

R에서 MLOps 시작하기 : 데이터 과학자 & 개발자/엔지니어 협력

Ian Choi

Field Developer Relations, Microsoft APAC

MLOps를 살펴보기 전에, DevOps란 무엇일까요?

사람, 프로세스, 제품.

“

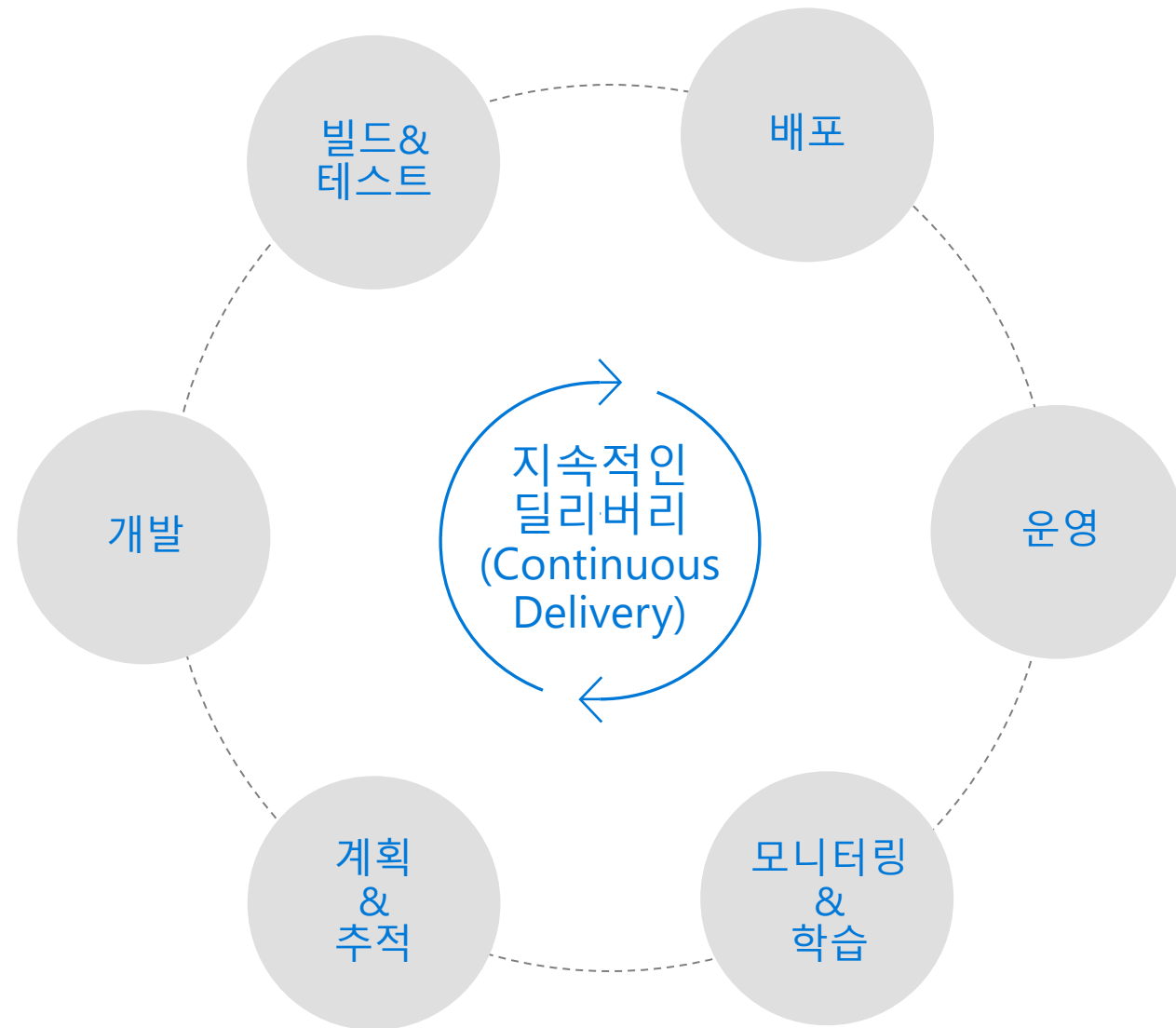
DevOps는 **사람**, **프로세스**, **제품**이 함께 결합하여 최종 고객에게 가치를 지속적으로 전달하게 해 줍니다.

