

SAMSUNG SDS

Realize your vision

Techtonic 2019

Partner



Foresee



Disrupt

2019.11.14 • SAMSUNG SDS Tower B1F

{ Magellan Hall }

Track 1 | AI/Analytics

자동 레이블링과 분산 학습을 통해
딥러닝을 쉽고 빠르게!

오경진 프로 (AI 플랫폼 Lab) / 삼성SDS

Partho Ricky Datta (AI Engineering Team) / 삼성SDS

AGENDA

1. Brightics DL
2. Auto Labeling
3. Distributed Model Training
4. Use Cases

1

Brightics DL

Deep Learning Platform? Why?

대용량 영상 데이터를 처리하는 비즈니스 환경에서, 경험한 어려움들

- 데이터 -

“정제 작업, 답안지 만들기
..힘들어요...”

...

“나는 AI 를 하고 싶은데..
가내수공업을 하고 있는 느낌..ㅠㅠ”



- 모델 개발 -

“이 많은 실험을 ..
일일이 다 해봐야 아나요?”

...

“언제 다 하지...ㅠㅠ?”

Enterprise AI 실현을 위한 요건 (R&D에서 비즈니스로)

기업 비즈니스를 위한 대규모 AI로의 전환을 위한 고려 사항

R&D Scale



Operationalization of AI



Enterprise Scale

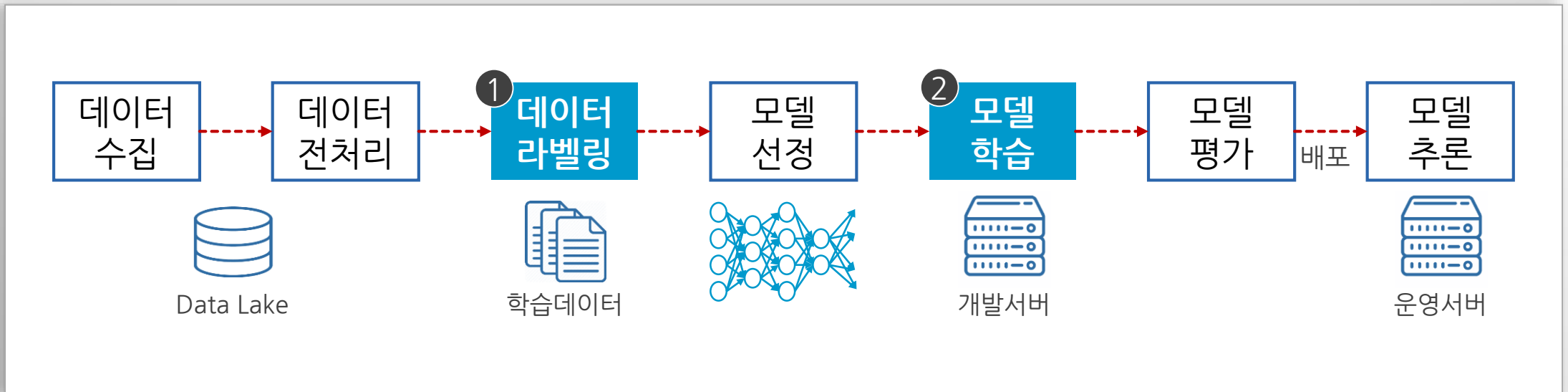


Brightics DL

Brightics DL

DL Model 개발 생산성 가속화 플랫폼

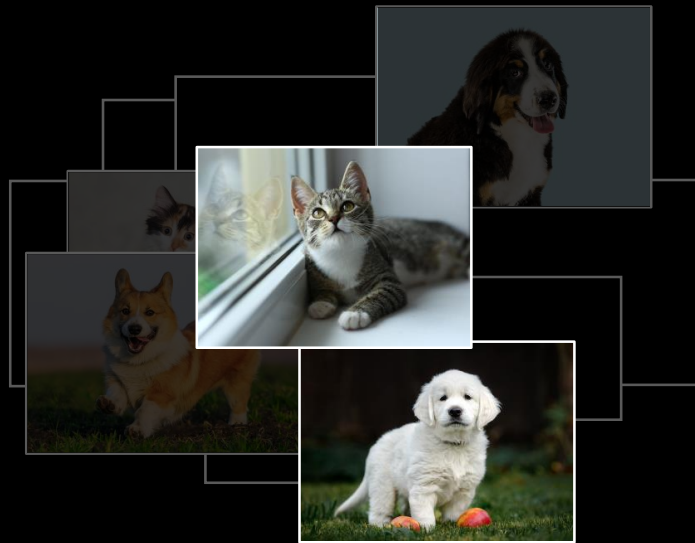
- 1 학습데이터 Labeling 과정을 자동화하는 기능 (Auto Labeling)
- 2 대용량 데이터 수집/전처리/학습과정을 분산처리 기능 (Distributed Model Training)



