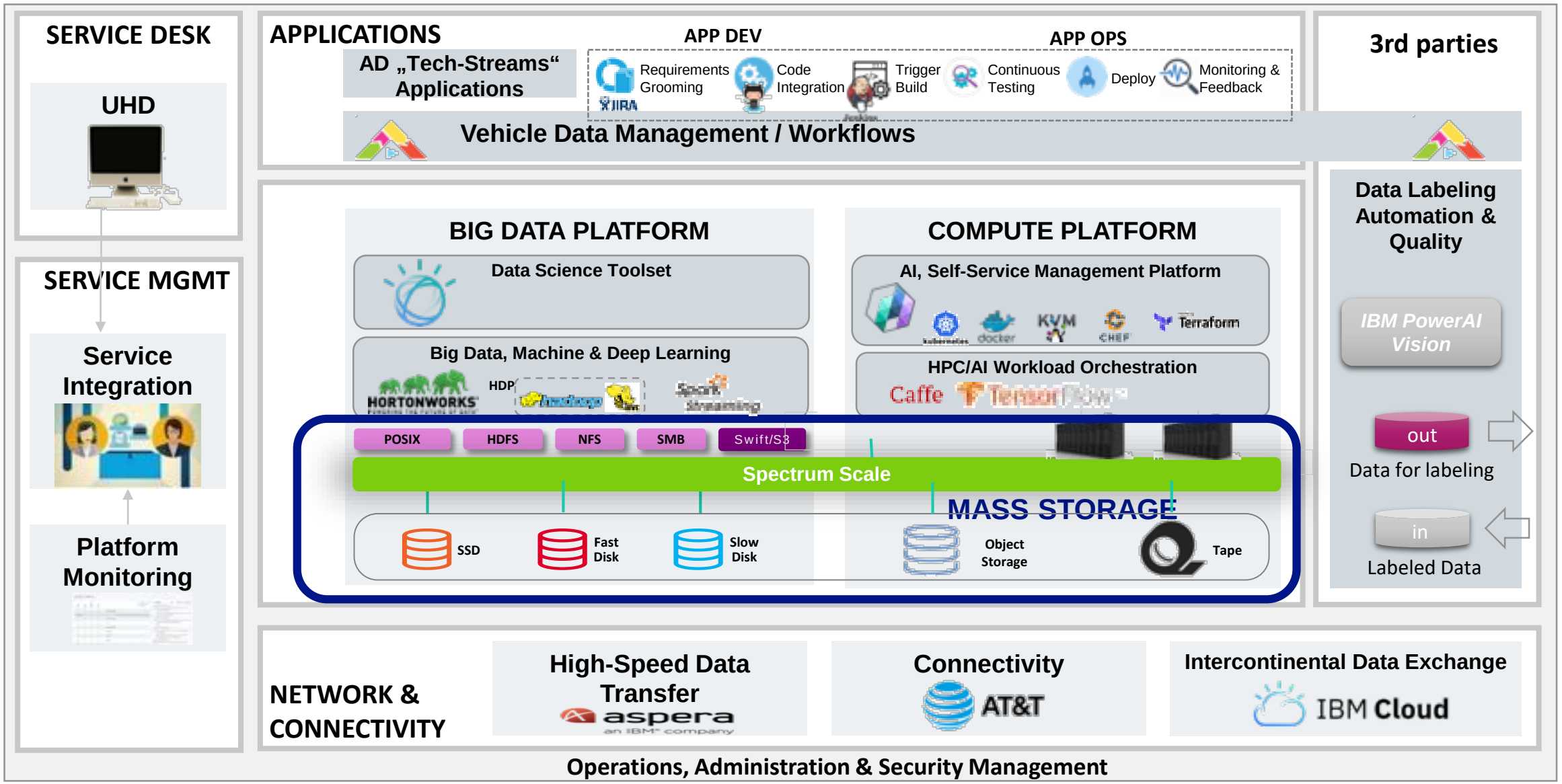
Source: Waymo LLC 2018

Agenda

Autonomous Driving overview and data flows

Data Platform for ADAS / autonomous driving development

Overall ADAS & AD solution overview

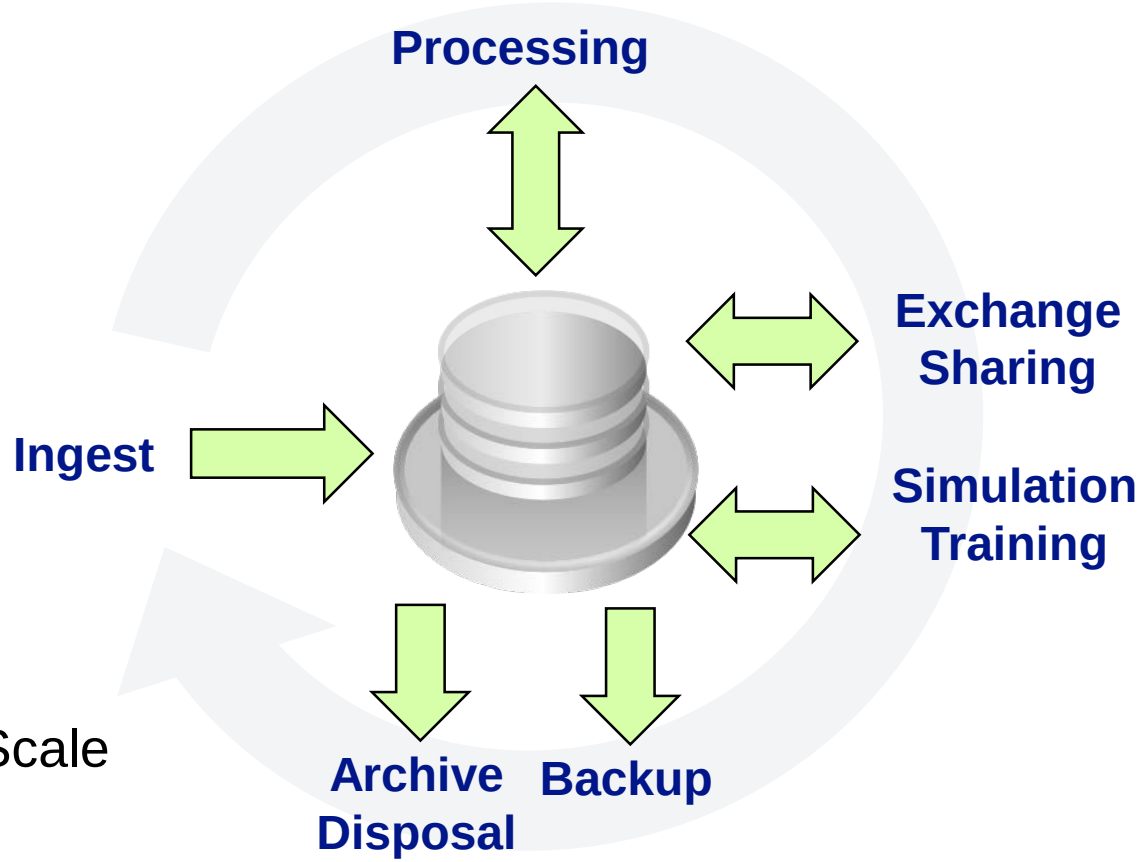


Storage use cases for automotive driving development

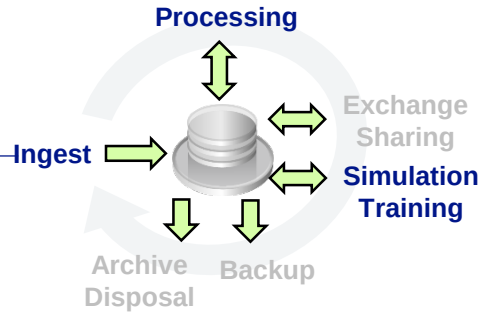
— Storage for AD is used to store data from ingestion to deletion and archival
  - Data is being generated by sensors installed in cars and in other devices

— Main storage use cases for AD development are:
  - Data Ingest
  - Data exchange / sharing
  - Data processing (tagging, cleanup)
  - Simulation & Testing
  - Training
  - Backup
  - Disposal & Archive

This presentation focuses on solutions for these data management use cases based on IBM Spectrum Scale

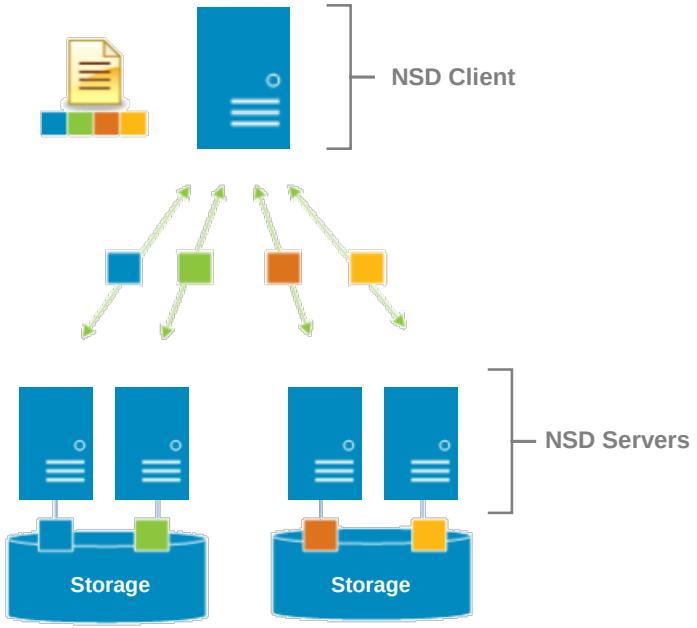


Parallel I/O Architecture



No Hot Spots

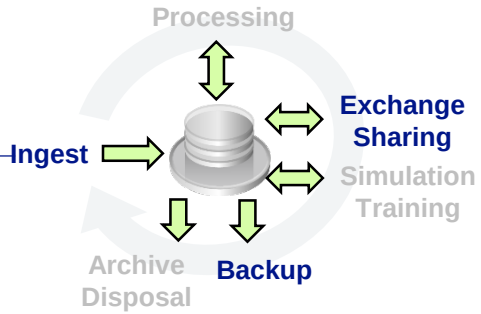
- All NSD* servers export disk to all clients in active-active mode
- Spectrum Scale stripes files across NSD servers and NSDs in units of file-system block-size
- File-system load spread evenly
