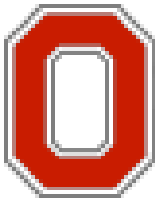


THE OHIO STATE UNIVERSITY

Improving Thor Data Loading using Parallel Format-agnostic Direct Spraying

Mohammad Rashti – RNET Technologies

Outline

- Data Loading in HPCC Systems
  - Potential Optimizations
- iNFORMER - The Big Picture of the SBIR project
  - iNFORMER Modules
- The HPCC on AWS Case – S3 Parallel Data Access
  - Feasibility study and experiments
- Ongoing and Future Work

Data Loading in HPCC Systems

Potential Optimizations

Data Loading in HPCC Systems

- In Thor, the data needs to be transferred into a single drop zone (Landing Zone)
- Landing zone node sprays the data over the DFS on the Thor nodes
- Imposes overhead on data loading
- Landing zone and spraying can become a bottleneck
  - Especially for large data sets or partial loads

Optimizations for Spraying Process

- Eliminate the landing zone bottleneck
  - Read data records directly from their original data format in their original place
- Use distributed data handling
  - Use multiple Thor nodes to load (and potentially decompress) the data
  - Each Thor node can read its own data
  - Load the data directly into destination files on DFS
- Optimize partial data loads
- Add support for in-memory processing
  - Skip trips to disk for iterative computing

