THE GOOGLE FILE SYSTEM SUMMERY ANALYSIS

By Brian Powers

SOSP'03, October 19–22, 2003, Bolton Landing, New York, USA.

The Main Idea of the Google File System

1

Innovation through simplicity

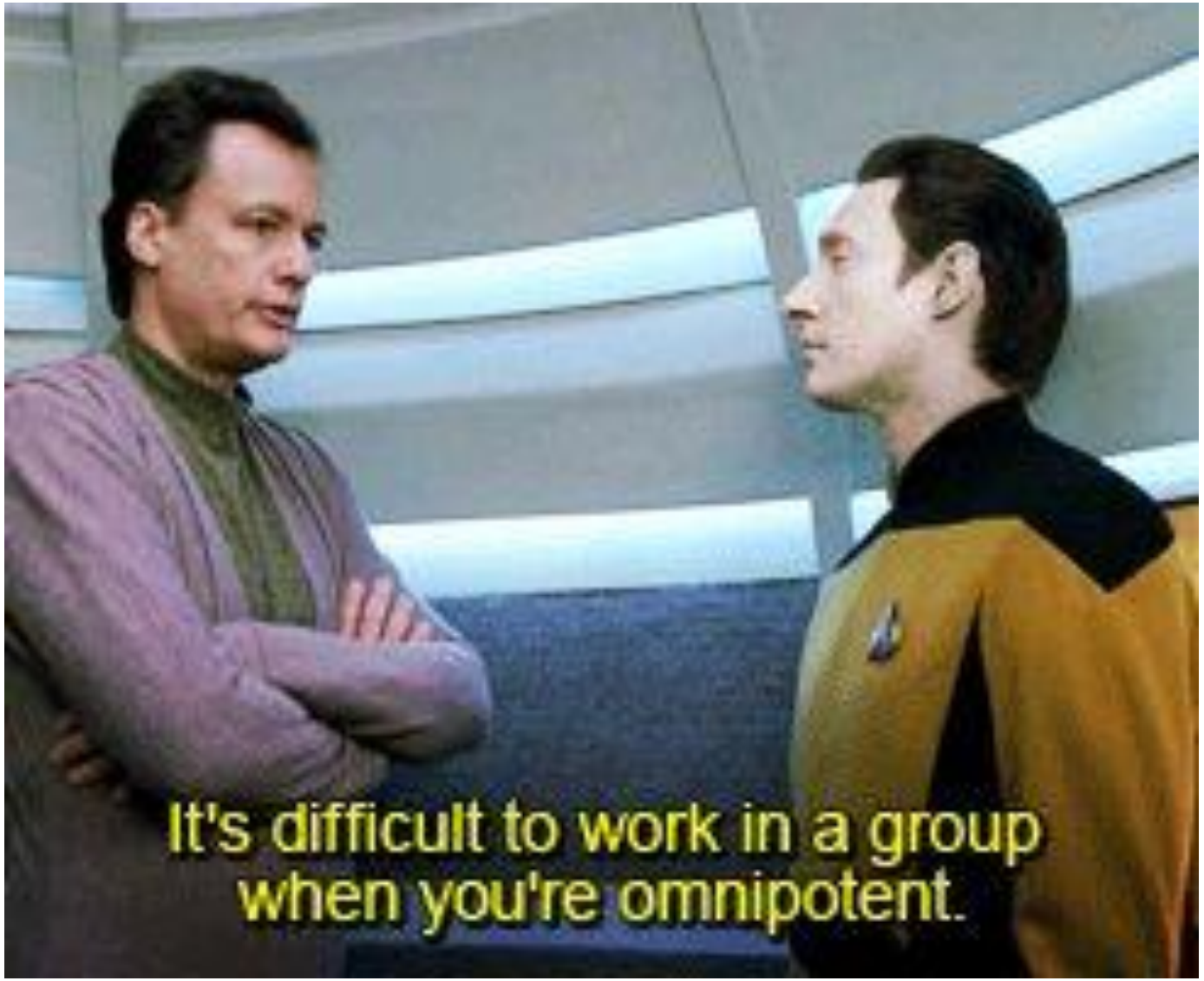
2

They have the resources and the vision to allow themselves to create any set of circumstances they need to optimize performance

3

Prepare for the worst

One master to rule them



One Master to rule them all

- The idea of one master is to simplify the entire process
- To do so they must use the master for as little as possible
- For example the client never reads or writes data through the master
- The master doesn't have to keep track of everything it just needs to keep track of who would know where the data is

A Chunk's Purpose

- When the Chunk begins its life by being created by the Master it is given a globally unique "Chunk Handle"
- Every chunk is made with siblings, and out of the siblings(replicas) one is selected to be the primary
- Chunks are made for one of three reasons: chunk creation, re-replicaition, and rebalancing
- Google attempts to keep a well balanced system by finding Chunkservers that have below average disk space utilized and prioritize placing chunks in them

Why Google is Smart

- Google constantly analyses problems and finds the most ideal way to handle that problem and then says can we make this happen any way that is simpler for efficiency sake
- Google is able to make everything three times (even the master) and still be efficient
- By having a mass load of inexpensive commodity components but preparing for and being able to handle their unreliable nature and still get

Main idea of the Comparison paper

1

To show the complexities of database design

2

To show there are so many ways to efficiently store data

3

How finding the best system for you is an art form

1

The amount of complex ideologies that are described in detail are astounding especially because a lot of them have such minor differences

2

The graphic and visual work helped clearly expand on points

How those Ideas are Implemented

The Effectiveness of the Implementations

- Form a professional stand point it was very impressive because it communicated sophisticated ideas
- But it was so bland and hard to keep focused on it
- I think capitalizing on the differences in the systems and concepts would have benefited them more
- Also this paper could have benefited from more visual aids, not graphs but a picture of the architecture of the systems for further understanding

Comparing Papers

Ideas From A Comparison of Approaches to Large-Scale Data Analysis

- Parallel DBMS
- MapReduce

How they are implemented in The Google File System

- Google operates on a number of inexpensive commodity components that have all their data as Linux files
- The master "Maps" and "splits" all chunks on creation with chunk handles and placing chunks where it finds unutilized diskspace

The Main idea of the Stonebreaker talk

- "No size fits all"
- Most things you could work with in this industry will be outdated by a better system because of a simple realization in the next 5-10 years

What I have learned from comparing all three works

- Google isn't that special
- Google only takes what everyone else does and uses their vast resources to over kill everything
- Just because it isn't special doesn't mean it isn't impressive, they are too big to fail as of now and are the product of thriving off of innovation