1. 广播变量

Spark 可将数据广播到执行当前任务的所有 Executor 中。
场景：大表匹配小表时，可将小表数据（匹配规则）广播到 Executor 中缓存起来。

如果匹配规则存储在分布式文件系统中，则先收集到 Driver 端，由 Driver 保存全量匹配规则；然后通过网络将全量的匹配规则广播出去，即从 Driver 端复制多份到当前任务的所有 Executor 中。

类似于 mapreduce 的大表 join 小表，将小表数据缓存到 map 端，匹配时在 map 端进行 join。而 hive 中的大表 join 小表，则将大表放在前面。