# The Google File System & A Comparison of Approaches to Large-Scale Data Analysis

MICHAEL SHERSHIN

12/11/2014

# Main Idea

- Building a system from inexpensive commodity components that often fail
- Can handle large amounts of querys at any given time
- The masters maintains all the file systems
- Each application implements the file system API
- Hierarchical organization

# Implementation

- Master contains
  - All file system metadata
    - Name space
    - Access control information
    - Mapping from files to chunks
    - All current locations of chunks
  - Garbage collection
- Connects via API in every applicaton

# My Thoughts

- Likes
  - Centralized master
  - APIs
- Dislikes
  - Cheep components that could fail often

# Comparisons

- Poor fault-tolerance
- Efficient
- Higher-level language
- Data feeds through fixed standard queries
- Large amount of users per hour

# Advantages and Disadvantages

## Advantages

- User defined programming
- Elastic

## Disadvantages

- Data feeds though fixed standard queries
- Poor fault-tolerance