Advanced Data Engineering: Assignment 9

NGUYEN T. Hoang - SID: 15M54097
(ホアン)

Fall 2015, W831 Tue. Period 5-6
Due date: 2015/12/22

Problem

- Roughly estimate the execution time for parallel block 2way merge sort using 8 and 16 processors.
- Roughly estimate the execution time for parallel block bitonic sort using 8 and 16 processors.
  - Cardinality of a relation R: 100,000.
  - Total length of a tuple: 100B.
  - Disk transfer bandwidth: 10MB/s.
  - Network bandwidth: 10MB/s.
  - *Ignore CPU costs, disk access latency, and network setup time.*

Question 1: Parallel block 2way merge sort

Roughly estimate the execution time using 8 processors.

Answer: Time for sorting fraction of data set in a processor is assumed equal with the number of read and write of disk.

$$T_{8\_sort} = \frac{1/N \times \{R\} \times 2}{\text{Disk Rate}} = \frac{1/8 \times 100,000 \times 100 \times 2}{10,000,000}$$
$$= 0.2 \ second$$

Since the number of tuples is much larger than the number of PEs. We have the time for transfering data throughout the network:

$$T_{8\_transfer} = \frac{(1 - 1/N) \times \{R\}}{\text{Transfer Rate}} = \frac{7/8 \times 100,000 \times 100}{10,000,000}$$
$$= 0.875 \ second$$

Time for disk write and read. Assuming no pipeline effect:

$$T_{8\_disk} = \frac{6 \times (1 - 1/N) \times |R|}{\text{Disk Rate}} = \frac{6 \times (1 - 1/8) \times 100,000 \times 100}{10,000,000}$$
$$= 5.25 \ second$$

Total time to sort for 8 processors is: 6.325 *second*.

Roughly estimate the execution time using 16 processors.

Answer: Time for sorting fraction of data set in a processor is assumed equal with the number of read and write of disk.

$$T_{16\_sort} = \frac{1/N \times \{R\} \times 2}{\text{Disk Rate}} = \frac{1/16 \times 100,000 \times 100 \times 2}{10,000,000}$$
$$= 0.1 \ second$$

Since the number of tuples is much larger than the number of PEs. We have the time for transfering data throughout the network:

$$T_{16\_transfer} = \frac{(1 - 1/N) \times \{R\}}{\text{Transfer Rate}} = \frac{15/16 \times 100,000 \times 100}{10,000,000}$$
$$= 0.9375 \ second$$

Time for disk write and read. Assuming no pipeline effect:

$$T_{16\_disk} = \frac{6 \times (1 - 1/N) \times |R|}{\text{Disk Rate}} = \frac{6 \times (1 - 1/16) \times 100,000 \times 100}{10,000,000}$$
$$= 5.625 \ second$$

Total time to sort for 16 processors is: 6.66 *second*. The reason for decreasing of sorting speed between 8 processors and 16 processor is that the network bandwidth and the disk bandwidth have same data rate. In 16 processors scheme, the time of data transfer in the network dominates other time.

Question 2: Parallel block bitonic sort

Roughly estimate the execution time using 8 processors.

Answer: Time for sorting fraction of data set in a processor is assumed equal with the number of read and write of disk.

$$T_{8\_sort} = \frac{1/N \times \{R\} \times 2}{\text{Disk Rate}} = \frac{1/8 \times 100,000 \times 100 \times 2}{10,000,000}$$
$$= 0.2 \ second$$

Since the number of tuples is much larger than the number of PEs. We have the time for transfering data throughout the network:

$$T_{8\_transfer} = \frac{n \times (n+1) \times \{R\}}{2 \times \text{Transfer Rate}} = \frac{3 \times 4 \times 100,000 \times 100}{2 \times 8 \times 10,000,000}$$
$$= 0.75 \; second$$

Time for disk write and read. Assuming no pipeline effect:

$$T_{8\_disk} = \frac{|R|/N \times n \times (n+1)}{\text{Disk Rate}} = \frac{100,000/8 \times 100 \times 3 \times 4}{10,000,000}$$
$$= 1.5 \; second$$

Total time to sort for 8 processors is: 2.45 *second*.

Roughly estimate the execution time using 16 processors.

Answer: Time for sorting fraction of data set in a processor is assumed equal with the number of read and write of disk.

$$T_{16\_sort} = \frac{1/N \times \{R\} \times 2}{\text{Disk Rate}} = \frac{1/16 \times 100,000 \times 100 \times 2}{10,000,000}$$
$$= 0.1 \; second$$

Since the number of tuples is much larger than the number of PEs. We have the time for transfering data throughout the network:

$$T_{16\_transfer} = \frac{n \times (n+1) \times \{R\}}{2 \times \text{Transfer Rate}} = \frac{4 \times 5 \times 100,000 \times 100}{2 \times 16 \times 10,000,000}$$
$$= 0.625 \; second$$

Time for disk write and read. Assuming no pipeline effect:

$$T_{16\_disk} = \frac{|R|/N \times n \times (n+1)}{\text{Disk Rate}} = \frac{100,000/16 \times 100 \times 4 \times 5}{10,000,000}$$
$$= 1.25 \; second$$

Total time to sort for 16 processors is: 1.975 *second*.