A Unified Data Platform for Big Data & Deep Learning

Gong Wei (gongwbj@cn.ibm.com)

IBM Spectrum Scale Development and Client Adoption

March 26, 2018
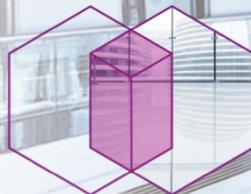
Agenda

- Challenges for Storage System in Big Data and Analytics

- Spectrum Scale Big Data and Analytics Solution

- Spectrum Scale for Deep Learning

- Customer Case Sharing

Challenges for Storage System in Big Data and Analytics
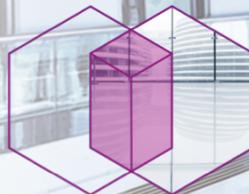
Typical Challenges in Big Data Analytics (1/4)

Inefficient data movements and many copies

Diagram labels: NFS/SMB — Storage System #1; HDFS — Storage System #2; POSIX — Storage System #3; Objects (Object 1, Object 2, Object 3, Object 4) — Storage System #4
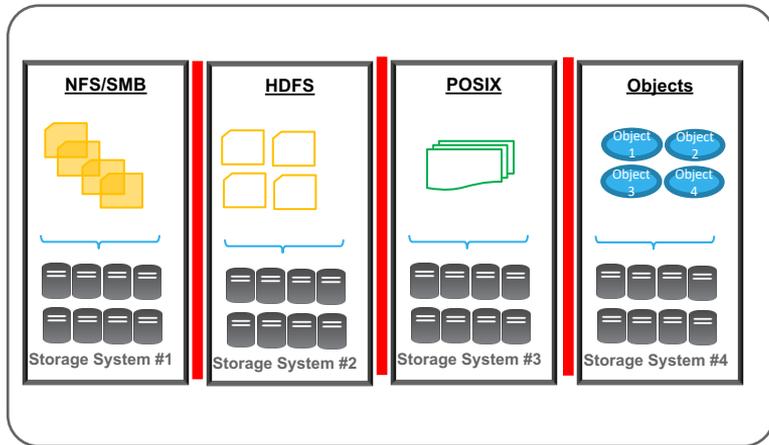
- ❖ Different analytics workloads might need different interface
  - ✓ map/reduce jobs need HDFS interface
  - ✓ Data ingestion clients need NFS/SMB interface
  - ✓ Traditional workloads(such as data warehouse, HPC etc) requires POSIX interface

- ❖ Unrich data accessing interfaces from storage system make customers have to set up siloed infrastructure

- ❖ Siloed data lakes bring inefficient data analysis(e.g. long time analysis because of data movement)

- ❖ Siloed data lakes bring inefficient storage space utilization(e.g. many copies on different storage systems)

- ❖ Multiple siloed storage systems bring further issues on cost, management and scaling.

Typical Challenges in Big Data Analytics (2/4)

Inefficient data lifecycle management
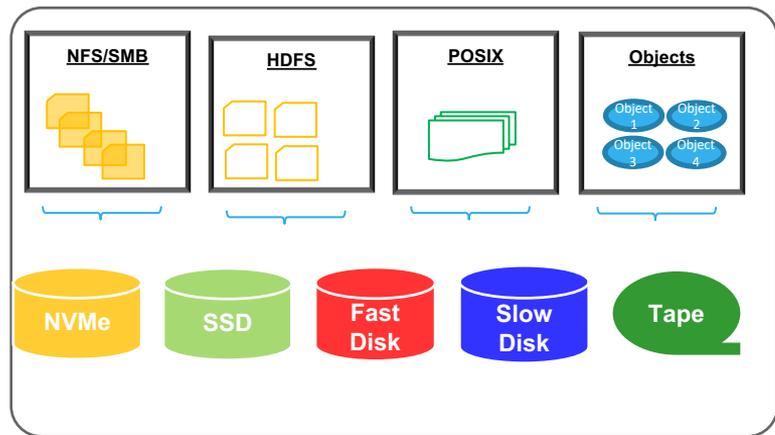
- ❖ Different tiers for different data(hot/warm/cold data)

- ❖ Leverage different disks from NVMe to Tape according to performance/cost

- ❖ Policy-based data movement automatically

- ❖ Easily move cold data back when needed

Typical Challenges in Big Data Analytics (3/4)