”

“

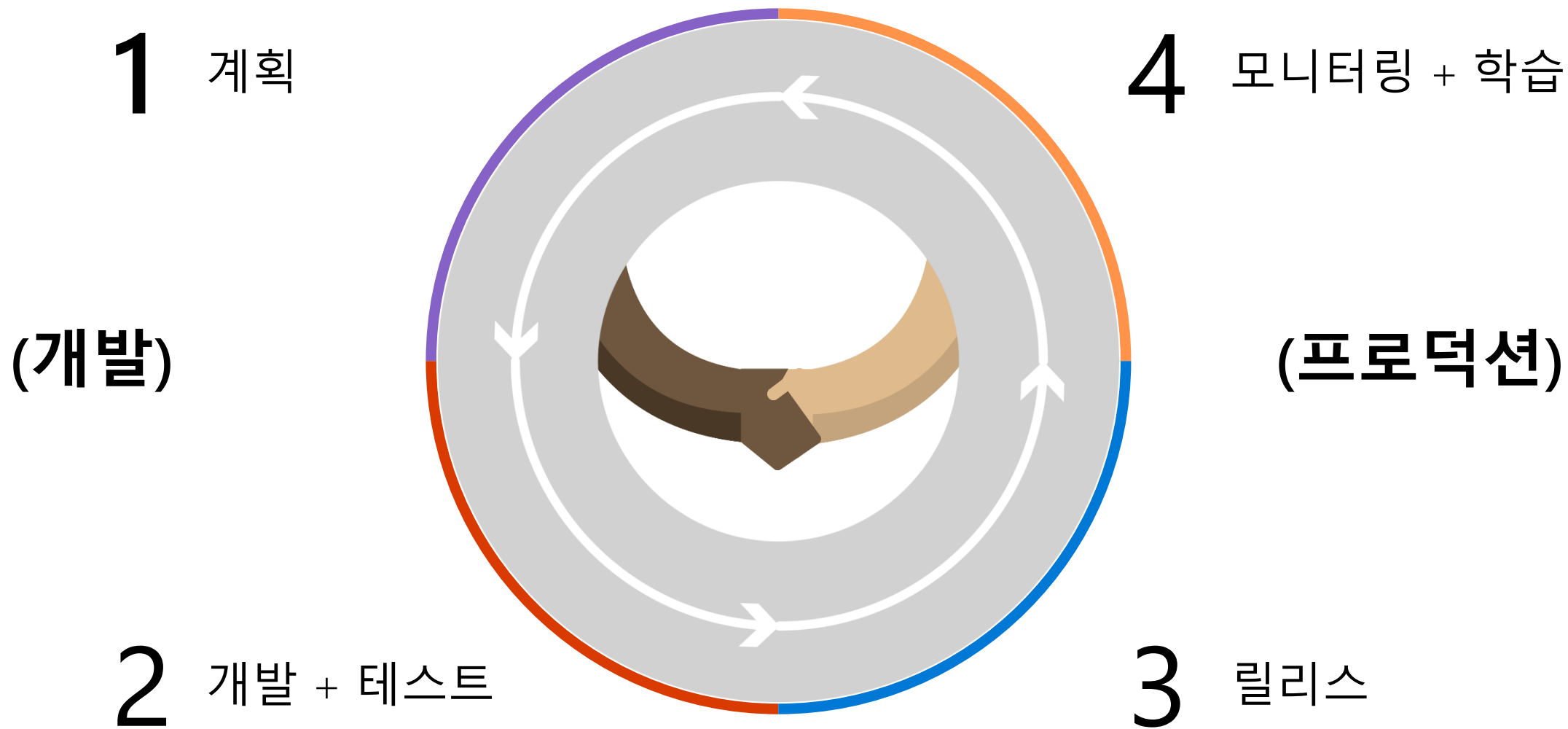
DevOps is the union of **people**, **process**, and **products** to enable continuous delivery of value to your end users.

”

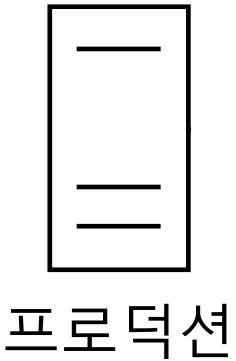
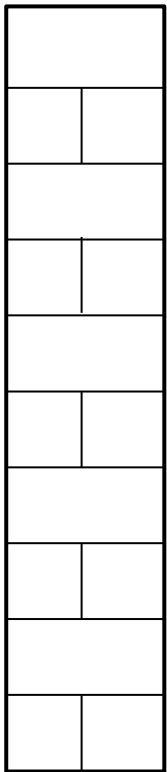
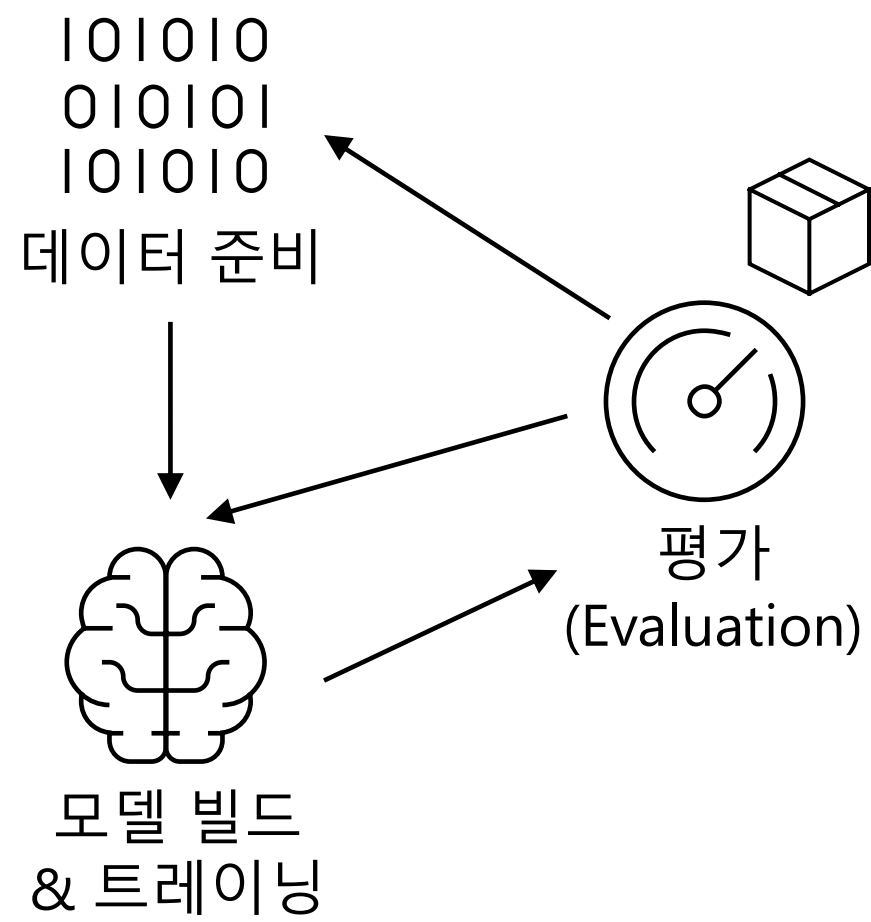


