

SAMSUNG SDS

Realize your vision

Techtonic 2019

Partner



Foresee



Disrupt

2019.11.14 • SAMSUNG SDS Tower B1F
{ Magellan Hall / Pascal Hall }

Track 1 | AI/Analytics

GPU 클러스터링을 통한 공유 및 분산 처리 플랫폼
R&D Cloud

신욱수 Lab장 (클라우드선행연구Lab) / 삼성SDS

AGENDA

1. 왜 GPU가 중요한가?
2. GPU 사용시 이슈
3. AI용 R&D Cloud 소개
4. Use Case
5. Summary

1

왜 GPU가 중요한가?

1. 왜 GPU가 중요한가?
2. GPU 사용시 이슈
3. AI용 R&D Cloud 소개
4. Use Case
5. Summary

왜 GPU가 중요한가

그 동안 CPU를 많이 썼을텐데? 왜 지금은 GPU를 이야기하는 것일까?

CPU ? GPU

왜 GPU가 중요한가

Core(≡ ALU) 는 기본적으로 Thread를 1개 처리 (hyperthread 기술을 통해 2개)

CPU <= 18 cores

- ▶ CPU = Control Unit(CU) + Arithmetic-Logic Unit(ALU) + Register + ...
- ▶ 명령어가 입력된 순서대로 처리하는 순차방식에 특화

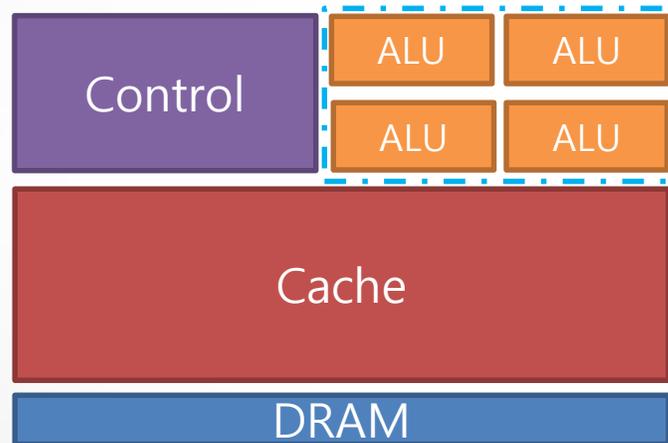
GPU <= 5760 cores

- ▶ 숫자 데이터를 빠르고 반복적으로 처리하는 병렬 방식에 특화
- ▶ 데이터 처리를 위한 메모리 용량 중요

왜 GPU가 중요한가

GPU는 기본적으로 단순 계산을 잘 하도록 계산에 특화되어 설계됨

CPU \leq 18 cores



<Central Processing Unit>

- Low Compute Density
- Complex Logic
- Large Cache

GPU \leq 5760 cores



<Graphics Processing Unit>

- High Compute Density
- Built for Parallel Operations
- High Throughput

왜 GPU가 중요한가

GPU가 연산을 할 수 있는 ALU 개수가 많음

-> CPU로 오래 걸리는 계산 작업을 GPU로 빠르게 병렬 수행 가능

CPU <= 18 cores

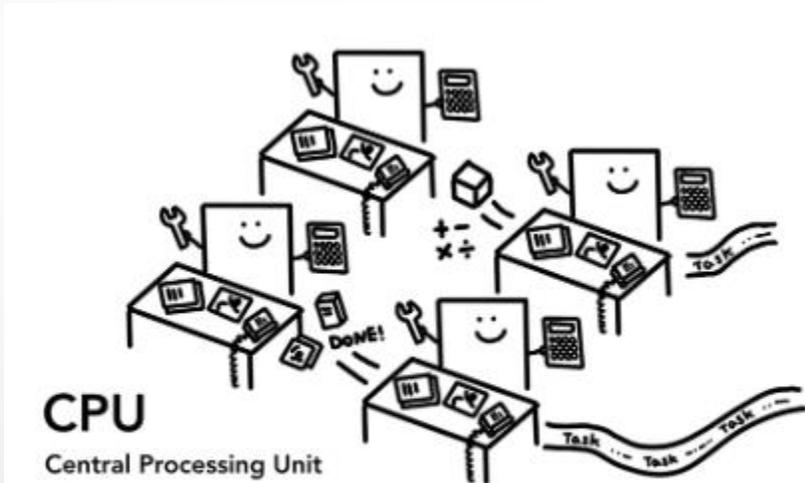


Figure 1: 4 CPU cores as office workers sitting at each desk handling tasks as they come in

GPU <= 5760 cores

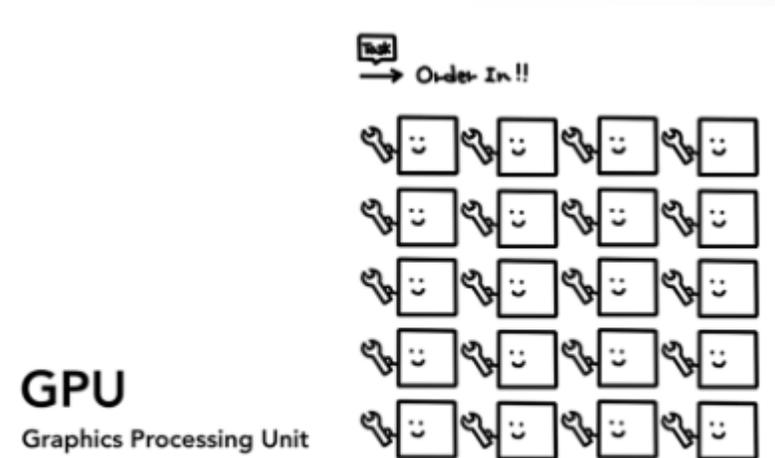


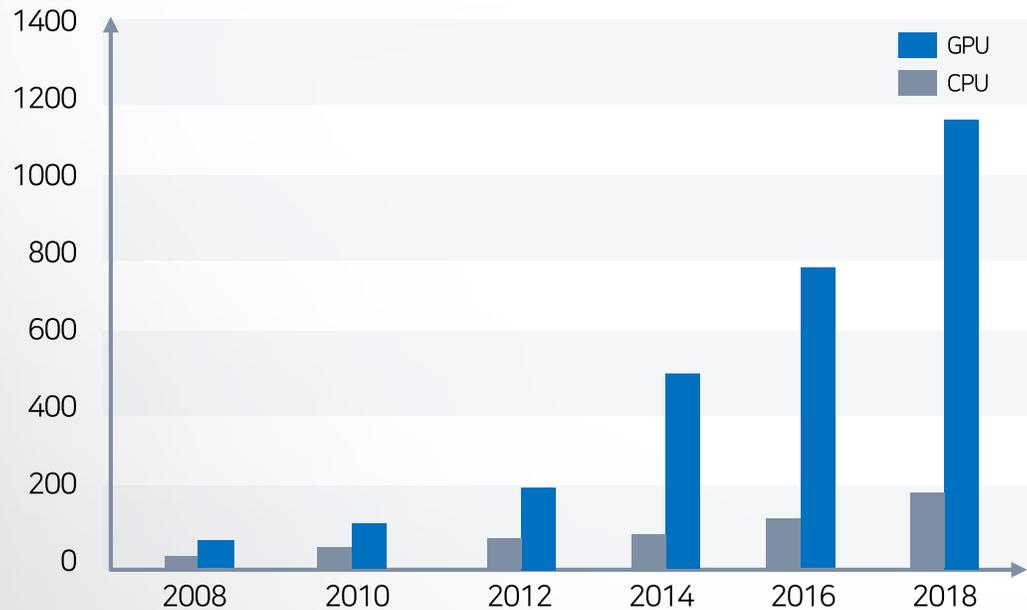
Figure 2: Many GPU cores with wrench suggesting they handle a limited task

