

Writing Top Notch Programs

- Software framework.

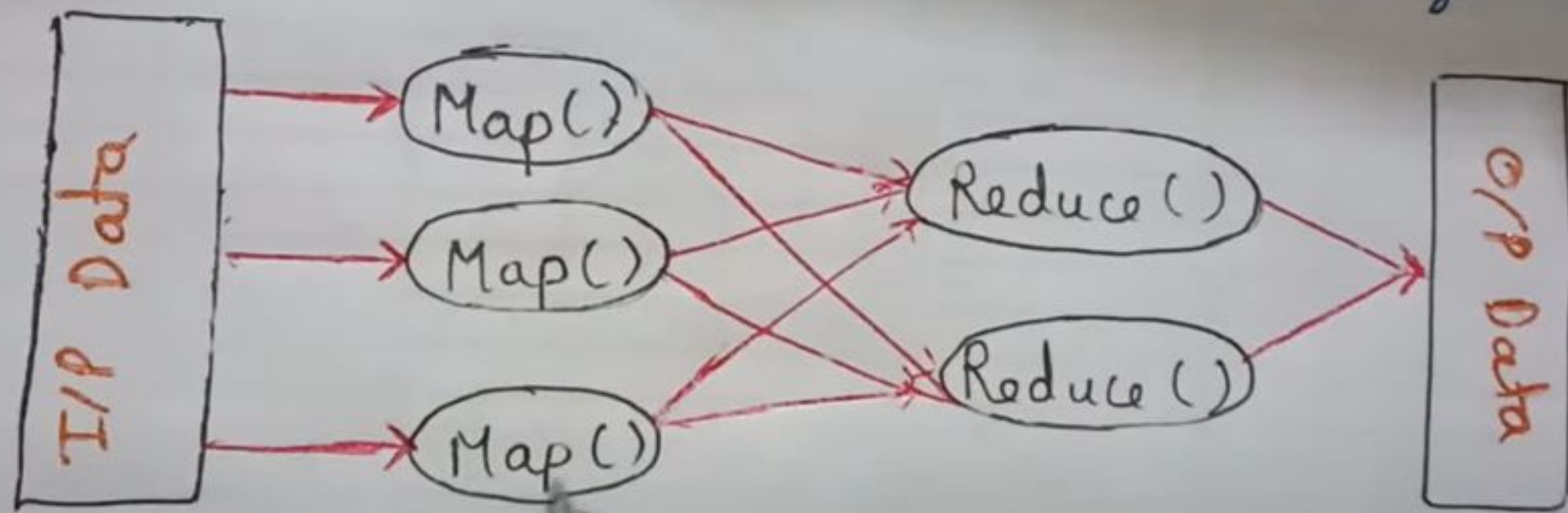


0:05 / 8:42



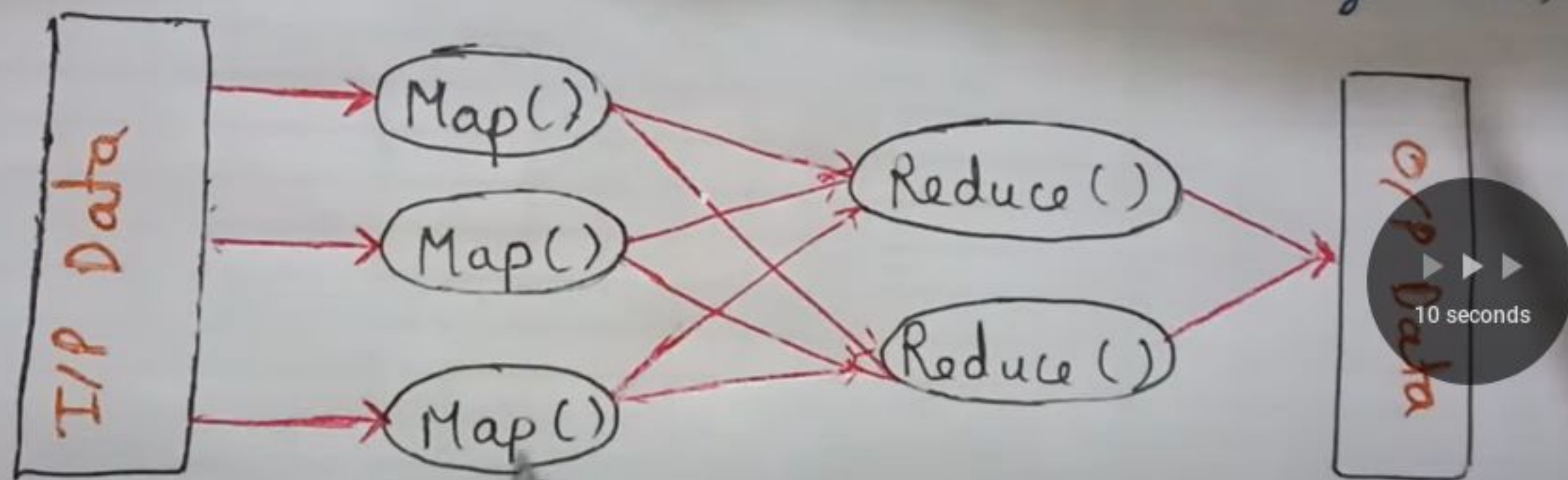
Writing Map Reduce Programs

- Software framework.
