

# Команда 1

Рожков Александр, Перевалов Ефим, Нурмухаметов  
Рафик

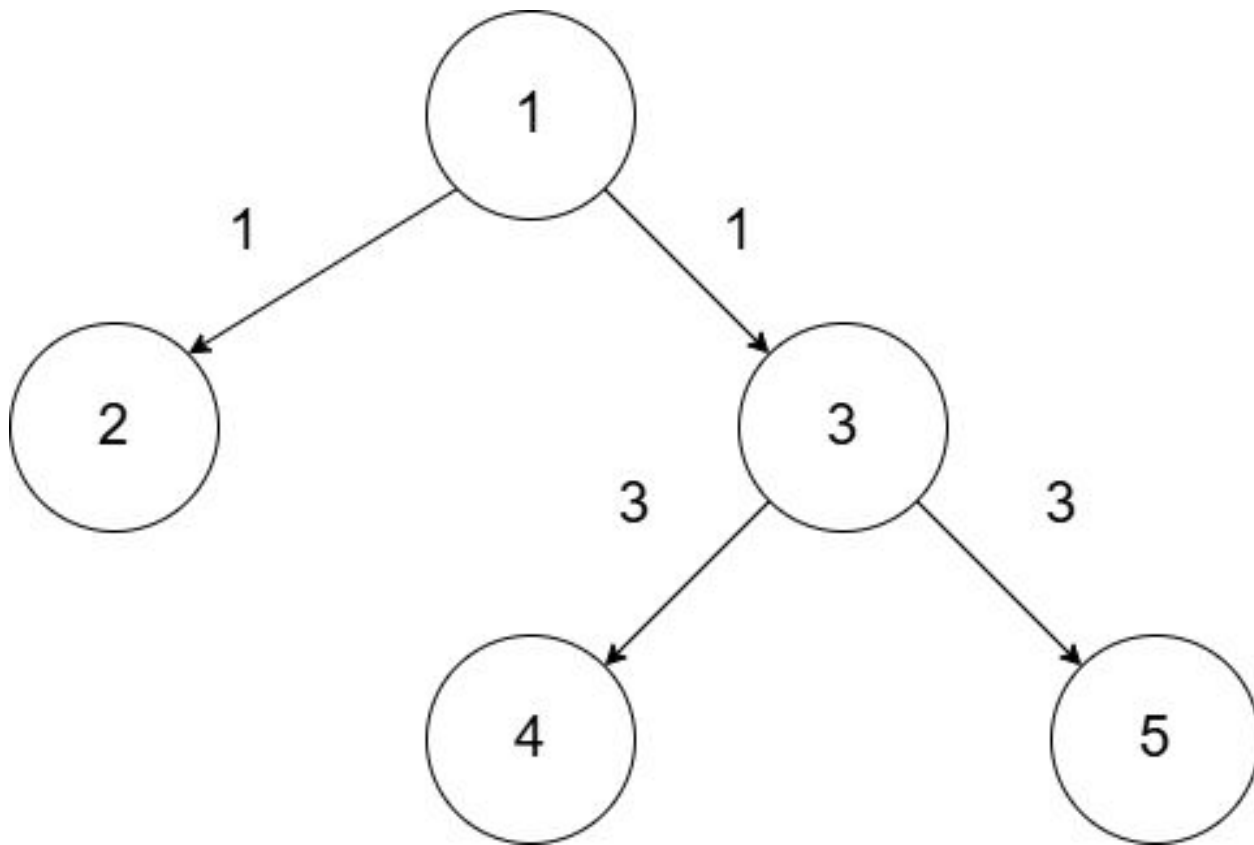
# BFS Single-source Parent

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



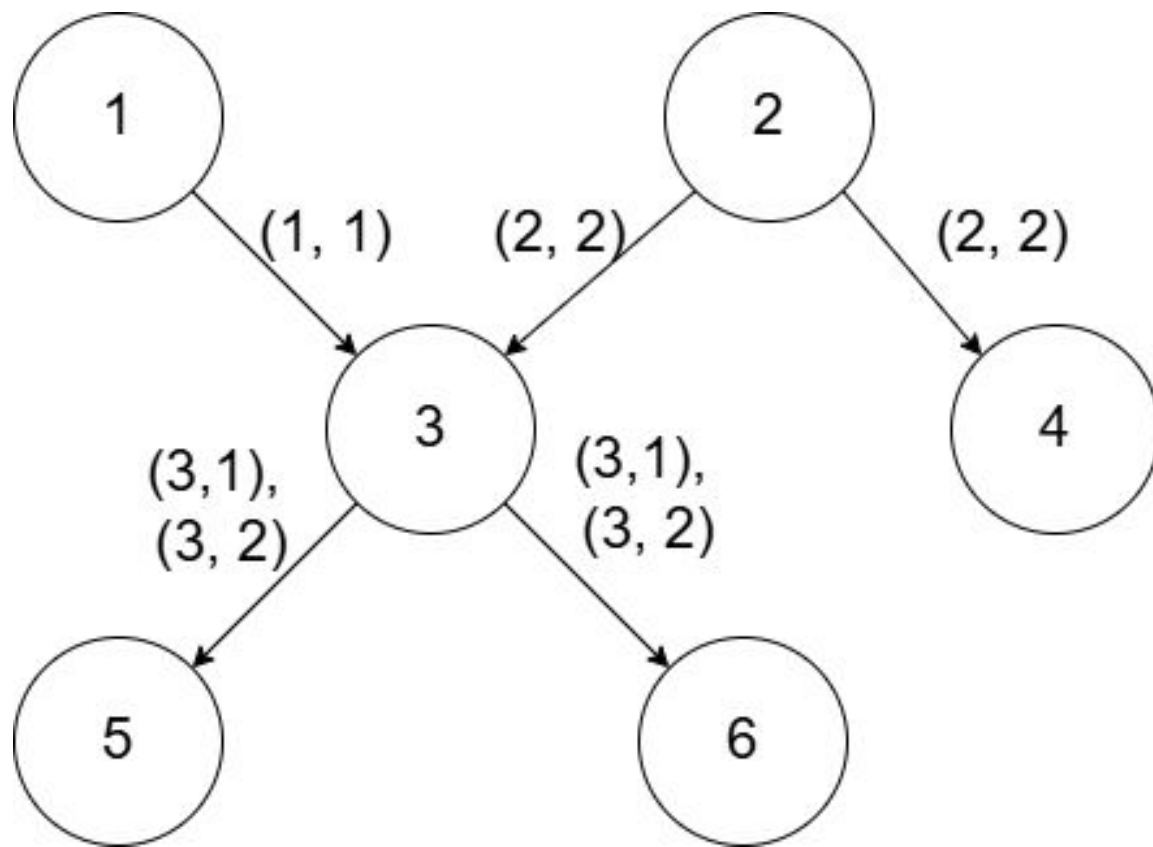
# BFS Multiple-source Parent

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

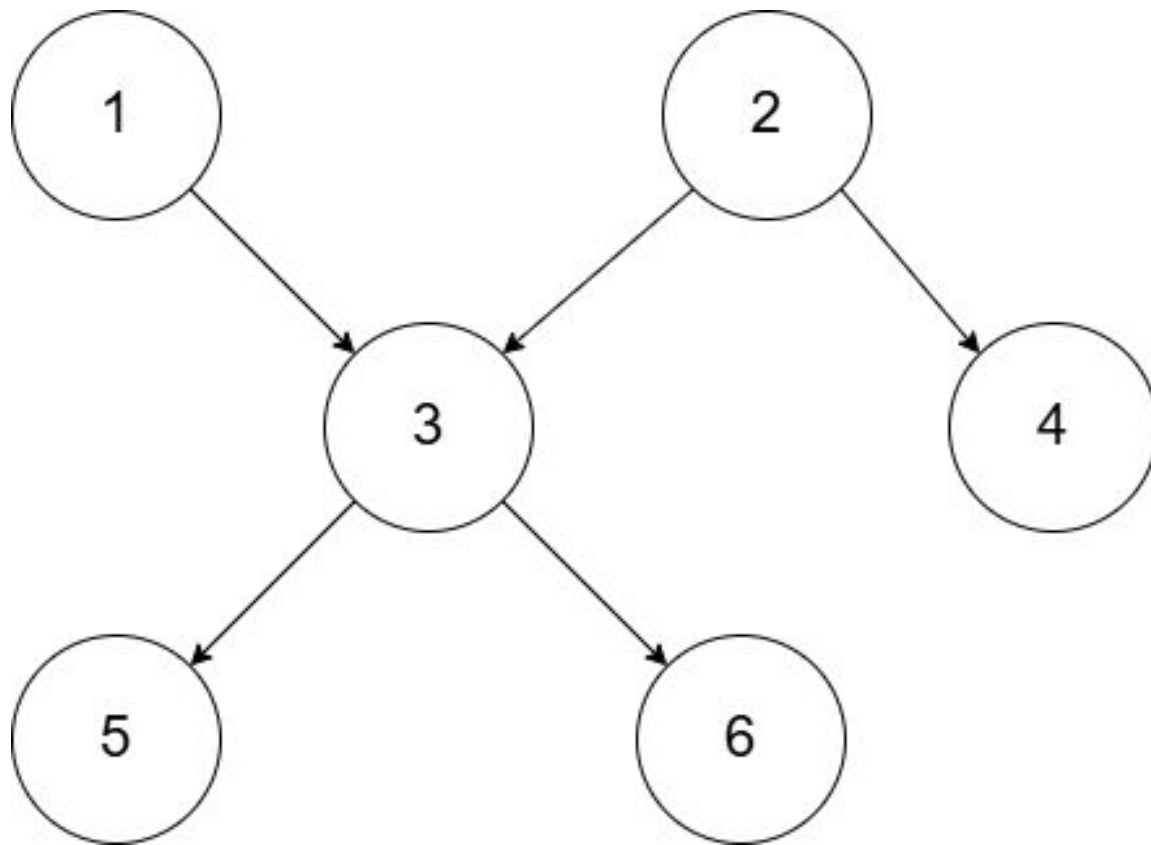


## BFS Multiple-source Parent в GraphBLAS

- ❑ 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      |
|------|------------|------------|
| USA  | 23 947 347 | 58 333 344 |
| CTR  | 14 081 816 | 34 292 496 |
| W    | 6 262 104  | 15 248 146 |
| E    | 3 598 623  | 8 778 114  |
| LKS  | 2 758 119  | 6 885 658  |
| CAL  | 1 890 815  | 4 657 742  |

# Эксперимент



[Датасет](#) - USA road networks

# Характеристика машины

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