2

Auto Labeling

with Brightics AI



Brightics AI

학습효과가 가장 큰 이미지
20% 선별

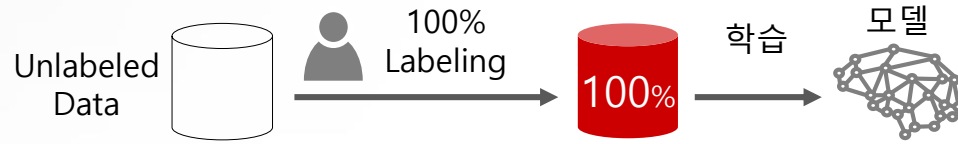
나머지
80% 데이터를 라벨링함

Click to continue →

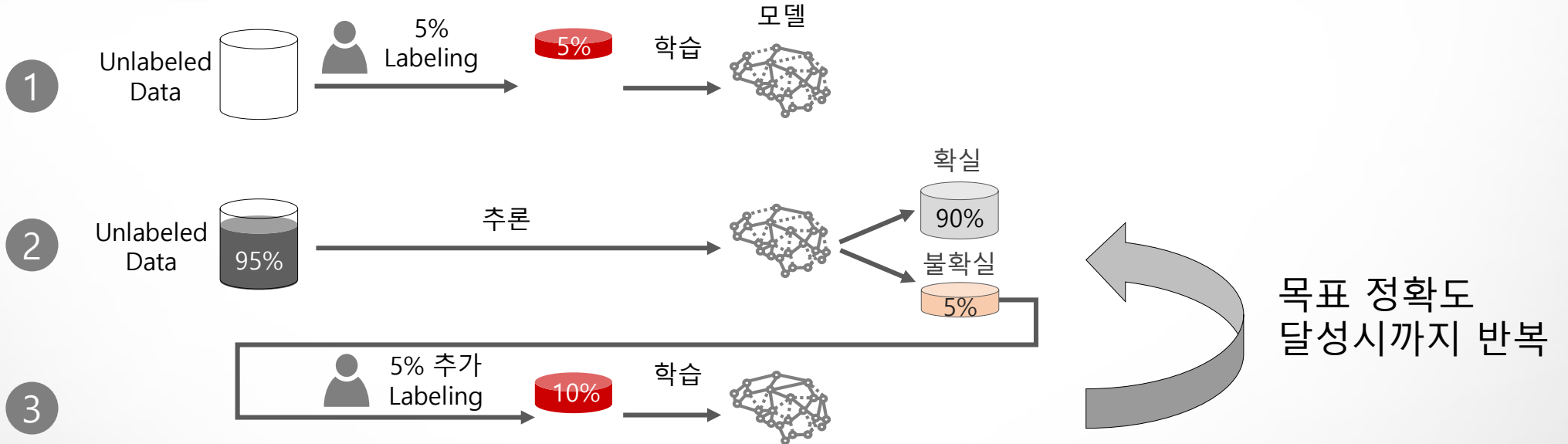
Auto Labeling 원리 이해하기 (1/2)