- Easy to scale file-system capacity and performance while keeping the architecture balanced



NSD Client does real-time parallel I/O
to all the NSD servers and storage volumes/NSDs

* NSD = Network Shared Disk, a representation of a storage LUN or disk drive to Spectrum Scale

High speed data transfer across long distances



— **IBM Aspera provides WAN transport at wire speed**

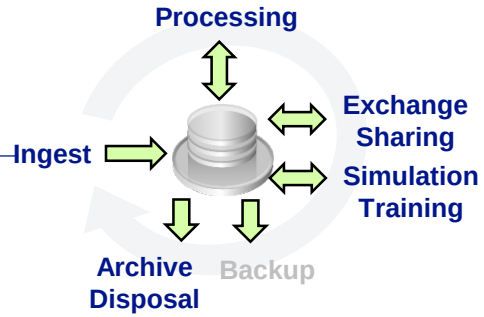
- **Efficiently utilize** available network bandwidth
- **Distance Independent**: Transfer speeds remain constant as distances increase
- **Predictable & Reliable**: Transfer times decrease linearly as bandwidth increases.
- **Versatile**: Supports massive file sizes (500 GB+) as easily as very large sets (millions) of small files

— IBM Aspera **support many storage systems and protocols**

- Including BM Spectrum Scale, NAS (NFS, SMB), Cloud Object Storage and HDFS

MOVING A 10GB FILE				
	Network Bandwidth	Across US	US - Europe	US - Asia
Legacy Transport	100 Mbps	10-20 Hours	15-20 Hours	Impractical
	1 Gbps			
	10 Gbps			
Aspera FASP®	100 Mbps	14 Min	14 Min	14 Min
	1 Gbps	1.4 Min	1.4 Min	1.4 Min
	10 Gbps	8.4 Sec	8.4 Sec	8.4 Sec

