Команда 1

Рожков Александр, Перевалов Ефим, Нурмухаметов Рафик

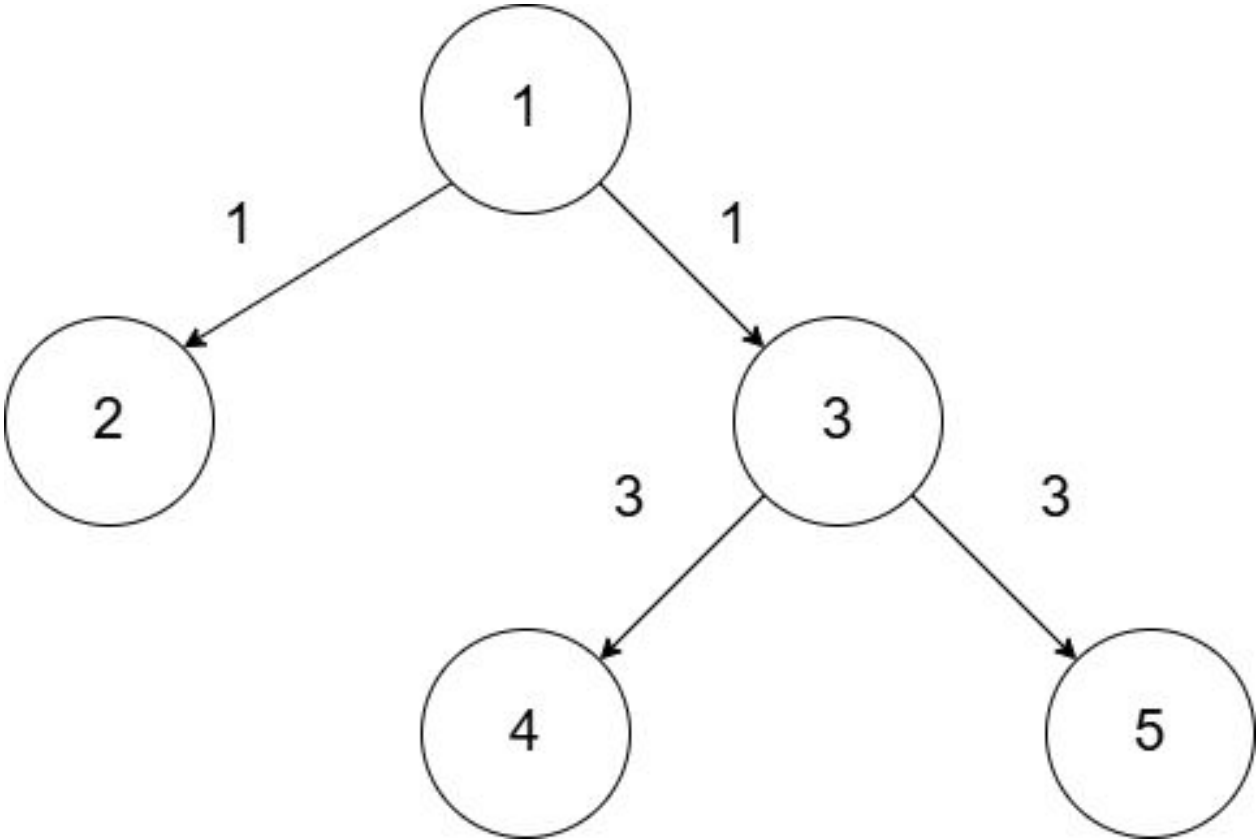
BFS Single-source Parent

BFS Single-source Parent - это алгоритм поиска в ширину, который обходит граф от заданной начальной вершины, сохраняя информацию о "родителях" каждой вершины

BFS Single-source Parent в PySpark

- Отправляем из стартовой вершины сообщение соседям (номер вершины)
- Обновляем таблицу при получении сообщения вершинами
- Дальше отправляем сообщение от “детей”. В сообщение меняем значение номера вершины

