CS 441 – Engineering distributed objects for cloud

ASSIGNMENT 2 – MAP REDUCE DEPLOYMENT

Objective

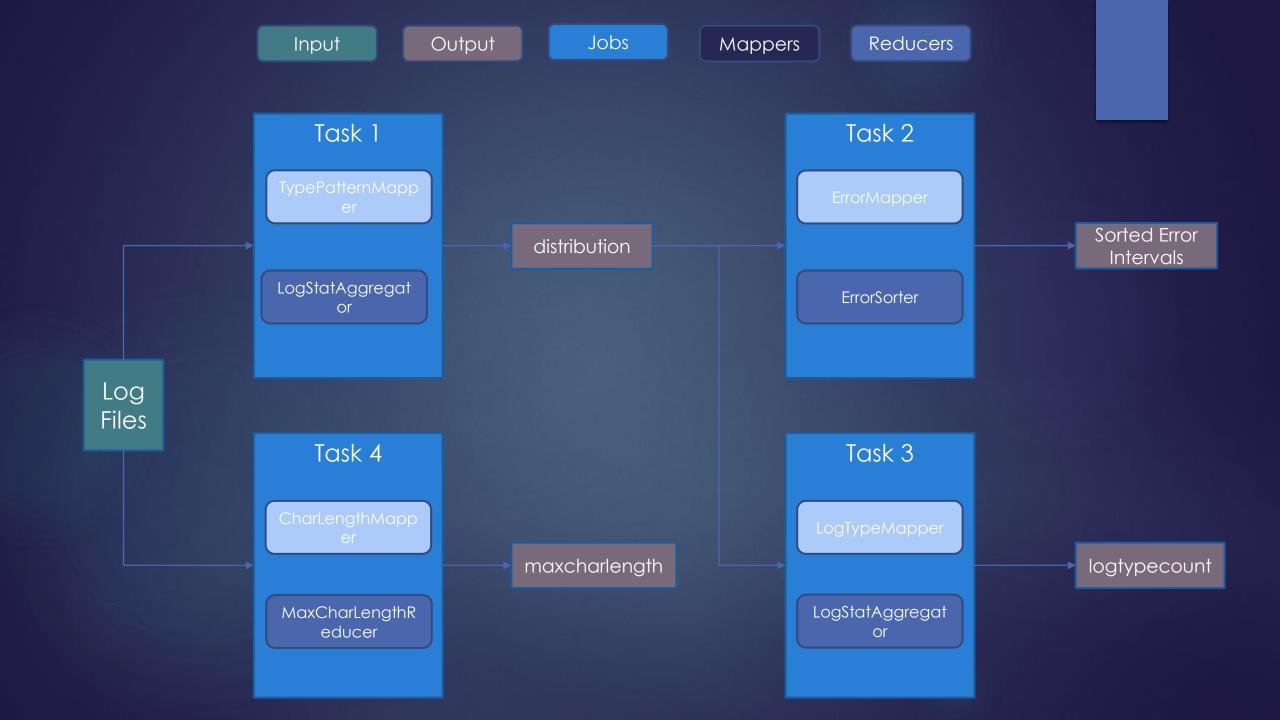
- The objective of this project is to use Map Reduce to achieve distributed processing of Log files to extract required information
- In practice, large scale applications generally produce huge volume of log files that have to be analyzed for various purposes such as
 - ▶ Post-error investigation
 - Risk analysis
 - Performance optimization etc.

Tasks

- We implement 4 tasks to extract information from the log files
 - ▶ Log Distribution: Compute distribution of logs across log types (WARN, DEBUG, ERROR, INFO etc.), time intervals (pre-defined), and presence of desired text patterns
 - ► Error distribution: Compute Intervals sorted descending by count of Error messages that contain a desired text pattern
 - Log count by type: Compute count of log messages for each log type
 - Message Length: Compute the length of log messages with the most number of characters that contain a desired text pattern for each log type

Implementation

- Mapper and reducer classes were implemented for each of these tasks in Scala and deployed as a JAR file
- Since task 2 and 3 are a natural extension of task 1, we will use the output distribution of task 1 to compute the results for tasks 2 and 3 to save processing time
- ▶ The flow of the tasks is given in the next slide



Deployment Demo



Output

Task 1

Intorval	Log Type	Pattern	Count
IIIIEIVUI	Log Type	runem	COULI
13:17:30	DEBUG	has_pattern	50
13:18:00	INFO	has_pattern	1067
13:18:30	INFO	has_pattern	977
13:19:00	DEBUG	has_pattern	88
13:19:00	ERROR	has_pattern	41
13:19:00	WARN	has_pattern	718
13:19:30	ERROR	has_pattern	46
13:19:30	WARN	has_pattern	706
13:20:00	DEBUG	has_pattern	101
13:20:00	WARN	has_pattern	678
13:20:30	WARN	has_pattern	749
13:21:00	WARN	has_pattern	691
13:21:30	ERROR	has_pattern	37
13:21:30	INFO	has_pattern	1041

Task 2

	Log
Interval	Count
13:18:00	47
13:23:00	47
13:19:30	46
13:23:30	43
13:22:30	43
13:25:30	41
13:19:00	41
13:22:00	40
13:26:30	38
13:26:00	38
13:21:30	37
13:24:00	35
13:24:30	35
13:18:30	34

Task 3

Log	Log
Type	Count
DEBUG	3673
INFO	41093
WARN	28634
ERROR	1496

Task 4

Log	Max
Туре	Length
DEBUG	105
INFO	105
WARN	105
ERROR	105



THANK YOU