Open data access



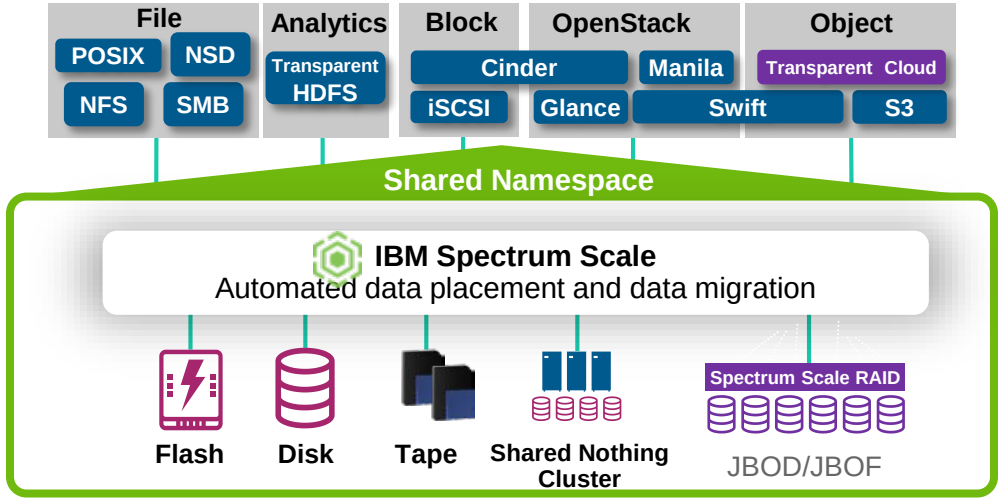
— **IBM Spectrum Scale provides open data access through a variety of open and standardized protocols (NFS, SMB, Swift, S3, POSIX)**

- Also includes a open API that overcomes limitations of standardized protocols (e.g. parallel transfer)

— Integration with **Big Data** environments such as Hadoop and Docker

- Through HDFS transparency layer and Ubiquity providing persistent storage for Docker container

— Facilitates support for **OpenStack components** (Cinder, Manila, Glance)

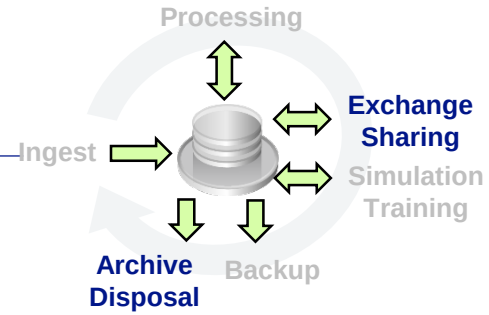


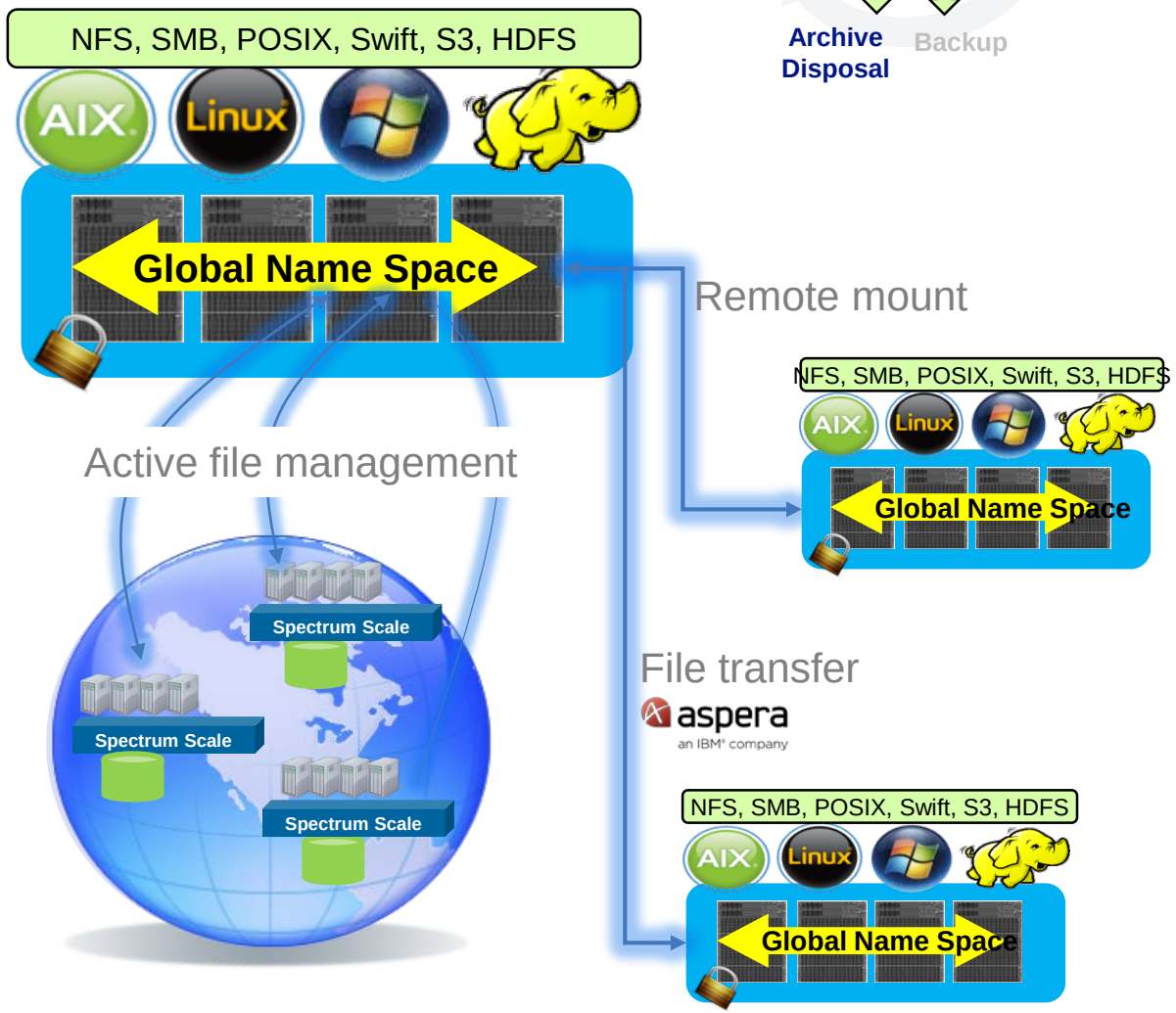
Global sharing

— IBM Spectrum Scale **Active File Management** (AFM) allows to share files asynchronously between sites
  - Files are globally visible and only locally present when accessed or pre-fetched
  - Tolerates reliability and latency of WAN connections

