

Deconstructing Commodity Storage Clusters – ISCA'05

1 Introduction

- What is the motivation for this study? What assumptions do they make about their system to be analyzed?

2 Methodology

- What is the basic architecture of the EMC Centera? What types of workloads does it target?
- The analysis in this paper uses both *observations* and *delay* to infer the underlying protocol structure between nodes of the distributed system. They analyze both packets and disk traffic. What are the strengths and weaknesses of only using *observation* to understand the system (as is done in previous related work)?
- What are the complications that arise when *delaying* events to infer causality?

3 Deducing System Structure

- From observation alone: what happens when a client performs a write (that is, can you walk through Figure 1)? At a high level, what is going on during phases A1, B1 and C1? What is going on during phases B12 (and C7)?
- From delaying packets (and disk activity) what new information was learned about the write protocol? How is this result shown in the figures?
- Often in a distributed system, not only does a node need to know something, but it needs to know that someone else knows that too. Determining who knows what, and who knows that someone else knows something, can be important in distributed protocols. What does message va4 mean? What does bc6 mean? bc7? What does ab6 mean? ab7? What does the secondary storage node never know?
- What do you think are other reasonable ways in which the protocol could have been structured? Why is Centera structured as it is?
- Do you think that there are other controlled workloads that would have been interesting to use for analysis?

4 Inferring Policies

- What is a BLOB? How is the size of a BLOB determined? Do you think you could determine BLOB size from outside the Centera?
- What dynamic load factors influence which primary and secondary storage nodes selected for data writes? (How was this determined?) How is the load information shared across the cluster? (How was this shown?) Are there any static factors that influence replica locations? Do you think there are other factors that might be considered that were not uncovered?
- For reads, is any caching performed at the access or storage nodes? Any prefetching? What is the load balancing policy for reads? Do you think these decisions are reasonable?

5 Conclusions

- What are the strengths of this paper and type of analysis?
- What are the weaknesses of this paper and type of analysis?