A Neat MapReduce Animation!

Courtesy of: Profs. Michael Franklin and Dan Nicolae
CS-Stat 119 (Intro to Data Science II)
The University of Chicago
The Map Reduce Abstraction

Example: Word-Count

**Map** (docRecord) {
    for (word in docRecord) {
        emit (word, 1)
    }
}

**Reduce**(word, counts) {
    emit (word, SUM(counts))
}

Map: Idempotent
Reduce: Commutative and Associative

[Dean & Ghemawat, OSDI'04]
The Map Reduce System

[Dean & Ghemawat, OSDI'04]
The Map Reduce System

[Dean & Ghemawat, OSDI'04]
The Map Reduce System

[Dean & Ghemawat, OSDI'04]
The Map Reduce System

Mapper 1
- Record
- Reduce Buffer
- Record
- Reduce Buffer
- Record

Mapper 2
- Record
- Reduce Buffer
- Record
- Reduce Buffer
- Record

Mapper 3
- Record
- Reduce Buffer
- Record
- Reduce Buffer
- Record

Reducer 1: $n$ color = key value, $n$ = count for key

Reducer 2

[Dean & Ghemawat, OSDI'04]
The Map Reduce System

Mapper 1
Record
Record
Record

Reducer Buffer

Mapper 2
Record
Record
Record

Reducer Buffer

Mapper 3
Record
Record
Record

Reducer Buffer

[Dean & Ghemawat, OSDI'04]

n color = key value,
\( n \) = count for key
The Map Reduce System

Mapper 1
Record
Record
Record

Mapper 2
Record
Record
Record

Mapper 3
Record
Record
Record

Reduce Buffer

Reducer 1

Reducer 2

color = key value, n = count for key

[Dean & Ghemawat, OSDI'04]
The Map Reduce System

Mapper 1
- Record
- Record
- Record

Mapper 2
- Record
- Record
- Record

Mapper 3
- Record
- Record
- Record

Reducer 1
- Reduce Buffer
  - 2

Reducer 2
- Reduce Buffer
  - 2
  - 1

n color = key value, n = count for key

[Dean & Ghemawat, OSDI'04]
The Map Reduce System

Mapper 1
Record
Record
Record

Mapper 2
Record
Record
Record

Mapper 3
Record
Record
Record

Reduce Buffer
2
1 1

Reduction Buffer
1

Reducer 1
2
1
1
2

Reducer 2
2
1
1
2

n
color = key value, n = count for key

[Dean & Ghemawat, OSDI'04]
The Map Reduce System

Mapper 1
Record
Record
Record

Mapper 2
Record
Record
Record

Mapper 3
Record
Record
Record

Reduce Buffer
Reduce Buffer
Reduce Buffer

Reducer 1

Reducer 2

n
color = key value, n = count for key

[Dean & Ghemawat, OSDI'04]
The Map Reduce System

Mapper 1
- Record
- Record
- Record

Mapper 2
- Record
- Record
- Record

Mapper 3
- Record
- Record
- Record

Reduce Buffer
- 1
- 1

Reducer 1
- 1
- 1
- 1
- 2

Reducer 2
- 2
- 2
- 1
- 1

n color = key value, n = count for key

[Dean & Ghemawat, OSDI'04]
The Map Reduce System

Mapper 1
- Record
- Record
- Record

Mapper 2
- Record
- Record
- Record

Mapper 3
- Record
- Record
- Record

Reducer 1
- Color = key value, n = count for key

Reducer 2

[Dean & Ghemawat, OSDI'04]
The Map Reduce System

Mapper 1
- Record
- Record
- Record

Mapper 2
- Record
- Record
- Record

Mapper 3
- Record
- Record
- Record

Reduce Buffer
- 1
- 1
- 2

Reduce Buffer
- 1
- 1
- 2

Reducer 1
- 1
- 1
- 2

Reducer 2
- 2
- 1
- 3

Output 1

Output 2

[Dean & Ghemawat, OSDI'04]

n color = key value, n = count for key
The Map Reduce System

Mapper 1
- Record
- Record
- Record

Mapper 2
- Record
- Record
- Record

Mapper 3
- Record
- Record
- Record

Reduce Buffer
- 2
- 1
- 1

Reduce Buffer
- 1
- 2

Reduce Buffer
- 1
- 1

Reducer 1
- 1
- 1
- 2

Reducer 2
- 2
- 1
- 1

Output
- 2
- 3

Dean & Ghemawat, OSDI'04

n = count for key

color = key value,
The Map Reduce System

Mapper 1
- Record
- Record
- Record

Mapper 2
- Record
- Record
- Record

Mapper 3
- Record
- Record
- Record

Reduce Buffer
- 2
- 1
- 1

Reducer 1
- 1
- 1
- 2

Reducer 2
- 2
- 1
- 1
- 3
- 2

Output 1
- 2

Output 2
- 3

[Dean & Ghemawat, OSDI'04]

n
color = key value, n = count for key
The Map Reduce System

Mapper 1
- Record
- Record
- Record

Mapper 2
- Record
- Record
- Record

Mapper 3
- Record
- Record
- Record

Reduce Buffer
- 2
- 1
- 1

Reducer 1
- 1
- 1
- 1
- 2

Reducer 2
- 2
- 1
- 1
- 2

Output 1
- 2
- 3

Output 2
- 3
- 3

[Dean & Ghemawat, OSDI'04]
The Map Reduce System

Mapper 1
- Record
- Record
- Record

Mapper 2
- Record
- Record
- Record

Mapper 3
- Record
- Record
- Record

Reduce Buffer
- 2
- 1
- 1

Reduce Buffer
- 1
- 2

Reduce Buffer
- 2
- 1
- 1

Reducer 1
- 1
- 1
- 1

Reducer 2
- 2
- 1
- 1

Output 1
- 2
- 3

Output 2
- 3
- 3

[Dean & Ghemawat, OSDI'04]

n = count for key

- color = key value,
The Map Reduce System

Mapper 1
Record
Record
Record
Reduce Buffer
2
1
1

Mapper 2
Record
Record
Record
Reduce Buffer
1
2

Mapper 3
Record
Record
Record
Reduce Buffer
1
1
2

Reducer 1
1
1
1
2

Reducer 2
2
1
1
2

Output1
2
3

Output2
3
3

[n = count for key]

[Dean & Ghemawat, OSDI'04]
The Map Reduce System

[Dean & Ghemawat, OSDI'04]
The Map Reduce System

[Dean & Ghemawat, OSDI'04]