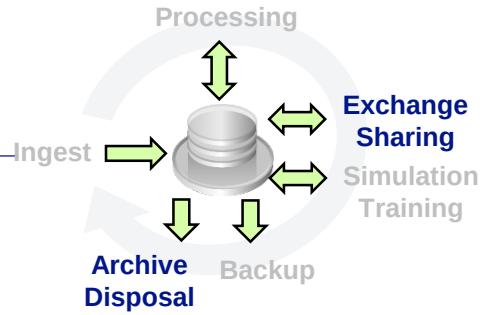
— **Remote file system mount** (cross-cluster-mount) allows to share files between site synchronously at high speeds

— **IBM Aspera** can be used for efficient long-distance file transfer





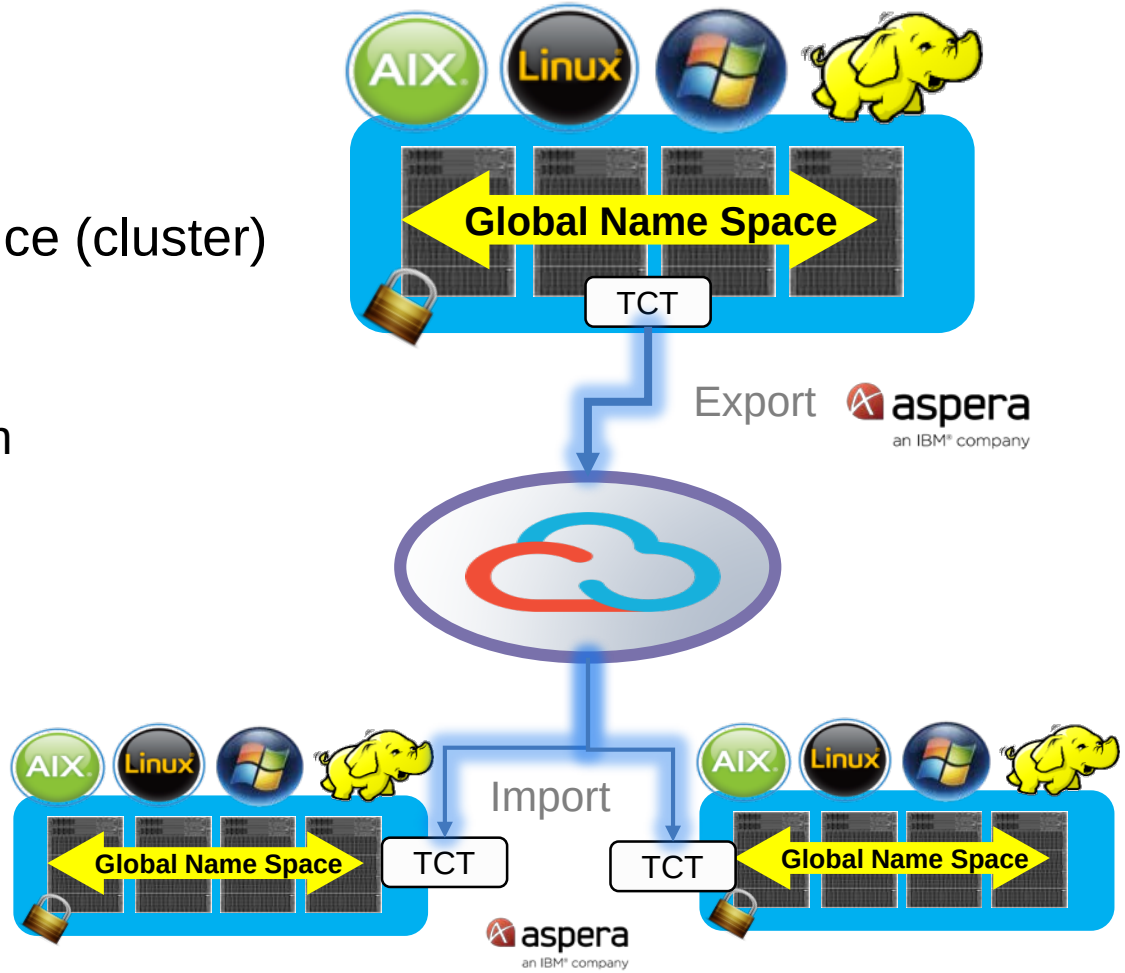
Cloud data exchange



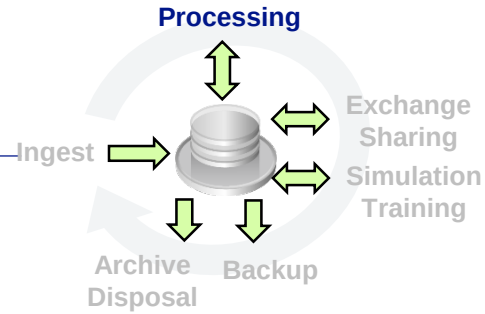
— With IBM Spectrum Scale **Transparent Cloud Tiering** files are copied to Object Storage
  - Object storage can be on or off-premises
  - Mapping of files and objects is included (manifest)

— **Objects can be imported as files** in other TCT instance (cluster)
  - Based on file to object mapping (manifest)
  - Import creates stub-files, does not transfer data
  - Objects are transferred upon access or pre-fetch operation