반쯤 학습한 모델에게 남은 unlabeled data를 추론해보라고 하고, 답이 불확실한 data만 labeling

기존

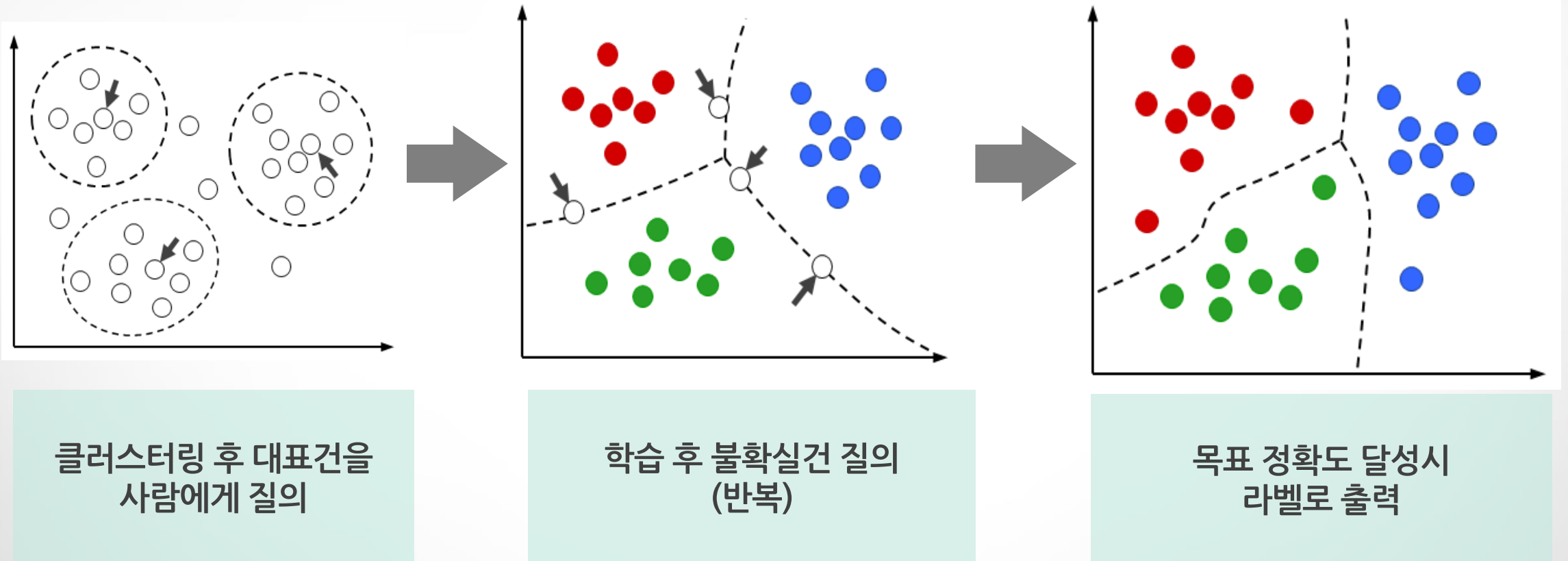


Active Learning



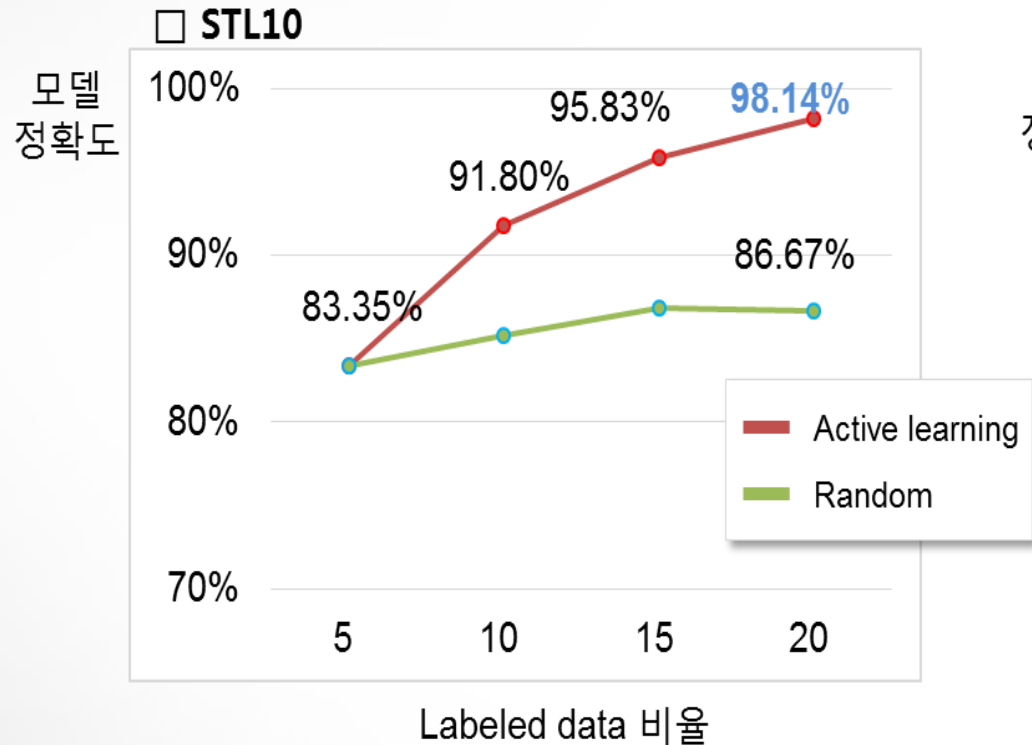
Auto Labeling 원리 이해하기 (2/2)

학습데이터를 분석하여 대표 데이터에 대해서만 사람에게 labeling을 요청한 후,
데이터를 학습해가면서 명확하지 않은 데이터건만 사람에게 labeling 추가 요청

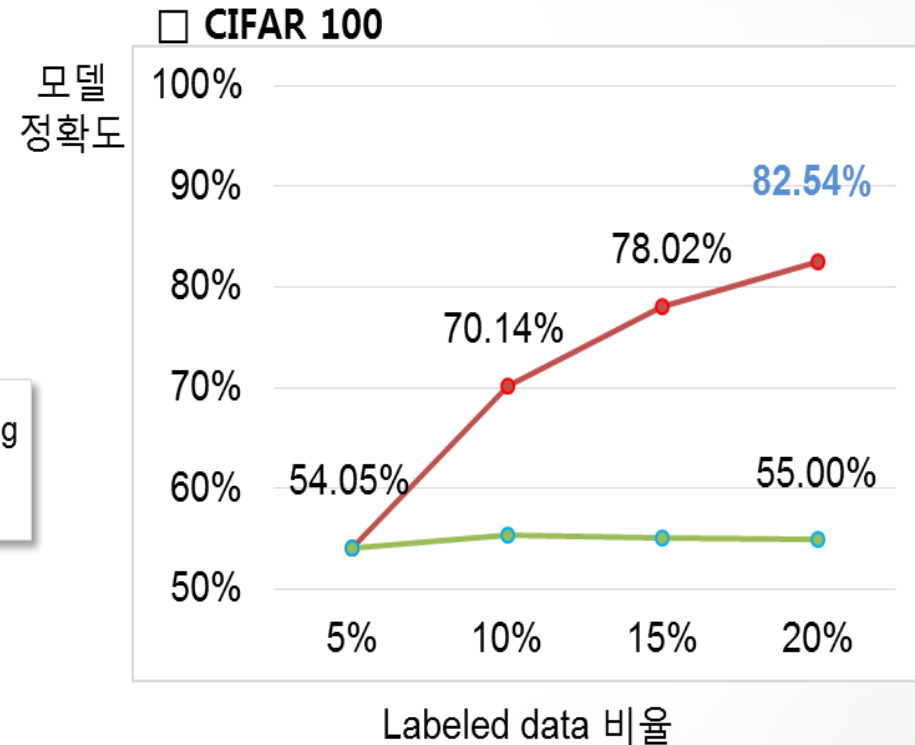


Auto Labeling 효과 이해하기 (Classification)

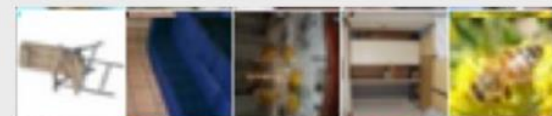
공개 데이터 (STL10) 대상 , 20%의 Label 정보로 Auto Label 정확도 98% 달성



- 10개 종류
- 13,000장 (종류별 1300장)



- 100개 종류
- 60,000장 (종류별 600장)



