Uni Freiburg, Web Science Group
Prof. Peter Fischer
Systems Infrastructure for Data Science - Winter 2012/13

Exercise Sheet #12: Web Databases

Feb 1, 2013

1 Social Networks and Map/Reduce

In previous exercises, we have studied how to determine common friends of users A and B on social graph. The assumption has been that these users A and B are already friends. To understand the limitations of Map/Reduce based approaches (including Pig), let us consider the following computations:

A. Compute common friends of users that are not directly connected (e.g., A is friend with C, B is friend with C, but A and B are not friends)

B. Determine the set of friends that are also connected by an intermediary friend (e.g. A is friend with B, B is friend with C; result includes B and C)

C. Is it possible to compute (with Map-Reduce) if two arbitrary users are connected, i.e. they have a path of friends between them? How would you approach this problem?

2 Web Databases

One important issue for Web Databases is the ability to always perform writes, despite involving large, massively distributed data sets

A. Which use case require such an ability. What additional requirements might they have, and which compromises might application designers want to make

B. Discuss why there are inherent limits in updating large, massively distributed data set in a consistent manner. One aspect to consider is the CAP theorem. Which other constraints might be an issue.

C. Describe the general idea of eventual consistency, and what systems and applications can do get specific, consistent versions.