



COMP 30520-41110

Example of Map/Reduce



Outline

- Problem
- Map/Reduce functions
- Implementation



Problem: Sort

- Sort the temperature data from scientific experiments

INPUT

| Year | Month | Day | Temperature |
|------|-------|-----|-------------|
| 2015 | 10 | 29 | 12 |
| 2015 | 10 | 30 | 11 |
| 2015 | 10 | 31 | 14 |
| 2015 | 11 | 01 | 12 |
| 2015 | 11 | 02 | 13 |
| ... | ... | ... | ... |

OUTPUT

| Year | Month | Temperature |
|------|-------|-------------|
| 2015 | 10 | 11, 12, 14 |
| 2015 | 11 | 12, 13 |



Map Function

- Input:?
- Output:?

INPUT

| Year | Month | Day | Temperature |
|------|-------|-----|-------------|
| 2015 | 10 | 29 | 12 |
| 2015 | 10 | 30 | 11 |
| 2015 | 10 | 31 | 14 |
| 2015 | 11 | 01 | 12 |
| 2015 | 11 | 02 | 13 |
| ... | ... | ... | ... |



Map Function

- Input: $\langle \#line, \{year, month, day, temperature\} \rangle$
- Output: $\langle \{year, month\}, \{temperature\} \rangle$

Ex: $\langle \{2015, 10\}, 12 \rangle, \langle \{2015, 10\}, 11 \rangle$

INPUT

| Year | Month | Day | Temperature |
|------|-------|-----|-------------|
| 2015 | 10 | 29 | 12 |
| 2015 | 10 | 30 | 11 |
| 2015 | 10 | 31 | 14 |
| 2015 | 11 | 01 | 12 |
| 2015 | 11 | 02 | 13 |
| ... | ... | ... | ... |

OUTPUT

| Year | Month | Temperature |
|------|-------|-------------|
| 2015 | 10 | 11, 12, 14 |
| 2015 | 11 | 12, 13 |



Reduce Function

- Input: $\langle \{\text{year, month}\}, \{\text{temperature}\} \rangle$
- Output: ?

INPUT

| Year | Month | Day | Temperature |
|------|-------|-----|-------------|
| 2015 | 10 | 29 | 12 |
| 2015 | 10 | 30 | 11 |
| 2015 | 10 | 31 | 14 |
| 2015 | 11 | 01 | 12 |
| 2015 | 11 | 02 | 13 |
| ... | ... | ... | ... |

OUTPUT

| Year | Month | Temperature |
|------|-------|-------------|
| 2015 | 10 | 11, 12, 14 |
| 2015 | 11 | 12, 13 |



Reduce Function

- Input: $\langle \{\text{year, month}\}, \{\text{temperature}\} \rangle$
- Output: $\langle \{\text{year, month}\}, \text{lists of } \{\text{temperature}\} \rangle$

INPUT

| Year | Month | Day | Temperature |
|------|-------|-----|-------------|
| 2015 | 10 | 29 | 12 |
| 2015 | 10 | 30 | 11 |
| 2015 | 10 | 31 | 14 |
| 2015 | 11 | 01 | 12 |
| 2015 | 11 | 02 | 13 |
| ... | ... | ... | ... |

OUTPUT

| Year | Month | Temperature |
|------|-------|-------------|
| 2015 | 10 | 11, 12, 14 |
| 2015 | 11 | 12, 13 |



Map Function Implementation

- Input: $\langle \#line, \{year, month, day, temperature\} \rangle$
- Output: $\langle \{year, month\}, \{temperature\} \rangle$

Ex: $\langle \{2015, 10\}, 12 \rangle, \langle \{2015, 10\}, 11 \rangle$

```
map ( key, value) {  
    convert(value, reducekey, temperature);  
    emit(reducekey, temperature);  
}
```




Map Function Implementation

- Input: $\langle \#line, \{year, month, day, temperature\} \rangle$
- Output: $\langle \{year, month\}, \{temperature\} \rangle$

Ex: $\langle \{2015, 10\}, 12 \rangle, \langle \{2015, 10\}, 11 \rangle$

```
map ( key, value) {  
    //convert(value, reducekey, temperature);  
    String [] tokens = value.split(',');  
    String reducekey = tokens[0] + tokens[1];  
    int    temperature = Integer.parseInt(tokens[3]);  
    emit(reducekey, temperature);  
}
```



Reduce Function Implementation

- Input: $\langle \{\text{year, month}\}, \{\text{temperature}\} \rangle$
- Output: $\langle \{\text{year, month}\}, \text{lists of } \{\text{temperature}\} \rangle$

```
reduce ( key, value) {  
    int temperature = value;  
    temperatureList.add(temperature) ;  
  
    emit(key, temperatureList) ;  
}
```