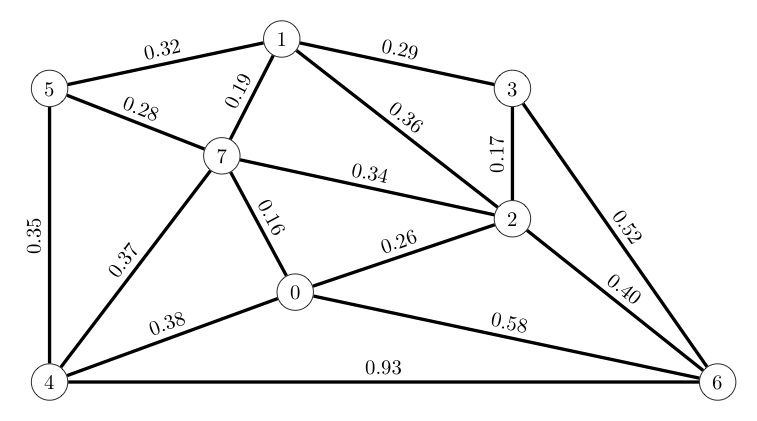
考试试题：

1. 体系结构：
   1. 假设一个事务是将SQL嵌入到C语言中实现的，在运行过程中，有大约90%的时间是花在SQL语句执行上的，如果只对SQL语句实施并行，那么期望能够获得多大的加速比，请说明理由。
   2. 一些数据库操作，比如连接操作，在数据（比如，连接涉及的其中一个关系表中的数据）存放在内存中和不在内存中这两种情况下，速度会有明显差异。请利用上述现象，解释超线性加速比（superliner speedup）现象，即一个程序的加速比高于分配给它的资源数量。
2. 并行数据库
   1. 请描述流水线并行的优点和缺点。
   2. 对于下面这样一个图，要找一个从节点5到节点6的路径，使得路径上的权重和最小。请设计一个并行策略完成这个任务。  
      
3. 分布式数据库
   1. 请列举一个读一次、写所有可用（read one, write all available）方法导致错误状态的一个例子。
   2. 讨论集中式数据库和分布式数据库的优缺点。
4. 时空数据和移动性
   1. 数据库中，时间有哪两种类型，如何区分。
   2. 如果通过增加一个时间属性来将关系转换为时态关系，函数依赖能保持吗，为什么？
5. 大数据
   1. 假设你希望将一个大学的schema建模成一个图（graph），对于下面的每一个关系，他们是建模成节点还是边？这个模型能捕捉sections和courses之间的连接吗？  
      (1) student (2) instructor (3) course (4) section (5) takes (6) teaches
   2. 假如你需要保存很多的小文件（每个大小2K），从分布式文件系统和分布式键值存储中选择，你会选择哪种，为什么？