Fast Convergence PageRank in Hadoop

Karthik Bellur - kvb26
Prashanth Basappa - pb476
Surya Sumukh - ss3345

Edges Preprocessing

We process the raw input data from edges.txt to create filtered_edges.txt. The filtered input file contains all the nodes present in the original file (graph) but retains only a subset of the edges from the original file.

The procedure (along with parameters) used to perform this filtering is shown below:

```
double fromNetID = 0.5433;              // using netid ss3345
double rejectMin = 0.9 * fromNetID;     // =0.48897
double rejectLimit = rejectMin + 0.01;  // =0.49897

boolean selectInputLine(double x) {
   return ( ((x >= rejectMin) && (x < rejectLimit)) ? false : true );
}
```

We then perform some additional processing on the filtered_edges.txt file in order to bring it to a format which is convenient for the Mapper process to read from.

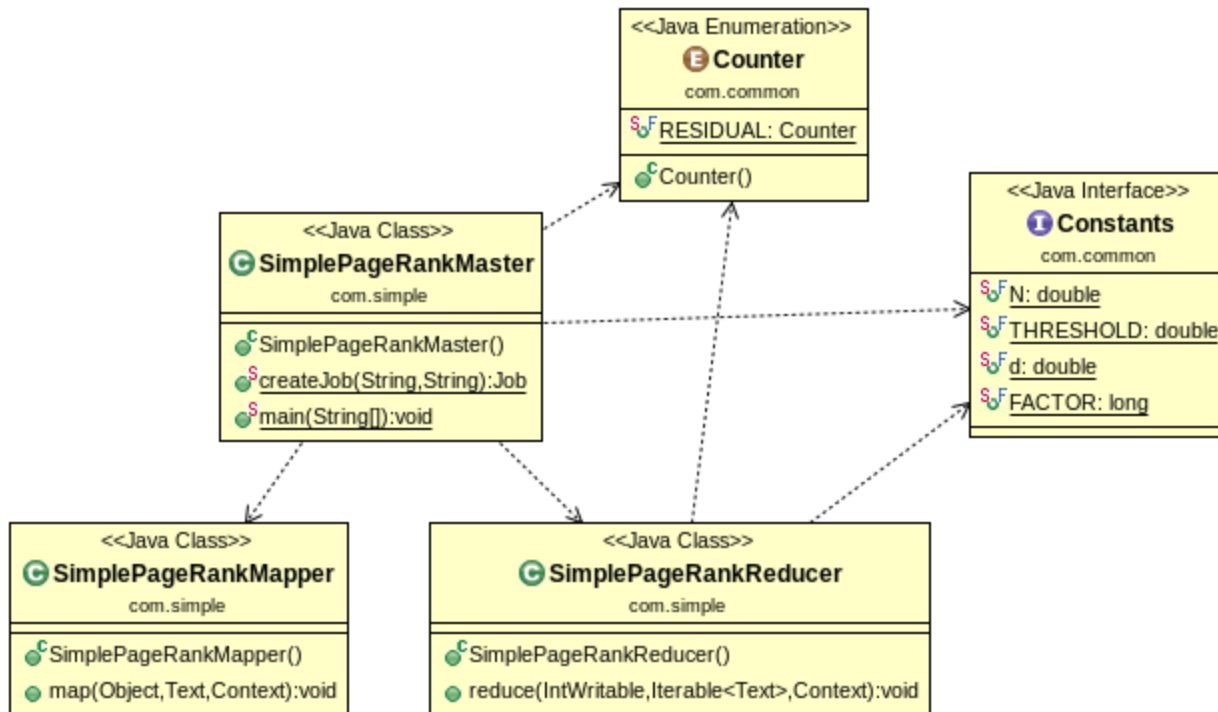**Mapper Input File Format:**
<NodeID PageRank D1 D2 D3 ….. Dn>

NodeID = Node ID of a source node
PageRank = Page Rank of the source node
D1 D2 D3 … Dn = List of destination nodes

Simple Computation of PageRank

SimplePageRankMaster
- Bootstrapper for the MapReduce process
- Set up the Mapper & Reducer jobs; configure the input/output paths
- Create a MapReduce counter to store "residual"
- Run 5 iterations of the MapReduce
- Output the residual (so we can conclude that PageRank is still far from convergence)

SimplePageRankMapper
Input:

        <NodeID PageRank D1 D2 D3 ….. Dn>

        NodeID = Node ID of a source node
        PageRank (PR) = Page Rank of the source node
        D1 D2 D3 … Dn = List of destination nodes
        n = Total number of nodes in the graph

Output:

        For each node in <D1, D2, D3, … Dn>

             emit(NodeID, PageRank/n)

        For each node in the graph

             emit(NodeID, "PR_" + <NodeID PageRank D1 D2 D3 ….. Dn>)

SimplePageRankReducer

Input:

        <NodeID, Incoming PageRank mass>

        <NodeID, "PR_" + <NodeID PageRank D1 D2 D3 ….. Dn>>

Output:

        <NodeID, <NodeID NewPageRank D1 D2 D3 ….. Dn>>

NewPageRank (NPR) = (1-d) / N + $\Sigma\ Incoming\ PageRank\ mass$

The residual PageRank is calculated by the Reducer as shown below and added to a MapReduce counter.

        Residual PageRank = Math.abs(NewPageRank - PageRank) / NewPageRank

The average residual PageRank over all the nodes in the graph is calculated as shown below:

        Avg Residual PageRank = Residual PageRank / n;

Results

The output from running Simple MapReduce is shown below. As we can clearly see, at the fifth iteration, the PageRank computation is nowhere near completion.

==

Iteration 1, Avg Error: 2.3390409059731767
Iteration 2, Avg Error: 0.3228536403835209
Iteration 3, Avg Error: 0.19199830713774937
Iteration 4, Avg Error: 0.09401514819841512
Iteration 5, Avg Error: 0.06275703048611415
Iteration 6, Avg Error: 0.03391270084497176
Iteration 7, Avg Error: 0.027217138770923632
Iteration 8, Avg Error: 0.0165739970520847
Iteration 9, Avg Error: 0.01436597930621835
Iteration 10, Avg Error: 0.009770442041358376
Iteration 11, Avg Error: 0.008454095705091721
Iteration 12, Avg Error: 0.006145381842592998
Iteration 13, Avg Error: 0.005370459553726486
Iteration 14, Avg Error: 0.00402346657326737
Iteration 15, Avg Error: 0.0035068517140230287

