



Figure 1: Gantt plot visualization of task distribution over 4 slave nodes, each with 16 available cores. Tasks are denoted by horizontal lines, and the shaded regions show tasks managed by a single, common executor. The numbers running horizontally at the top denote the stage number, as formed in the execution graph. Stages 0 and 2 consist of the random initialization of two block vector RDDs,  $v_1$  and  $v_2$ , while stages 1 and 3 are the materializations of those vectors, respectively, triggered by a count operation. Stages 6 through 15, inclusive, consist of materializing the vector  $v_1 + v_2$  via a join operation. The block size was  $2^{20}$ , with  $\alpha = 1$  partitions per core and  $\gamma = 1$  blocks per partition.