왜 GPU가 중요한가

메모리와 통신 속도 & 숫자계산이 월등히 빠름

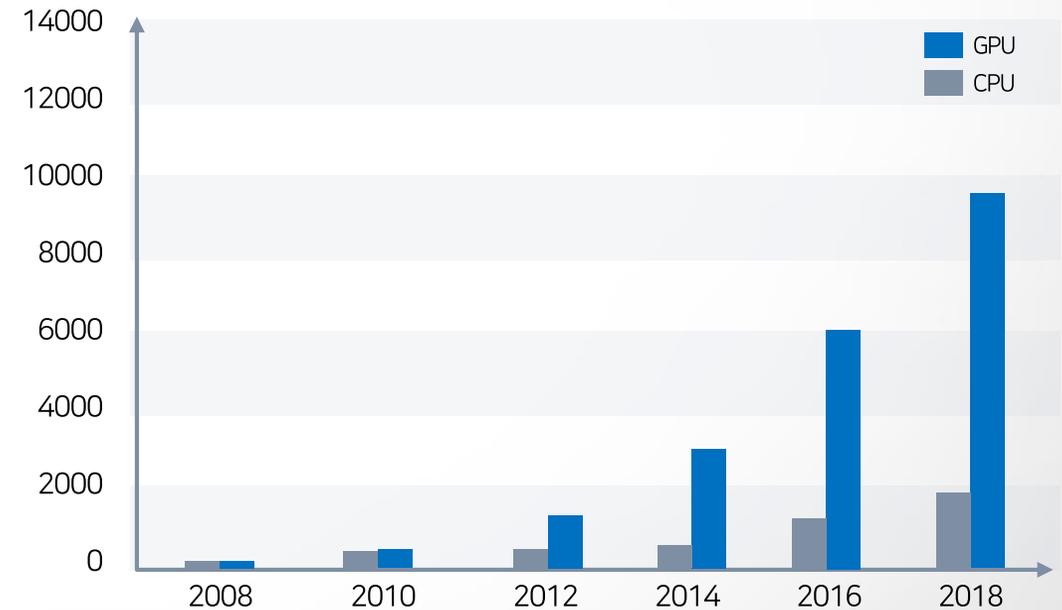
메모리 통신 속도 비교

Peak Memory Bandwidth (GB/s) 2008-2018



숫자 계산 속도 비교

Peak Double Precision (GFLOPs) 2008-2018



※ ※ Source : <https://www.nextplatform.com/2019/07/10/a-decade-of-accelerated-computing-augurs-well-for-gpus/>

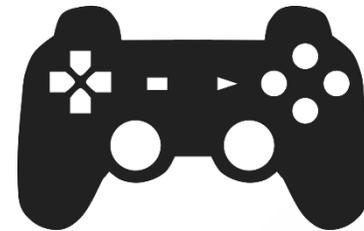
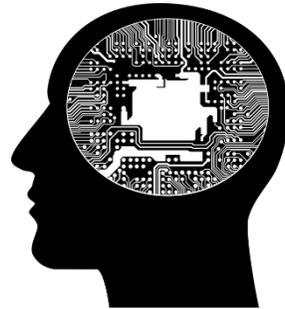
왜 GPU가 중요한가

CPU가 수행하던 작업을 GPU가 할 수 있을까?

왜 GPU가 중요한가

GPGPU(General Purpose computing in GPU)

CPU가 수행하던 작업을 GPU가 대신 수행하게 하는 기술
복잡한 연산을 처리해야 하는 작업에 적합



왜 GPU가 중요한가

Mythbusters Demo GPU versus CPU by NVIDIA (2009)



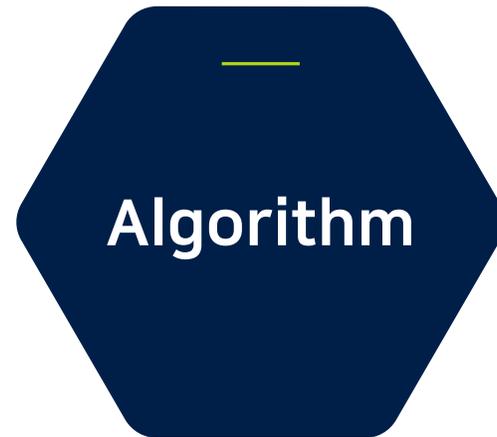
2

GPU 사용시 이슈

1. 왜 GPU가 중요한가?
2. GPU 사용시 이슈
3. AI용 R&D Cloud 소개
4. Use Case
5. Summary

GPU 사용시 이슈 -1. 복잡해진 알고리즘

Large Scale Visual Recognition Challenge



고양이



강아지



GPU 사용시 이슈 -1. 복잡해진 알고리즘

Large Scale Visual Recognition Challenge



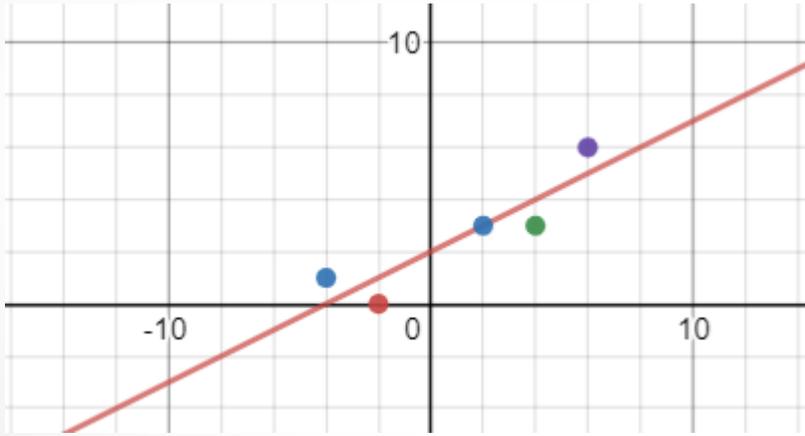
▶ 2012
Deep Neural Network

Figure 1: The performance and size of the DNNs in ILSVRC'10-15 [17, 22, 41, 43].

GPU 사용시 이슈 -1. 복잡해진 알고리즘

Deep Neural Network 이전에는...

