2 a) Which part of the US has the most Data Engineer jobs for each year?

(using pig)

Command:

group1 = foreach L1 generate $4,$8,$7;

group2 = FILTER group1 BY $0 == 'DATA ENGINEER';

group2011 = FILTER group2 BY $2 == '2011';

group2012 = FILTER group2 BY $2 == '2012';

group2013 = FILTER group2 BY $2 == '2013';

group2014 = FILTER group2 BY $2 == '2014';

group2015 = FILTER group2 BY $2 == '2015';

group2016 = FILTER group2 BY $2 == '2016';

group3\_2011 = group group2011 by ($0,$1,$2);

group3\_2012 = group group2012 by ($0,$1,$2);

group3\_2013 = group group2013 by ($0,$1,$2);

group3\_2014 = group group2014 by ($0,$1,$2);

group3\_2015 = group group2015 by ($0,$1,$2);

group3\_2016 = group group2016 by ($0,$1,$2);

group4\_2011 = foreach group3\_2011 generate group,COUNT(abc2011.$0);

group4\_2012 = foreach group3\_2012 generate group,COUNT(abc2012.$0);

group4\_2013 = foreach group3\_2013 generate group,COUNT(abc2013.$0);

group4\_2014 = foreach group3\_2014 generate group,COUNT(abc2014.$0);

group4\_2015 = foreach group3\_2015 generate group,COUNT(abc2015.$0);

group4\_2016 = foreach group3\_2016 generate group,COUNT(abc2016.$0);

g\_2011 = order group4\_2011 by $1 DESC;

g\_2012 = order group4\_2012 by $1 DESC;

g\_2013 = order group4\_2013 by $1 DESC;

g\_2014 = order group4\_2014 by $1 DESC;

g\_2015 = order group4\_2015 by $1 DESC;

g\_2016 = order group4\_2016 by $1 DESC;

ans\_2011 = LIMIT g\_2011 1;

ans\_2012 = LIMIT g\_2012 1;

ans\_2013 = LIMIT g\_2013 1;

ans\_2014 = LIMIT g\_2014 1;

ans\_2015 = LIMIT g\_2015 1;

ans\_2016 = LIMIT g\_2016 1;

h1b\_ans = UNION ans\_2011,ans\_2012,ans\_2013,ans\_2014,ans\_2015,ans\_2016;

dump h1b\_ans;