Cross-site HA or DR for business continuity
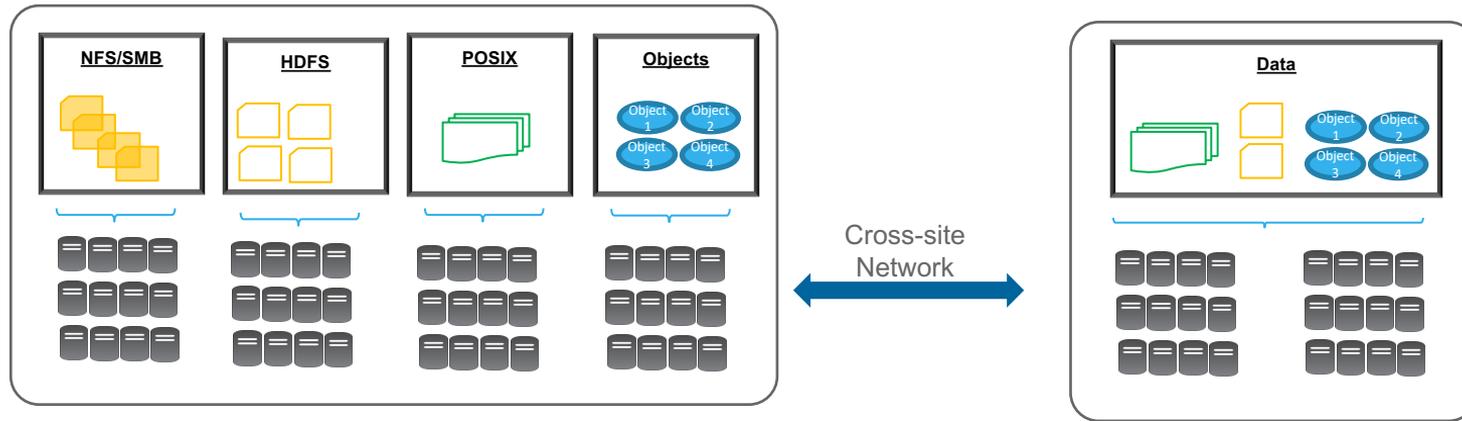
- ❖ Flexible options for different business continuity requirements
  - ✓ Different RTO(Recovery Time Object)/RPO(Recovery Point Object)
- ❖ Different options for low/medium/high cost for cross-site HA or DR

Typical Challenges in Big Data Analytics (4/4)

Inefficient Data Archive and Share to Cloud Storage
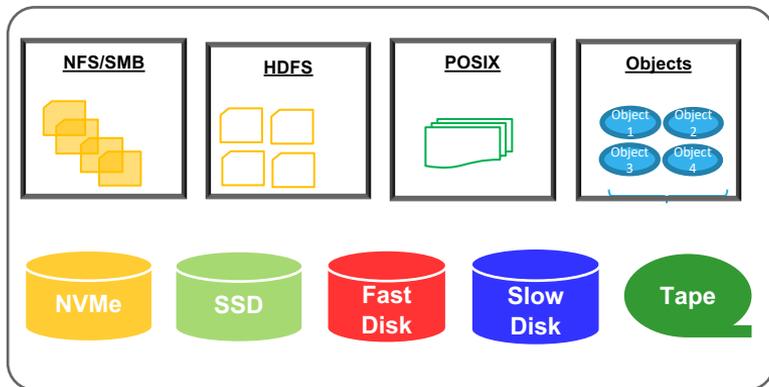
- ❖ Efficient data sharing cross different sites
- ❖ Efficient data sharing between off-premise and on-premise
- ❖ Flexible options considering different cost/requirements

# Spectrum Scale Big Data and Analytics Solution

Spectrum Scale Supports Big Data and Analytics

- Spectrum Scale completely transparent to Hadoop
  - Spectrum Scale HDFS Transparency Connector makes Spectrum Scale transparent to Hadoop
  - HDFS Transparency Connector Works with Spectrum Scale 4.1.X/ 4.2.x/5.0.x
  - Shipped with Spectrum Scale 4.2.x/5.0.x
    - Download the latest from IBM developerWorks GPFS wiki

- HDFS Transparency and Spectrum Scale services could be managed from HortonWorks HDP Ambari GUI