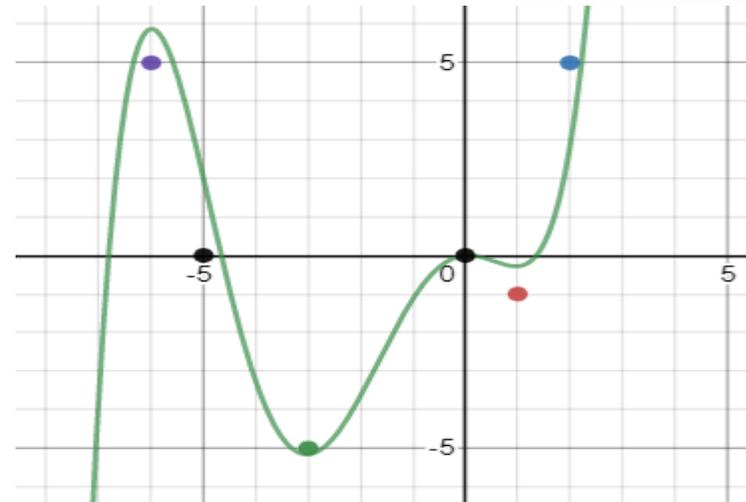
1차식 예측



$$y = \frac{1}{2}x + 1$$

파라미터(미지수)
=> 2개

5차식 예측

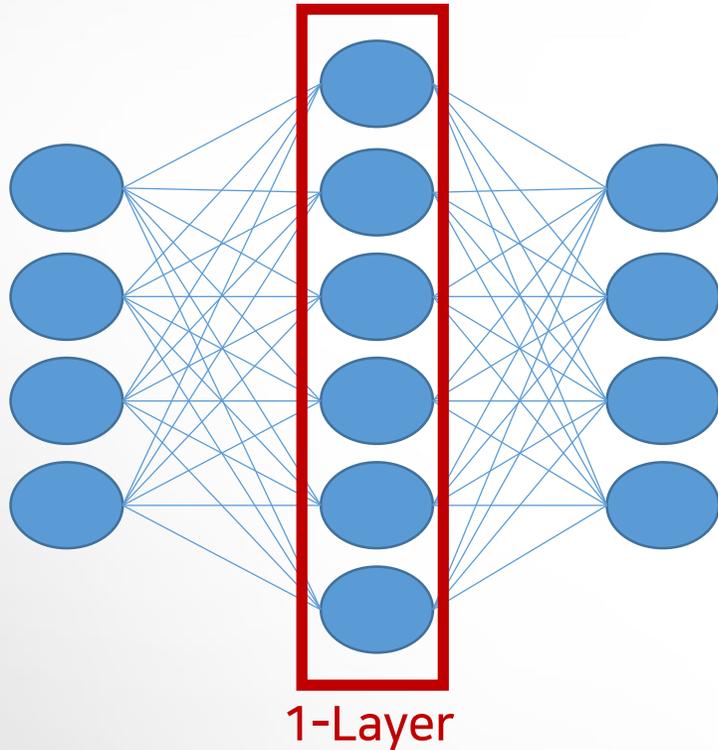


파라미터 개수는?

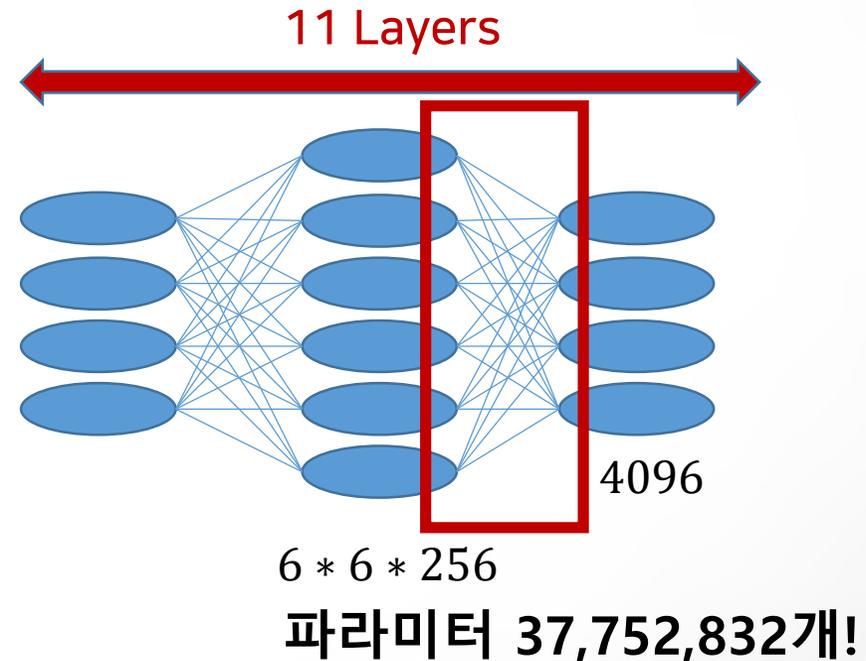
GPU 사용시 이슈 -1. 복잡해진 알고리즘

Deep Neural Network(DNN)가 무엇이기에?

간단한 DNN



AlexNet (8 Layers)



GPU 사용시 이슈 -1. 복잡해진 알고리즘

Large Scale Visual Recognition Challenge

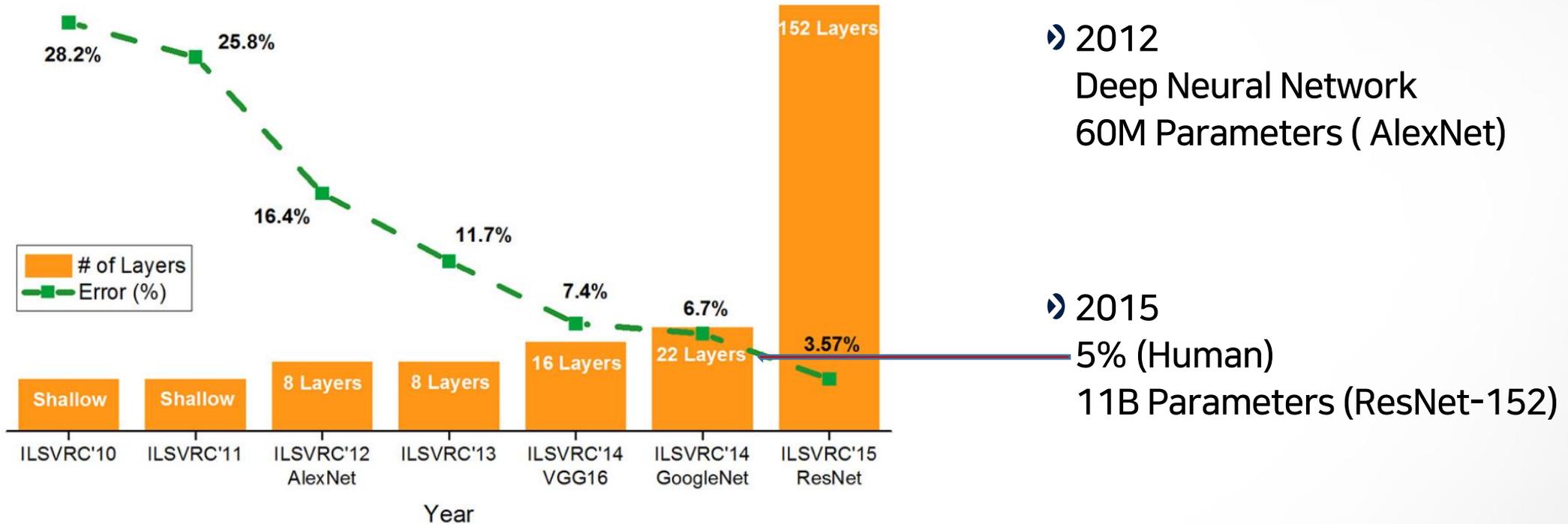
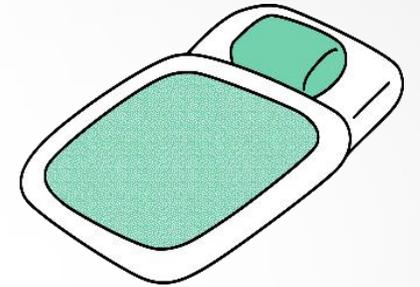