Iteration 16, Avg Error: 0.002715876421055704
Iteration 17, Avg Error: 0.0023408198707003487
Iteration 18, Avg Error: 0.0018533922916393036
Iteration 19, Avg Error: 0.0015907067699896386
Iteration 20, Avg Error: 0.0012754841440100403
Iteration 21, Avg Error: 0.00109014491484611
Iteration 22, Avg Error: 8.858339535630373E-4
===

Blocked Computation of PageRank

Node to Block Mapping:

The MapReduce master process, on startup, reads from the blocks.txt file and creates a list consisting of the prefix sums of the values contained in the blocks.txt file. A binary search of this list would then give us the corresponding Block for the Node under consideration in O(log(numberOfBlocks)) time. The Block PageRank works by converging the rank of each node within the block first and later converges across the blocks.
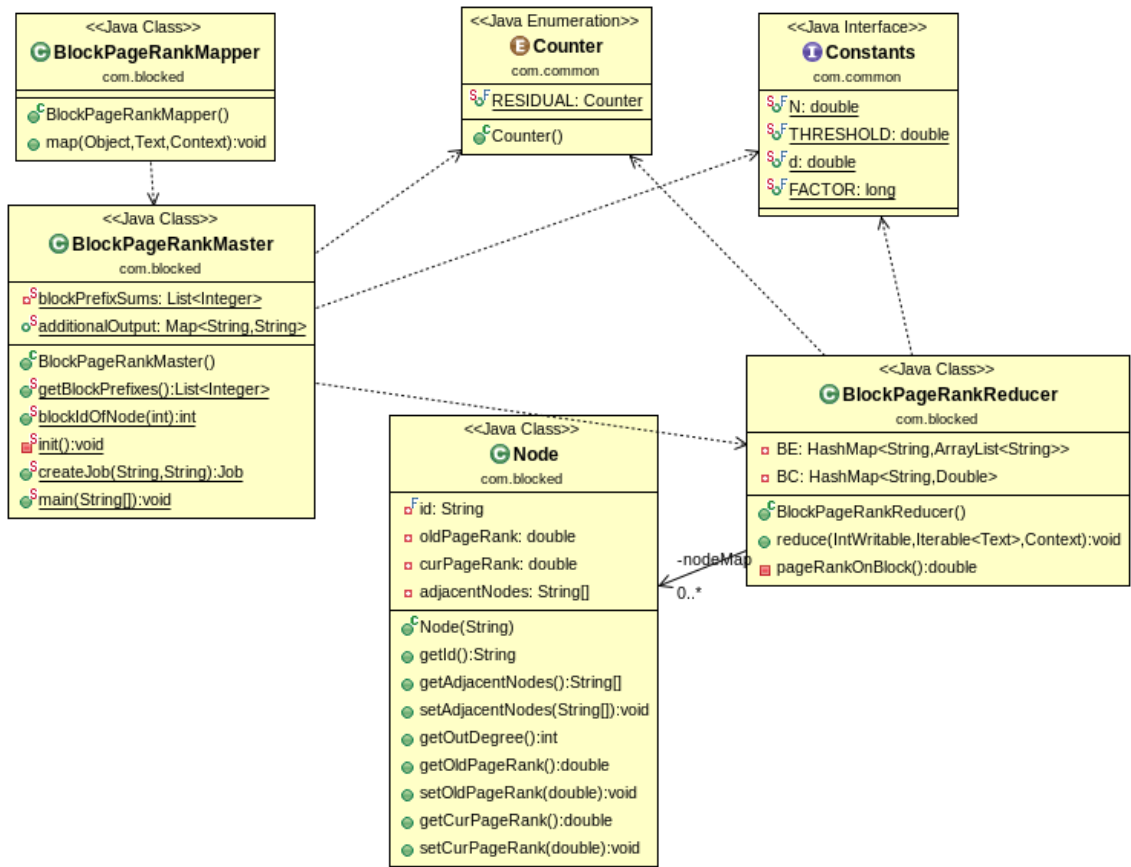
BlockedPageRankMapper

Input:

<NodeID PageRank D1 D2 D3 ….. Dn>

NodeID = Node ID of a source node
PageRank (PR) = Page Rank of the source node
D1 D2 D3 … Dn = List of destination nodes
n = Total number of nodes in the graph

Output:

For each node in the graph
emit(BlockID, "PR " + <NodeID PageRank D1 D2 D3 ….. Dn>)
For each node in <D1, D2, D3, … Dn>
If(blockIdOfNode(Di) == blockIdOfNode(NodeID))
emit(blockIdOfNode(Di), "BE " + NodeID + " " + Di)
else
emit(blockIdOfNode(Di), "BC " + NodeID + " " + Di + " " + PR / n)

BlockedPageRankReducer

do until convergence:

For each Value for block b
BE <- Set of edges in the same block b
BC <- Set of incoming edges to block b

For each node v in block b
For each node u: edge(u,v) is in b
sum incoming page rank for v

For node u : edge(u,v) is an incoming edge for a different block b'
sum incoming page rank for v

newPageRank(v) = (1-d)/n + d*(sum of incoming page ranks to v)

Residual PageRank = Math.abs(NewPageRank - PageRank) / NewPageRank

emit (v , new PageRank(v), n , D1,D2,....Dn)

Convergence Condition:

Average Residuals / N <= 0.001

Results

The in-block PageRank computation runs until convergence is achieved in the block. The following are the result of Block(Jacobi)-PageRank implementation.