iNFORMER

The Big Picture of the SBIR Project

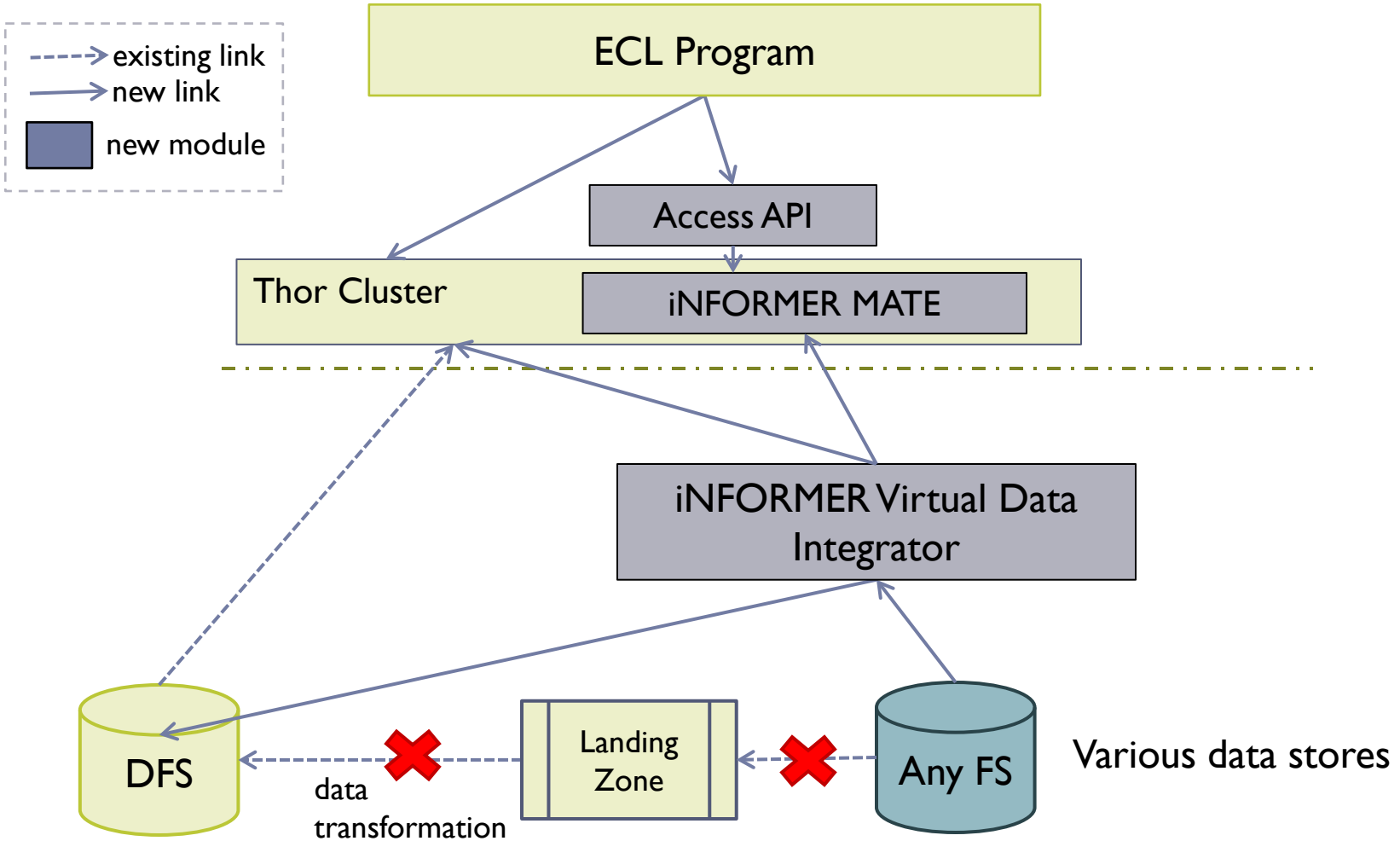
SBIR Project

- This work is part of an SBIR project from the Department of Energy
- Carried on in two phases
  - Phase I (9 months)
    - Feasibility study of the ideas and commercial potential
    - Design of the prototype
  - Phase II (2 years)
    - Prototype implementation
    - Commercialization of the product
    - Competitive renewal proposal submission
- The Phase I is currently in progress (ending in Dec.)
- Phase II may start in 2Q 2015

iNFORMER's General Architecture

- iNFORMER is the ultimate product developed in this project
- It is designed to provide:
  - Efficient batch and in-situ data analytics
  - Low-overhead native-format data access through a unified API
- iNFORMER components can be integrated into HPCC Systems
- Next slide shows a big-picture diagram of:
  - iNFORMER components (the dark boxes)
  - Where it can fit in HPCC systems.