Figure 1: The performance and size of the DNNs in ILSVRC'10-15 [17, 22, 41, 43].

GPU 사용시 이슈 -1. 복잡해진 알고리즘



오래 걸린다



시간 단축을 위해 GPU
서버를 더 사자!



비싸다.....



GPU 사용시 이슈 -1. 복잡해진 알고리즘

오래 걸리고 비싸고...
해결 할 수 없을까요?

GPU 사용시 이슈 -2. 사용률

집에 있는 자전거를 생각해보자



- ▶ 하루 대부분 멈춰있는 자전거
- ▶ 부피도 커서 보관도 어려움
- ▶ 다같이 타려면 가족수대로 구매 필요

사긴 아깝고
보관도 귀찮은데?

GPU 사용시 이슈 -2. 사용률

다른 사람이 안 쓸 때 내가 쓰면 되지!!

GPU 사용시 이슈 -2. 사용률

서울시 공유 자전거 따릉이



▶ 시간당 천원으로 저렴하게 이용

▶ 보관 문제 해결

▶ 여러대가 필요할 때도 OK

▶ GPU에도 이런 시스템을 적용하면 어떨까?

3

AI용 R&D Cloud 소개

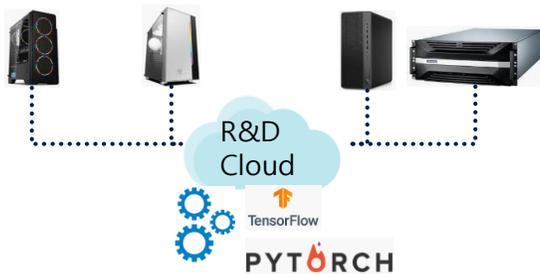
1. 왜 GPU가 중요한가?
2. GPU 사용시 이슈
3. AI용 R&D Cloud 소개
4. Use Case
5. Summary

R&D용 AI 플랫폼에서 필요한 것은?

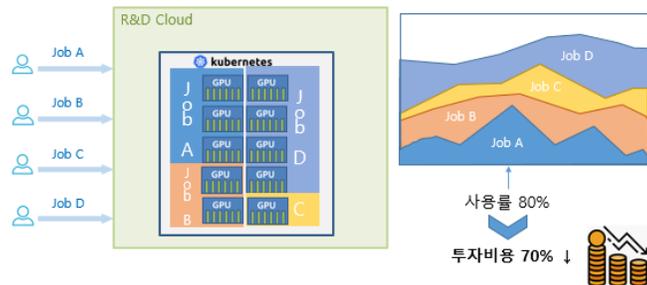
GPU → 효율성(사용할 때만 점유) & 생산성(여러 장비 동시 사용)

GPU 리소스를 효율적으로 사용

- ▶ 기존 및 저가 장비로 대규모 머신러닝 환경 구축

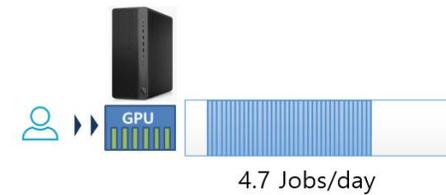


- ▶ 조직 내 전체 리소스 사용률을 높여 투자비용을 절감

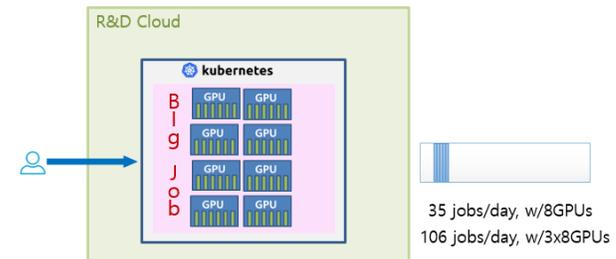


분산처리 기능을 통해 생산성 향상

- ▶ AS-IS : 단일 머신으로 하루 4건 처리



- ▶ TO-BE : GPU 클러스터를 이용해 필요한 만큼 가속



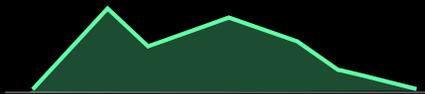
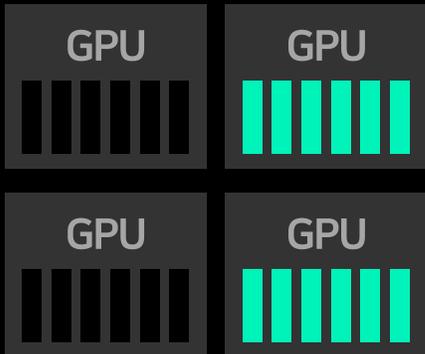
멀티노드 분산학습 플랫폼을 탑재한 GPU Cluster에서 딥러닝 학습 시 실시간 GPU 점유 및 완료 시 자동반환



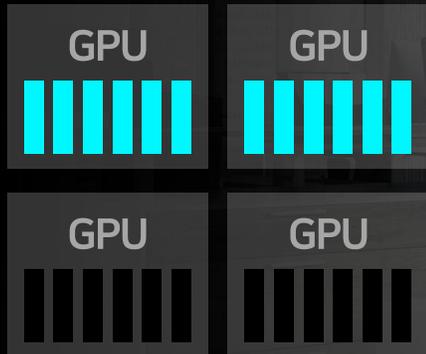
GPU 여러 용도의 작업으로 인해 시스템을 이용한 성능 향상과 자원 분배