- Spectrum Scale and Hortonworks Data Platform are certified on both IBM Power and X86.
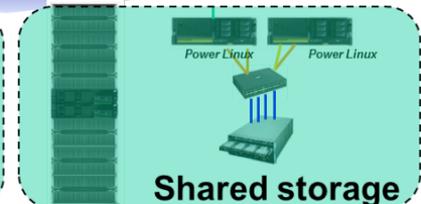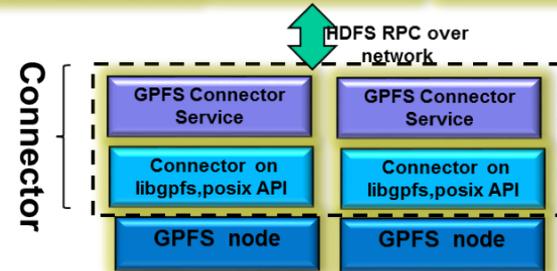
Unified Data Platform for In-lace Analytics
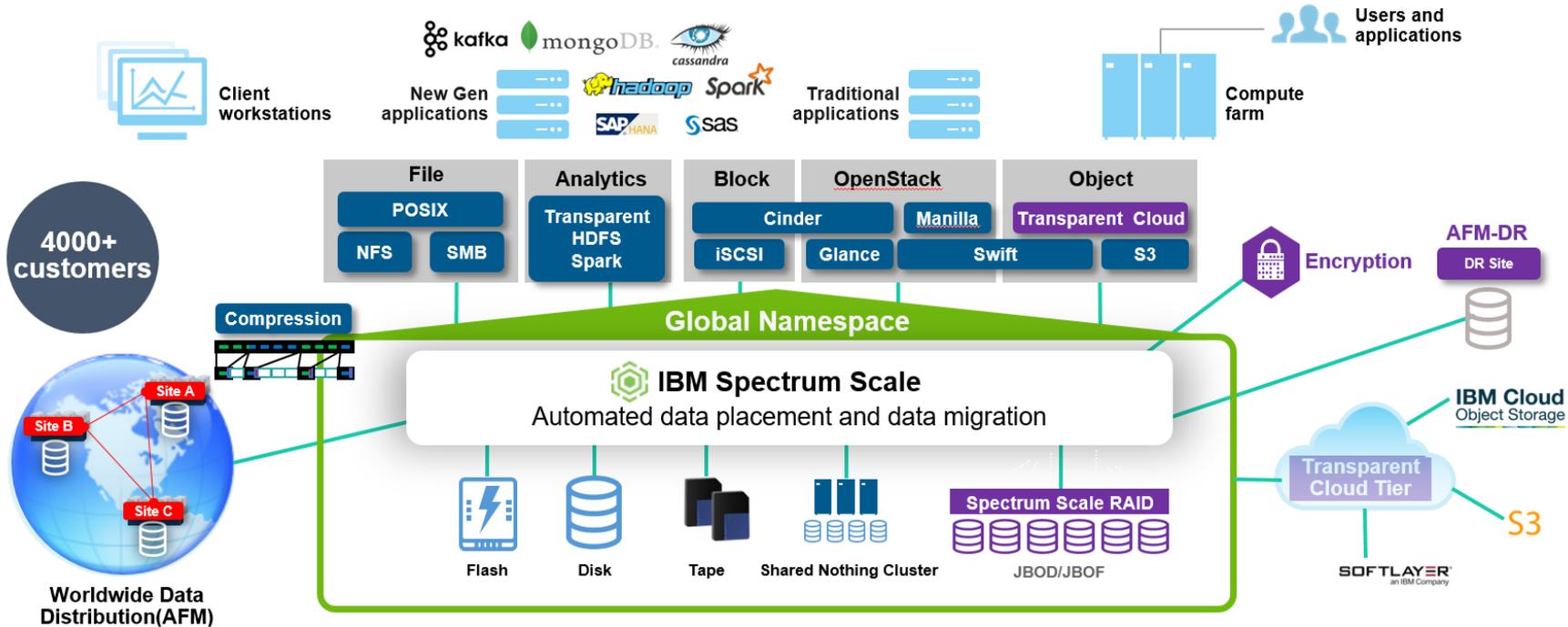
Spectrum Scale In-Place Analytics (1/2)

Traditional Analytics Solution

All-in-place Analytics Solution

Existing System

Analytics System

Data ingest

Export result

Analytics System

Existing System

HDFS Transparency

Spectrum Scale File System

File   Object
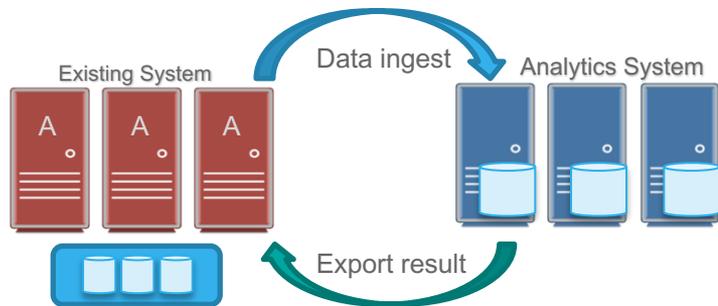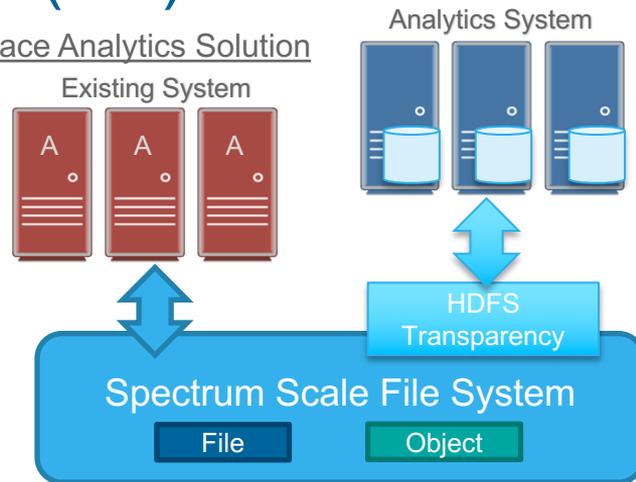
- Build analytics system from scratch, not only for compute but also for storage
- Add storage and compute resource at the same time no matter it's required
- Native HDFS doesn't support native POSIX
- Lacks of enterprise data management and protection capability

