

Data Analytics Frameworks

CS 838, Spring 2012

Compare either: (A) FlumeJava with MapReduce, (B) Dremel with MapReduce, or (C) FlumeJava with Dryad.

1. What are the differences in the primitives offered by the framework?
2. How do these differences influence the flexibility of the framework?
3. How do these differences influence the performance of data analysis conducted using the framework?
4. How does FlumeJava or Dremel influence the data center? What are its effects on the network? Storage services? Compute resources?
5. If other applications or tasks are running in the data center how does FlumeJava or Dremel influence their performance?
6. Does MapReduce or Dryad have less of an effect on other applications or tasks running in the data center?