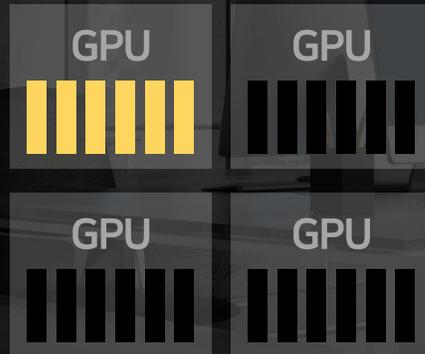
Job A



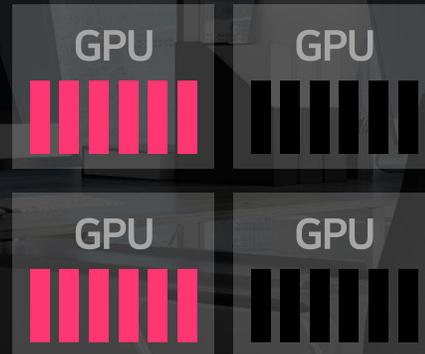
Job B



Job C



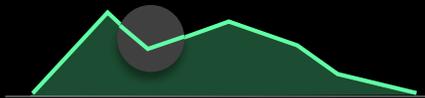
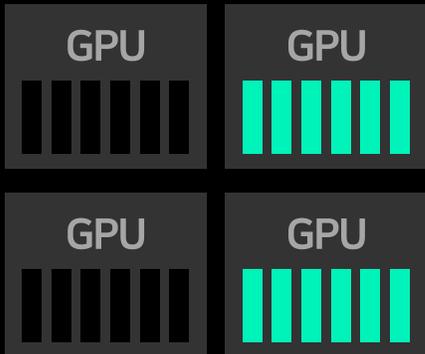
Job D



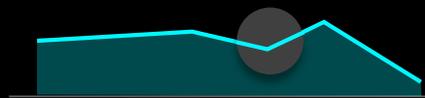
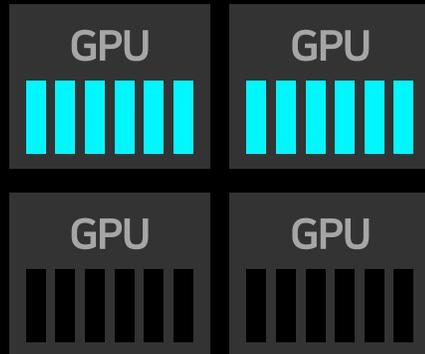
GPU 이외의 작업으로 인해 사용률이 낮은 상황이 자주 발생함



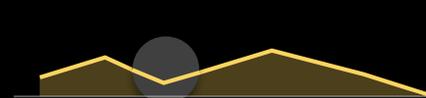
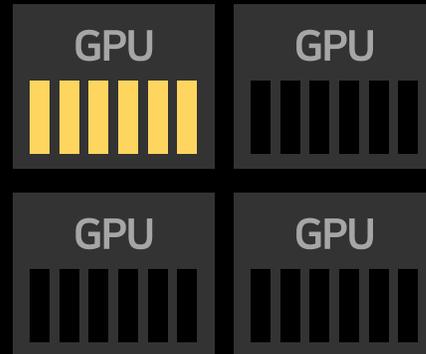
Job A