- ✓ Can leverage existing Spectrum Scale storage
- ✓ Unified interface for File and Object analytics
- ✓ POSIX compatibility
- ✓ Mature enterprise data management and protect solutions derived from Spectrum Storage family and 3rd part components

Spectrum Scale In-Place Analytics (2/2)

NFS interface from HDFS

Spectrum Scale In-place Analytics Solution
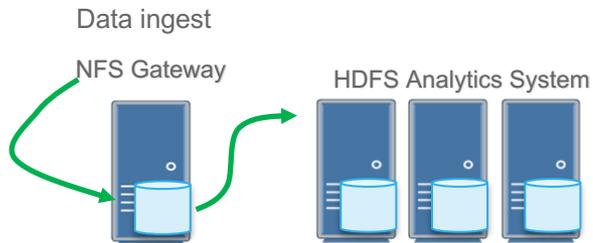
- HDFS doesn't support random read/write, only append mode
- HDFS NFS Gateway has to write data from clients to the local disks first and then move it to HDFS to handle the out-of-order write requests
- No HA for NFS Gateway so far
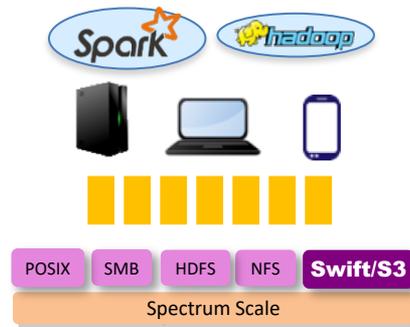
- ✓ Rich data ingest interface(SMB, NFS, HDFS, swift/S3 etc)
- ✓ Spectrum Scale Protocol HA
- ✓ Random read/write support
- ✓ Efficient data ingest because of no data movement from local disk to Spectrum Scale
- ✓ Only one data copy and all data are visible immediately from all interface

Spectrum Scale ILM for Data Analytics

Client workstations

hadoop

Compute Farm

Users and applications

Single name space

HDFS NFS POSIX SMB Swift/S3

Spectrum Scale

Storage pool1 *Storage pool2* *Storage poolx* *External Storage poolx*

cleversafe

amazon web services | S3 SOFTLAYER

- Data migration between different storage pool
- Policy-based Auto Tiering
- TSM+Tape based enterprise data backup for Hadoop

Flash **Disk** **Storage rich servers** **Tape**

IBM TSM/LTFS

- Sharing object data
  - Copy data from object storage to Spectrum Scale for analytics
- Store cold data in object storage
  - Object storage as one tier
  - Policy-driven object migration

Spectrum Scale DR Solutions for Analytics

- Hbase Cluster Replication
  - WAL-based asynchronous way
  - All nodes in both cluster should be accessible for each other
  - Both clusters could provide Hbase service on the same time
  - Only available for Hbase
 - Supported over HDFS Transparency + Spectrum Scale

- Spectrum Scale AFM IW-based replication
  - AFM IW-based replication from production cluster(cache site) to standby cluster(home site)
  - Both sites should be Spectrum Scale cluster for Hadoop application failover
  - Only one cluster can provide Hbase service(conflict in assigning region servers if Hbase is up on both cluster)