==
Iteration 1, Avg Error: 2.815142360959094, Avg iterations per block = 17.58823529411765
Iteration 2, Avg Error: 0.03704449600863943, Avg iterations per block = 7.132352941176471
Iteration 3, Avg Error: 0.02345781708331509, Avg iterations per block = 5.955882352941177
Iteration 4, Avg Error: 0.009277177006260672, Avg iterations per block = 3.911764705882353
Iteration 5, Avg Error: 0.003460152065729755, Avg iterations per block = 2.411764705882353
Iteration 6, Avg Error: 0.001113494738992747, Avg iterations per block = 1.5294117647058822
Iteration 7, Avg Error: 6.041766997942297E-4, Avg iterations per block = 1.1323529411764706
==
Block ID: 0, PageRank of Node: 0 is 1.5703839457269233E-5;PageRank of Node: 1 is 4.4810107849610366E-5;
Block ID: 1, PageRank of Node: 10329 is 2.4334061068205395E-7;PageRank of Node: 10330 is 1.7989342095629917E-5;
Block ID: 2, PageRank of Node: 20374 is 2.1890460137472094E-7;PageRank of Node: 20375 is 2.1890460137472094E-7;
Block ID: 3, PageRank of Node: 30630 is 2.1890460137472094E-7;PageRank of Node: 30631 is 2.1890460137472094E-7;
Block ID: 4, PageRank of Node: 40646 is 2.3279937614313317E-5;PageRank of Node: 40647 is 2.3279937614313317E-5;
Block ID: 5, PageRank of Node: 50463 is 0.0050780831624499395;PageRank of Node: 50464 is 2.3279937614313317E-5;
Block ID: 6, PageRank of Node: 60842 is 3.0812195406941668E-6;PageRank of Node: 60843 is 2.4444389171297474E-5;
Block ID: 7, PageRank of Node: 70592 is 2.3279937614313317E-5;PageRank of Node: 70593 is 2.3279937614313317E-5;
Block ID: 8, PageRank of Node: 80119 is 9.507256519539887E-7;PageRank of Node: 80120 is 7.821966515173198E-7;
Block ID: 9, PageRank of Node: 100000 is 4.83500517331085E-7;PageRank of Node: 100001 is 3.131097432191495E-7;
Block ID: 10, PageRank of Node: 100502 is 1.2040829037978832E-6;PageRank of Node: 100503 is 1.0594484001979645E-6;
Block ID: 11, PageRank of Node: 110568 is 9.520418827193556E-5;PageRank of Node: 110569 is 1.2262078484181837E-6;
Block ID: 12, PageRank of Node: 120946 is 6.674079363885249E-7;PageRank of Node: 120947 is 8.490160949828883E-7;
Block ID: 13, PageRank of Node: 131000 is 2.2871734180573333E-7;PageRank of Node: 131001 is 2.2871734180573333E-7;
Block ID: 14, PageRank of Node: 140575 is 2.2871734180573333E-7;PageRank of Node: 140576 is 2.2871734180573333E-7;
Block ID: 15, PageRank of Node: 150954 is 0.0018552242045197444;PageRank of Node: 150955 is 0.0016861519180218177;
Block ID: 16, PageRank of Node: 161333 is 2.288470375690345E-7;PageRank of Node: 161334 is 2.288470375690345E-7;
Block ID: 17, PageRank of Node: 171155 is 2.3000468357014282E-7;PageRank of Node: 171156 is 2.3000468357014282E-7;
Block ID: 18, PageRank of Node: 181515 is 7.407730801964816E-4;PageRank of Node: 181516 is 3.7218586610636445E-7;
Block ID: 19, PageRank of Node: 191626 is 2.1890460137472094E-7;PageRank of Node: 191627 is 2.1890460137472094E-7;
Block ID: 20, PageRank of Node: 202005 is 1.1031918242692894E-6;PageRank of Node: 202006 is 9.235563645483702E-7;
Block ID: 21, PageRank of Node: 212384 is 2.1890460137472094E-7;PageRank of Node: 212385 is 2.1890460137472094E-7;
Block ID: 22, PageRank of Node: 222763 is 2.2860962210788993E-7;PageRank of Node: 222764 is 2.2860962210788993E-7;
Block ID: 23, PageRank of Node: 232594 is 3.0353288141651023E-7;PageRank of Node: 232595 is 2.7299769039798887E-7;
Block ID: 24, PageRank of Node: 242879 is 5.841702020693696E-7;PageRank of Node: 242880 is 2.3783816055548745E-7;
Block ID: 25, PageRank of Node: 252939 is 5.753219596459093E-7;PageRank of Node: 252940 is 5.753219596459093E-7;
Block ID: 26, PageRank of Node: 263150 is 2.81691464014372E-6;PageRank of Node: 263151 is 3.660842271616116E-6;
Block ID: 27, PageRank of Node: 273211 is 2.7756562628603804E-6;PageRank of Node: 273212 is 3.0895371439466736E-6;
Block ID: 28, PageRank of Node: 283474 is 2.8752749114044366E-7;PageRank of Node: 283475 is 2.8752749114044366E-7;
Block ID: 29, PageRank of Node: 293256 is 9.609980362426563E-6;PageRank of Node: 293257 is 9.377918481998591E-6;
Block ID: 30, PageRank of Node: 303044 is 5.509374540492791E-6;PageRank of Node: 303045 is 4.814529091451341E-7;
Block ID: 31, PageRank of Node: 313371 is 2.3403306394032815E-5;PageRank of Node: 313372 is 5.915306792149116E-5;
Block ID: 32, PageRank of Node: 323523 is 1.7876114285909575E-6;PageRank of Node: 323524 is 2.3925436597300035E-6;
Block ID: 33, PageRank of Node: 333884 is 4.808751262134438E-6;PageRank of Node: 333885 is 5.935603080086936E-7;
Block ID: 34, PageRank of Node: 343664 is 3.735332149684028E-7;PageRank of Node: 343665 is 3.2013160104204537E-7;
Block ID: 35, PageRank of Node: 353646 is 9.398620477660855E-7;PageRank of Node: 353647 is 4.303041702719303E-6;
Block ID: 36, PageRank of Node: 363930 is 4.64417738515433376E-7;PageRank of Node: 363931 is 3.5056181956668417E-7;
Block ID: 37, PageRank of Node: 374237 is 9.26936925723965E-6;PageRank of Node: 374238 is 1.1207850789076085E-5;
Block ID: 38, PageRank of Node: 384555 is 5.223652644069767E-6;PageRank of Node: 384556 is 9.919346342694194E-7;
Block ID: 39, PageRank of Node: 394930 is 2.5666734171396126E-6;PageRank of Node: 394931 is 1.6805295065746466E-6;