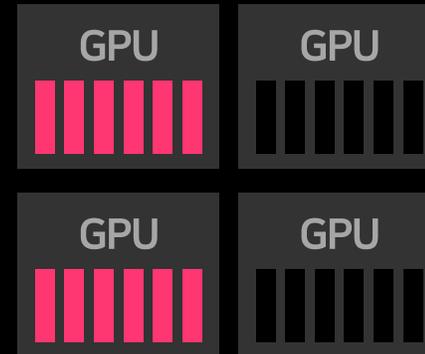
Job B



Job C



Job D



Click to continue →

R&D Cloud를 이용하여 GPU 자원을 공유할 수 있음



Job A



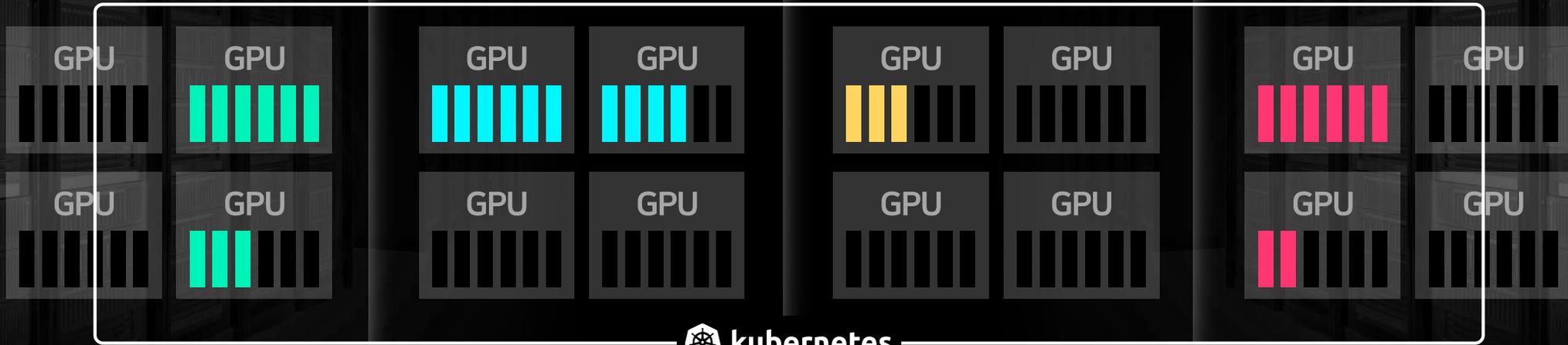
Job B



Job C



Job D



 kubernetes

사용자 요청 통합 및 양한 유/변 배정능을 통합하여 업무용 클러스터를 조정하여 투쟁 사용량 감소 사용



Job A



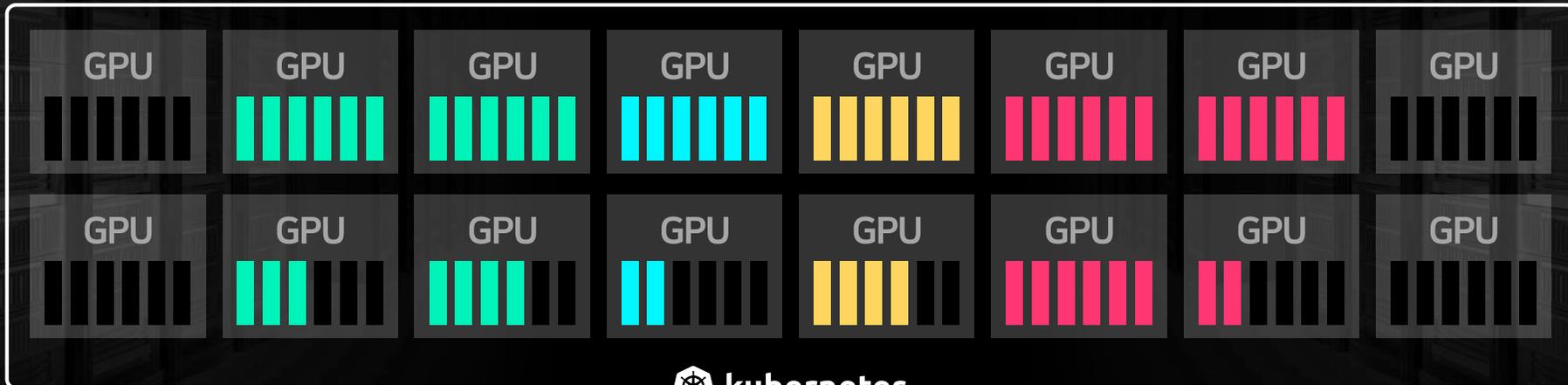
Job B



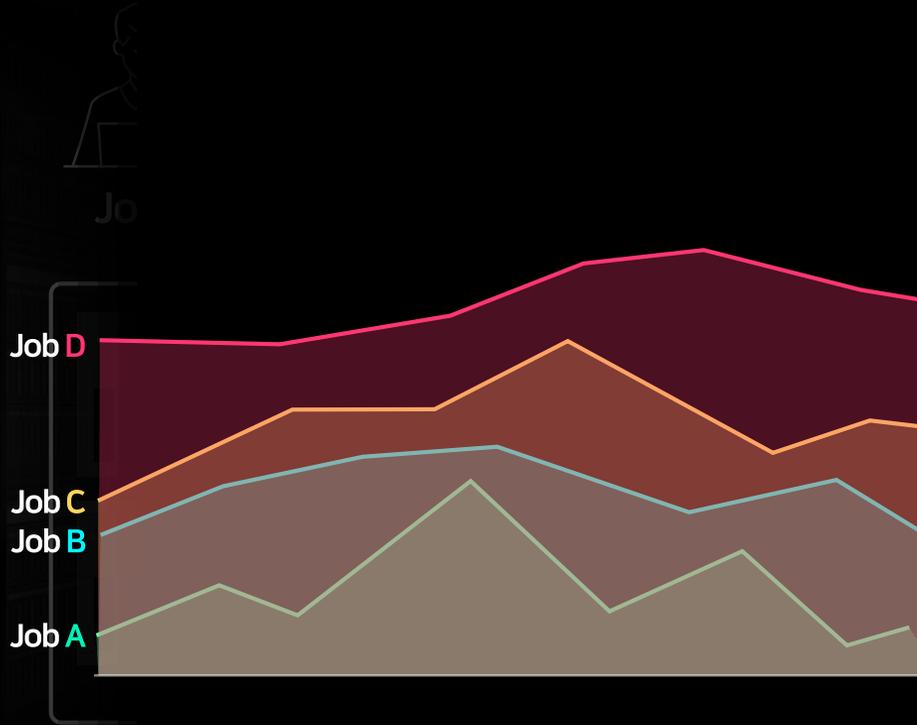
Job C



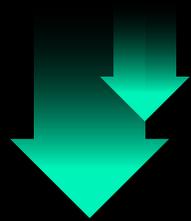
Job D



이러한 통합 및 공유/분배 기능을 통하여 사용률이 증가하여 투자비용은 감소



투자비용 **80%**

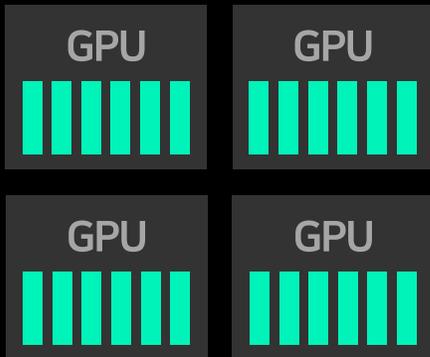


멀티 노드간 병렬 분산처리 기능을 활용하여 훨씬 빠르게 작업 수행 가능



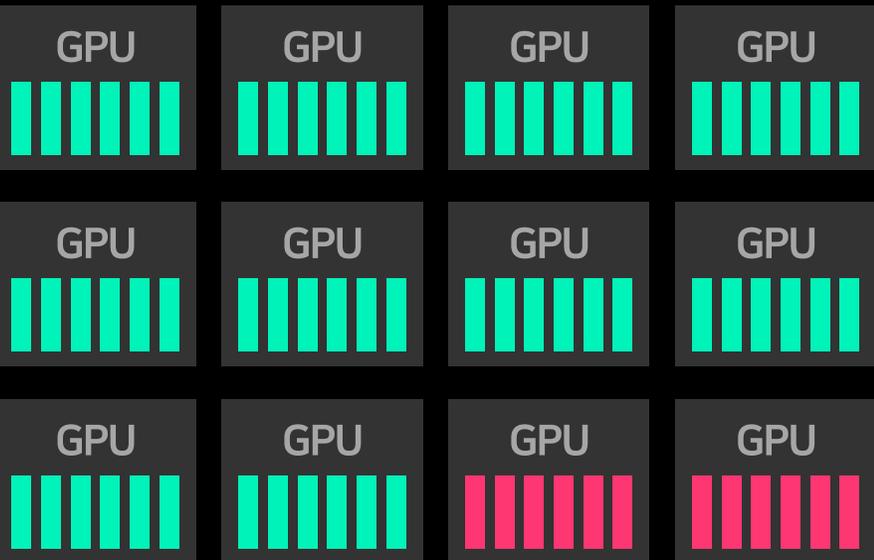
4.7 Jobs/day

Job E



35 Jobs/day, w/8GPUs

