HDF5 - AWS VOL Connector

Quincey Koziol
Principal Engineer
Amazon AWS - HPC
Agenda

• What are customer use cases & requirements?
• Solution
Customer Request – Cloud Native HDF5 Storage for HPC Applications in AWS
Request - Cloud Native HDF5 Storage for HPC Applications @ AWS

Many HPC applications use HDF5 for checkpoint+restart and for data analysis (visualization) output

- Want fast parallel I/O for HDF5 everywhere, including in the cloud

HPC applications running in the cloud also want to apply cloud services on their HDF5 data

- Leverage the rich cloud service ecosystem, e.g., at Amazon:
  - Redshift (data analytics), Sagemaker (AI/ML), Glacier (archive), etc.
Solution – AWS VOL Connector
High performance parallel I/O for HDF5 in the cloud

MPI applications running on AWS instances have native access to Amazon’s high-performance data center network & software

- Use tuned version of MPI for inter-node communication
  - Leverages the unique underlying network architecture

Applications running on AWS instances have high-performance access to Amazon’s services

- Use S3 objects for bulk data (dataset elements)
  - S3 bandwidth of ~100Gbs per node, e.g. aggregate performance of ~10Tbs for 100 nodes
- Use DynamoDB for metadata (groups / links / dataset description / attributes)
  - Low latency (<10ms), high throughput (>millions/sec) key-value tables
Access HDF5 information from AWS Cloud Services

Access HDF5 “raw data” from Cloud Analytics and ML Services:

- Redshift – Exascale data analytics
- Sagemaker – Hardware accelerated machine learning
- <anything that can access S3 objects>

Query HDF5 metadata with powerful, industry-standard operations

- DynamoDB queries are widely supported, and can be accelerated with indices
- Multi-container queries possible, to query HDF5 metadata across millions of “files”
Solution – AWS VOL Connector

High performance parallel I/O for HDF5, when running in the cloud

- Uses MPI, but not MPI-IO
  - Unless an MPI implementation would like to take up the idea of parallel I/O to S3 objects 😊
- Adds and exploits new properties as hints for applications to indicate their access behavior:
  - “Checkpoint”, “Streaming”, “Plot file”, “Subsetting”, ...

Cloud-native HDF5 containers, building with existing AWS capabilities and opening HDF5 data to broad cloud ecosystem

- Focus on unique aspects of HDF5, without wasting time recreating existing capabilities
Solution – AWS VOL Connector

HDF5 Core Library

HDF5 API and language bindings

Virtual Object Layer (VOL) Framework

Pass-through VOL connectors (e.g., async IO, caching, etc.)

Native Connector

VFDs

POSIX, MPI, I/O, SWMR, S3, HDFS

REST, DAOS, ....

Data Elevator

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.
Solution – AWS VOL Connector

HDF5 Core Library

HDF5 API and language bindings

Virtual Object Layer (VOL) Framework

Pass-through VOL connectors (e.g., async IO, caching, etc.)

Native Connector

VFDs

POSIX
MPI
I/O
SWMR
S3
HDFS

REST
DAOS
Data Elevator
AWS

VOL Connectors
Solution – AWS VOL Connector

HDF5 VOL Callbacks
- Metadata Operation
- Raw Data Operation

AWS VOL Connector

DynamoDB
S3

HDF5 “Container”

Redshift, Sagemaker, more!

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.
Thank you!
EXTENSIBLE ARRAYS
EXTENSIBLE ARRAYS
EXTENSIBLE ARRAYS
Thank you!
Thank you!
Thank you!
Thank you!
Thank you!