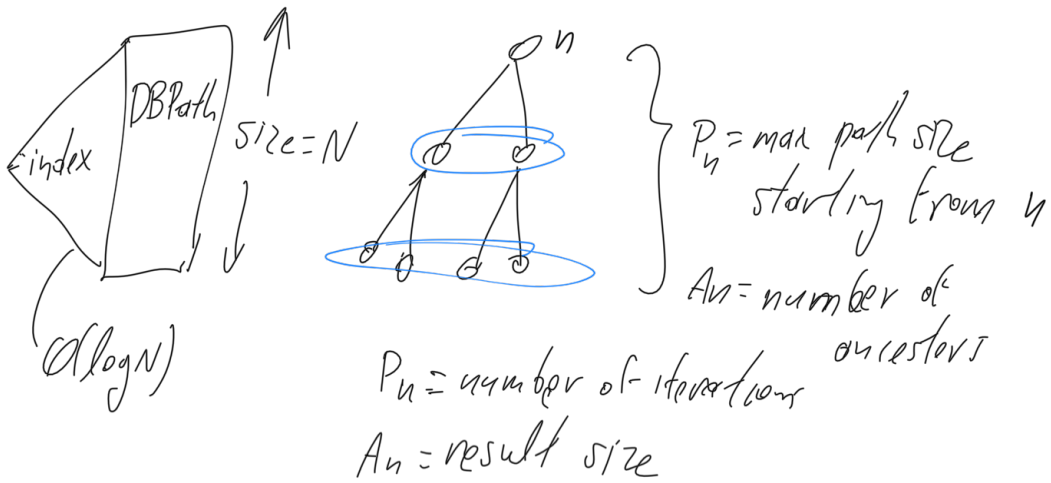
AiiDA paper - DbPath section

Comments on the dbpath:



- For a binary tree (branching factor 2) the number of nodes N= 2^(h+1)-1
  - https://www.geeksforgeeks.org/relationship-number-nodes-height-binary-tree
- For a more general case, for a tree with branching factor b, the equation is N= b^(h+1)-1

Useful commands

- to reindex the full database: REINDEX DATABASE broken_db;
- to see the definition of a function: \df+ name_of_function
- To dump a single table named mytab: pg_dump -t mytab mydb > db.sql
- 1: P?
  - select depth, count(distinct(parent_id)) from db_dbpath group by depth;
- **1': P**
  - **Various max_depths and number of parents that have it**
    - **Inner query: find the max_depth for every parent_id. Outer query, group by max_depth and count occurrences**
    - select md as max_depth , count(pid) as occurrences from (select max(depth) as md, parent_id as pid from db_dbpath group by parent_id) as sq group by md order by max_depth;
  - **Various max_height and number of children that have it**
    - **Inner query: find the max_height for every child_id. Outer query, group by max_height and count occurrences**
    - select mh as max_height , count(cid) as occurrences from (select max(depth) as mh, child_id as cid from db_dbpath group by child_id) as sq group by mh order by max_height;
- **2: A**

    - **Number of descendants and parent's occurrences**
        - select dsc_cnt as number_of_descendants, count(pid) as parent_occurrences from (select parent_id as pid, count(child_id) as dsc_cnt from db_dbpath group by pid) as sq group by dsc_cnt order by number_of_descendants;
    - **Number of ancestors and child's occurrences**
        - select ans_cnt as number_of_ancestors, count(cid) as child_occurences from (select child_id as cid, count(parent_id) as ans_cnt from db_dbpath group by child_id) as sq group by ans_cnt order by number_of_ancestors;
- Rename database
    - alter database aiidadb_davide_tmp_dj rename to aiidadb_davide_dj_mini;
- Output to a file
    - \o /tmp/aiidadb_davide_dj_1m_pquery.txt
- How to clean the database
    - delete from db_dbattribute where dbnode_id > 1000000;
    - delete from db_dblink where input_id > 1000000 or output_id > 1000000;
    - delete from db_dbcalcstate where dbnode_id > 1000000;
    - delete from db_dbextra where dbnode_id > 1000000;
    - delete from db_dbgroup_dbnodes where dbnode_id > 1000000;
    - delete from db_dbworkflowdata where aiida_obj_id > 1000000;
    - delete from db_dbworkflowstep_calculations where dbnode_id > 1000000;
    - ALTER TABLE db_dbnode drop constraint db_dbnode_dbcomputer_id_2195c2d4d9b222ff_fk_db_dbcomputer_id;
    - ALTER TABLE db_dbnode drop constraint db_dbnode_user_id_43fd81cadf67f183_fk_db_dbuser_id;
    - ALTER TABLE db_dbattribute DROP CONSTRAINT db_dbattribute_dbnode_id_783fe2b9b1ee948f_fk_db_dbnode_id;
    - ALTER TABLE db_dbcomment DROP CONSTRAINT db_dbcomment_dbnode_id_e225ac462eb8f6c_fk_db_dbnode_id;
    - ALTER TABLE db_dbextra DROP CONSTRAINT db_dbextra_dbnode_id_c556b194c79dec1_fk_db_dbnode_id;
    - ALTER TABLE db_dbgroup_dbnodes DROP CONSTRAINT db_dbgroup_dbnodes_dbnode_id_53a1829a1973b99c_fk_db_dbnode_id;
    - ALTER TABLE db_dblink DROP CONSTRAINT db_dblink_input_id_6feafb02380ed56f_fk_db_dbnode_id;
    - ALTER TABLE db_dblink DROP CONSTRAINT db_dblink_output_id_6345a663e713ed93_fk_db_dbnode_id;
    - ALTER TABLE db_dbworkflowdata DROP CONSTRAINT db_dbworkflowdata_aiida_obj_id_28130672924934ca_fk_db_dbnode_id;
    - ALTER TABLE db_dbworkflowstep_calculations DROP CONSTRAINT

        - db_dbworkflowstep_ca_dbnode_id_5ac7aa3704de0639_fk_db_dbnode_id;
        - delete from db_dbnode where id > 1000000;
        - REINDEX DATABASE aiidadb_davide_tmp_dj;
- Link export:
    - pg_dump -t db_dblink aiidadb_davide_dj_1m > /tmp/link_schema.sql
    - cp /tmp/link_schema.sql .
    - cp link_schema.sql link_schema_links_only.sql
    - cat link_schema_links_only.sql | sed "1,64d" > link_schema_links_only_2.sql
    - tac link_schema_links_only_2.sql | sed "1,54d" | tac > link_schema_links_only_3.sql
- Link import
    - DELETE FROM db_dblink;
    - psql aiidadb_davide_dj_1m < davide_db_dump/leo_dbpath/aiida_mat2d-dbpath-schema.psql
    - psql aiidadb_davide_dj_1m < davide_db_dump/leo_dbpath/trigger-statement.psql

Database from Davide:

- Full: 7351846
- Half: 3675923, will do experiments with 3’700’000
    - in 4 hours it inserted 4274000 out of 5153947 links
    - running for 13h (11am)
    - psql aiidadb_davide_dj_half < davide_db_dump/spyros_schema/3.7M/link_schema_links_only_4.sql
- Mini: 110’000
    - db_dbpath size: 602’157
- Small: 1’000’000
    - finished in around 3h-4h
    - db_dbpath size: 225’548’126