Windows 멀티스레드 프로그래밍 활용

유영천

https://megayuchi.com

tw:@dgtman

멀티 스레드 프로그래밍의 목적

- 처리량 향상
- 비동기 처리

데모

- 메모리 카피 테스트
- 이미지 프로세싱 테스트
- 라이트맵 계산 테스트
- voxelization

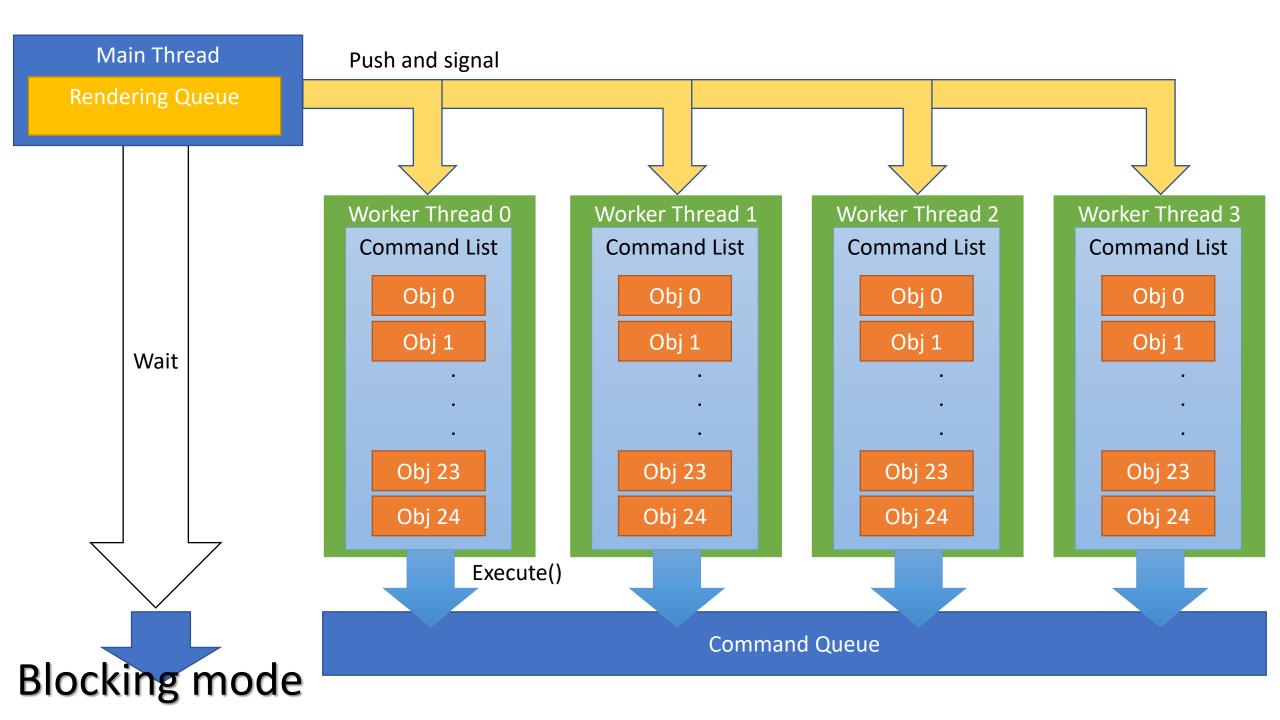
용도별패턴

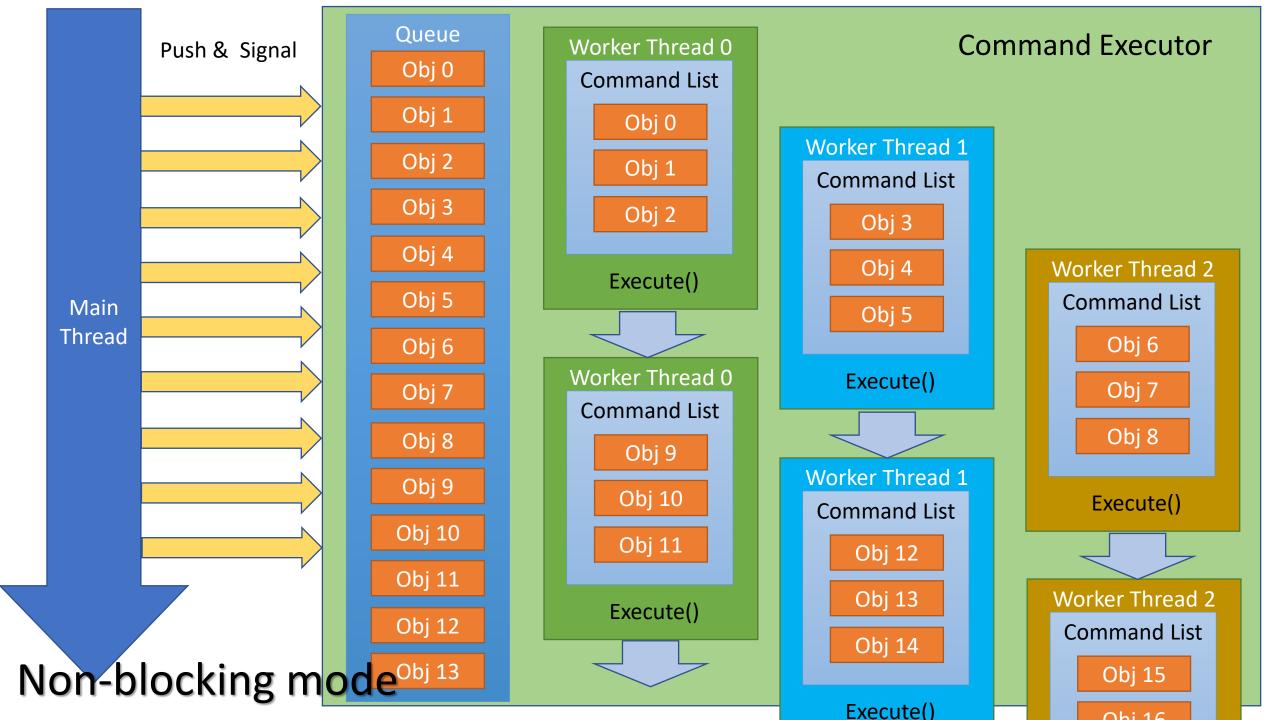
용도별패턴

- 처리량 향상
 - 응답시간이 중요하지 않을때(or 처리량 부하 >>> 응답성 저하)
 - 이미지 프로세싱 동기화가 거의 필요하지 않음.
 - Lightmap baking등 각종 baking 동기화가 거의 필요하지 않음. Or 아주 약한 동기화.