BFS Single-source Parent в PySpark



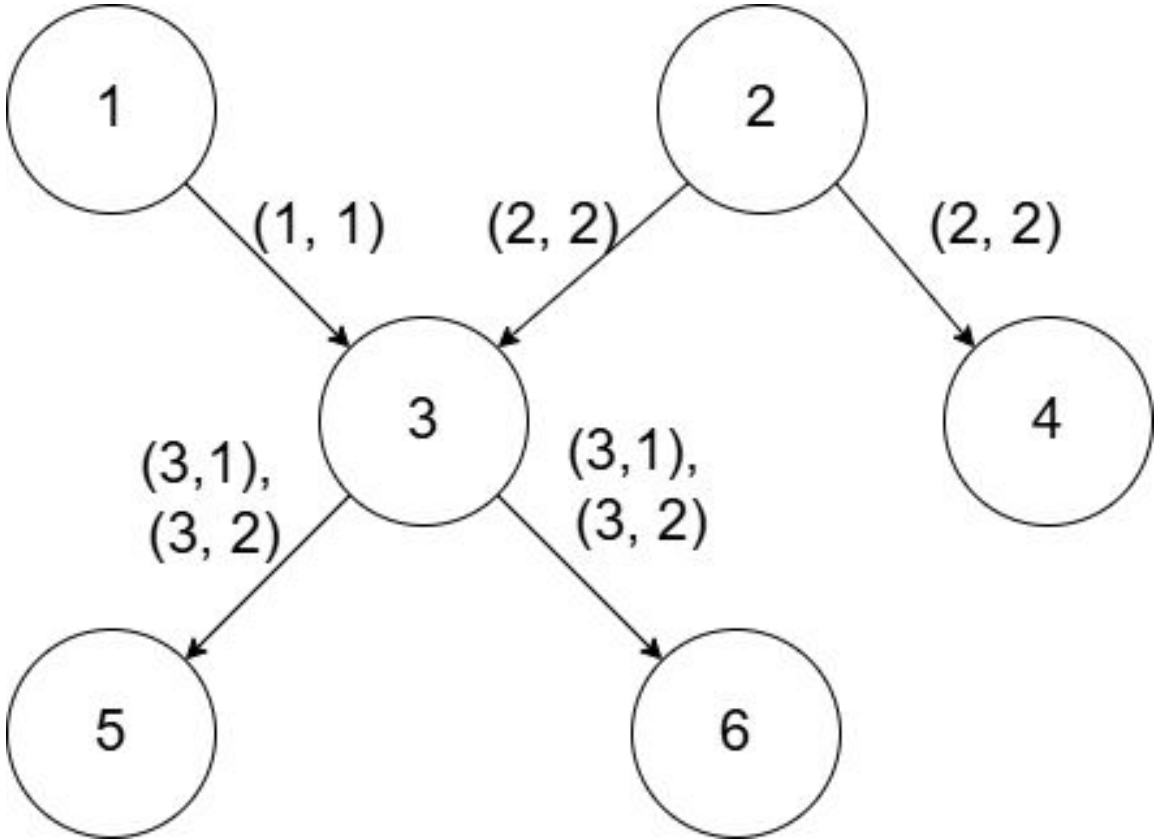
BFS Multiple-source Parent

Multiple-source Parent BFS - это алгоритм обхода графа, который начинается с нескольких исходных вершин одновременно, а не с одной исходной вершины, как в Single-source Parent BFS

BFS Multiple-source Parent в PySpark

- Отправляем из стартовых вершин сообщение соседям (номер вершины и номер стартовой вершины)
- Обновляем таблицу при получении сообщения вершинами
- Дальше отправляем сообщение от “детей”. В сообщение меняем значение номера вершины