— **No global locking**, last writer wins



Big Data integration



— **IBM Spectrum Scale integrates with Hadoop File System (HDFS)**

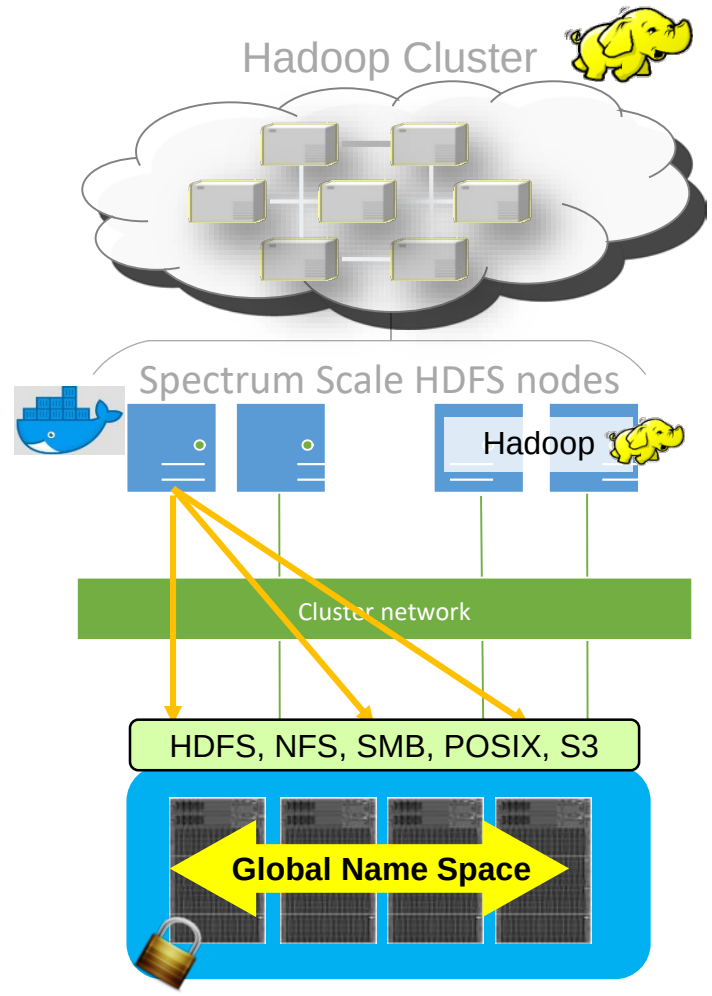
- Acts as HDFS data and name nodes
- Fully supported with Hortonworks Data Platform (HDP)

— **Hadoop workload can run directly on Spectrum Scale file system** via HDFS transparency connector

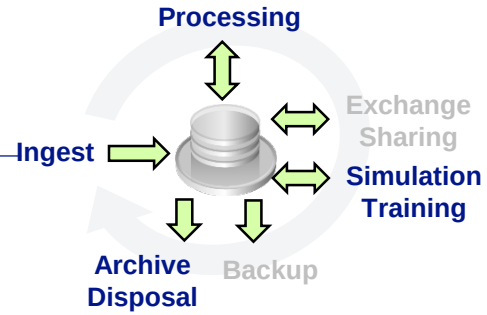
- Allows separation of compute and storage, providing
  - One storage for all big data and mass storage
  - Less storage required (Erasure encoding instead of 3 copies)
  - No data copy operations between Hadoop cluster and storage

— **Spectrum Scale can be used as persistent storage for Docker Container** through Storage Enabler for Containers

- Leverages open source software “Ubiquity”



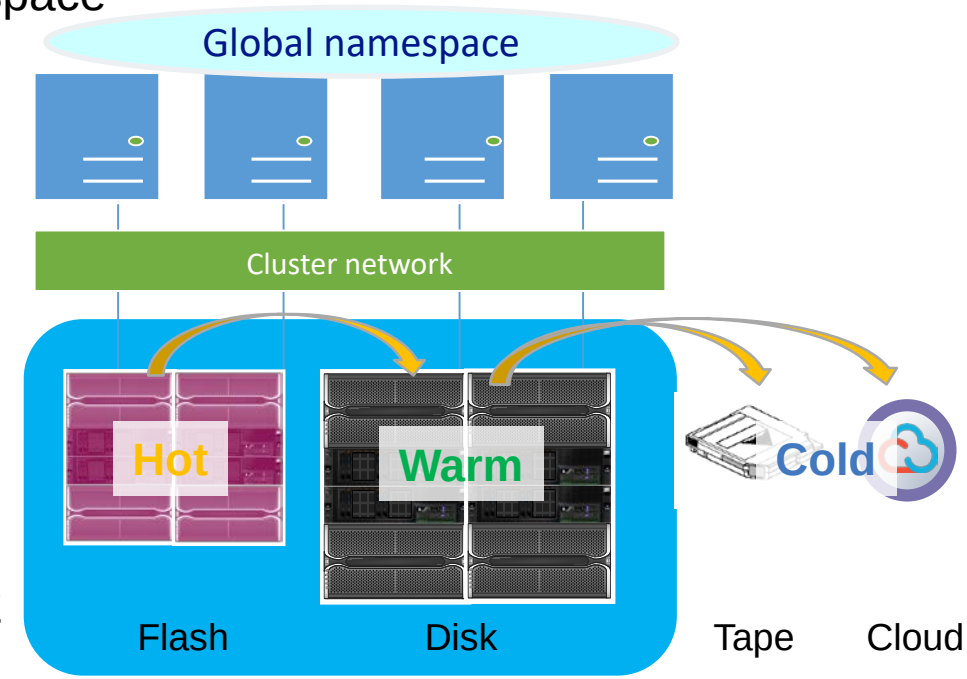
Storage tiering



— **IBM Spectrum Scale supports different performance tiers based on different storage media (SSD / Flash, Disk, Cloud, Tape)**
  - Allows for hot, warm and cold storage under the global names space

— Files are placed on **configurable performance tier** and **migrated** through storage hierarchy
  - Migration can be based on thresholds and selection criteria for files (such as age, size, type)

— **Files can be migrated to lower cost media** such as Cloud Object Storage using Transparent Cloud Tiering or Tape using IBM Spectrum Archive or IBM Spectrum Protect for Space Management



