

## BIEN 3 - Task #859

### fix slowness in `_taxonlabel_set_parent_id()`

01/20/2014 09:14 PM - Aaron Marcuse-Kubitza

<b>Status:</b>	New	<b>Start date:</b>	01/20/2014
<b>Priority:</b>	Normal	<b>Due date:</b>	
<b>Assignee:</b>	Aaron Marcuse-Kubitza	<b>% Done:</b>	0%
<b>Category:</b>		<b>Estimated time:</b>	0.00 hour
<b>Target version:</b>			
<b>Activity type:</b>			
<b>Description</b>			
occurs when doing a full import of NCBI:			
<pre>(make inputs/.NCBI/nodes/import_scrub by_col=1 continue=1; make inputs/.NCBI/publish) &amp;</pre>			
to reproduce problem:			
1. import NCBI:			
<pre>unset n; export log= n=40000 version=NCBI_test; inputs=(inputs/test_taxonomic_names/) # NCBI w ill be imported at beginning . bin/import_all</pre>			
2. search log output for <code>INSERT INTO "_taxonlabel_set_parent_id(parent_id=taxonlabel_pkeys.out.taxonl"</code>			
stats:			
<pre>n=10000: Took 0:00:13.805218 sec n=20000: Took 0:01:08.731089 sec -&gt; Took 0:00:31.031543 sec w/o nested xact; &gt;6 min w/o indexes so not using the indexes isn't the problem n=30000: Took 0:04:32.584036 sec -&gt; Took 0:01:19.635390 sec n=40000: -&gt; Took 0:04:18.476969 sec n=50000: very long (did not finish) n="" (all 163907) @r11549: Took 0:05:31.246567 sec -&gt; now: Took &gt;3:16:50.253774 sec (&gt;3:25:26.6390 58 sec on vegbiendev) (did not finish) &gt;O(n<sup>2</sup>) growth -&gt; likely seq scan? should be O(n*(logn)<sup>2</sup>) = # nodes * # ancest/node * O(logn) inde x scan/each</pre>			