Auto-Labeling

+ Create

Project TEST

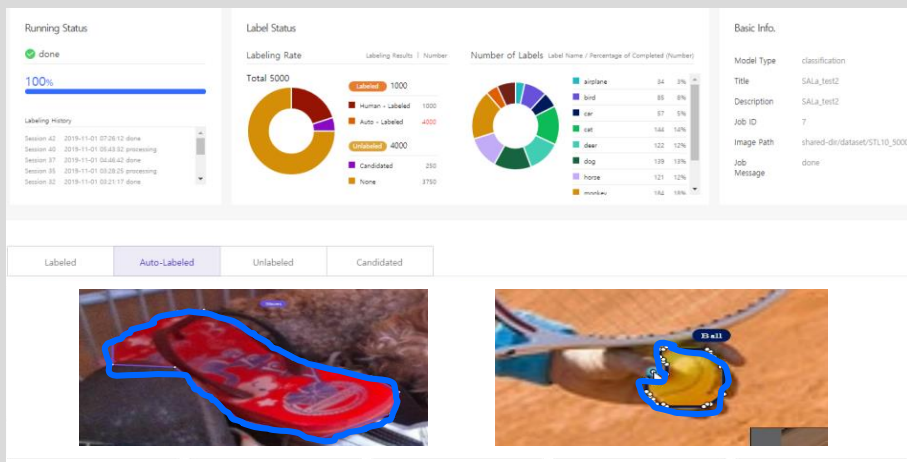
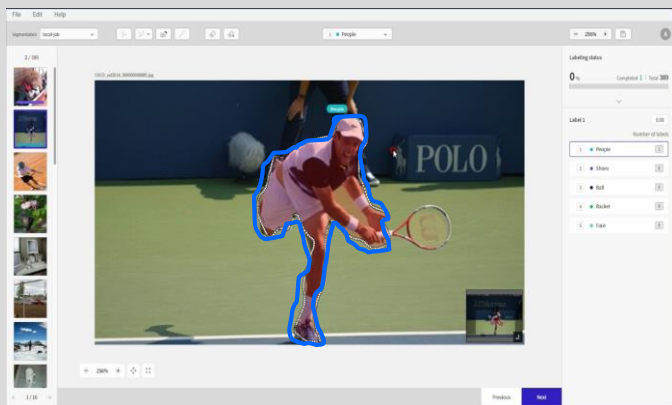
User testuser00@samsung.com



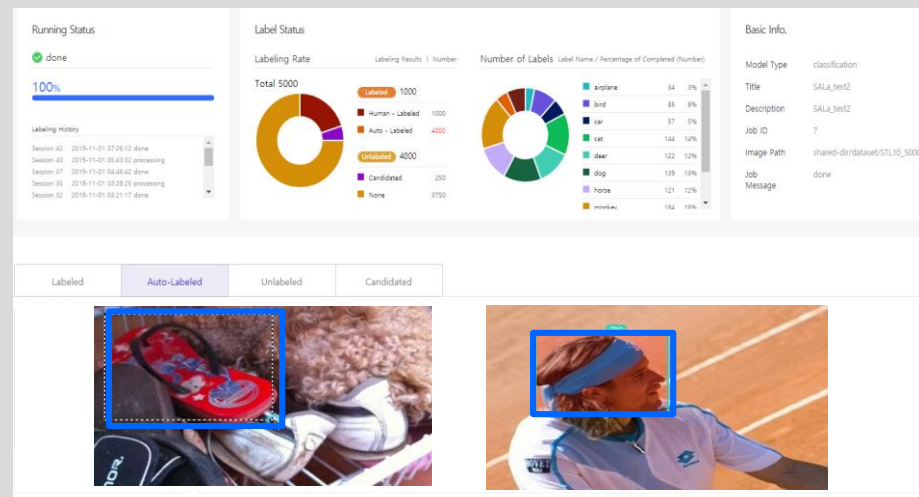
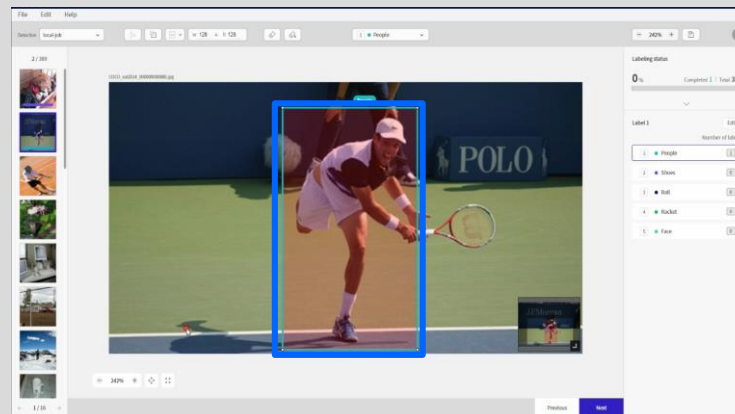
Running Status	Model Type	Title	Label list
----------------	------------	-------	------------

Coming Soon...