iNFORMER in HPCC – The Big Picture



Big Picture view of iNFORMER components and potential integration with HPCC

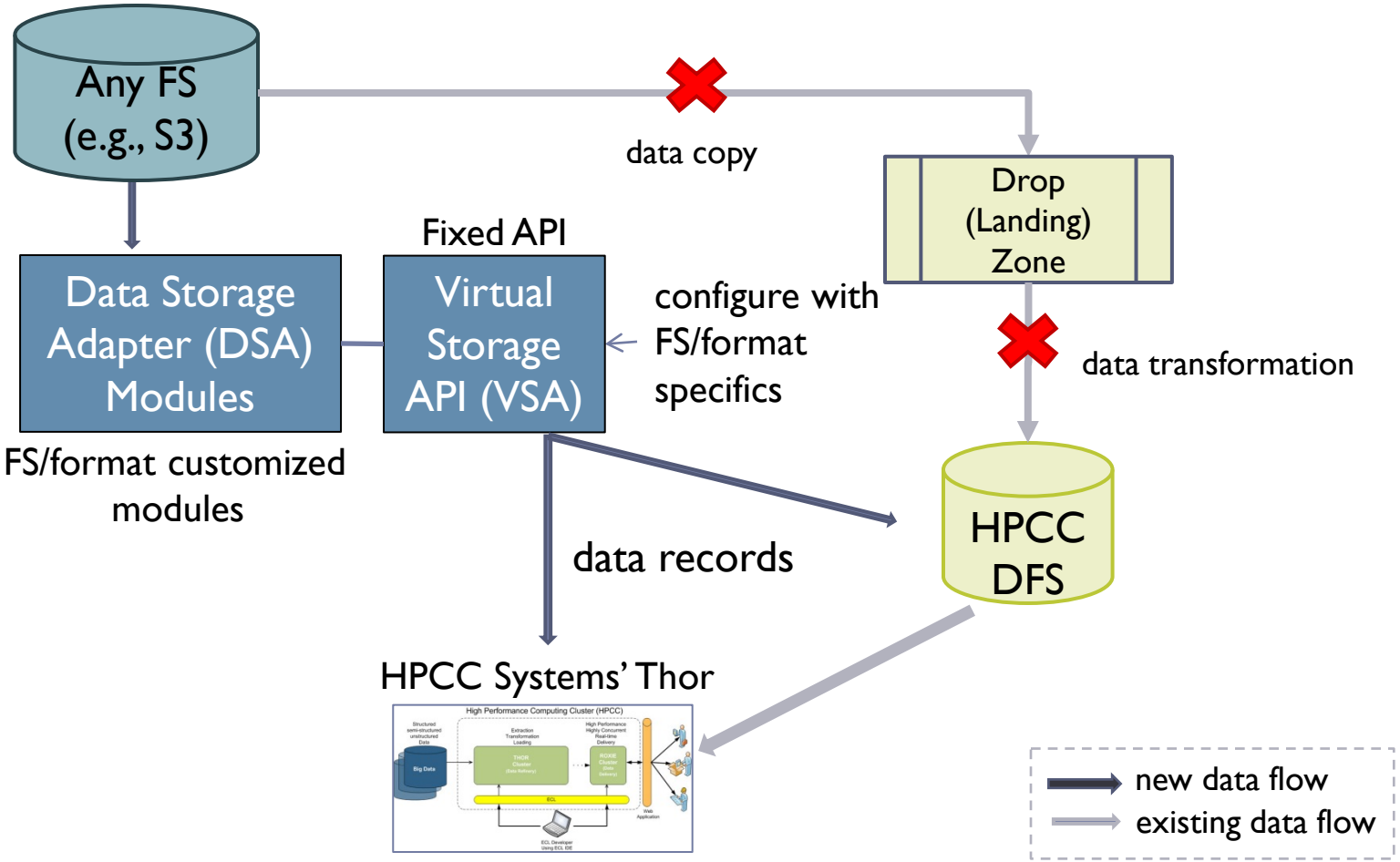
iNFORMER MATE

- MATE is meant to replace MapReduce
  - Using the concept of Reduction Object
  - Replace map and reduce with generalized reduction
  - Avoid intermediate shuffle/sort of key/value pairs
- Main benefits of the Reduction Object
  - Not tied to the limitations of the key/value model
  - More efficient data processing with less data movement
  - Lower memory requirements
  - Performance improvement: 1.50 to 20 fold

Virtual Data Integrator (VDI)

- VDI offers a unified API for native-format data access
  - Access the data in their original storage format
- Improves data access experience over various file-systems/data formats
  - Parallel File Systems: Thor DFS, HDFS, PVFS, GPFS, etc.
  - Data Formats: Plain Text, XML, NetCDF, HDF5, etc.
- Eliminates the need to transform the entire data into a platform-specific format (e.g., DFS)
  - Especially for in-memory processing
- VDI has two parts:
  - Data Storage Adapter (DSA)
  - Virtual Storage API (VSA)

