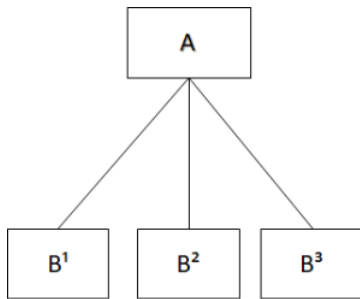


# DAG Jobs

If a user has to manage multiple jobs in a direct acyclic graph (DAG) order it is possible to organize the submission with a *DAG input file*. In the graph below, the vertices, generally tagged with a letter, represent the jobs and the edges represent the degree of relationship.



In this case the A job is a parent of the B¹, B² and B³ jobs, namely the children. This implies that the B\* jobs start once the A job finished. This automatic implementation of successive submissions could be very useful.

## Example

To properly perform a DAG job it occurs to write a *DAG input file* with the specific tags of the jobs and the degree of relationship, for instance:

```
-bash-4.2$ cat simple.dag
JOB A sleep.sub
JOB B snore.sub
PARENT A CHILD B
```

In this case the `sleep.sub` job is called A, whereas the `snore.sub` job is B. Moreover, A is a parent of B, so the B job starts only once the A job finished.

In order to submit properly the DAG job, the *DAG input file* has to be into a folder reachable from the schedd. At tier-1, in general this folder could be a directory into the *gpfs\_data* file system. After this, to submit the DAG job it is enough to issue the following commands

```
-bash-4.2$ export _condor_SCHEDD_HOST=sn-02.cr.cnaf.infn.it
-bash-4.2$ condor_submit_dag simple.dag
Renaming rescue DAGs newer than number 0
-----
File for submitting this DAG to HTCondor           : simple.dag.condor.sub
Log of DAGMan debugging messages                  : simple.dag.dagman.out
Log of HTCondor library output                    : simple.dag.lib.out
Log of HTCondor library error messages            : simple.dag.lib.err
Log of the life of condor_dagman itself           : simple.dag.dagman.log

Submitting job(s).
1 job(s) submitted to cluster 1271844.
-----
```

The submission produces the log files shown in the output. Then, to check the job status a user can launch the `condor_q` command

```
-bash-4.2$ condor_q

-- Schedd: sn-02.cr.cnaf.infn.it : <131.154.192.42:9618?... @ 01/26/22 17:47:19
OWNER      BATCH_NAME          SUBMITTED   DONE    RUN    IDLE  TOTAL JOB_IDS
arendinajuno simple.dag+1271844    1/26 16:51      _      1      _        1 1271846.0

Total for query: 1 jobs; 0 completed, 0 removed, 0 idle, 1 running, 0 held, 0 suspended
Total for arendinajuno: 1 jobs; 0 completed, 0 removed, 0 idle, 1 running, 0 held, 0 suspended
Total for all users: 26968 jobs; 18806 completed, 1 removed, 4984 idle, 3132 running, 45 held, 0 suspended
```

and after the A job is done, the child job B is queued:

```
-- Schedd: sn-02.cr.cnaf.infn.it : <131.154.192.42:9618?... @ 01/26/22 17:48:56
OWNER      BATCH_NAME      SUBMITTED  DONE   RUN    IDLE  TOTAL JOB_IDS
arendinajuno simple.dag+1271947  1/26 17:46      _     1      _      1 1271949.0
```

Total for query: 1 jobs; 0 completed, 0 removed, 0 idle, 1 running, 0 held, 0 suspended

Total for arendinajuno: 1 jobs; 0 completed, 0 removed, 0 idle, 1 running, 0 held, 0 suspended

Total for all users: 28900 jobs; 18810 completed, 1 removed, 6840 idle, 3204 running, 45 held, 0 suspended