 - Voxelization 동기화가 거의 필요하지 않음.
 - 서버에서의 3D오브젝트 충돌처리/picking 동기화가 다소 필요함.
 - 응답시간이 중요할때 (or처리량 부하 >= 응답성 저하)
 - D3D12 멀티 스레드 렌더링 동기화가 다소 필요함.
 - IOCP 워커 스레드 동기화가 필요함.
- Blocking 모드 API를 non-block모드로 사용하고 싶을때
 - 소켓 accept, connect 동기화가 약간 필요함
 - DB Query 동기화가 약간 필요함.

D3D12 멀티 스레드 렌더링

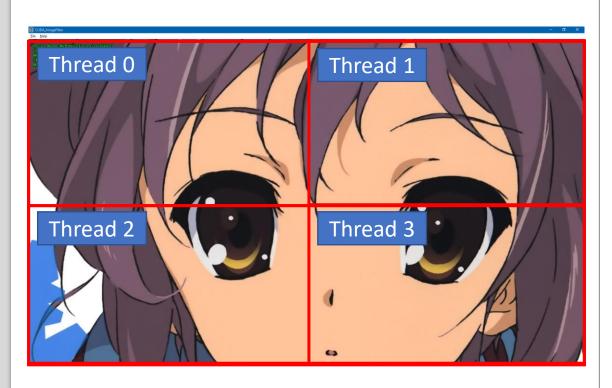
- 멀티 스레드를 사용하지 않으면 일단 GPU가 무조건 논다.
- 그렇지만 처리량을 극대화 시키려다 보면 응답성이 떨어진다.