iNFORMER Virtual Data Integrator (VDI) with HPCC

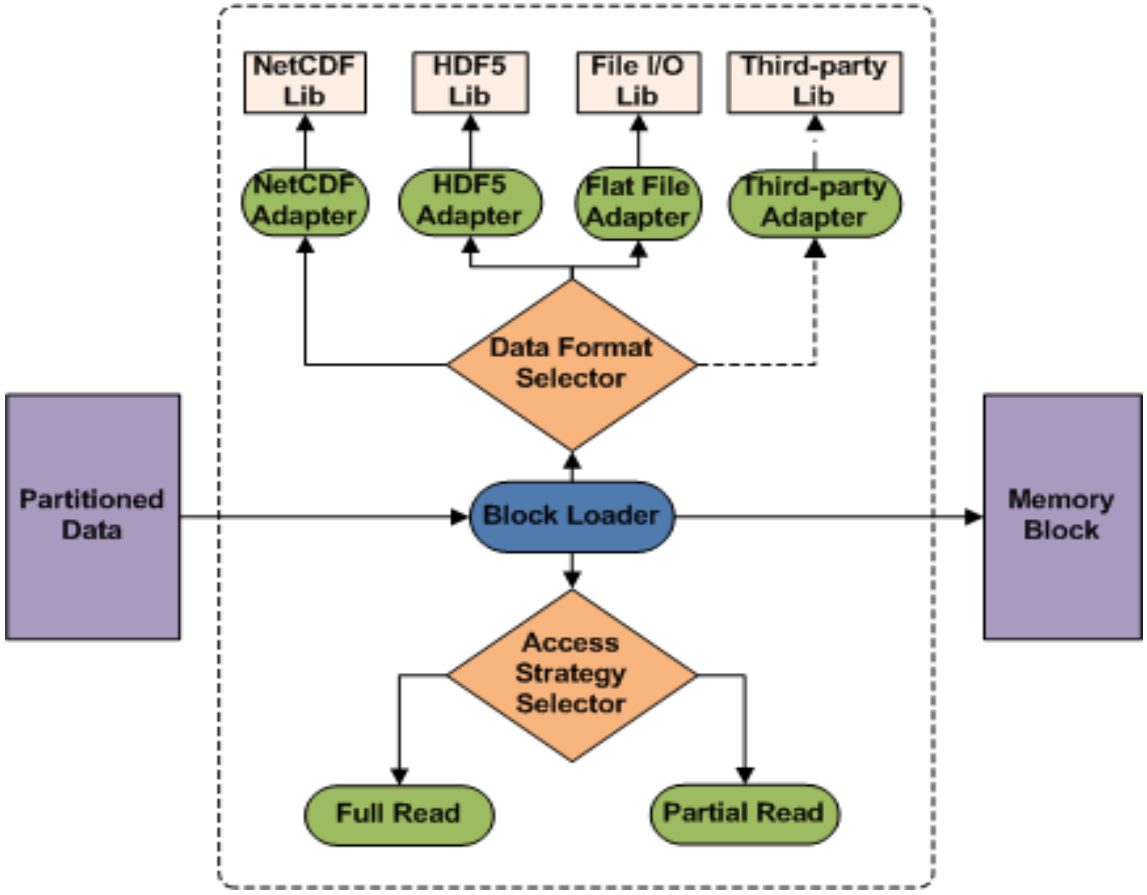


iNOFORMER VDI and potential integration with HPCC Systems

Data Storage Adapter (DSA)

- A set of adapters that interface with various storage formats
  - Access the respective data records in their original format
- Consolidates data integration process
- Allows for partial data load/access
- Eliminates the need for excessive data transformation
- Data records on the distributed FS will be extracted and forwarded to the analysis engine and back.

Some Internals of the DSA Component



Internal Architecture of the DSA File Format Adapter

Virtual Storage API (VSA)

- VSA is a unified FS/format agnostic interface:
  - The data access calls are not bound to a format
  - The storage format specifics will guide the VSA implementation to choose the right DSA module
- VSA will have well-defined hooks for easy implementation of modules for new formats
- A set of pre-implemented FS/format adapters will be developed.