Block ID: 40, PageRank of Node: 404713 is 8.217271839725033E-7;PageRank of Node: 404714 is 6.822667893264482E-7;
Block ID: 41, PageRank of Node: 414618 is 3.700472954103684E-7;PageRank of Node: 414619 is 3.700472954103684E-7;
Block ID: 42, PageRank of Node: 424748 is 3.2386348496106725E-6;PageRank of Node: 424749 is 6.086528600757208E-6;
Block ID: 43, PageRank of Node: 434708 is 3.1970806877005528E-6;PageRank of Node: 434709 is 3.20048354247453E-6;
Block ID: 44, PageRank of Node: 444490 is 2.405712971306168E-7;PageRank of Node: 444491 is 5.047205149882555E-6;
Block ID: 45, PageRank of Node: 454286 is 3.1844619655270145E-7;PageRank of Node: 454287 is 3.51322484397602E-7;
Block ID: 46, PageRank of Node: 464399 is 2.34410343972097E-7;PageRank of Node: 464400 is 2.34410343972097E-7;
Block ID: 47, PageRank of Node: 474197 is 8.405457685724451E-7;PageRank of Node: 474198 is 3.267385270165729E-7;
Block ID: 48, PageRank of Node: 484051 is 1.2320062818137596E-5;PageRank of Node: 484052 is 4.4980608123026855E-7;
Block ID: 49, PageRank of Node: 493969 is 1.3854788181991435E-6;PageRank of Node: 493970 is 5.133145252741315E-7;
Block ID: 50, PageRank of Node: 503753 is 8.169806738466181E-7;PageRank of Node: 503754 is 5.754251253096591E-7;
Block ID: 51, PageRank of Node: 514132 is 2.7905810677520255E-5;PageRank of Node: 514133 is 4.14046349313061E-6;
Block ID: 52, PageRank of Node: 524511 is 5.5764569296479834E-5;PageRank of Node: 524512 is 2.4463191196245015E-6;
Block ID: 53, PageRank of Node: 534710 is 2.1890460137472094E-7;PageRank of Node: 534711 is 2.34410343972097E-7;
Block ID: 54, PageRank of Node: 545089 is 3.277602181969144E-5;PageRank of Node: 545090 is 3.277602181969144E-5;
Block ID: 55, PageRank of Node: 555468 is 7.517550141623096E-7;PageRank of Node: 555469 is 6.828197346262626E-7;
Block ID: 56, PageRank of Node: 565847 is 2.1890460137472094E-7;PageRank of Node: 565848 is 2.1890460137472094E-7;
Block ID: 57, PageRank of Node: 576226 is 2.1890460137472094E-7;PageRank of Node: 576227 is 4.8367544433378445E-5;
Block ID: 58, PageRank of Node: 586605 is 4.0897187176505313E-7;PageRank of Node: 586606 is 3.012109567354415E-7;
Block ID: 59, PageRank of Node: 596586 is 1.5191925387297244E-6;PageRank of Node: 596587 is 1.3095648578454645E-6;
Block ID: 60, PageRank of Node: 606368 is 2.2474607253047052E-7;PageRank of Node: 606369 is 2.2474607253047052E-7;
Block ID: 61, PageRank of Node: 616149 is 4.3605371450769775E-7;PageRank of Node: 616150 is 1.9114335523072644E-6;
Block ID: 62, PageRank of Node: 626449 is 2.5755477587069427E-7;PageRank of Node: 626450 is 2.882201131527775E-7;
Block ID: 63, PageRank of Node: 636241 is 2.2871734180573333E-7;PageRank of Node: 636242 is 2.2871734180573333E-7;
Block ID: 64, PageRank of Node: 646023 is 4.15677871907854E-7;PageRank of Node: 646024 is 2.0642056491784535E-6;
Block ID: 65, PageRank of Node: 655805 is 2.1890460137472094E-7;PageRank of Node: 655806 is 2.2892438241237505E-7;
Block ID: 66, PageRank of Node: 665667 is 3.839142733396038E-7;PageRank of Node: 665668 is 5.5165634029124885E-7;
Block ID: 67, PageRank of Node: 675449 is 5.729935495091042E-6;PageRank of Node: 675450 is 8.561736757811034E-7;

==

Gauss-Siedel Page Rank Computation

The Gauss-Siedel version of PageRank computation is very similar to the Block PageRank implementation except that the updated pagerank of a node is used in calculating the pagerank of successive nodes in within the iteration as opposed to simultaneous updates of pagerank of all nodes after the iteration. The following are the results from Gauss-Siedel implementation.

