Parallel

The parallel clause can be attached to any relational operator in Pig Latin. However,

it controls only reduce-side parallelism, so it makes sense only for operators that force

a reduce phase. These are: group\*, order, distinct, join\*, limit, cogroup\*, and cross.

--parallel.pig

daily = load 'NYSE\_daily' as (exchange, symbol, date, open, high, low, close,

volume, adj\_close);

bysymbl = group daily by symbol parallel 10;

- Parallel clauses apply only to the statement to which they are attached; they

do not carry through the script.

- So if this group were followed by an order, parallel would need to be set for that order separately.

‘Grouped’ will reduce your data size significantly and you will want to change the parallelism:

--parallel.pig

daily = load 'NYSE\_daily' as (exchange, symbol, date, open, high, low, close,

volume, adj\_close);

bysymbl = group daily by symbol parallel 10;

average = foreach bysymbl generate group, AVG(daily.close) as avg;

sorted = order average by avg desc parallel 2;