- Spectrum Scale Active-Active DR
  - 2 replica in production cluster; another replica in standby cluster
  - Dedicated network for two clusters(10Gb bandwidth)
  - Distance between two clusters is less than 100Km
  - Can achieve RPO=0 in DR
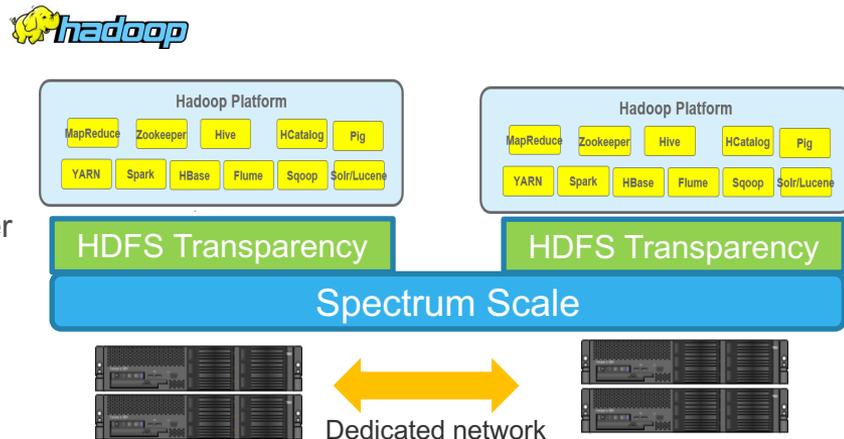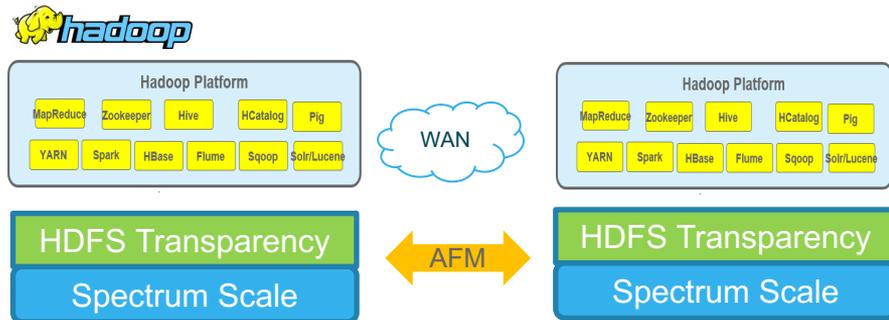
Spectrum Scale BD&A Solution New Features

2+ Hadoop clusters over the same file system
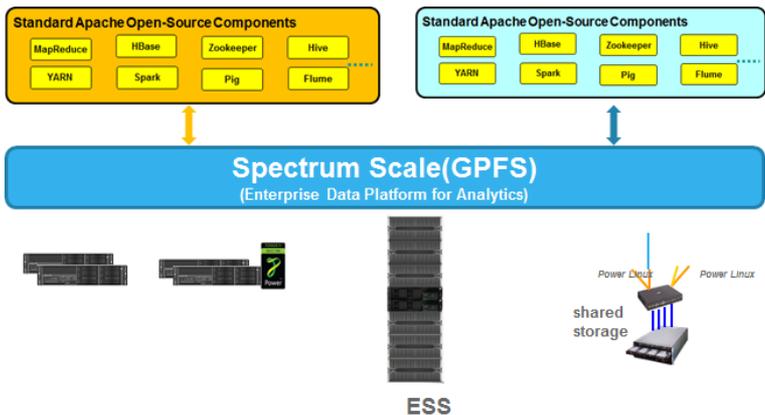
Natively Support 2+ File Systems

Standard Apache Open-Source Components

| MapReduce | HBase | Zookeeper | Hive |
| YARN | Spark | Pig | Flume |

Standard Apache Open-Source Components

| MapReduce | HBase | Zookeeper | Hive |
| YARN | Spark | Pig | Flume |

Spectrum Scale(GPFS)
(Enterprise Data Platform for Analytics)

Power Linux   Power Linux

shared storage

ESS

Hadoop Platform

| MapReduce | Zookeeper | Hive | HCatalog | Pig |
| YARN | Spark | HBase | Flume | Sqoop | Solr/Lucene |

hdfs://namenode:8020   /
/fs2

HDFS Transparency

/data/fs1   /data/fs2

GPFS   **GPFS**

ESS

Spectrum Scale BD&A Solution New Features

Short Circuit Write
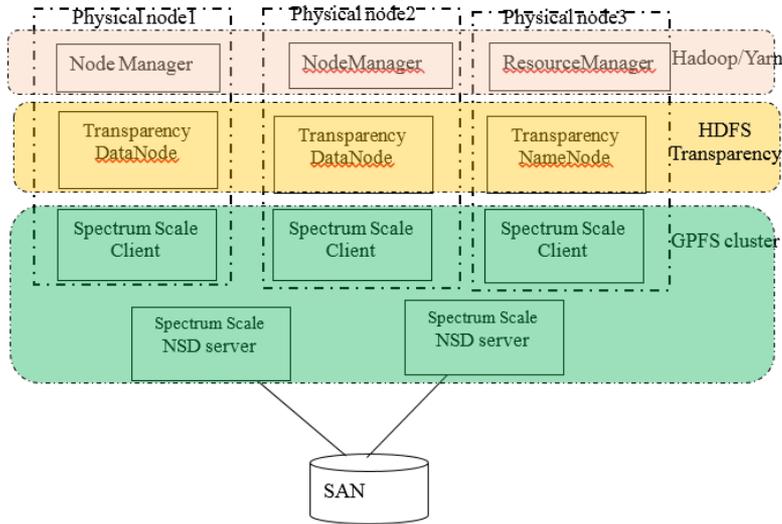
- All Hadoop nodes are Spectrum Scale nodes/HDFS Transparency nodes