- used for writing app which process big amounts of data, in parallel on large cluster.

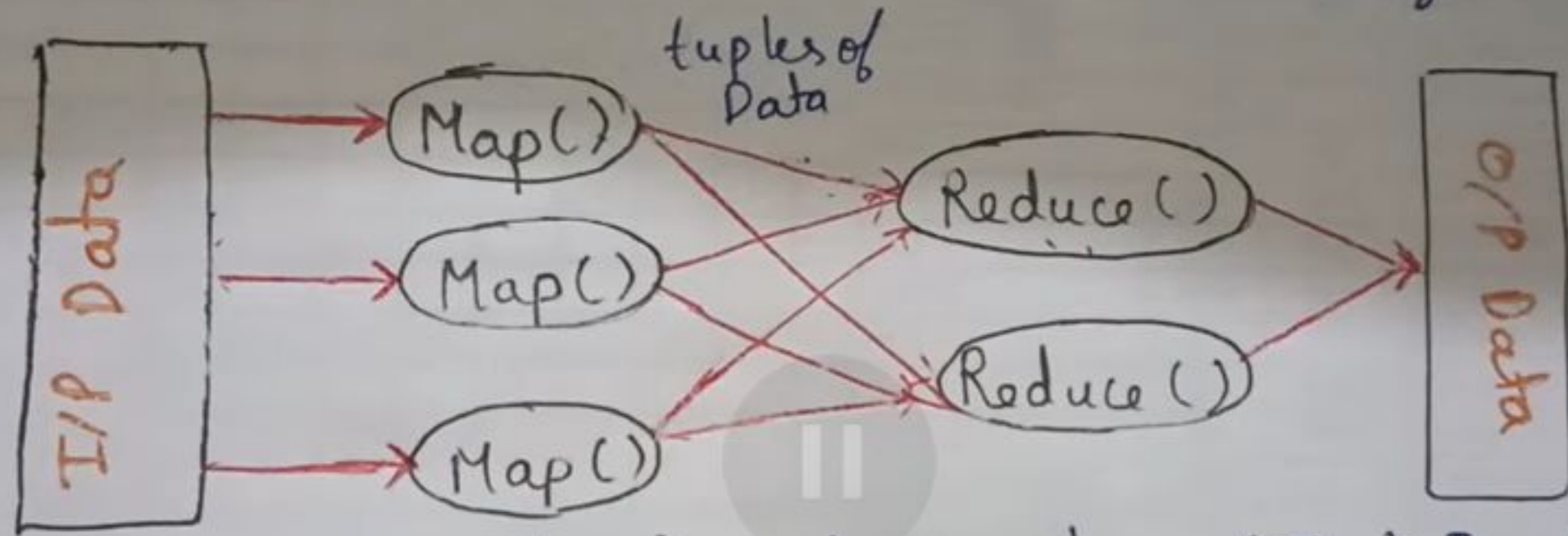


Writing Map Reduce Programs

- Software framework.
- used for writing app which process big amounts of data, in parallel on large cluster.