이미지 프로세싱

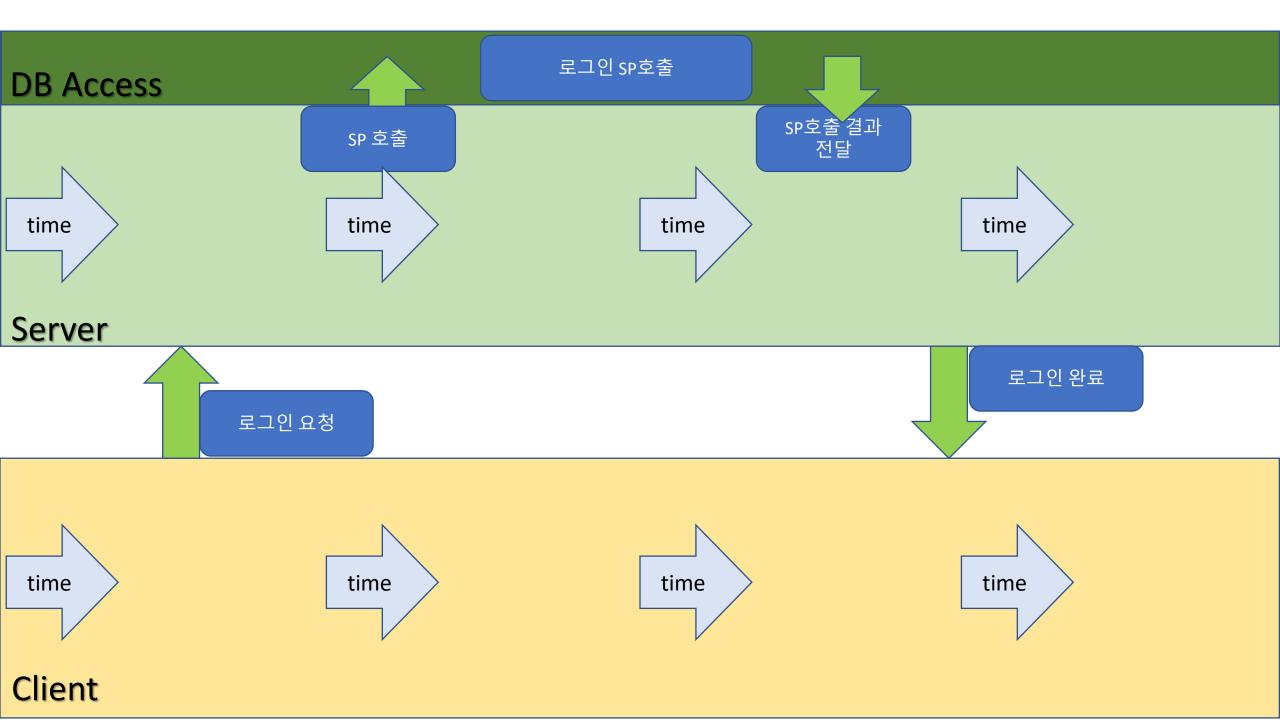
- 멀티 스레드로 가장 확실한 성능 향상을 볼 수 있는 예
- 이미지를 스레드 개수만큼 분할해서 처리한다.
- 정적 이미지, 동영상 모두 같은 방식으로 적용 가능
- 하지만 CUDA가 출동하면 어떨까?





