

RDD → DAG

Spark traduit les transformations sur les RDDs en DAG
(Directed Acyclic Graph)

DAG scheduler

Un *DAG scheduler* divise les opérations en étapes (*Stages*) de tâches (*Tasks*). Une étape est composée de *Tasks* basées sur les partitions du RDD (données distribuées).

Task Scheduler

Les étapes sont transmises au planificateur de tâches (*Task Scheduler*). Le *Task Scheduler* lance les tâches via le gestionnaire du cluster (*Cluster Manager*).