The HPCC on AWS Case

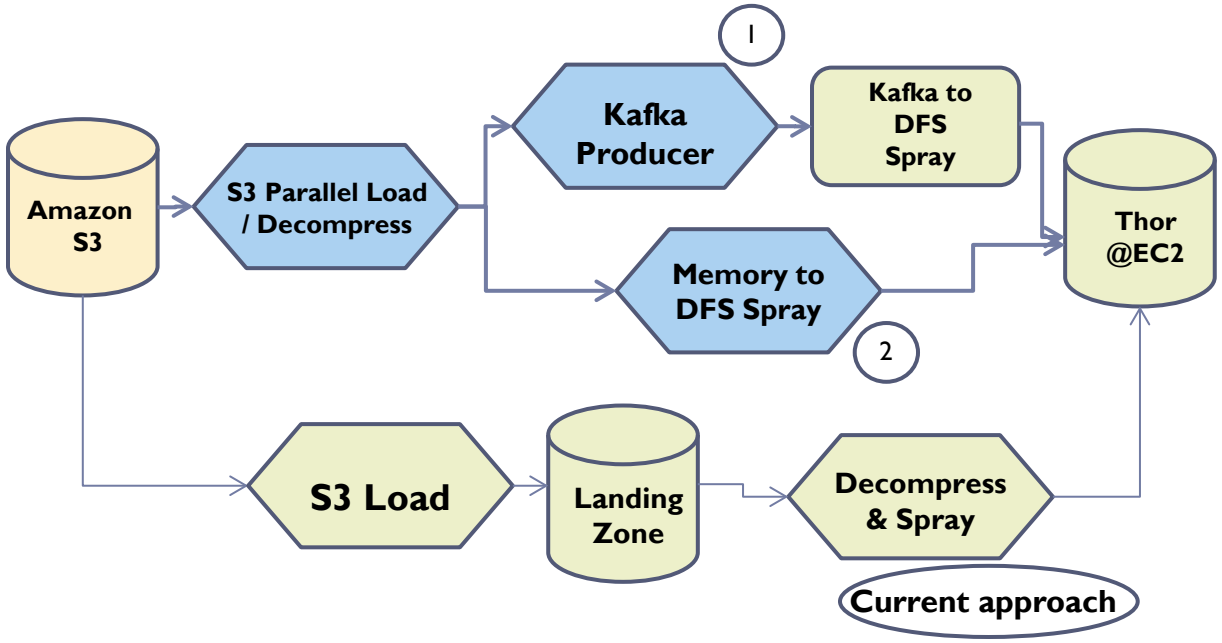
S3 Parallel Data Access

Phase I - HPCC Feasibility Study

- Efficient loading of AWS S3-resident compressed files into Thor
- Data loading / spraying optimizations in this project
  - Distributed data load / decompress / read from S3 into the memory of Thor nodes
  - Distributed data Spray to DFS
- Each node should load its own portion of data to avoid shuffling (communication)
  - Use a zip index file to facilitate this
- Study the potential for in-memory processing
  - Avoid multiple disk accesses

S3 to EC2 Distributed Data Load

- Two potential plans for the proposed distributed S3 loading module
- Test data: Elsevier sample data sets hosted on S3
  - XML-based records
  - Multi-entry zip files



Development Steps

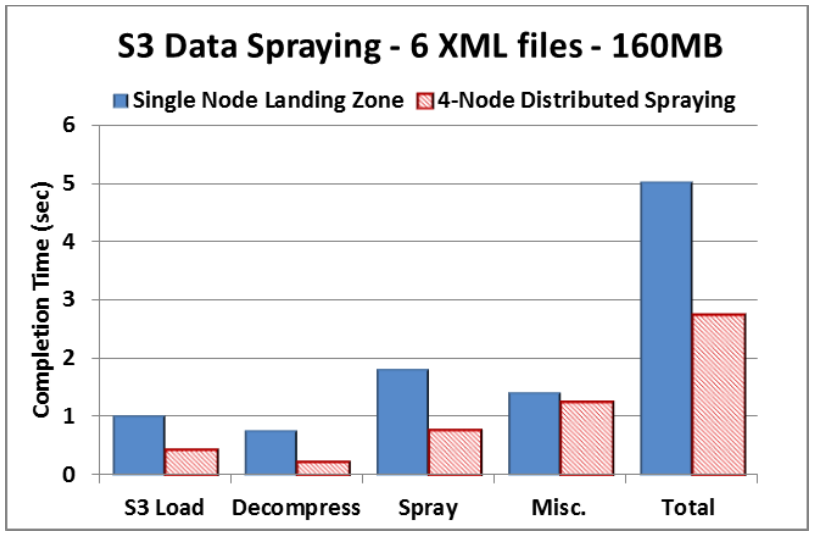
- Step I: Distributed Spraying (currently implemented)
  - Parallel loading, decompression and spraying of XML files from an S3 archive
  - Each process extracts an XML file data directly from the archive
    - One process per node
  - Load balancing applied among the processes
  - Message Passing Interface (MPI) used for inter-node communication and shuffling
- Step II: Skipping the Shuffling and the Landing Zone (next step)
  - Each process reads it own relevant portion of the file directly from the archive and decompresses on the fly.
  - Allows for partial data load from S3
  - May need full compression index for the archives
    - A separate module to be developed to do this

