

# Comparison of High-Performance Distributed File Systems on two Platforms: Linux and Windows

The superiority of EOS-based Comtrade Distributed File System (CDFFS) for Earth Observation Data Storage

Gregor Molan

Comtrade Group / Comtrade360



# The Rising Quantity of Earth Observation Data



- The proliferation of Earth observation satellites has led to a surge in data generation, posing significant data storage and management challenges.
- History of Earth observation data collection
- The exponential growth of new satellites, including nanosatellites

# Need for High-Performance Data Storage in Earth Observation

---

The demand for high-performance data storage in Earth observation is essential for:

- Extracting,
- Storing,
- Processing,
- Analysing



Insights from the vast volumes of satellite data.

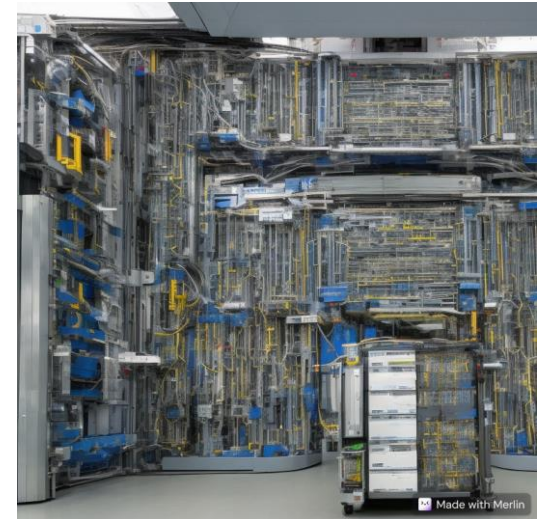
# Data Volume Challenge

---

- Increasing number of satellites
- Increasing data collection size
- Increasing data requests from users

The consequence:

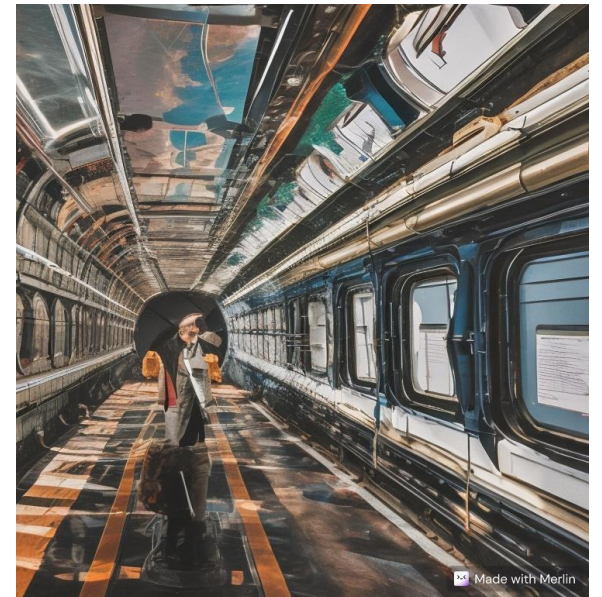
- Managing the sheer volume of Earth observation data efficiently is a critical challenge.



# Data Accessibility Challenge

---

- Crucial needs for timely analysis and decision-making are:
  - Fast access to space data
  - Reliable access to space data.



# Comparison of High-Performance File Systems

---

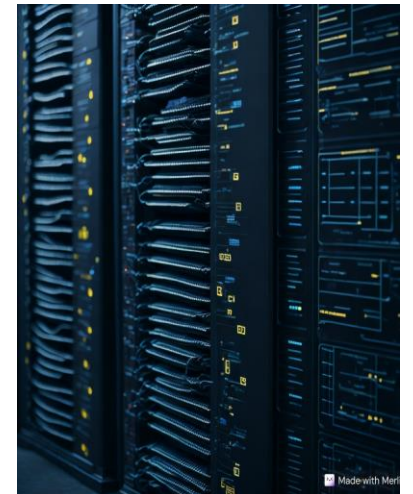
- Comparison of the following aspects in rigorous speed-testing
  - EOS based [CDFS](#)
- EOS distributed file system has been subjected alongside other high-performance file systems, such as [Ceph](#), [Lustre](#), [IBM Spectrum Scale](#), and [Hadoop](#), demonstrating its superiority in various aspects.



# Advantages of EOS (CDFS) for Earth Observation Data Storage

---

- CDFS offers seamless scalability to accommodate the growing volume of Earth observation data.
- Reliability for Earth Observation Data Storage.
- CDFS ensures data integrity and availability, which is critical for mission-critical applications.
- CDFS provides high-performance storage capabilities, enabling rapid data access and analysis.



# CDFS's Superiority in Handling Earth Observation Data

---

- CDFS's architecture and features
- Well-suited for handling
  - The unique requirements of Earth observation data
  - Handling large volumes
  - Handling diverse formats
  - Satisfying real-time processing needs



# CDFS: EOS Implementation Prepared to Collect Earth Observation Data

---

The initial set-up prepared at Comtrade as a PoC to collect and handle Earth observation data:

- Management Nodes:
  - 32 threads, 2x Intel® Xeon® Silver 4208 Processor
  - 384 GB RAM, 2x 2 TB SSDs
- Storage Nodes:
  - 32 threads, 2x Intel® Xeon® Silver 4208 Processor
  - 64 GB RAM, 1x 2 TB SSD, 6x 2 TB HDDs
- Client Node Nodes:
  - 12 threads, Intel® Core™ i5-12400 Processor
  - 16 GB RAM, 1 TB SSD

# Prospects and Enhancements for CDFS in Earth Observation

---

Potential future enhancements and developments in EOS that could further improve its capabilities for Earth observation data storage and management:

- Enhance EOS installation.
- Specify the EOS parameters to scale between throughput speed, latency, and reliability.
- Specify the EOS source code files and header files to additionally finetune the dependences between throughput speed, latency, and reliability.

# Key Takeaways and Conclusion

---

- Overview of the key takeaways of EOS to emphasise EOS's role as a superior Earth observation data storage solution.
  - Advantages of spreading EOS and CDFS to Earth observation data storage.
  - Recommendations for the EOS team considering extending EOS for Earth observation data needs.
- 

Gregor Molan

`gregor.molan@comtrade.com`  
`gregor.molan@cern.ch`