Integration with AI systems

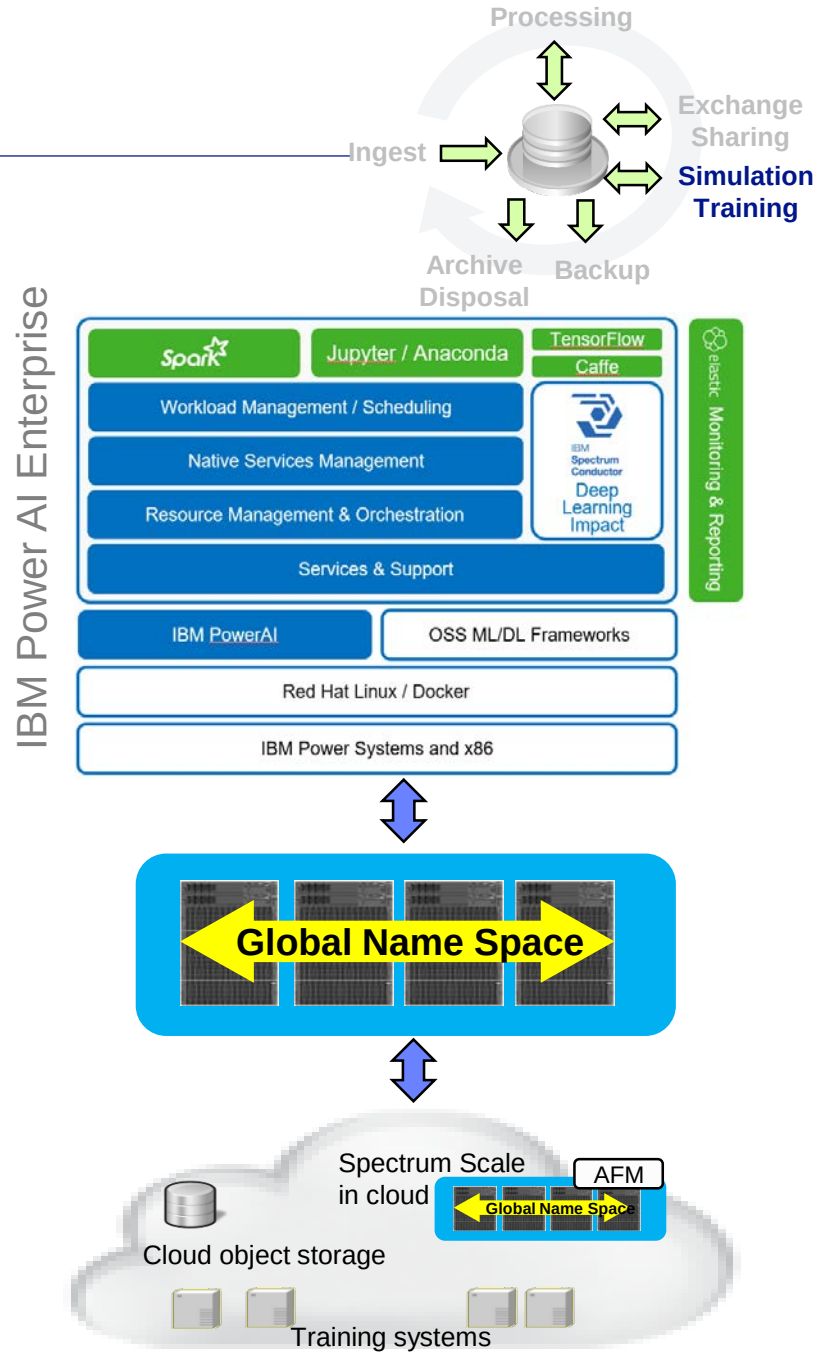
— IBM provides **infrastructure and reference architecture** supporting training use cases

[IBM Systems AI Infrastructure reference architecture]

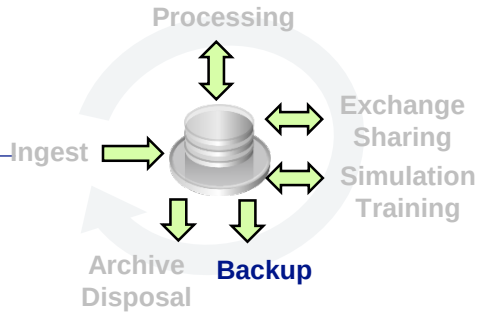
— AI training stack can run on **Nvidia DGX or IBM Power AC922**

- Both stacks integrate with IBM Spectrum Scale **directly**
  - Leverage parallel I/O for high performance
  - Benefit from scalability and reliability

— **Cloud extension** is possible through Spectrum Scale TCT, AFM and Aspera transfer acceleration



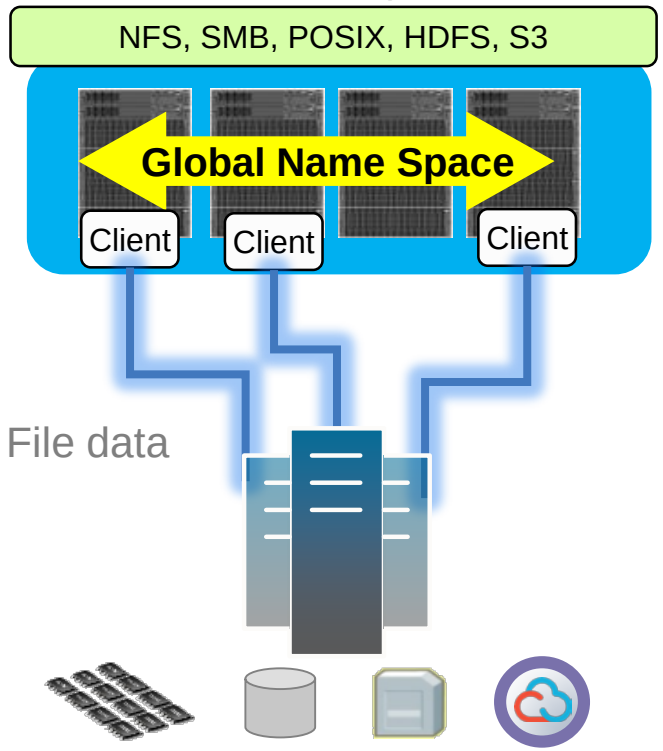
Scalable backup and restore



— IBM Spectrum Scale **integrates with** IBM Spectrum Protect **backup infrastructure**
  - Spectrum Protect backup client runs on Spectrum Scale servers
  - Spectrum Protect is placed in different location

— Backup process scales with amount of data
  - **Fast file identification**: based on policy engine
  - **Fast data transfer**: from multiple Spectrum Scale servers
  - **Scalable backup performance**: to multiple backup servers
  - **File and directory level restore**