Results
We observe that the average number of iterations per block decreases compared to Block(Jacobian) Pagerank computation. The number of outer iterations remains the same.
==
Iteration 1, Avg Error: 2.815524714329495, Avg iterations per block = 10.088235294117647
Iteration 2, Avg Error: 0.038112750463348076, Avg iterations per block = 5.073529411764706
Iteration 3, Avg Error: 0.02474789486741678, Avg iterations per block = 4.470588235294118
Iteration 4, Avg Error: 0.010522014506078252, Avg iterations per block = 3.2205882352941178
Iteration 5, Avg Error: 0.004513812880346745, Avg iterations per block = 2.3676470588235294
Iteration 6, Avg Error: 0.001625731506209593, Avg iterations per block = 1.6176470588235294
Iteration 7, Avg Error: 7.413569166557215E-4, Avg iterations per block = 1.2647058823529411

```
================================================================
Block ID: 0, PageRank of Node: 0 is 1.548160560045399E-5;PageRank of Node: 1 is 4.4966013831598596E-5;
Block ID: 1, PageRank of Node: 10329 is 2.4335764937139057E-7;PageRank of Node: 10330 is 1.7982111004408104E-5;
Block ID: 2, PageRank of Node: 20374 is 2.1890460137472094E-7;PageRank of Node: 20375 is 2.1890460137472094E-7;
Block ID: 3, PageRank of Node: 30630 is 2.1890460137472094E-7;PageRank of Node: 30631 is 2.1890460137472094E-7;
Block ID: 4, PageRank of Node: 40646 is 2.3265909829604348E-5;PageRank of Node: 40647 is 2.3265909829604348E-5;
Block ID: 5, PageRank of Node: 50463 is 0.005079731388521509;PageRank of Node: 50464 is 2.3308592731017942E-5;
Block ID: 6, PageRank of Node: 60842 is 3.0630256805963164E-6;PageRank of Node: 60843 is 2.43947995768889515E-5;
Block ID: 7, PageRank of Node: 70592 is 2.3265909829604348E-5;PageRank of Node: 70593 is 2.3265909829604348E-5;
Block ID: 8, PageRank of Node: 80119 is 9.527905553013807E-7;PageRank of Node: 80120 is 7.849792540253267E-7;
Block ID: 9, PageRank of Node: 100000 is 4.825139418936263E-7;PageRank of Node: 100001 is 3.130787396233434E-7;
Block ID: 10, PageRank of Node: 100502 is 1.2038058322703632E-6;PageRank of Node: 100503 is 1.0588397718447831E-6;
Block ID: 11, PageRank of Node: 110568 is 9.523910366925085E-5;PageRank of Node: 110569 is 1.22611651658326678E-6;
Block ID: 12, PageRank of Node: 120946 is 6.66058484864697E-7;PageRank of Node: 120947 is 8.467751589895255E-7;
Block ID: 13, PageRank of Node: 131000 is 2.2871757526556838E-7;PageRank of Node: 131001 is 2.2871757526556838E-7;
Block ID: 14, PageRank of Node: 140575 is 2.2871732227366062E-7;PageRank of Node: 140576 is 2.2871732227366062E-7;
Block ID: 15, PageRank of Node: 150954 is 0.0018574079093089287;PageRank of Node: 150955 is 0.001690832996871611;
Block ID: 16, PageRank of Node: 161333 is 2.2884699633576878E-7;PageRank of Node: 161334 is 2.2884699633576878E-7;
Block ID: 17, PageRank of Node: 171155 is 2.3000630487102756E-7;PageRank of Node: 171156 is 2.3000630487102756E-7;
Block ID: 18, PageRank of Node: 181515 is 7.407532826540156E-4;PageRank of Node: 181516 is 3.7218586610636477E-7;
Block ID: 19, PageRank of Node: 191626 is 2.1890460137472094E-7;PageRank of Node: 191627 is 2.1890460137472094E-7;
Block ID: 20, PageRank of Node: 202005 is 1.1032339197181905E-6;PageRank of Node: 202006 is 9.23995581181258E-7;
Block ID: 21, PageRank of Node: 212384 is 2.1890460137472094E-7;PageRank of Node: 212385 is 2.1890460137472094E-7;
Block ID: 22, PageRank of Node: 222763 is 2.2862117200890484E-7;PageRank of Node: 222764 is 2.2862117200890484E-7;
Block ID: 23, PageRank of Node: 232594 is 3.0353219236588353E-7;PageRank of Node: 232595 is 2.7299722238152118E-7;
Block ID: 24, PageRank of Node: 242879 is 5.851212573916004E-7;PageRank of Node: 242880 is 2.3781372774574877E-7;
Block ID: 25, PageRank of Node: 252939 is 5.822743879862634E-7;PageRank of Node: 252940 is 5.822743879862634E-7;
Block ID: 26, PageRank of Node: 263150 is 2.7437668912778875E-6;PageRank of Node: 263151 is 3.733599629524312E-6;
Block ID: 27, PageRank of Node: 273211 is 2.760317824253951E-6;PageRank of Node: 273212 is 3.070211956148776E-6;
Block ID: 28, PageRank of Node: 283474 is 2.872058006036832E-7;PageRank of Node: 283475 is 2.872058006036832E-7;
Block ID: 29, PageRank of Node: 293256 is 9.805246328812845E-6;PageRank of Node: 293257 is 9.547576615249111E-6;
Block ID: 30, PageRank of Node: 303044 is 5.474699684634694E-6;PageRank of Node: 303045 is 4.802423514272277E-7;
Block ID: 31, PageRank of Node: 313371 is 2.3427081868451828E-5;PageRank of Node: 313372 is 5.929558353311483E-5;
Block ID: 32, PageRank of Node: 323523 is 1.7911818239265616E-6;PageRank of Node: 323524 is 2.401455328959501E-6;
Block ID: 33, PageRank of Node: 333884 is 4.7975390883990835E-6;PageRank of Node: 333885 is 5.9194793763271151E-7;
Block ID: 34, PageRank of Node: 343664 is 3.6120448255397827E-7;PageRank of Node: 343665 is 3.193943286035716E-7;
Block ID: 35, PageRank of Node: 353646 is 9.398559046970281E-7;PageRank of Node: 353647 is 4.30065224034303E-6;
Block ID: 36, PageRank of Node: 363930 is 4.634618048878554E-7;PageRank of Node: 363931 is 3.50218779426627996E-7;
Block ID: 37, PageRank of Node: 374237 is 9.217069592626992E-6;PageRank of Node: 374238 is 1.1144181550961304E-5;
Block ID: 38, PageRank of Node: 384555 is 5.1855990645516335E-6;PageRank of Node: 384556 is 9.490651610757545E-7;
Block ID: 39, PageRank of Node: 394930 is 2.5795919855897288E-6;PageRank of Node: 394931 is 1.6876220724526874E-6;
Block ID: 40, PageRank of Node: 404713 is 8.291872559373211E-7;PageRank of Node: 404714 is 6.876660771363363E-7;
Block ID: 41, PageRank of Node: 414618 is 3.7004729540842117E-7;PageRank of Node: 414619 is 3.7004729540842117E-7;
Block ID: 42, PageRank of Node: 424748 is 3.829918974782872E-6;PageRank of Node: 424749 is 7.017244683738107E-6;
Block ID: 43, PageRank of Node: 434708 is 3.199823459164142E-6;PageRank of Node: 434709 is 3.204092981155874E-6;
Block ID: 44, PageRank of Node: 444490 is 2.4060617877203743E-7;PageRank of Node: 444491 is 5.046304749352563E-6;
Block ID: 45, PageRank of Node: 454286 is 3.1844661566629544E-7;PageRank of Node: 454287 is 3.513711972479123E-7;
Block ID: 46, PageRank of Node: 464399 is 2.34410343972097E-7;PageRank of Node: 464400 is 2.34410343972097E-7;
Block ID: 47, PageRank of Node: 474197 is 8.408876277767034E-7;PageRank of Node: 474198 is 3.2678039023530417E-7;
Block ID: 48, PageRank of Node: 484051 is 1.2261606955313E-5;PageRank of Node: 484052 is 4.49135194851699E-7;
Block ID: 49, PageRank of Node: 493969 is 1.3855027384345468E-6;PageRank of Node: 493970 is 5.133239332920622E-7;
Block ID: 50, PageRank of Node: 503753 is 8.098878929804291E-7;PageRank of Node: 503754 is 5.631069558914033E-7;
Block ID: 51, PageRank of Node: 514132 is 2.786385420547346E-5;PageRank of Node: 514133 is 4.13122264863156435E-6;
Block ID: 52, PageRank of Node: 524511 is 5.7175901041760505E-5;PageRank of Node: 524512 is 2.6503055565698326E-6;
Block ID: 53, PageRank of Node: 534710 is 2.1890460137472094E-7;PageRank of Node: 534711 is 2.34410343972097E-7;
Block ID: 54, PageRank of Node: 545089 is 3.274721184201871E-5;PageRank of Node: 545090 is 3.274721184201871E-5;
Block ID: 55, PageRank of Node: 555468 is 7.517378651507393E-7;PageRank of Node: 555469 is 6.828154836938133E-7;
```

