Project Nebula - A Distributed File System

Trung Nguyen

College of Engineering and Applied Science, University of Cincinnati

What is it?

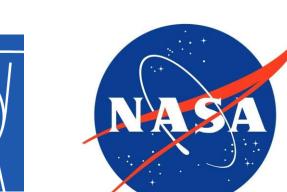
- A file system that is distributed on multiple file servers or multiple locations.
- Allows to access or store isolated files as they do with the local ones, access files from any network or computer

Why it exist?

- Cloud storage amazon s3
 Big Data

- CDNs
- ML Model Storage





Metadata

DB

Storage Server

Health Monitor

Chunk

Checksum

Chunk Manager

Storage Node 1

Storage Node 2

Storage Node N

Video Streaming







Fault

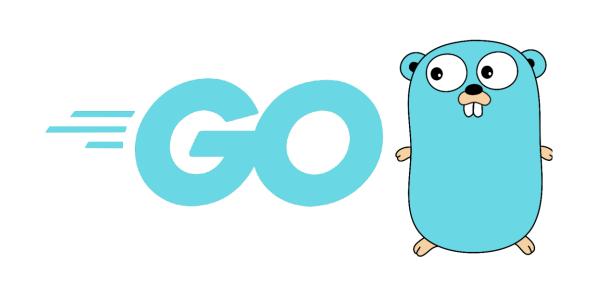
Detector

Key Features

- Scalability: Handles increasing storage needs efficiently.
- Fault Tolerance: Data replication and recovery mechanisms.
- **High Performance:** Optimized read/write operations with caching.
- Security: Access control and encryption.

Client Master Server Metadata request Application Chunk Access gRPC Locator Control API Metadata response Replication Control Chunk Storage File Data Replication Status

Tech Stack











Project Link

Replica

Selection

Replication Manager

Failure

Recovery



Advisor:

Heartbeat

Replica Operation

- Khang Nguyen Software Engineer @Amazon
- Pacific Mugisha Software Engineer @Apple