

CSE427s - 2a [P] MR1: MR
on Weather Data

1. How would you compute the **maximum temperature per year** from *all* 1.4 billion daily weather records in the National Climatic Data Center database?
group the data, then sort them separately and then combine the result
2. Okay, let's do it! First we extract **year** and **temperature** for each record. That is the *Map phase*. What is the Map phase output for the 6 weather records in our example?
(1901,206)
(1904,39)
3. Now, we **group** the temperatures by year and **sort** the results by year. That is the *Group & Sort phase*. What is the Group & Sort phase output for our example data?
4. And now? ...well let's **get the maximum** per year! This step is called the *Reduce phase*. What is the Reduce phase output for our example data?
(1901,206)
(1904,350)