Block ID: 56, PageRank of Node: 565847 is 2.1890460137472094E-7;PageRank of Node: 565848 is 2.1890460137472094E-7;
Block ID: 57, PageRank of Node: 576226 is 2.1890460137472094E-7;PageRank of Node: 576227 is 4.8367094585672885E-5;
Block ID: 58, PageRank of Node: 586605 is 4.083348882324875E-7;PageRank of Node: 586606 is 3.012109563053017E-7;
Block ID: 59, PageRank of Node: 596586 is 1.5278722412025084E-6;PageRank of Node: 596587 is 1.310577662622058E-6;
Block ID: 60, PageRank of Node: 606368 is 2.2474429398114086E-7;PageRank of Node: 606369 is 2.2474429398114086E-7;
Block ID: 61, PageRank of Node: 616149 is 4.338829635230041E-7;PageRank of Node: 616150 is 1.910244032692483E-6;
Block ID: 62, PageRank of Node: 626449 is 2.5727093874323233E-7;PageRank of Node: 626450 is 2.8702628466160575E-7;
Block ID: 63, PageRank of Node: 636241 is 2.2871757526556838E-7;PageRank of Node: 636242 is 2.287175949827025E-7;
Block ID: 64, PageRank of Node: 646023 is 4.121941431746162E-7;PageRank of Node: 646024 is 2.0383304275119755E-6;
Block ID: 65, PageRank of Node: 655805 is 2.1890460137472094E-7;PageRank of Node: 655806 is 2.2892437188187577E-7;
Block ID: 66, PageRank of Node: 665667 is 3.8409457047586454E-7;PageRank of Node: 665668 is 5.506185074170191E-7;
Block ID: 67, PageRank of Node: 675449 is 5.2219163542683965E-6;PageRank of Node: 675450 is 8.558964795368689E-7;

Random Blocked Computation of PageRank

Randomly assign nodes to blocks instead of using the METIS graph partition.
We use a simple function: (nodeId * 541) % 68. We observe that Random Blocked Computation converges much slower than Blocked Pagerank computation because there are are many edges going in between blocks.
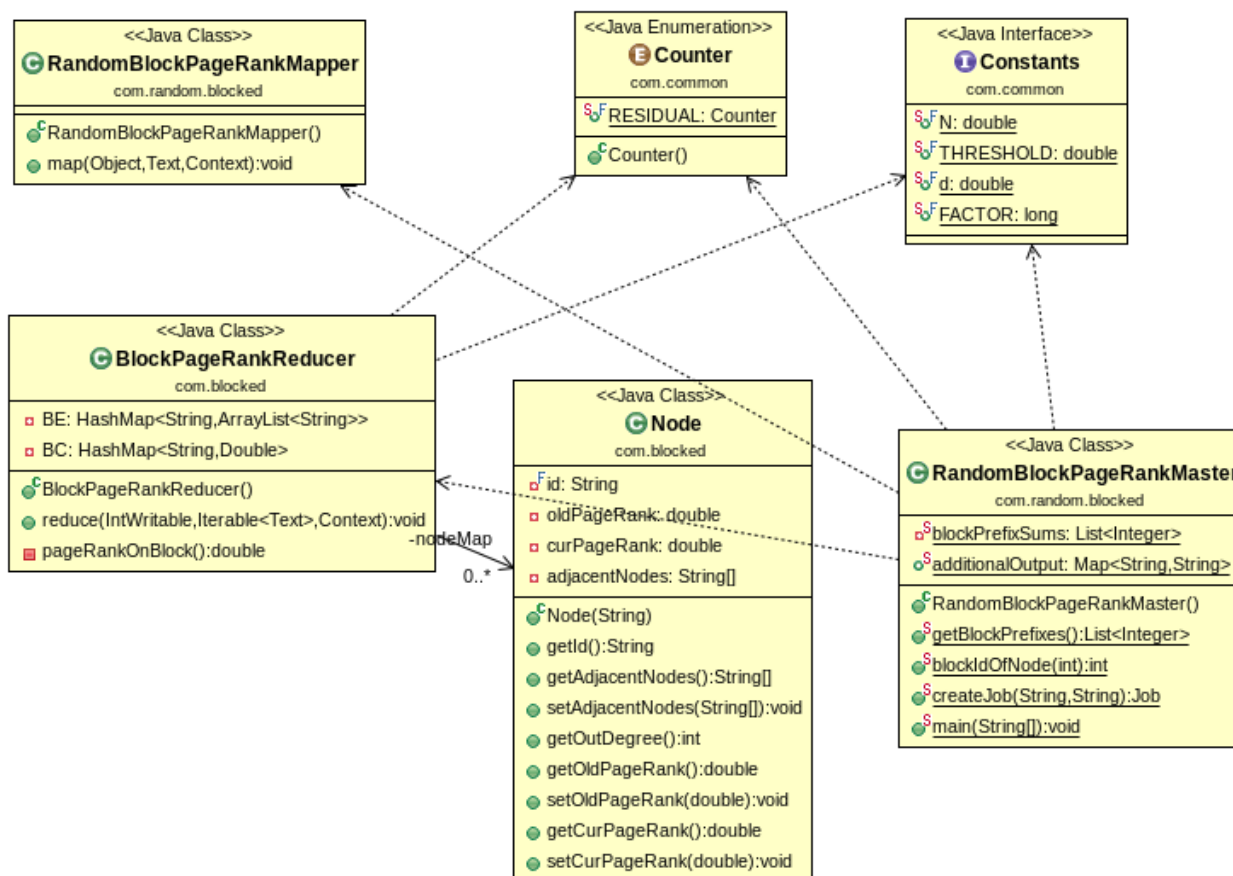
Results: (limited to 10 iterations)

