Why is current scheduling a problem?

- Current cloud scheduling strategies are too naive, focused on satisfying industry needs.
- Schedulers are too simple: no priorities, static partitioning, static quotas, simplistic queuing.
- From a resource provider point of view:
  - Difficult to partition resources.
  - Compromise between using the infrastructure as much as possible and give enough share to each group.
  - Difficult to get a good utilization of the infrastructure (e.g. backfilling not possible).
  - Difficult to prioritize workloads from different users/groups.
- From a user’s point of view:
  - Difficult to exploit resources: limited quota even if free resources.
  - Opportunistic usage not possible.
Three complementary approaches:

**Fair Sharing scheduling** to implement resource allocation based on fair-share approach rather than a static quota.

**Partition Director** to easily move resources between different infrastructures and adjust the user shares.

**Preemptible instances** to allow a better utilization of idle resources.