Segmentation



Detection

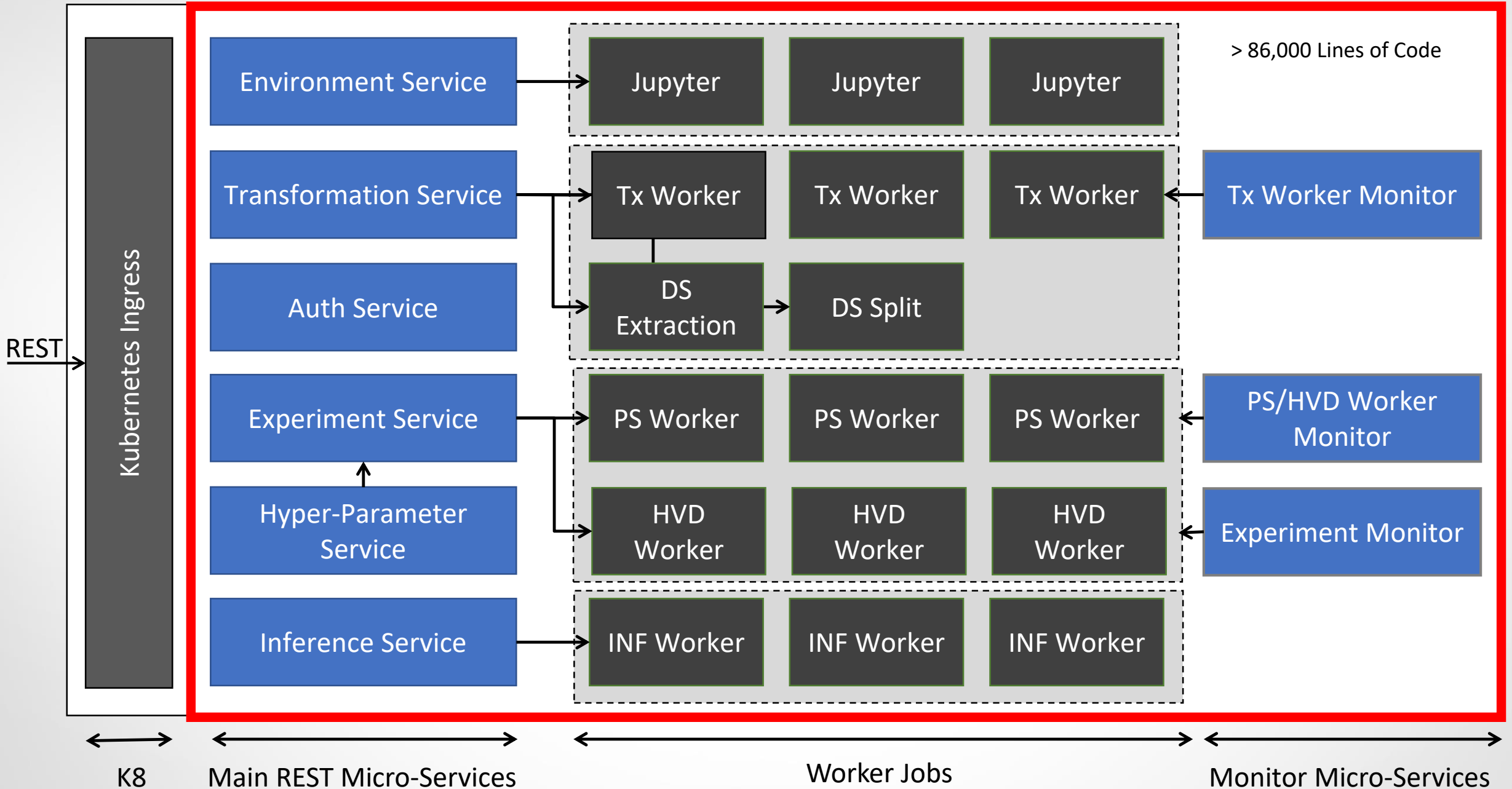


3

Distributed Model Training

Brightics DL Architecture Diagram

Samsung SDS Proprietary
Intellectual Property



Brightics DL - Usage

```
In [ ]: estimator = Estimator.NewClassificationEstimator(model_fn=my_model_function_with_mnist)
name = "BestModelEstimator" + str(time.time()).replace(".", "")
estimator = Estimator.create(name, "Hello", estimator)

hyper_parameters = HParams(iterations=50000, batch_size=10)
rc = RunConfig(no_of_ps=5, no_of_workers=20, summary_save_frequency=5000, run_eval=True, use_gpu=False,
checkpoint_frequency_in_steps=500)