==

Iteration 1, Avg Error: 2.33966405440509

Iteration 2, Avg Error: 0.3223165944281482

Iteration 3, Avg Error: 0.19119565693270874

Iteration 4, Avg Error: 0.0934357806867767

Iteration 5, Avg Error: 0.06209010113392584

Iteration 6, Avg Error: 0.0335172131984881

Iteration 7, Avg Error: 0.02690045678093487

Iteration 8, Avg Error: 0.016401792099003255

Iteration 9, Avg Error: 0.014182099441063585

Iteration 10, Avg Error: 0.009666827196707676

==

Block ID: 0, PageRank of Node: 0 is 1.6132736272253524E-5;PageRank of Node: 100028 is 7.357287339385182E-7;
Block ID: 1, PageRank of Node: 100005 is 3.3291609107124394E-7;PageRank of Node: 100073 is 3.658885712398831E-7;
Block ID: 2, PageRank of Node: 100050 is 4.742701633682869E-7;PageRank of Node: 100118 is 5.266046035979042E-7;
Block ID: 3, PageRank of Node: 100027 is 7.423427414294689E-7;PageRank of Node: 100095 is 8.19816471384248E-7;
Block ID: 4, PageRank of Node: 100004 is 2.8811895821474963E-7;PageRank of Node: 100072 is 3.674876830434679E-7;
Block ID: 5, PageRank of Node: 100049 is 4.563730062657801E-7;PageRank of Node: 100117 is 7.835418342828125E-7;
Block ID: 6, PageRank of Node: 100026 is 7.081903764269271E-7;PageRank of Node: 100094 is 5.328008638137532E-7;
Block ID: 7, PageRank of Node: 100003 is 3.1408733970929386E-7;PageRank of Node: 100071 is 3.6363920732250115E-7;
Block ID: 8, PageRank of Node: 100048 is 4.56764924035006E-7;PageRank of Node: 100116 is 6.421723606879457E-7;
Block ID: 9, PageRank of Node: 100025 is 2.993879549628395E-7;PageRank of Node: 100093 is 4.524287906755052E-7;
Block ID: 10, PageRank of Node: 100002 is 3.1451787421536125E-7;PageRank of Node: 100070 is 3.620693338305336E-7;
Block ID: 11, PageRank of Node: 100047 is 4.6509098920266325E-7;PageRank of Node: 100115 is 4.077055263390968E-7;
Block ID: 12, PageRank of Node: 100024 is 2.966735114800896E-7;PageRank of Node: 100092 is 9.770899456471675E-7;
Block ID: 13, PageRank of Node: 100001 is 3.1356376328543025E-7;PageRank of Node: 100069 is 3.672765492562904E-7;
Block ID: 14, PageRank of Node: 100046 is 4.668513267445341E-7;PageRank of Node: 100114 is 4.436338637742729E-7;
Block ID: 15, PageRank of Node: 100023 is 3.00995514766538E-7;PageRank of Node: 100091 is 8.191830141270522E-7;
Block ID: 16, PageRank of Node: 100000 is 5.019690881435088E-7;PageRank of Node: 100068 is 6.204360470588751E-7;
Block ID: 17, PageRank of Node: 100045 is 4.657273187881427E-7;PageRank of Node: 100113 is 3.8549742711467263E-7;
Block ID: 18, PageRank of Node: 100022 is 3.298967337460649E-7;PageRank of Node: 100090 is 4.917892918885077E-7;
Block ID: 19, PageRank of Node: 100067 is 5.908311032107682E-7;PageRank of Node: 100135 is 5.419684467859606E-7;
Block ID: 20, PageRank of Node: 100044 is 4.742701633682869E-7;PageRank of Node: 100112 is 9.304104046693284E-7;
Block ID: 21, PageRank of Node: 100021 is 2.8622977906617104E-7;PageRank of Node: 100089 is 9.806006213024643E-7;
Block ID: 22, PageRank of Node: 100066 is 5.895011766678486E-7;PageRank of Node: 100134 is 4.613515819999864E-7;
Block ID: 23, PageRank of Node: 100043 is 4.563730062657801E-7;PageRank of Node: 10011 is 7.07159879315721E-7;
Block ID: 24, PageRank of Node: 100020 is 3.028454462805806E-7;PageRank of Node: 100088 is 8.294239072743935E-7;
Block ID: 25, PageRank of Node: 100065 is 5.860308996149302E-7;PageRank of Node: 100133 is 8.823757043632676E-7;
Block ID: 26, PageRank of Node: 100042 is 4.6405101266035006E-7;PageRank of Node: 10010 is 4.6503044117088317E-7;
Block ID: 27, PageRank of Node: 100019 is 7.881580840438685E-7;PageRank of Node: 100087 is 5.479596942117885E-7;
Block ID: 28, PageRank of Node: 100064 is 5.647332992799622E-7;PageRank of Node: 100132 is 8.110826086154812E-7;
Block ID: 29, PageRank of Node: 100041 is 4.6509098920266325E-7;PageRank of Node: 10009 is 4.4656635311389037E-7;
Block ID: 30, PageRank of Node: 100018 is 2.906285685998754E-7;PageRank of Node: 100086 is 1.0051692062410326E-6;
Block ID: 31, PageRank of Node: 100063 is 5.836642895023551E-7;PageRank of Node: 100131 is 4.884677635979917E-7;
Block ID: 32, PageRank of Node: 100040 is 4.247080893937501E-7;PageRank of Node: 10008 is 4.5183911317757946E-7;
Block ID: 33, PageRank of Node: 100017 is 3.259298550121742E-7;PageRank of Node: 100085 is 2.5881384340150017E-7;
Block ID: 34, PageRank of Node: 100062 is 7.08308865050247E-7;PageRank of Node: 100130 is 4.69092515975572165E-7;
Block ID: 35, PageRank of Node: 100039 is 4.657273187881427E-7;PageRank of Node: 10007 is 6.444263585270303E-7;
Block ID: 36, PageRank of Node: 100016 is 2.882031259176187E-7;PageRank of Node: 100084 is 2.590159000990478E-7;
Block ID: 37, PageRank of Node: 100061 is 6.583477422583891E-7;PageRank of Node: 100129 is 1.0174860954149257E-6;