비동기 DB쿼리

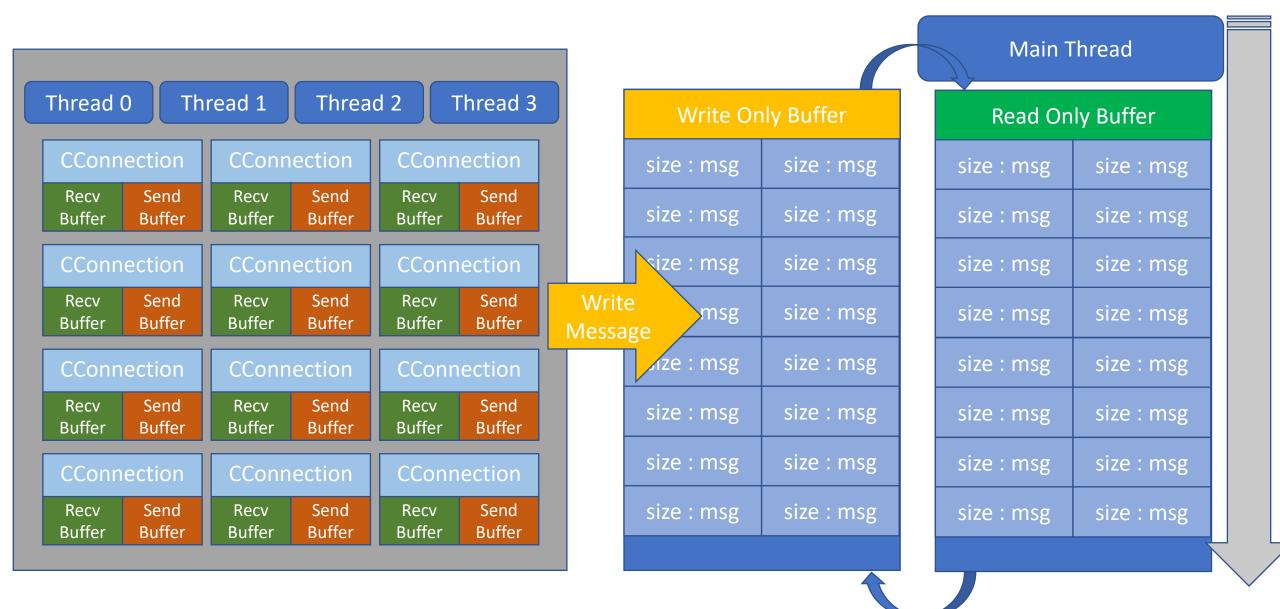
- DB에 쿼리 후 응답 수신 후 메모리에 업데이트
 - 로그인, 아이템 획득 등
- 서버의 메모리 업데이트 후 DB에 쿼리(저장)
 - 총알 소모, 캐릭터 데이터 세이브, 기타 등등
- 어느쪽이든 비동기 처리



패킷 수신 -> 메시지 처리

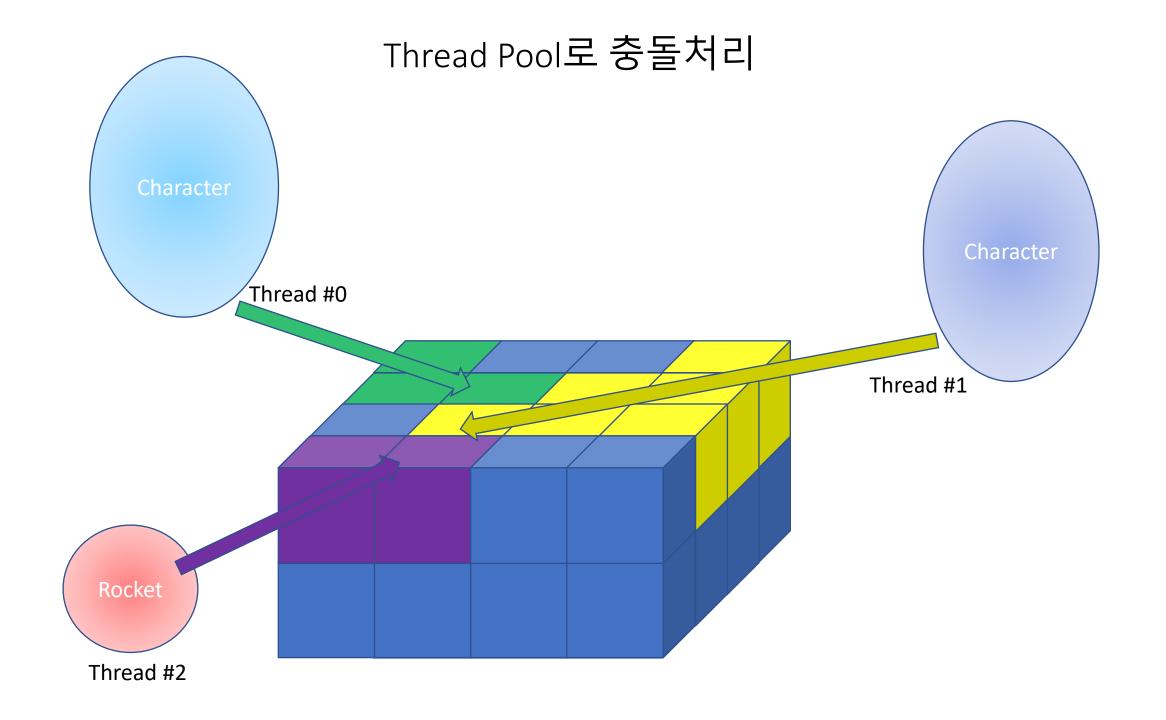
- 최소 패킷 구조 = size(4 bytes) + body(N bytes)
- I/O 워커 스레드의 메시지 수집과 메인 스레드의 경쟁 상태를 줄인다.
- Double buffering
 - 하나 이상의 패킷이 수집되면 Network측 Worker thread가 쓰기 버퍼에 수집된 패킷을 써넣는다.
 - Main Thread는 패킷 수신이 통보되면 쓰기 버퍼와 읽기 버퍼의 포인터를 swap한다(가벼운 lock사용).
 - Main Thread는 읽기 버퍼의 쌓인 패킷을 처리한다.
 - 처리가 완료되면 쓰기 버퍼와 읽기 버퍼의 포인터를 swap한다.