106 Jobs/day, w/3x8GPUs



Distributed Model Training



DevOps팀 분석전문가

Cluster Provisioning

- Specify CPU/GPU/Memory
- Assign specs
- Use Kubernetes



Brightics DL

(Deep Learning 플랫폼)

```
Estimator (...)  
Experiment.run(..., 4)
```

4 is the number of nodes

모델아키텍처

- Build the computational graph



TensorFlow

분산학습 준비

- Set the Parameters Server and Master
- Set the logs to view in TensorBoard



 **kubernetes**


TensorFlow

학습

- Start the model training




TensorFlow

4

Use Case

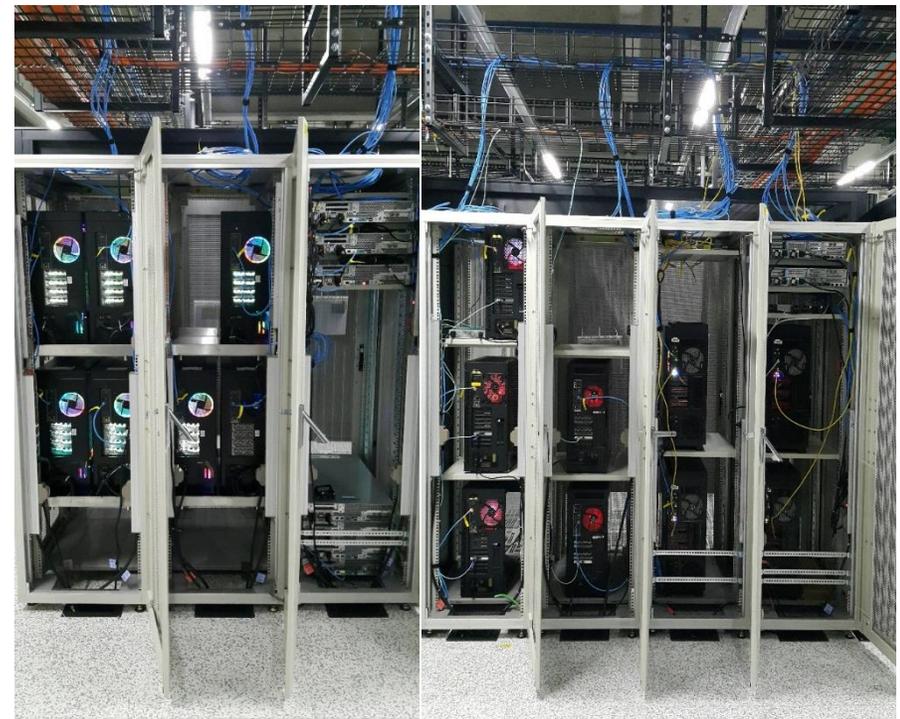
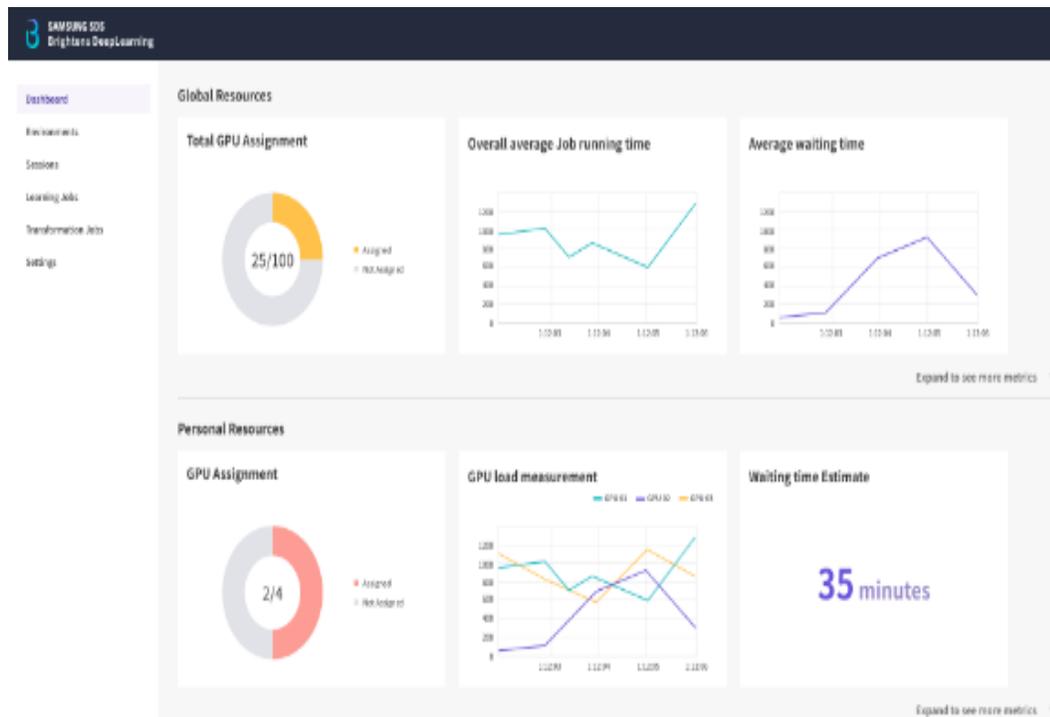
1. 왜 GPU가 중요한가?
2. GPU 사용시 이슈
3. AI용 R&D Cloud 소개
4. Use Case
5. Summary