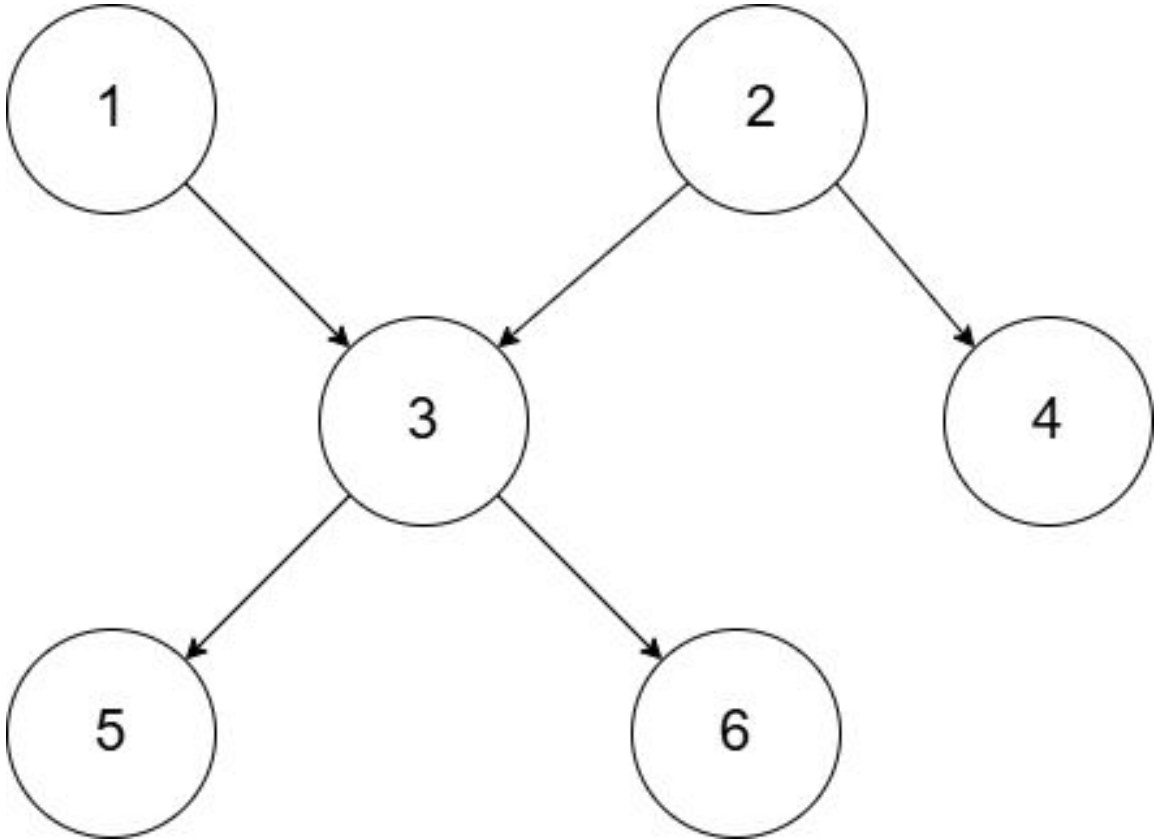
BFS Multiple-source Parent в PySpark



BFS Multiple-source Parent в GraphBLAS

- Граф представляется в виде матрицы смежности (строки/столбцы – вершины, ненулевое значение – наличие ребра между вершинами)
- Матрично-векторное умножение может выполнять обход BFS путем распространения вектора вдоль матрицы смежности

BFS Multiple-source Parent в GraphBLAS



BFS Multiple-source Parent в GraphBLAS

CB1	1	0	0	0	0	0
CB2	0	1	0	0	0	0
CB1	0	0	1	0	0	0
CB2	0	0	1	1	0	0
CB1	0	0	0	0	1	1
CB2	0	0	0	0	0	0

Эксперимент

На какой из 2 библиотек (PySpark, GraphBLAS) быстрее будет реализация BFS Single-source Parent, BFS Multiple-source Parent. Для MS BFS количество стартовые вершины: 2, 4, 8, …, 32. На каждый вариант по 30 запусков

Граф	Вершин	Ребер
Math_overflow	55 863	858 490
wiki-Voute	7 115	103 689
Email-Enron	36 692	183 831
CA-AstroPh	18 772	396 160
p2p-Gnutella31	62 586	147 892
soc-sign-Slashdot090216	81 871	545 671

Эксперимент

Датасет - Stanford Large Network Dataset Collection

Характеристика машины

- ОС Ubuntu 20.04 LTS
- Intel Core i5-10210U
  - 1.6 GHz
  - 4 ядра
- 16 GB RAM