패킷 수신 -> 메시지 처리



서버에서의 이동(충돌)처리

- 클라이언트와 마찬가지로 30fps or 60fps로 처리.
- 키 입력에 대한 순서가 보장되어야 한다.
- 플레이어간 입력 순서가 보장되어야 한다.
- 따라서 완전 동기식으로 처리하되 개별 이동처리 시간을 줄이는 수밖에 없다.
- 처리해야할 캐릭터 수에 비례해서 성능 하락.
- 처리량이 늘어서 응답성이 떨어지는 경우이므로 Thread Pool을 이용하여 병렬처리 한다.



스레드간 동기화

동기화 객체

- Event
- SRWLOCK
- Critical Section
- Custom Spin Lock

Event

- WaitFor...() SetEvent()
- Kernel object
- 대기 시간 길어짐-응답성 떨어짐
- CPU자원의 낭비 적음

Critical Section

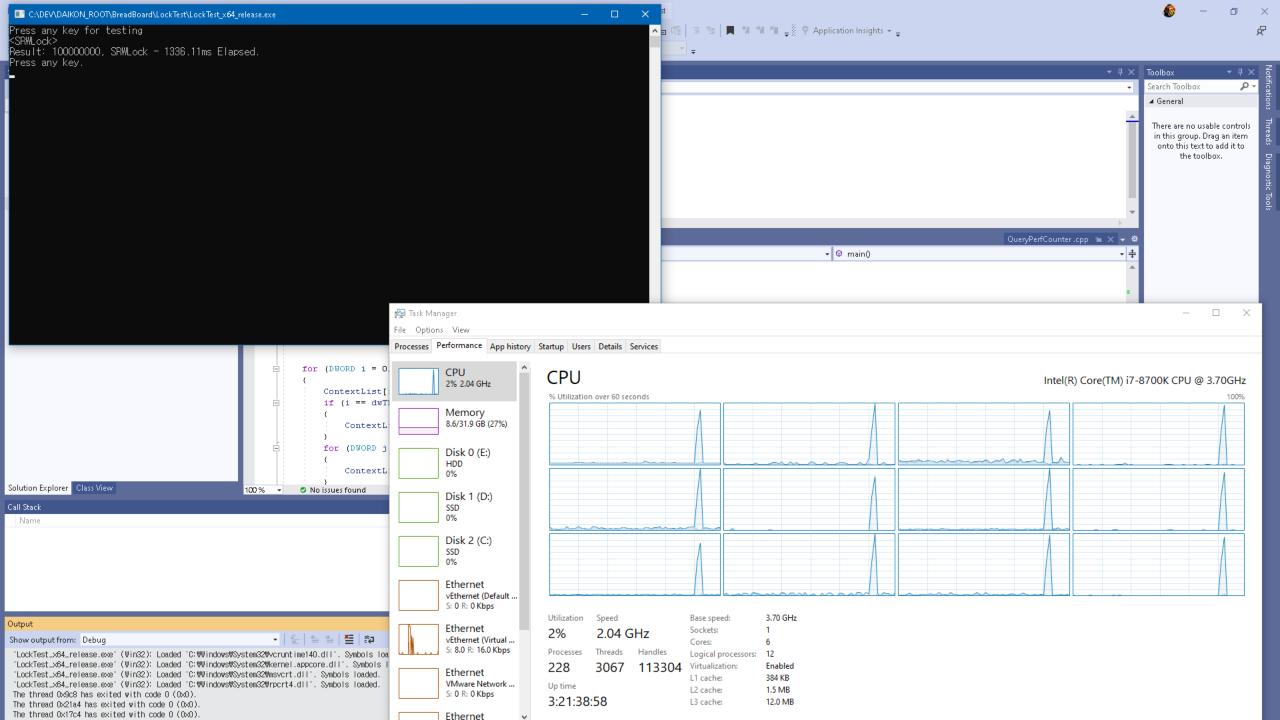
- 동기화가 필요한 코드블럭을 감싸는 형태
- Kernel Object보단 응답성이 빠르다고 알려져있음.
- 재진입 가능
- 진입 시도 실패시 wait상태로 진입
- Spin lock을 일부 도입한 형태로 사용 가능

