

# SeqStar 技术介绍

崔延宣

March 2, 2021

SeqStar 是一个支持现实场景复杂查询语句的属性图匹配系统。SeqStar 由两部分组成：“以节点为中心”的存储引擎和属性图查询引擎。

SeqStar 的存储引擎能更好地支持属性图。为此我们提出了一种“以节点为中心”的存储模式：我们将与节点有关的全部边都存在邻接点上，而不是以出边、入边的方式分开存储。在进行数据图扫描时，SeqStar 可以直接快速访问到和节点相连的所有边信息，不需要再从出边入边中分别查找。SeqStar 设计了空间占用很小的索引加速节点的查找：全局索引提供了标签到节点的映射，给定节点标签，它可以快速定位到具有这些标签的节点（标签相同的节点连续存放）；局部索引为每个节点提供了标签到它的邻接点的映射，标签相同的邻接点也是连续存放。在接口设计上，SeqStar 为上层提供了两个迭代器来访问数据图：一是 VertexIter，给定标签使用它可以遍历所有具有这个标签的节点；另一个是 NeighborIter，给定一个节点和节点标签通过它可以遍历具有给定标签的所有邻接点。通过这两个迭代器可以避免对磁盘的顺序访问。

SeqStar 的查询引擎可以高效地执行复杂的图查询语句。SeqStar 首先分析查询语句，从 WHERE 语句中提取出一系列“局部约束”下推到图扫描过程中用来加速过滤不必要的匹配。SeqStar 将原始的模式图分解为一系列的星形子图（每个星形子图包含一个根节点和一系列与根节点相邻的叶子节点），通过存储引擎使用一次顺序扫描完成星形子图的匹配。对于匹配结果，SeqStar 通过推迟笛卡尔积、合并等价节点实现了一种压缩率高达百万的压缩算法。压缩数据可以在匹配星形子图过程中以追加的方式存储。为了进一步减少内存开销，我们设计了一种流水线连接（pipeline join）算法，在不产生额外内存占用的情况下直接对压缩结果进行连接操作。最终结果可以流式地对连接操作结果做笛卡尔积输出。