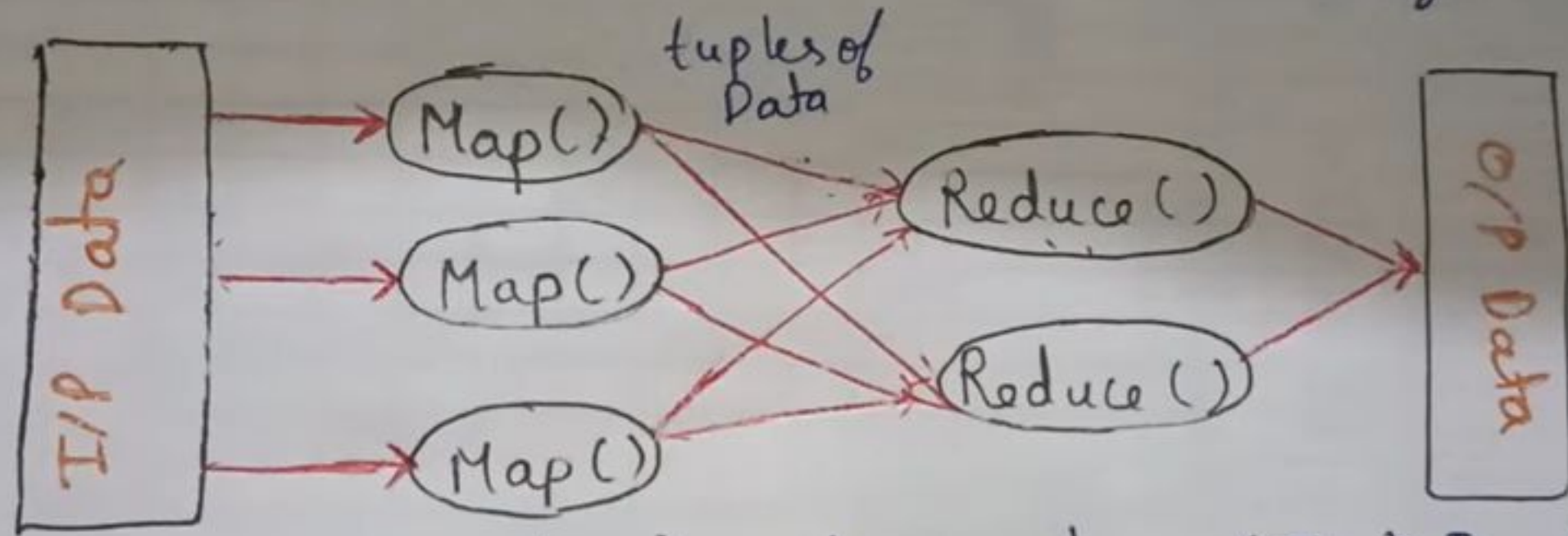
- Software framework.
- used for writing app which process big amounts of data, in parallel on large cluster.



- Single Master (JobTracker) and one slave (TaskTracker) per cluster.

Weather Dataset

- Software framework.
- used for writing app which process big amounts of data, in parallel on large cluster.



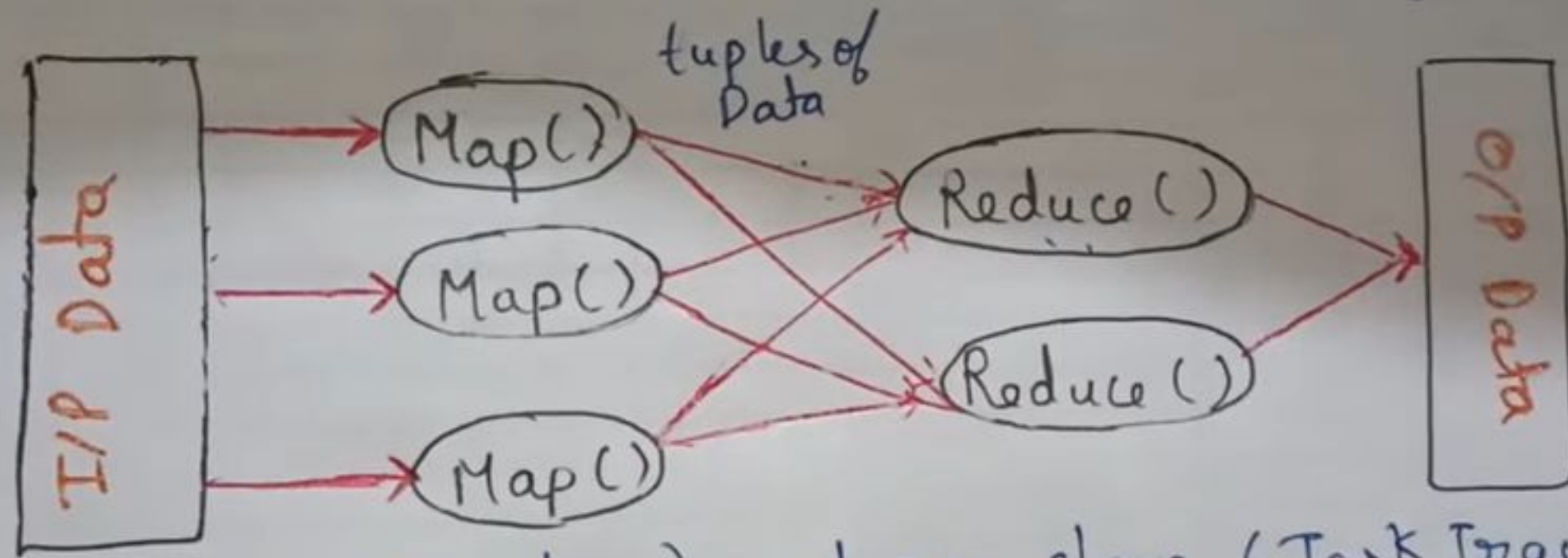
- Single Master (JobTracker) and one slave (TaskTracker) per cluster.