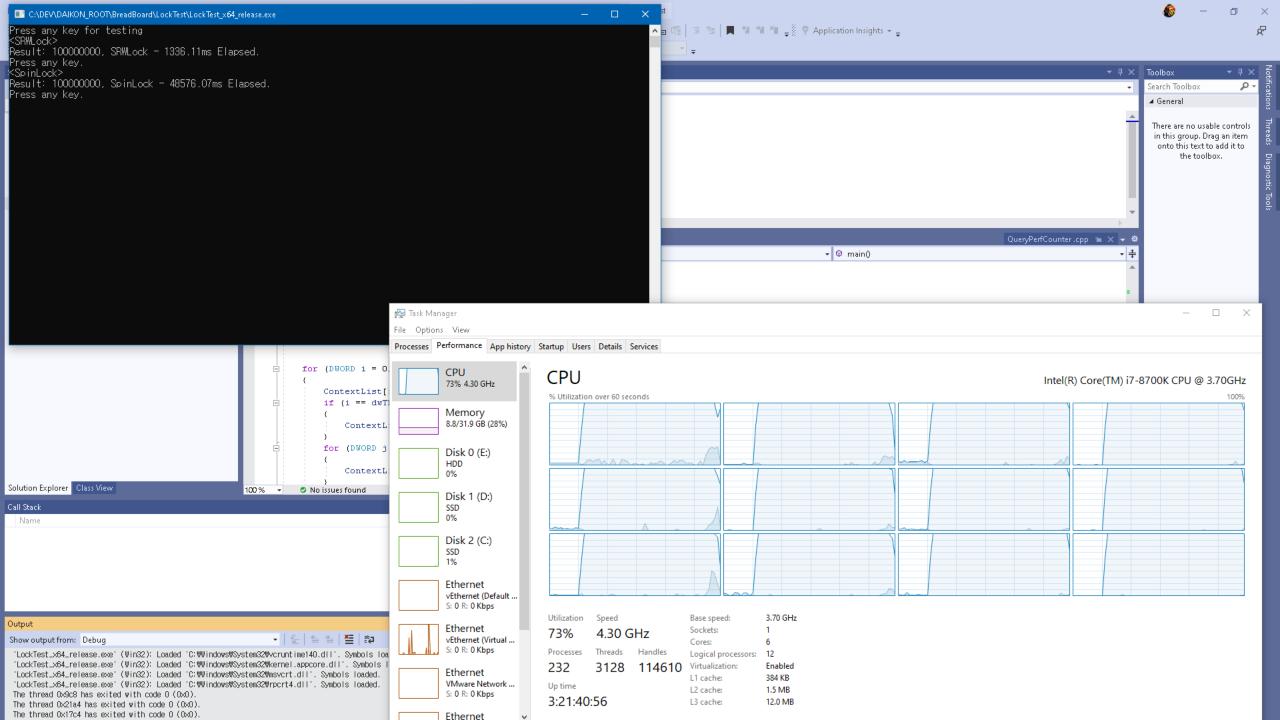
SRWLOCK

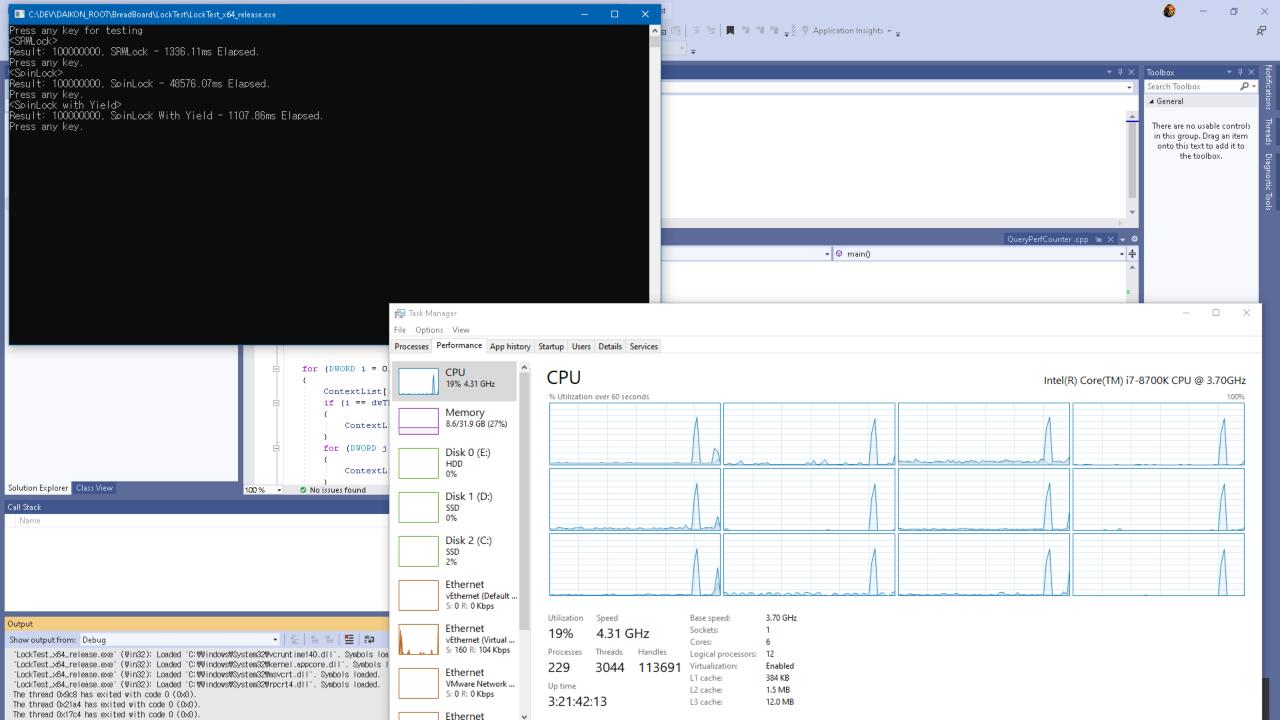
- 접근하는 스레드가 모두 Read모드인 경우 비용0로 억세스
- 접근하는 스레드가 read/ write 이거나 write /write인 경우 비용 발생
- Spin lock을 돌다가 일정 회수 이상 반복하면 wait상태로 진입
- 재진입 불가
- 재진입이 필요치 않다면 Critical Section을 완전히 대체 가능.