— Backup data can be stored on Flash, Disk, Tape or cloud



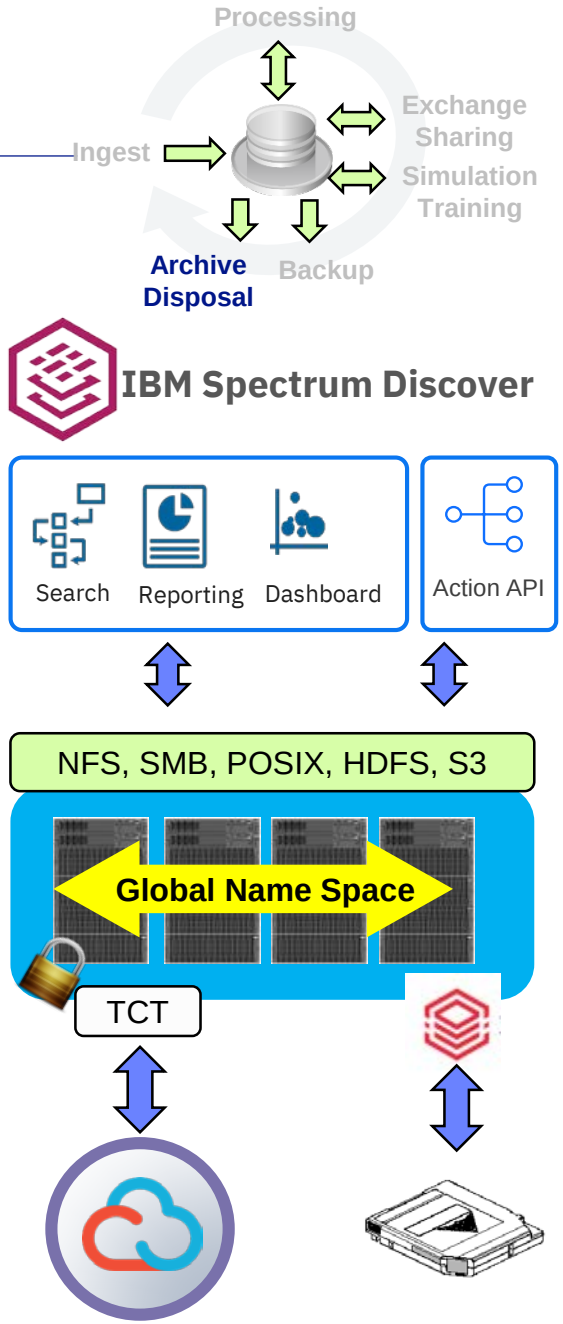
IBM Spectrum Protect server with backup storage infrastructure

Archiving

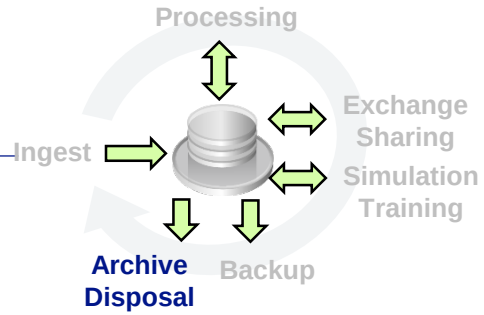
- Archiving means to **move data** that is still needed but less frequently accessed to a **lower cost storage medium**
  - Save storage cost over long period of time
  - Adhere to compliance requirements

- IBM Spectrum Scale provides full automation to move aged data to lower cost storage tier
  - **Archiving to tape** using IBM Spectrum Archive
  - **Archiving to cloud object storage** using TCT (IBM Cloud, AWS)

- IBM Spectrum Discover **automates the cataloging** of unstructured data by capturing system metadata as it is created
  - Enable comprehensive insight by combining system metadata with custom tags to increase storage admin and data consumer productivity
  - Provides search and reporting capabilities



Archival: compliance features

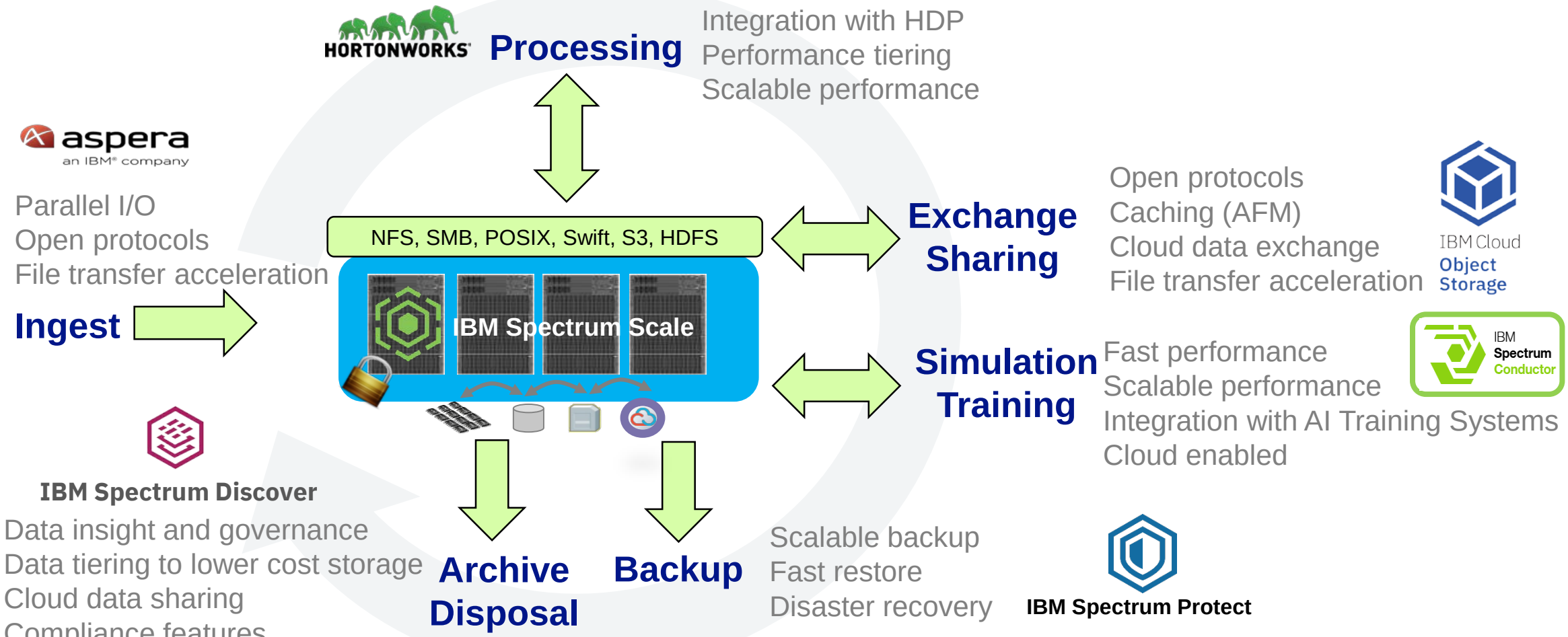


- Compliance means to comply to laws and regulations

- IBM Spectrum Scale provides functions supporting compliance requirements:
  - **File audit logging**: keeps records for each file access
  - **Immutability**: inhibits deletion of files during retention period
  - **Encryption**: encrypts individual files based on file attributes
  - **Authentication**: authenticates users prior to accessing files
  - **Authorization**: authorized operations on files based on access control lists
  - **In flight encryption**: encrypts files during transfer to Spectrum Scale file system

