1. 图的并行计算

图的分布式或者并行处理其实是把图拆分成很多的子图，然后分别对这些子图进行计算，计算时可以分别迭代进行分阶段的计算，即对图进行并行计算。

2. 图的存储模式

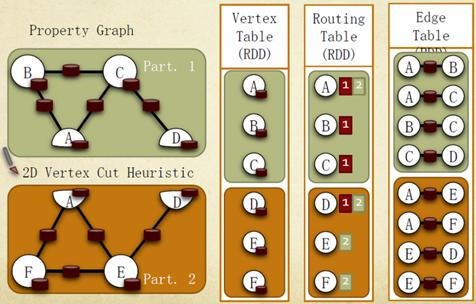
一般点分割：每条边只存储一次，都只会出现在一台机器上。邻居多的点会被复制到多台机器上，增加了存储开销，同时会引发数据同步问题。优点是可以大幅减少内网通信量。

1. GraphX的存储模式

l**VertexTable(id, data)**：id为Vertex id，data为Edge data

l**EdgeTable(pid, src, dst, data)**：pid为Partion id，src为原定点id，dst为目的顶点id

l**RoutingTable(id, pid)**：id为Vertex id，pid为Partion id



2. 图并行计算

均基于Bulk Synchronous Parallell，即整体同步并行，它将计算分成一系列的超步（superstep）的迭代（iteration）。从纵向上看，它是一个串行模式，而从横向上看，它是一个并行的模式，每两个superstep之间设置一个栅栏（barrier），即整体同步点，确定所有并行的计算都完成后再启动下一轮superstep。

语言选择：

scala写graphX方便，scala和java无缝连接，旁边新建java文件import就可以调用scala，spark就是scala写的，提供了各种接口，因此，应该也可以提供python接口。