Weather Dataset



- Software framework.

- used for writing app which process big amounts of data, in parallel on large cluster.



- Single Master (JobTracker) and one slave (TaskTracker) per cluster.

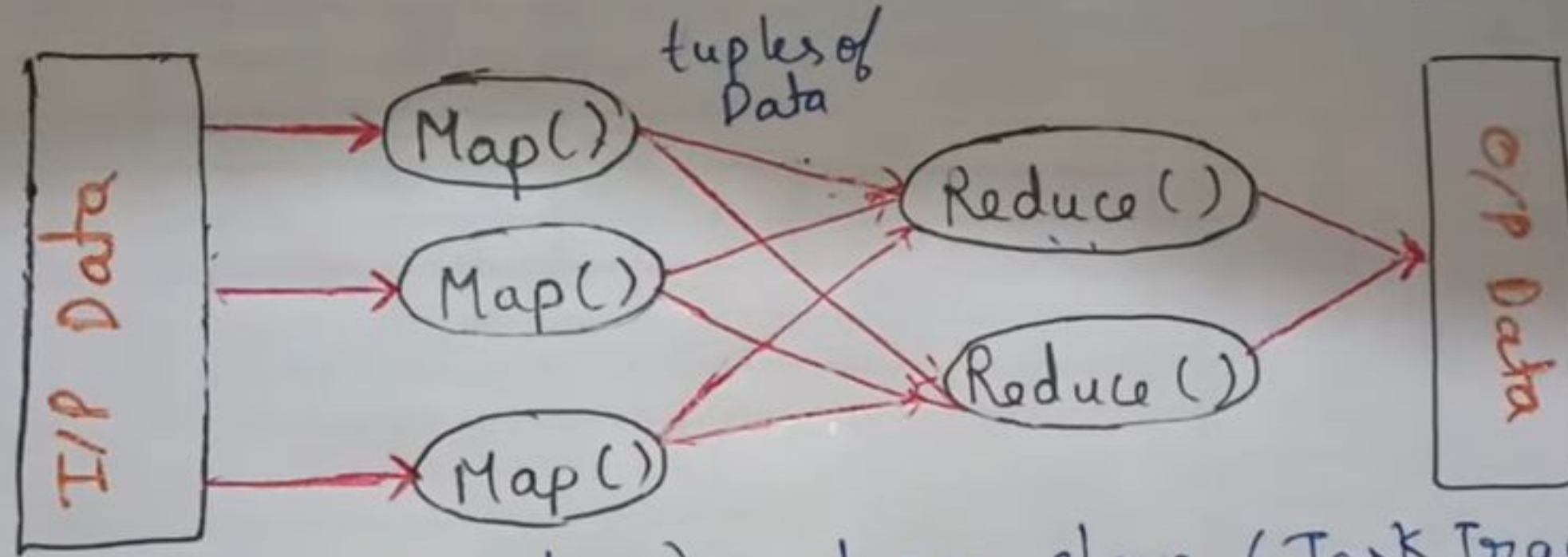
Weather Dataset

Data Format

- from NCEP (National Climatic Data Centre).
- line oriented ASCII format.
- Each line is record.
- Data files are organized

- Software framework.

- used for writing app which process big amounts of data, in parallel on large cluster.



- Single Master (JobTracker) and one slave (TaskTracker) per cluster.

Weather Dataset

Data Format

- from NCDR (National Climatic Data Centre).
- line oriented ASCII format.
- Each line is record.
- Data files are organized

→ Analyzing the Data with Hadoop

→ Weather Dataset → MapReduce Programs → Hadoop



10 seconds



6:25 / 8:42



→ Analyze

→ L

→ I/P - Re
→ Pull out
→ Example

Map Reduce flow:

input

map

0067011990..	(0, 0067011..)	(1950, 0)
0043011990..	(106, 00430..)	(1950, 22)
0043011990..	(212, 00430..)	(1950, -11)
0043012650..	(318, 00430..)	(1950, 111)
0043012650..	(424, 00430..)	(1950, 78)

cat*

map.nb

shuffle

reduce

(1949, [111, 78])	1949
(1950, [0, 22, -11])	1950

cat

reduce.nb