- Mtech
  - DB Location: /data1/um/drich/Wheeler-Labs/benchmarks/profmark-benchmark/
    - 1k subset: (target, query)
      - matches: (0,0), (1,5), (2,4), (3,3), (4,1),(5,2)
      - O14727, L5MCZ0, I3MRA0, Q9EQZ6, D6THH4, Q08462
    - 5k subset: (target, query)
      - matches: (0,3),(1,1),(2,2),(3,0),(4,4),(5,5)
      - O30409_1, H0VZZ7_1, O01761_1, Q2FYJ6_1, M9PC48_1, P0C6X7_1
  - Tools Location: /data1/um/drich/tools/fb-pruner/
  - Testing Scripts:
    - slurm script: /data1/um/drich/Wheeler-Labs/benchmarks/profmark-benchmark/scripts/job-runners-slurm/cloud/run_cloud_search.sh
    - bash script: /data1/um/drich/Wheeler-Labs/benchmarks/profmark-benchmark/scripts/cloud/fb-pruner.sh
    - % sbatch run_cloud_search <my_id> <alpha> <alpha_max>
    - % bash runall_cloud_search <my_id>
    - % srun/bash fb-pruner.sh <job_id> <my_id> <alpha> <alpha_max>
  - Results Location:
    - /data1/um/drich/Wheeler-Labs/benchmarks/profmark-benchmark/results/cloud/<my_id>
  - DB Utility Scripts:
    - location: /data1/um/drich/Wheeler-Labs/benchmarks/profmark-benchmark/db/scripts/
    - fasta→hmm:
      - % bash convert_fasta_to_hmm.sh <fasta_file>
      - Outputs the hmm file as “<fasta_file>.hmm”
    - find fasta/hmm models by query:
      - % python3 fasta_getby.py <fasta_file> --name <model_name> --limit <max_results> --index <int> --range <begin_index, end_index> --length <min_length, max_length>
      - % python3 hmm_getby.py <hmm_file> --name <model_name> --limit <max_results> --index <int> --range <begin index, end index> --length <min_length, max_length>
      - (all search arguments optional)
  - Notes: In order to run a slurm script for a given query/target database, you need to change the paths in the bash script above (need to add an argument so it can be changed from the slurm file). The slurm script can be executed with % sbatch run_cloud_search.sh <my_id> <alpha> <alpha_max>. <my_id> just tells the script what to name the top directory in the results directory. The runall_cloud_search.sh can be called with % bash runall_cloud_search.sh <my_id> and will call % sbatch for a loop of inner.
  - Output to cloud search is currently being sent to stdout, which goes to: “output/<slurm_job_number>/”. You can extract the data by running % bash format_output.sh <output_folder>. This will put all results into that folder called “final.csv”.

- Visualization
  - fb-pruner makefile location: fb-pruner/fb-pruner/Makefile
    - % make DBG=1 (debug flag for vizualization output)
    - USAGE: % fb-pruner <pipeline> <target> <query>
    - % ./build/fb-pruner viz <target> <query>
    - NOTES: This will output the dp matrices to a folder named “test_output/” in the current folder. The “my.*.mx” files are the ones the have dp matrices, one for each method. Currently the visualization pipeline only runs for the first model in the query and target files (writing a loop for it currently).
  - Python visualization script location: fb-pruner/fb-pruner/scripts/build_matrix_heatmap.py
    - python3 build_matrix_heatmap.py <mx_file> -tr -4
    - -tr renders a heatmap of just the match state matrix with the viterbi traceback.
    - -4 renders a heatmap of match,insert,delete,special state matrices (no traceback).