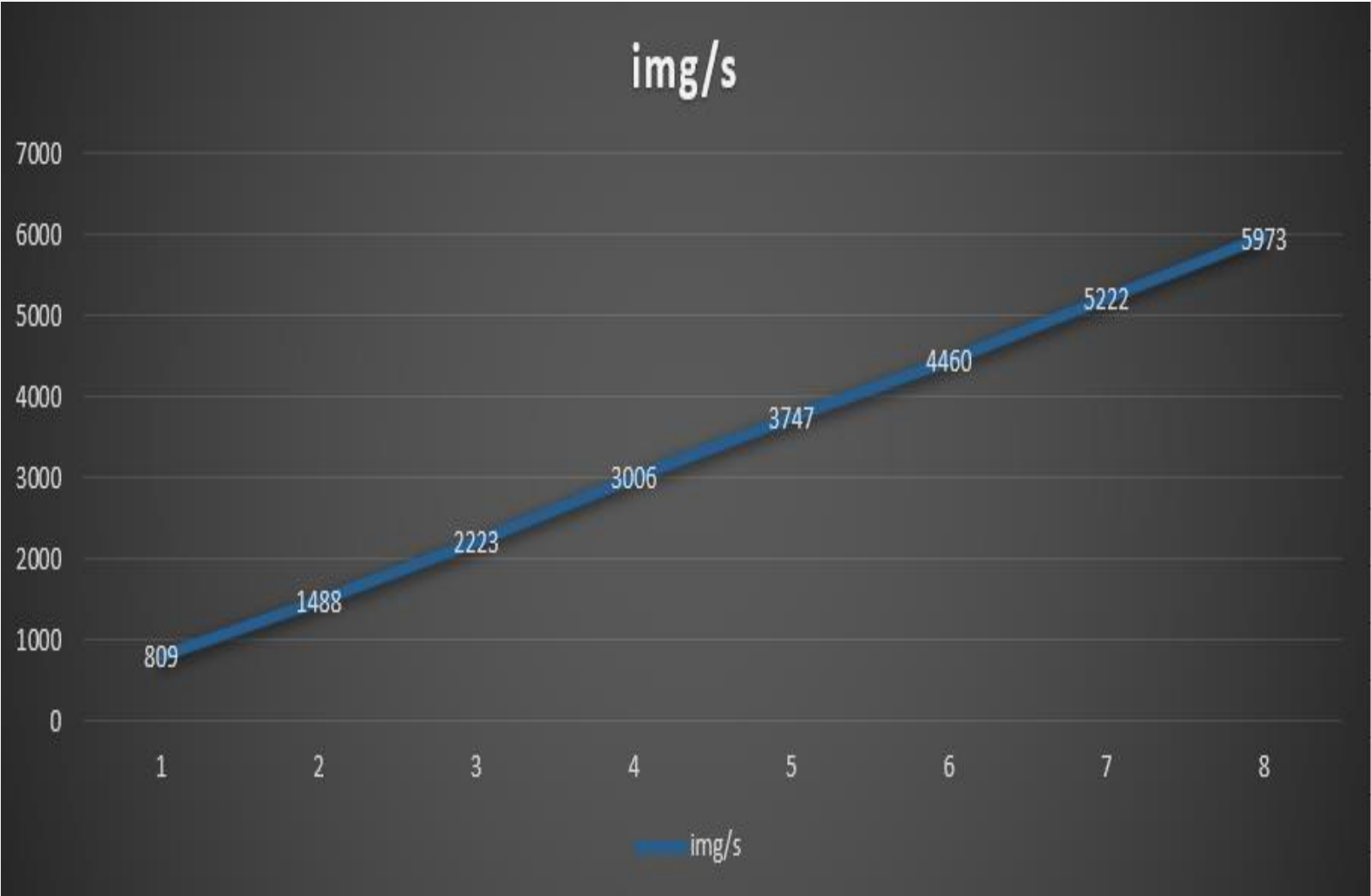
exper = Experiment.run(experiment_name="BestModelEstimator" + str(time.time()).replace(".", ""),
                      description="Really first model",
                      estimator=estimator,
                      hyper_parameters=hyper_parameters,
                      run_config=rc,
                      dataset_version_split=None,
                      input_function=input_function)

job = exper.get_single_job()
print(job.__dict__)
print("")
print("tensorboard url")
print(job.get_tensorboard_url())

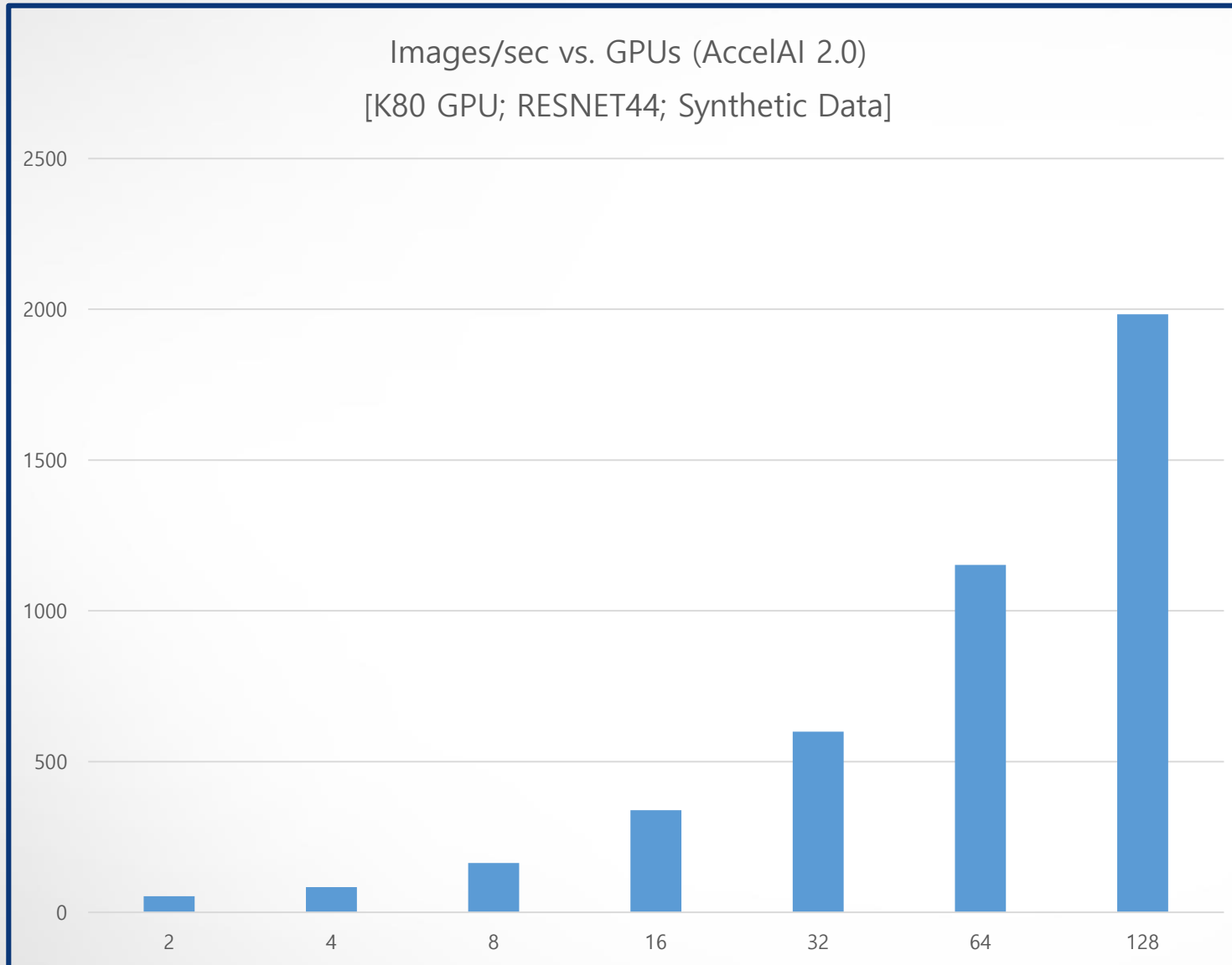
job.has_finished()

job.wait_until_finish()
```