Use Case-1. SDS 연구소의 R&D Cloud 적용 사례

사무환경 내 혼재 되어 있던 GPU 서버를 Cluster로 구성하여 통합 관리

GPU 서버 통합관리
조직별 / 서버별 GPU 사용량 관리

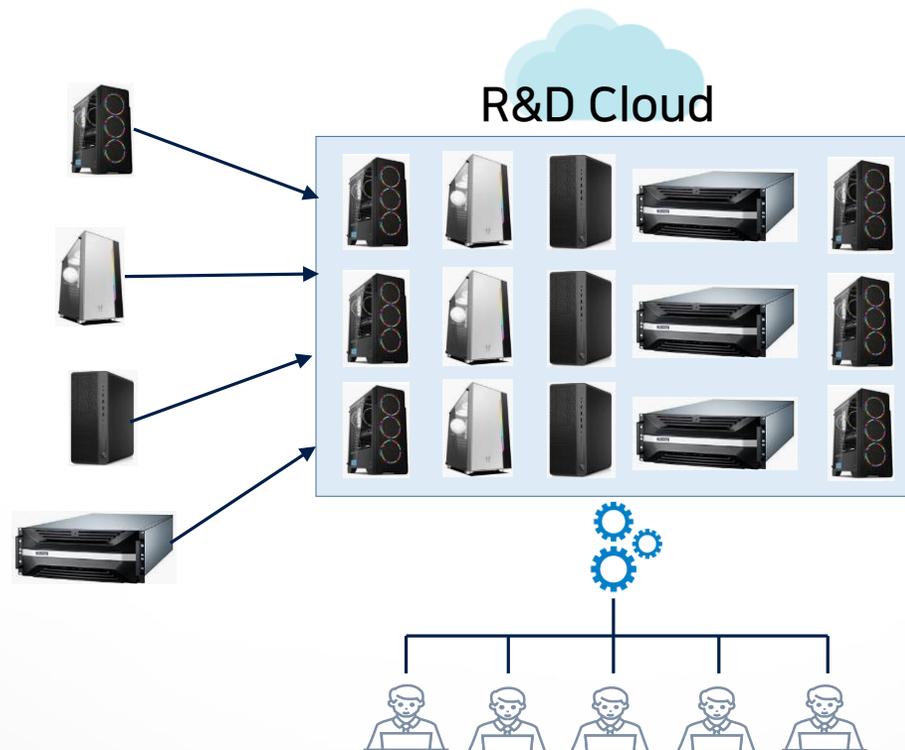
Colocation
데이터 센터 내 GPU Cluster 구성



Use Case-2. 신규 R&D Cloud 구축 사례

현장에 사용중인 장비가 있는 경우

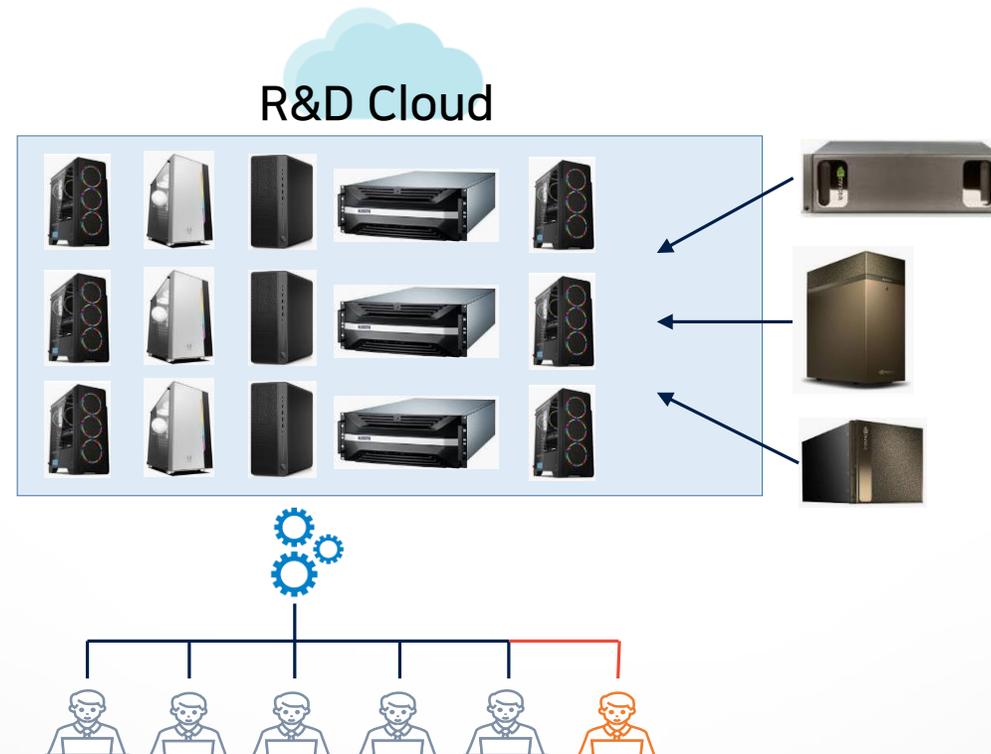
“ 데이터센터로 Colocation 후 솔루션을 적용하여 Cluster 구성 ”



Use Case-2. 신규 R&D Cloud 구축 사례

수요가 지속 증가하여 장비를 추가하는 경우

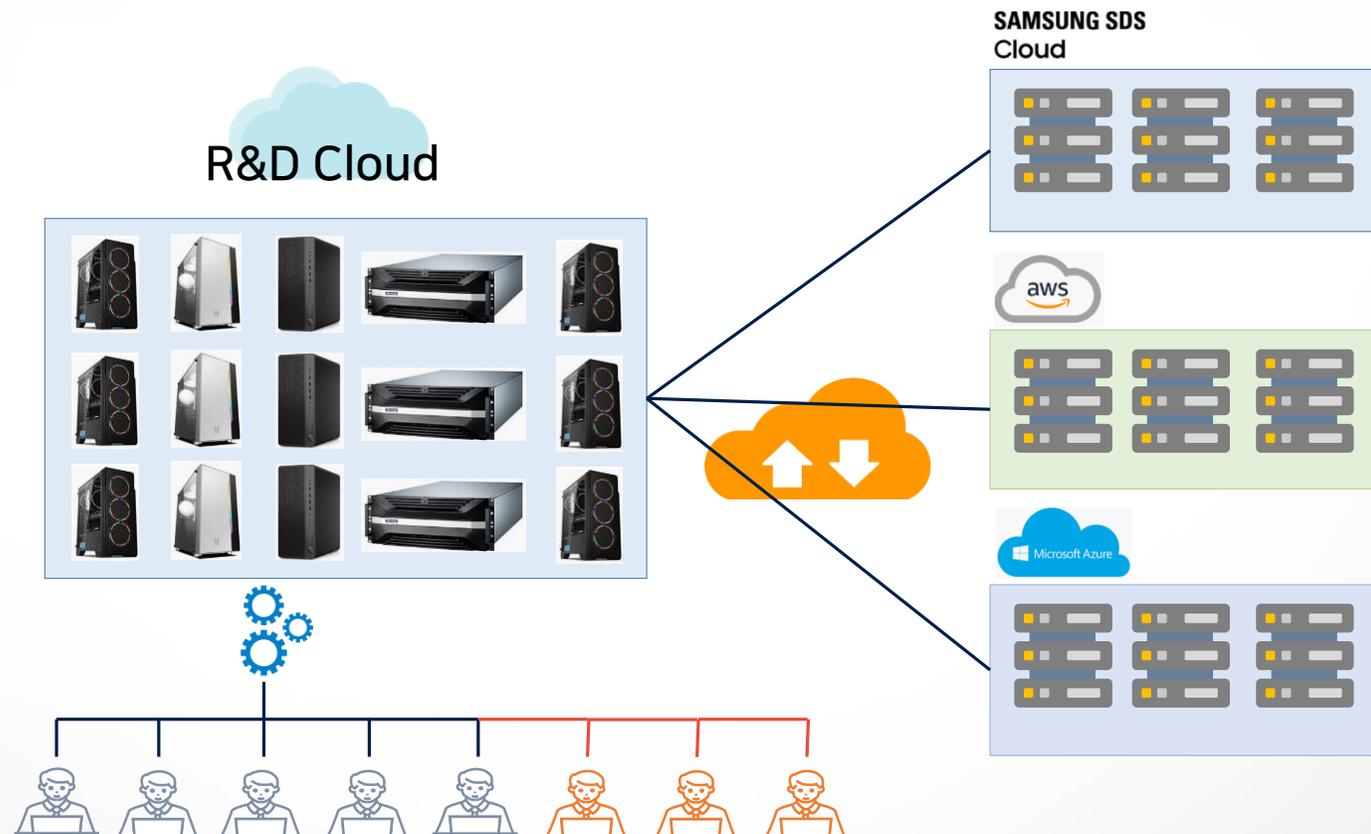
“ 신규 장비를 Colocation Cluster에 추가하여 확장 ”



Use Case-2. 신규 R&D Cloud 구축 사례

사용량이 증가하고 Managed Service를 이용하려는 경우

“ IaaS(Infra as a Service) 를 통한 즉시 확장 ”



5

Summary

1. 왜 GPU가 중요한가?
2. GPU 사용시 이슈
3. AI용 R&D Cloud 소개
4. Use Case
5. Summary

더 나은 R&D용 AI 플랫폼의 요구사항

→ 생산성, 효율성, 사용 및 관리 편의성

→ GPU 클러스터링을 통한 공유 및 분산처리 플랫폼

→ R&D Cloud

문의: rnd.cloud@samsung.com

Containers

Brightics DL (Deep Learning 플랫폼)

AI Services

라벨링자동화

모델추천

하이퍼파라미터
최적화

모니터링

데이터셋관리

실험관리

Core Services

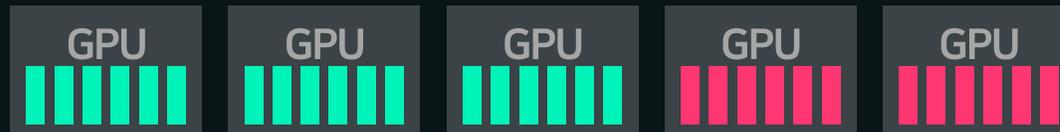
배포/인증

Infrastructure

Registry



kubernetes



분산 파일시스템





Thank You



The graphic features the text 'Q & A' in a clean, sans-serif font. The 'Q' and 'A' are white, while the ampersand is a vibrant lime green. To the right of the text, two orange triangles point towards a large, dark blue circle on the far right. The background is a dark blue gradient with a pattern of lighter blue circles on the left side.

Q & A

Partner Disrupt Foresee