

1. What is the theme of this research article?

2. In the introduction, authors mentioned two reasons why scale-up could be better than scale-out architecture. What are these two reasons?

3. What is the default assumption for Hadoop jobs today?

4. What are the four metrics used in this paper to compare scale-up with scale-out?

5. Why are the job sizes important for the comparison between scale-up and scale-out?

6. What are the five optimization categories for Hadoop in this paper and what are only applied in scale-up configuration?

7. What did authors do for storage optimization?

8. Why is increasing heap size helpful for improving the performance?

9. What are the three optimizations to speed up shuffle?

10. When authors deployed their jobs on Cloud using Azure, they found performance decrease in scale-out configuration. Why did that happen?

11. The authors used TeraSort to show the limitations of Scale-up. What were their conclusions?

12. Based on what you learned from this research article, what kind of scenario is best for scale-up?

13. Although authors did not explicitly mention in the article, what kind of scenario do you think may be good for scale-out?