- Short circuit write reduces the traffic between client and DataNode on the loop Io adapter
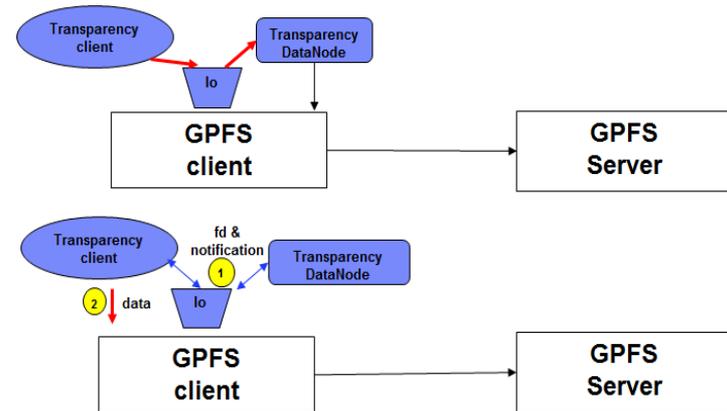
Spectrum Scale for Deep Learning

Spectrum Scale for Deep Learning

Spectrum Scale for Deep Learning

- ✓ SAN or internal SSD disks for entry level customers

- ✓ GSxS/SSD could be for customers that have large data size or whose data will increase in the near future

Advantages from Spectrum Scale

- ✓ Data read/write in parallel for performance

- ✓ Access the same data from all nodes

- ✓ SSD/NVMe

Requirements for Spectrum Scale

- ✓ High IOPS for small IO with low cost

- ✓ Take LROC to prefetch and cache the training data

- ✓ Workloads read the training data multiple times from local LROC(IO read acceleration)

Application

Computer Vision | Object Detection | Nature Language Processing | Voice Processing

Uniform Platform (AI & BDA)

BlueMind
Theano
MLLib
Graphx
Tensorflow
Caffe
Torch7

Spectrum Conductor with Spark

ESSL+(numpy, scipy, JBLAS)

IBM DB2 | SAP HANA | ORACLE | .....
Database

AFM (cross-site data sharing)
POSIX interface
Spectrum Scale
TCT (data sharing for Cloud)