Brightics DL (Shipping Today) - Parameter Server Performance

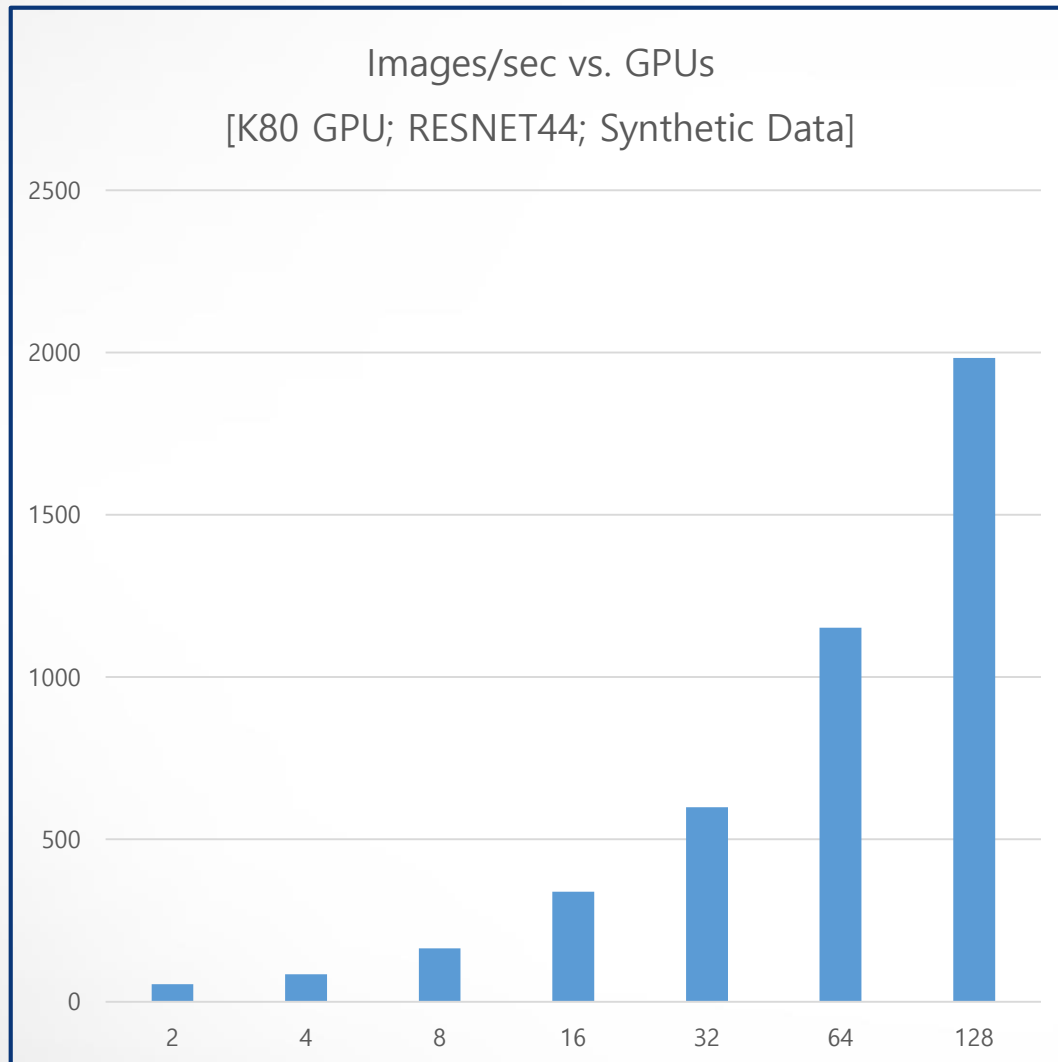


Brightics DL - Horovod Performance



No	GPUs	Images/sec	Scale factor
1	1	31	
2	2	54	1.75
3	4	84	1.57
4	8	164	1.95
5	16	339	2.07
6	32	599	1.77
7	64	1152	1.92
8	128	1984	1.72

Brightics DL - Available Now



4

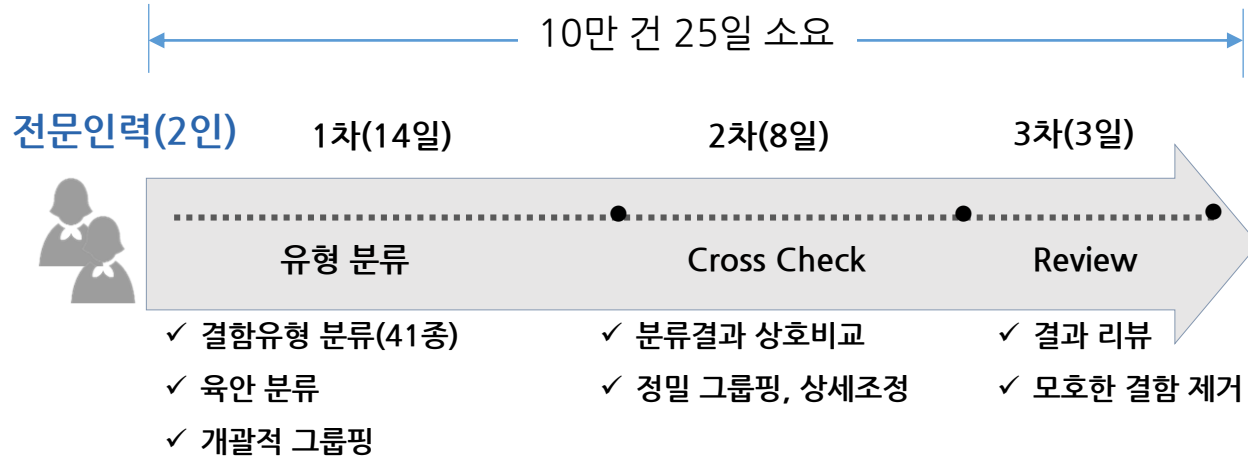
Use Cases

적용사례 1. A社, 결함검출 자동화(`18년)