Summary

IBM Infrastructure facilitates key data management use case for autonomous driving development




Thank You

IBM Systems Solution Brief

AI in action: Autonomous vehicles

https://www.ibm.com/blogs/systems/ai-in-action-autonomous-vehicles/

IBM Storage Solutions for ADAS and Autonomous Driving (AD)

https://www.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=34019934USEN

IBM Big Data for Autonomous Driving

https://www.youtube.com/watch?v=eGhiIHDJaqI

Play Video

IBM Systems
Solution Brief

IBM Storage Solutions for ADAS and Autonomous Driving

Extremely Scalable, Cost-Optimized ADAS Data Management and Autonomous Driving Development Infrastructure

IBM

Highlights

- Leverage no-bottleneck architecture to scale performance for extreme throughput and low-latency access.
- Best-in-class storage infrastructure that grows and shares while automatically and quickly moving file and object data to optimal storage tiers.
- Enable data-anywhere access that spans storage and locations to accelerate ADAS data management across the data center or around the world.
- Leverage built-in authentication, encryption, security and replication options to meet ADAS data management and regulatory requirements.
- Cut costs up to 90% with intelligent archiving and automatic policy-based storage tiering from flash to disk to tape.
- Eliminate costly replication, significantly lowering the total cost of ownership for storage systems at the petabyte level and beyond.

The automobile is quickly morphing from an isolated, largely mechanical piece of equipment to one of the most technically sophisticated and connected platforms on the planet. From entertainment and navigation to driver assistance and crash avoidance, today's car is vastly different from those of a few years ago. The huge opportunity to share the future of that connected car, especially around autonomous driving, is drawing the interest of both technology and auto manufacturing companies.

The one thing these initiatives all have in common is data – miles and miles of data. Each sensor and system on connected cars generates a steady stream of information. The research and development behind future systems requires analysis of massive files and data sets. Dealing with the volume, velocity and variety of all this data creates a unique challenge.

The automotive industry is entering a new, highly competitive, transitional period where demand for new conveniences, safety capabilities and selling models are driving dramatic change. Once an industry consisting of pure hardware and adrenalin, automotive design is increasingly differentiated by software – with many visits to the dealership replaced by over-the-air bug fixes. At the forefront is Advanced Driver Assistance Systems (ADAS), which introduce disruptive requirements on engineering IT infrastructure – particularly storage, where even entry-level capacities are measured in petabytes.

Extreme scalability demands for autonomous driving (AD) development

Autonomous vehicle development requires a lot of data which is generated by the vehicle's hardware, including: a camera that generates 20-60 MB/s, sonar at 10-100 KB/s, radar upwards of 10 KB/s, LiDAR systems that range between 10-70 MB/s, and GPS that runs at 50 KB/s.

To put that figure into perspective, self-driving cars will consume and generate approximately 50-70 terabytes (TB) of data for every eight hours of driving.

Notices and disclaimers

- © 2018 International Business Machines Corporation. No part of this document may be reproduced or transmitted in any form without written permission from IBM.
- **U.S. Government Users Restricted Rights — use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM.**
- Information in these presentations (including information relating to products that have not yet been announced by IBM) has been reviewed for accuracy as of the date of initial publication and could include unintentional technical or typographical errors. IBM shall have no responsibility to update this information. **This document is distributed "as is" without any warranty, either express or implied. In no event, shall IBM be liable for any damage arising from the use of this information, including but not limited to, loss of data, business interruption, loss of profit or loss of opportunity.** IBM products and services are warranted per the terms and conditions of the agreements under which they are provided.
- IBM products are manufactured from new parts or new and used parts.
  In some cases, a product may not be new and may have been previously installed. Regardless, our warranty terms apply."
- **Any statements regarding IBM's future direction, intent or product plans are subject to change or withdrawal without notice.**
- Performance data contained herein was generally obtained in a controlled, isolated environments. Customer examples are presented as illustrations of how those
- customers have used IBM products and the results they may have achieved. Actual performance, cost, savings or other results in other operating environments may vary.
- References in this document to IBM products, programs, or services does not imply that IBM intends to make such products, programs or services available in all countries in which IBM operates or does business.
- Workshops, sessions and associated materials may have been prepared by independent session speakers, and do not necessarily reflect the views of IBM. All materials and discussions are provided for informational purposes only, and are neither intended to, nor shall constitute legal or other guidance or advice to any individual participant or their specific situation.
- It is the customer's responsibility to insure its own compliance with legal requirements and to obtain advice of competent legal counsel as to the identification and interpretation of any relevant laws and regulatory requirements that may affect the customer's business and any actions the customer may need to take to comply with such laws. IBM does not provide legal advice or represent or warrant that its services or products will ensure that the customer follows any law.

Notices and disclaimers continued

- Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products about this publication and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products. IBM does not warrant the quality of any third-party products, or the ability of any such third-party products to interoperate with IBM's products. **IBM expressly disclaims all warranties, expressed or implied, including but not limited to, the implied warranties of merchantability and fitness for a purpose.**
- The provision of the information contained herein is not intended to, and does not, grant any right or license under any IBM patents, copyrights, trademarks or other intellectual property right.
- IBM, the IBM logo, ibm.com and [names of other referenced IBM products and services used in the presentation] are trademarks of International Business Machines Corporation, registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the Web at "Copyright and trademark information" at: www.ibm.com/legal/copytrade.shtml.
- .