

AIM III – Large Scale Data Analysis and Data Mining

Third Assignment

Data and topic for this assignment

In this assignment we'll have a look at the PACT programming model of the Nephele Execution Engine.

1. Average temperature per month

Perform the “average temperature per month” computation from Assignment 1 here using Map/Reduce in PACT instead of Hadoop.

Implement the missing parts in *AverageTemperaturePerMonthPact*.

2. Join books and authors using a PACT

We will perform the join of books and authors from Assignment 2 here using a “Match”-PACT. This PACT will read in the different tuple sets and supply all pairs having the same key to a user-defined function which provides a much easier way of joining than plain Hadoop code.

Implement the missing parts in *BookAndAuthorJoinPact*.

Deadline

Please send your solution as a patch file to ssc@apache.org with isabel@apache.org in CC until the 19th of May.