Auto Labeling 기술 적용

전문 인력 2인이 25일 동안 3차에 걸친 데이터 정제 작업 수행 1인/10일로 단축

도입 전



도입 후



[총 10일 소요]

✓ 4회 Iteration

✓ 회당 5% 추가 라벨링

→ 총수작업 2만여 장

투입 비용 절감 : 인력 투입 절감, 2인(100%) → 2인(50%)

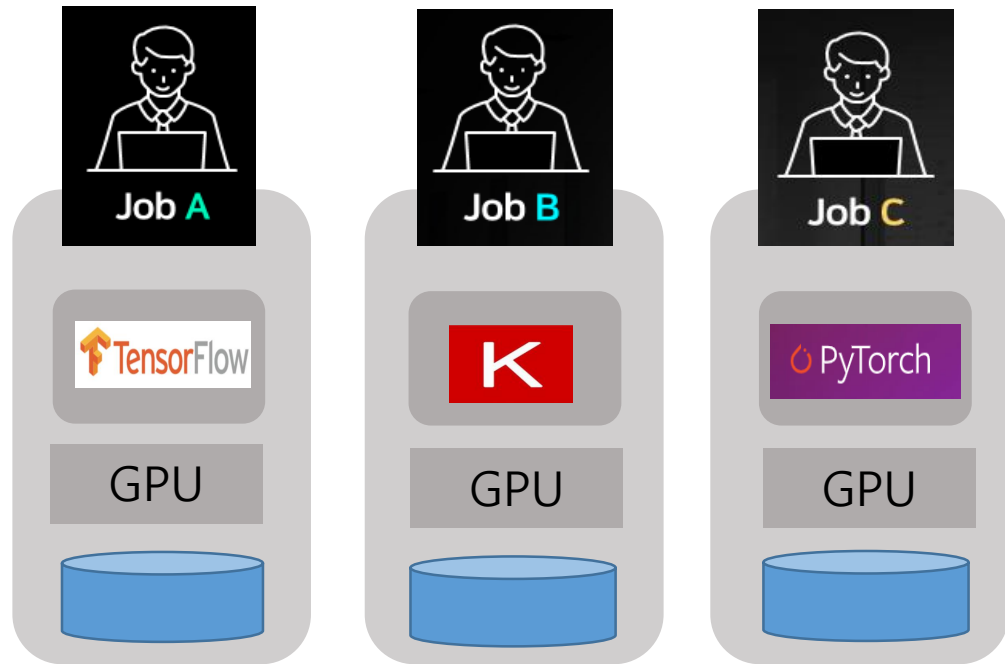
정제 품질 확보 : 분류 결과가 단일하고 재현가능

적용사례 2. B社, AI 모델 생성 가속화(` 19년)

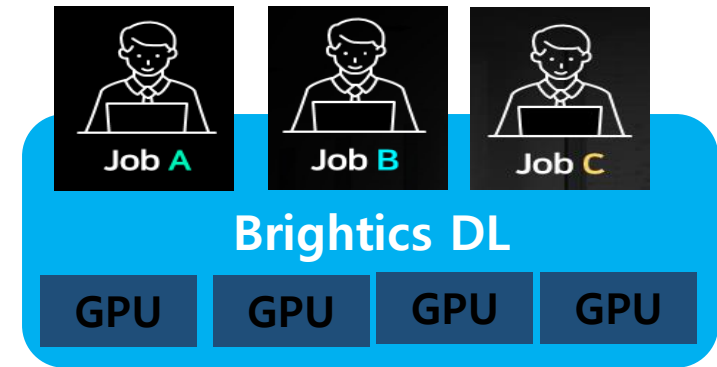
Brightics DL 적용

빠르고 쉽게 AI 모델을 생성하여 사용할 수 있는 지능화 환경 제공

도입 전

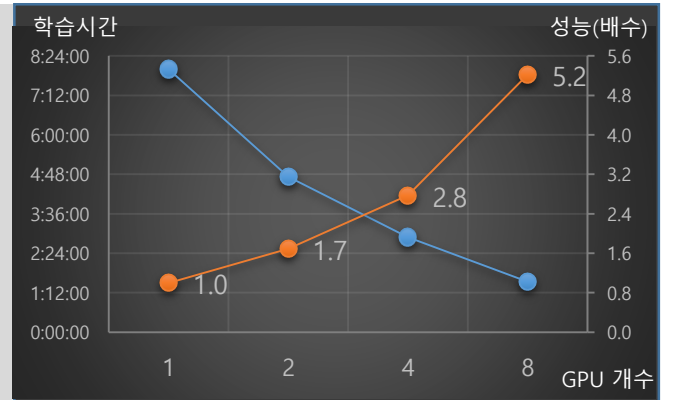


도입 후



[분산학습 속도측정 결과]

현장 이미지 데이터의
8개의 GPU로 학습 시,
5.2배 빠른 학습 성능
: 200 x 128 x 3 channel 50,000장



Summary

1. Brightics DL, *'탄생기'*
2. Auto Labeling, *'알고리즘부터 패키징까지'*
3. Distributed Model Training, *'원리부터 성능까지'*
4. Use Case, *'배운 점'*



Thank You





Q & A

Partner Disrupt Foresee