cleversafe | amazon S3 | SOFTLAYER

Hardware

POWER+GPU
Power Power Power Power Power Power Power Power

SAN Storage | Elastic Storage Server | Internal Disks

Spectrum Scale for Cognitive and BD&A Solution

Spectrum Scale Key Advantages:
- Support long-term rapid increasing big data with extremely scaling for file system
- Fast analytics results from in-place analytics without data movement
- Easy maintenance from centralized storage management for multiple Hadoop cluster
- Support internal disk based for entry level customer(less than 100TB data size) and scale to PB level in ESS
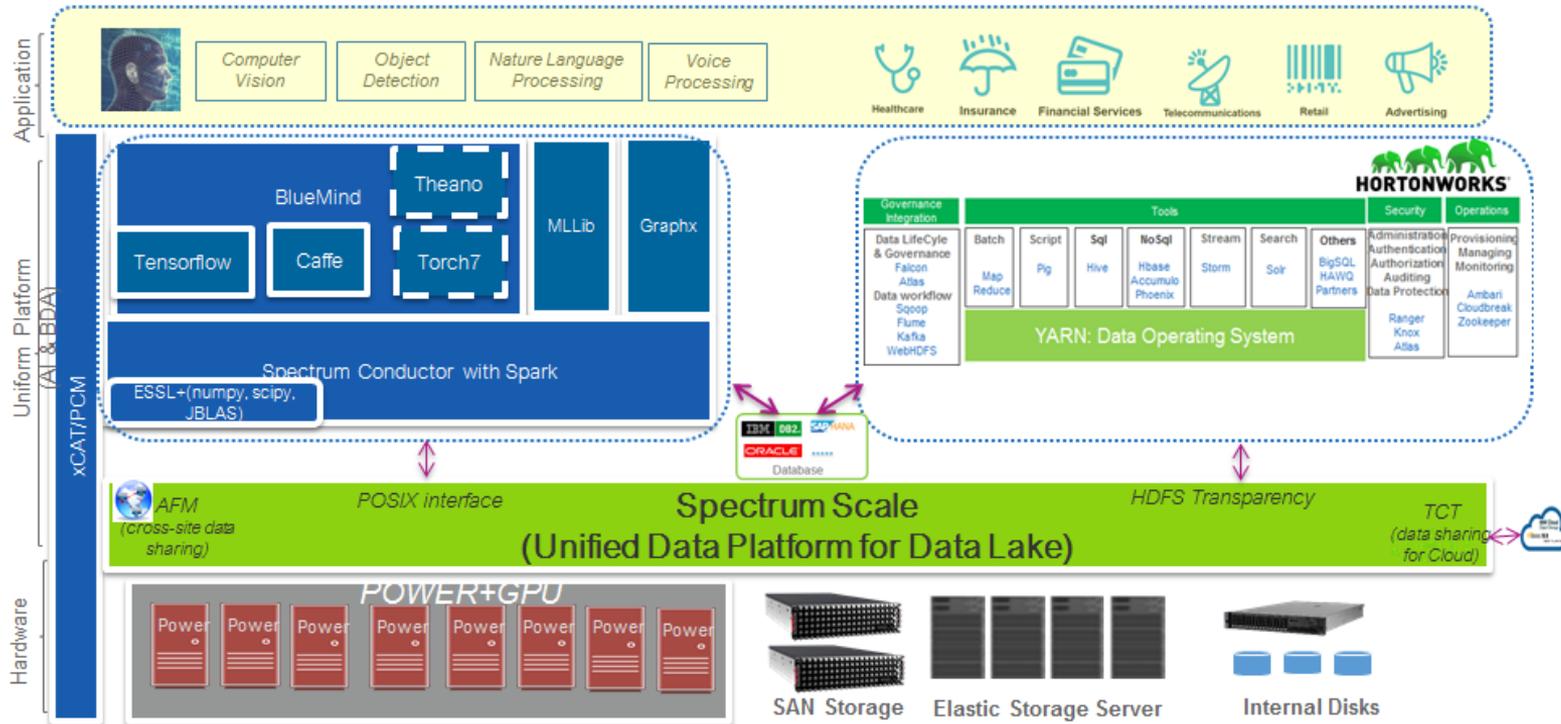
Customer Case Sharing
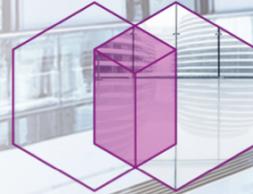
Case 1: Spectrum Scale Services Caffe

- Deep Learning Configurations
  - Take Caffe framework in the PoC
  - 16KB ~ 100KB per picture
  - 20GB totally(around 1M pictures)

- Spectrum Scale configuration
  - RDMA is not enabled yet
  - Sing SSD IOPS: 80K IOPS for 4K random read; 60K IOPS for 4K random write

- Requirements
  - 7000 pics/s for each single client with keeping GPU at full load

- POC Results
  - 7200 pics/s for single client
  - 50000 pics/s for 7 clients concurrently(only 40% Spectrum Scale disk bandwidth used)
  - Customer is satisfied with the results(They also tried other vendors solutions but cannot meet their performance requirements)

7 x86 compute node
E5-2650, 256GB, 4 P40 GPU

100Gb Infiniband Network

2 NSD Servers(replication 2)
2 Power 8 CPU
512 GB memory
6 800GB ssd disks
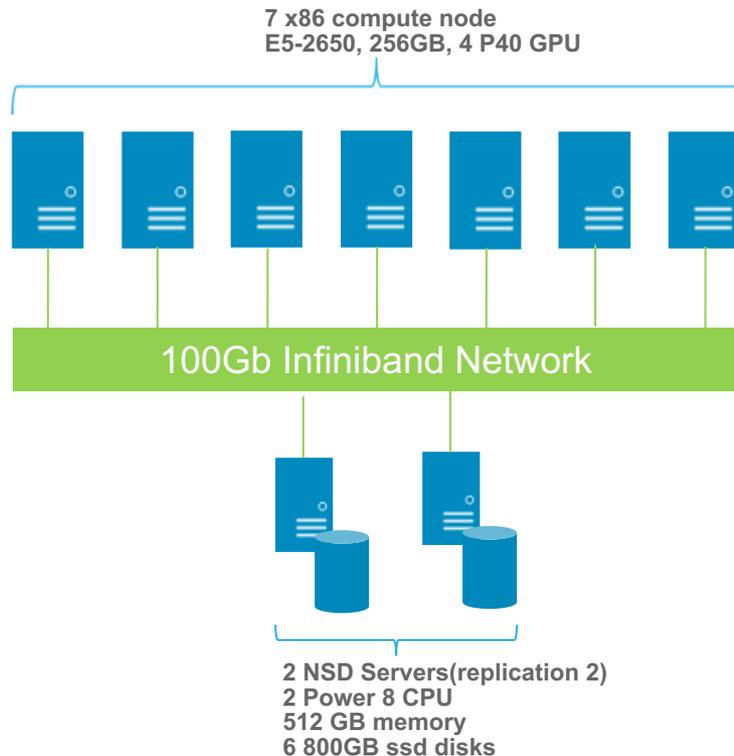
Case 2: Incremental Data Ingest from Native HDFS