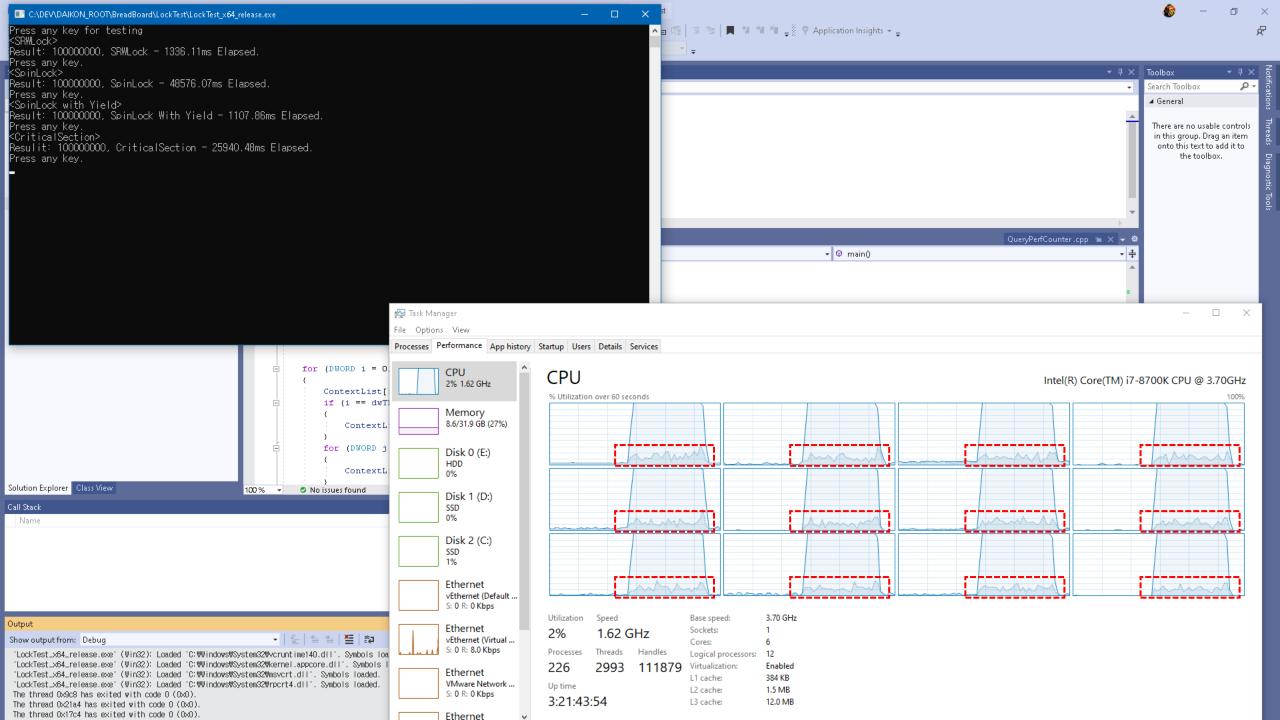
Spin lock

- Lock 접두어를 사용해서 간단히 만들 수 있음.
- Wait를 절대 하지 않을 경우 사용.
- 스레드간 경쟁이 긴 시간 발생할 경우 전체적인 성능을 떨어뜨릴 가능성이 있음.
- 짧은 시간동안의 lock이 필요할 경우 가장 빠르다.









동기화 tip

Lock으로 낭비되는 시간 줄이기

- Lock을 안걸면 가장 좋다.
- 써넣기할때 버퍼 전체의 lock을 거는 대신 써넣을 주소를 얻을때만 lock을 건다.
- 고정 사이즈 memory pool을 사용한다.
- Linked list와 spin lock을 사용한다.
- CRT Heap은 thread safe하지만 여러 스레드가 동시 진입할 경우성능이 떨어질 수 있다. 여러 스레드가 CRT heap을 동시 접근할일이 아예 없도록 할것. 미리 할당/ 스레드별로 다른 heap사용