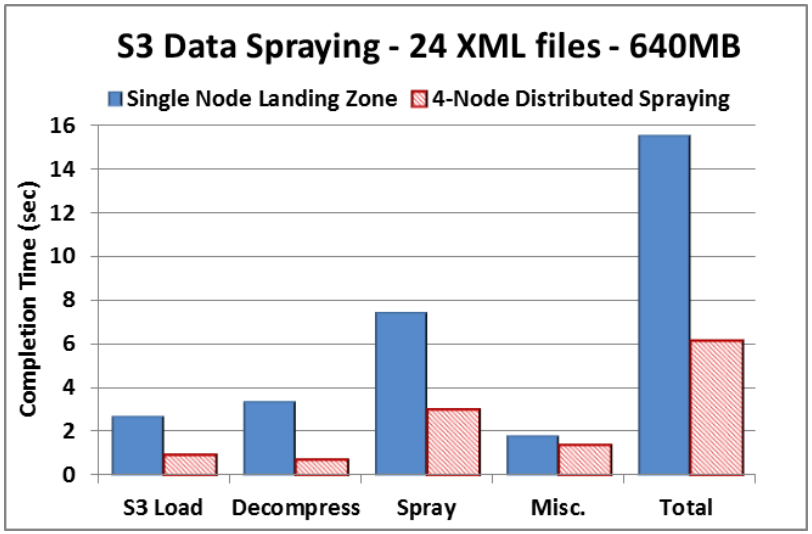
Step I – Distributed Spraying

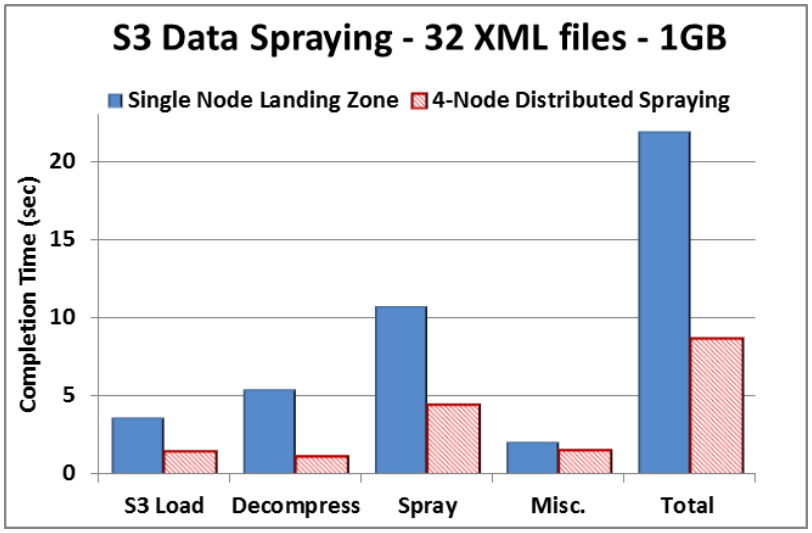
Processing on four (4)

EC2-based Thor Nodes

Step I Results







Comparing Sequential and Parallel Load/Spray from S3 into Thor

Ongoing and Future Work

Future Development Steps

- Next immediate steps (SBIR Phase I)
  - Complete spraying into Thor DFS
  - Evaluation using more data sets
  - Fully eliminate the scattering process by partial file reading
  - Design the zip indexing module
  - Full VDI module prototype design
- Future (SBIR Phase II) Plans
  - S3 adapter module as part of the VDI
  - Expansion of the adapter module to support general non-S3 cases
  - Full integration with HPCC to bypass the landing zone
  - Support for in-memory spraying and data processing

RNET Technologies, Inc.

Contact Info

Mohammad Rashti
240 W. Elmwood Dr., Dayton, OH
(937) – 433 - 2886
mrashti@RNET-Tech.com