Customer:

- Save data in HDFS
- Take IBM DLI with Spectrum Scale for their cognitive workloads
- They require a solution to do incremental data ingest from native HDFS
  - Snapshot diff based distcp
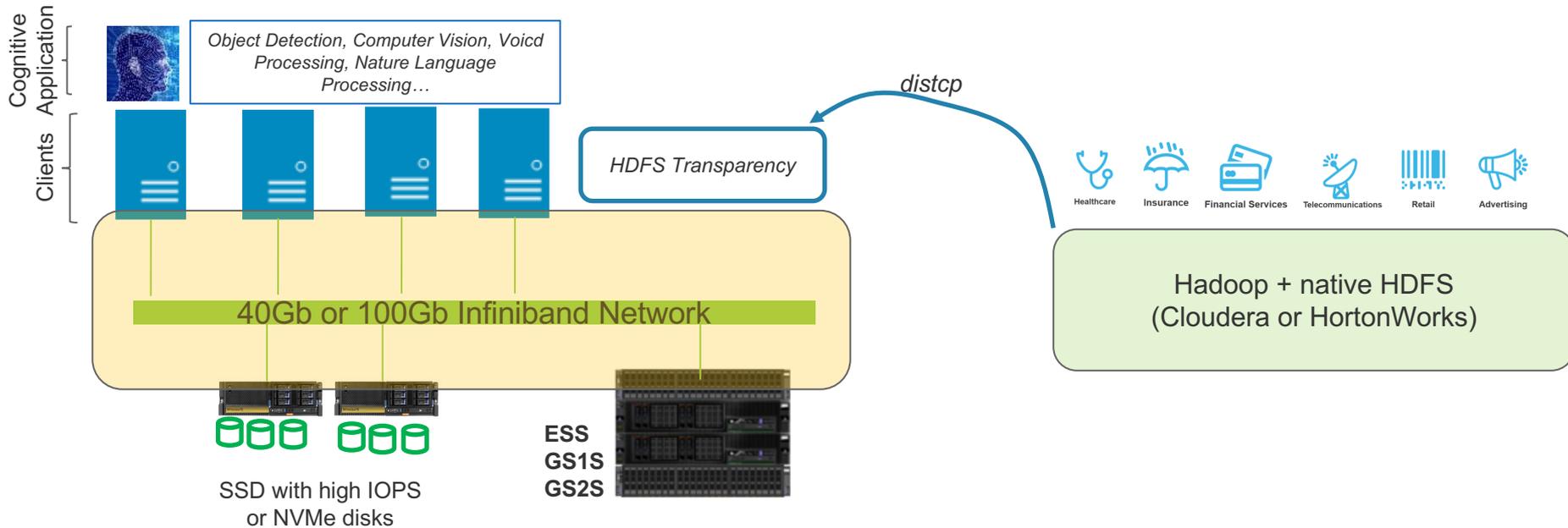
Bill Analytics System for a Telecom Operators

Network equipment's data collection

Data Collection Sub-system → **ETL System**

Internet activities capture and pre-processing (merge, etc)

✓ More than 100 Nodes
✓ Process 40 TB data each day
✓ Spectrum Scale FPO manages more than 2600 disks (2.6 PB raw capacity) in this system.

HBase	MapReduce

Spectrum Computing

Spectrum Scale FPO

| Power Linux | Power Linux | Power Linux |

Spectrum Scale Multi-cluster

| MapReduce | Spark | HBase | Hive |

Spectrum Computing

Spectrum Scale FPO

| Power Linux | Power Linux | Power Linux | Power Linux |

Built on 2013 runs Hadoop 1.x

Built on 2015 runs Hadoop 2.x

Import aggregated results into NoSQL database

NoSQL database

| Customer Behavior Subsystem | Real time sales/marketing sub-system | Other business sub-system |

Spectrum Scale and Spectrum Computing is the most stable Big Data system we have. Spectrum Scale POSIX access interface and Multi-cluster feature resolves data balance issue and provides 5X performance improvement in data loading phase.

-- Client IT Architecture

Thank You