

CUDA Programming Tutorial 4

Parallel Patterns – Scan

Sungjoo Ha

May 24th, 2017

Common Parallel Computing Scenarios

- ▶ 여러 쓰레드가 합심하여 하나의 결과를 생성
 - Reduce
- ▶ 여러 쓰레드가 주어진 데이터를 파티션
 - Split
- ▶ 여러 쓰레드가 쓰레드 별로 가변 길이의 결과를 내야 함
 - Compact

Split

T	F	F	T	F	F	T	F
---	---	---	---	---	---	---	---

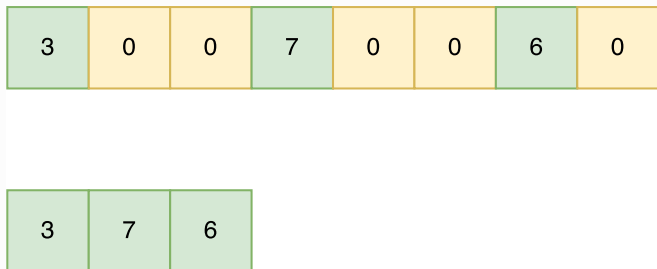
3	1	7	0	4	1	6	3
---	---	---	---	---	---	---	---

T	T	T	F	F	F	F	F
---	---	---	---	---	---	---	---

3	0	6	1	7	4	1	3
---	---	---	---	---	---	---	---

- ▶ 주어진 데이터를 조건에 따라 두 개의 그룹으로 분류
- ▶ 소팅, 트리 생성 등에 활용

Compact



- ▶ 주어진 데이터를 조건을 만족시키는 데이터만 남기기
- ▶ 충돌 체크 등에 사용

Prefix Sum a.k.a. Scan

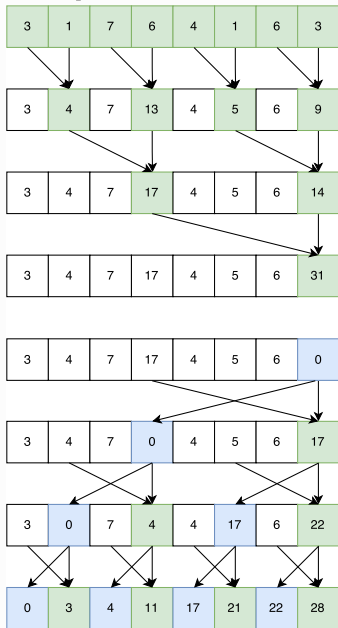
- ▶ 기초적인 non-trivial parallel pattern
- ▶ Prefix sum, scan, cumulative sum 등의 이름으로 불림

Scan

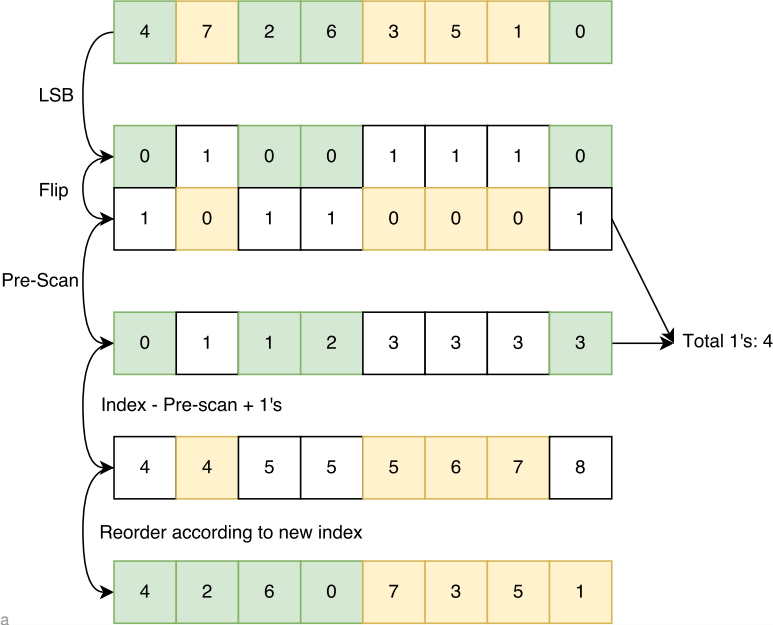


- ▶ 이항 연산을 순차적으로 적용하는 계산
- ▶ Identity로 시작하면 pre-scan, exclusive scan
- ▶ 첫 원소로 시작하면 scan, inclusive scan

Pre-Scan Computation



Parallel Radixsort



Scan

- ▶ 병렬 프로세서에서 $O(\log n)$ 타임 소트 등의 활용
- ▶ 병렬 연산이 아니더라도 벡터 연산을 통해 scan 계산이 직접 루프를 도는 것보다 빠르다면 유용
 - Numpy 등의 벡터 프로세싱 작업 시 유용
 - Scan 연산이 있으면 이동 평균 계산 등은 $O(1)$

References

- ▶ Prefix Sums and Their Applications, Blelloch, 1993
- ▶ Parallel Patterns I, CS 193G Lecture Slides