Block ID: 38, PageRank of Node: 10 is 1.623047397978942E-6;PageRank of Node: 100038 is 4.941651850757113E-7;
Block ID: 39, PageRank of Node: 100015 is 2.7023582141676795E-7;PageRank of Node: 100083 is 2.587278717872533E-7;
Block ID: 40, PageRank of Node: 100 is 5.594238248373409E-7;PageRank of Node: 100060 is 6.655224274194999E-7;
Block ID: 41, PageRank of Node: 100037 is 4.5451543015353077E-7;PageRank of Node: 10005 is 4.361536979112985E-7;
Block ID: 42, PageRank of Node: 100014 is 2.9065270558865634E-7;PageRank of Node: 100082 is 2.584633792519404E-7;
Block ID: 43, PageRank of Node: 100059 is 6.669837212458799E-7;PageRank of Node: 100127 is 5.344554286745352E-7;
Block ID: 44, PageRank of Node: 100036 is 4.4406630509726963E-7;PageRank of Node: 10004 is 5.372818278165498E-7;
Block ID: 45, PageRank of Node: 100013 is 2.8931841369044023E-7;PageRank of Node: 100081 is 2.5601481366362794E-7;
Block ID: 46, PageRank of Node: 100058 is 6.632733323362382E-7;PageRank of Node: 100126 is 9.438715382739723E-7;
Block ID: 47, PageRank of Node: 10003 is 4.304862989927522E-7;PageRank of Node: 100035 is 4.848665543974804E-7;
Block ID: 48, PageRank of Node: 100012 is 2.906113773849553E-7;PageRank of Node: 100080 is 2.5661647789458634E-7;
Block ID: 49, PageRank of Node: 100057 is 6.637320381687537E-7;PageRank of Node: 100125 is 8.231252563401243E-7;
Block ID: 50, PageRank of Node: 10002 is 1.2687729443564607E-5;PageRank of Node: 100034 is 4.803091374952448E-7;
Block ID: 51, PageRank of Node: 100011 is 3.349108148299533E-7;PageRank of Node: 100079 is 2.5481080712820956E-7;
Block ID: 52, PageRank of Node: 100056 is 5.144968757934449E-7;PageRank of Node: 100124 is 4.6481381398276024E-7;
Block ID: 53, PageRank of Node: 10001 is 2.898003077537218E-7;PageRank of Node: 100033 is 4.832603214911903E-7;
Block ID: 54, PageRank of Node: 100010 is 2.879414700805893E-7;PageRank of Node: 100078 is 2.547296883783245E-7;
Block ID: 55, PageRank of Node: 100055 is 4.994466765435137E-7;PageRank of Node: 100123 is 8.923209054003526E-7;
Block ID: 56, PageRank of Node: 10000 is 7.760589166718517E-7;PageRank of Node: 100032 is 7.872800810800995E-7;
Block ID: 57, PageRank of Node: 100009 is 2.9087772378385977E-7;PageRank of Node: 100077 is 2.54519537260443E-7;
Block ID: 58, PageRank of Node: 100054 is 4.6867732921941265E-7;PageRank of Node: 100122 is 8.315940976979923E-7;
Block ID: 59, PageRank of Node: 100031 is 7.287918652590525E-7;PageRank of Node: 100099 is 8.088661542855988E-7;
Block ID: 60, PageRank of Node: 1000 is 1.3324076658724398E-6;PageRank of Node: 100008 is 2.9529008336809535E-7;
Block ID: 61, PageRank of Node: 100053 is 5.072563237718504E-7;PageRank of Node: 100121 is 5.013848904351506E-7;
Block ID: 62, PageRank of Node: 100030 is 6.250359945185516E-7;PageRank of Node: 100098 is 5.263912642934322E-7;
Block ID: 63, PageRank of Node: 100007 is 2.9414708878385397E-7;PageRank of Node: 100075 is 2.543759394809893E-7;
Block ID: 64, PageRank of Node: 100052 is 4.888219027571832E-7;PageRank of Node: 100120 is 1.0577117293571058E-6;
Block ID: 65, PageRank of Node: 1 is 4.74491009947645E-5;PageRank of Node: 100029 is 7.388513320965503E-7;
Block ID: 66, PageRank of Node: 100006 is 6.817604538500863E-7;PageRank of Node: 100074 is 3.690442805478756E-7;
Block ID: 67, PageRank of Node: 100051 is 5.117006129851987E-7;PageRank of Node: 100119 is 3.05973189611517E-7;
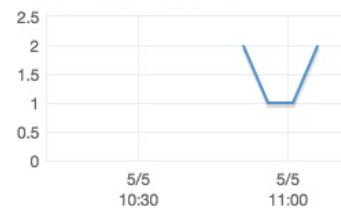
EMR Deployment

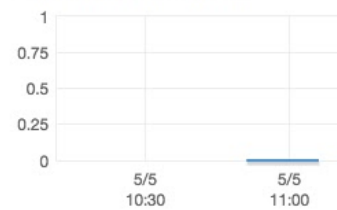We deployed the Simple PageRank MapReduce application on AWS EMR. Some screenshots from deployment shown below.
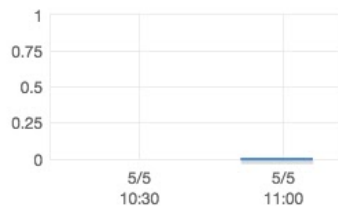
Is Idle? (Boolean)
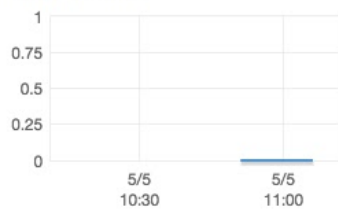
Container Allocated (Count)
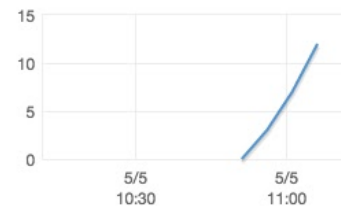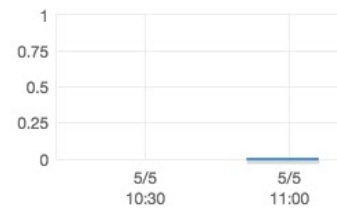
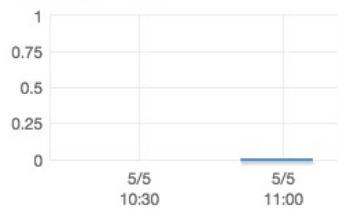**Container Reserved** (Count)

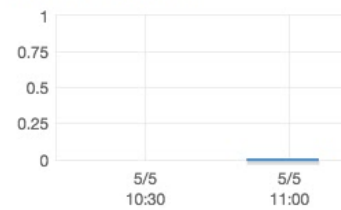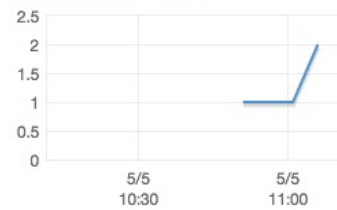**Container Pending** (Count)

Apps Completed (Count)

Apps Failed (Count)

Apps Killed (Count)

Apps Pending (Count)

Apps Running (Count)

▼ Steps

Add step   Clone step

Steps

View all interactive jobs | View all jobs

Filter: [All steps ▾] [Filter steps ...]   5 steps (all loaded)

		ID	Name	Status	Start time (UTC-4) ▾	Elapsed time	Log files
○	▶	s-3SBNCXUXV3HPK	Run Hue	Completed	2015-05-05 06:50 (UTC-4)	1 minute	View logs
○	▼ ●	s-IMR5906QS95T	SimplePageRank	Running	2015-05-05 06:52 (UTC-4)	13 minutes	View logs

    JAR location: s3://edu-cornell-cs-cs5300s15-ss3345/job/page_rank.jar
    Main class: None
    Arguments: com.simple.SimplePageRankMaster s3://edu-cornell-cs-cs5300s15-ss3345/input/ s3://edu-cornell-cs-cs5300s15-ss3345/output
    Action on failure: Terminate cluster

		ID	Name	Status	Start time (UTC-4)	Elapsed time	Log files
○	▶	s-29HY0YXPOFADO	Setup pig	Completed	2015-05-05 06:50 (UTC-4)	6 seconds	View logs
○	▶	s-2GPSLYW144EPI	Setup hive	Completed	2015-05-05 06:49 (UTC-4)	24 seconds	View logs
○	▶	s-2M5LK7D4L6702	Setup hadoop debugging	Completed	2015-05-05 06:49 (UTC-4)	5 seconds	View logs