·····worker·····

worker ····· Dask ····· OOM ··· spill ·····

Dask worker ····· Dask ····· worker·····
· OOM····· spill ····· worker·scheduler ·····
·····

····· worker

·· Dask · distributed ···worker ···dashboard ····· ID····· Slurm ·····Dask ·
managed memory ·····unmanaged ·· Python·NumPy ····· managed ···
····· worker ····· OOM · cgroup ····· Nanny ·····
·····

··	·····	····

managed ····· spill | ····· | ·····

·· RSS ·· managed ·· | ····· | ·····

··· cgroup · OOM ·· | ····/···· | ·· worker ·····

spill ····· | ····· | ····· I/O

·····

·····join · groupby ·····
·····join key · groupby key ····· ke
y····· Parquet ·····

·····worker RSS·spill ·····
····· join/groupby ····· schema·····
·· worker ····· worker ·····

·· spill ·····

spill ····· worker ····· worker ·····
····· worker ····spill ····· s
cratch ····· worker ·····

··· worker ····· scheduler·····worker ···· worker ·····Das
k · memory_limit ····· cgroup ····· worker ··
····· spill ·····

·····

·····worker ·· RSS·
managed/unmanaged·spill····· schema·····worker ·····
····· run_id ·····

····· join ·····
····· worker ·····spill ·····
····· OOM ···

····

Dask ·· worker ····· https://distributed.dask.org/en/stable/worker-memory.html

Dask ·· Parquet ····· https://docs.dask.org/en/stable/dataframe-parquet.html