잠시, **영상**을 함께 봅시다...

DevOps 프로세스



머신 러닝 프로세스



DEVOPS	MLOPS
코드 관리 (소스 파일)	코드 관리 (소스 파일) 데이터 파일, 노트북, README 등 문서 관리
인프라 관리 (as code)	인프라 관리 (as code) 환경 관리 (as code)
소스 코드/버전 제어	소스 코드 제어 실험 결과 추적 데이터셋 관리
실행 파일 빌드 빌드: 짧으면 수 분, 길게는 몇 시간 소요 (대부분) 상용 컴퓨팅 자원 또는 PaaS	모델 트레이닝 모델 트레이닝: 때로는 며칠/몇 주 소요 GPU 컴퓨팅
빌드 버전 관리	모델 버전 관리 재현 가능한 환경 관리
테스트 (deterministic) 코드 버그 수정	테스트 (probabilistic) 코드 버그 수정 and/or 데이터 모델 변경 / 모델 재트레이닝 등

Accident Fatality Probability Estimator

Occupant Age:

16

69

95

16

24

32

40

48

56

64

72

80

88

95

Occupant gender:

f

Occupant role:

pass

Vehicle Year:

1,955

2,002

1,955

1,960

1,965

1,970

1,975

1,980

1,985

1,990

1,995

2,000

2,005

Seatbelt:

belted

Airbag:

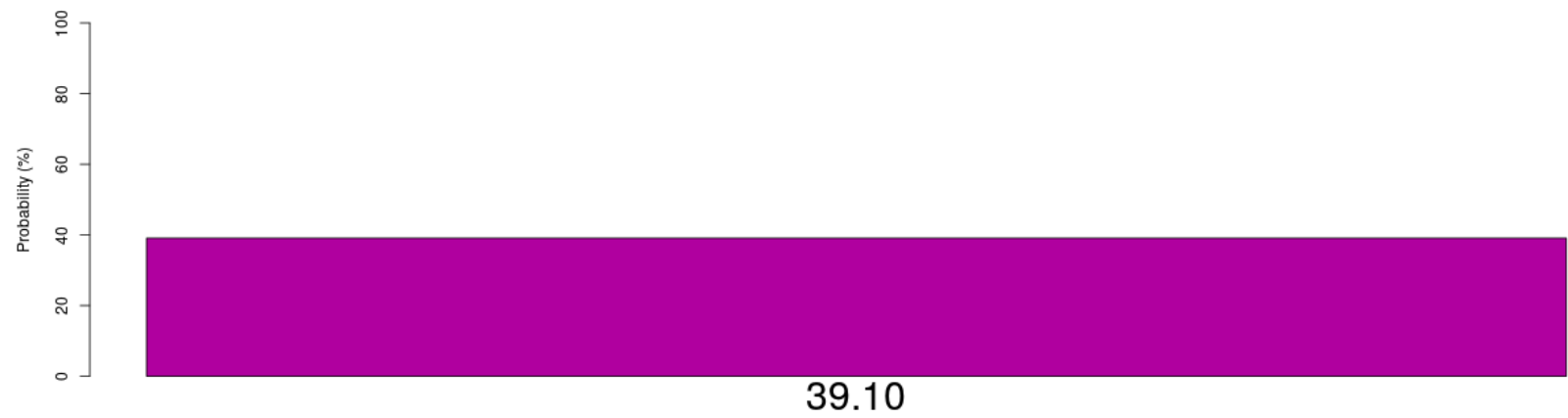
none

Impact speed:

40-54

Collision type:

notfrontal



Azure Machine Learning 서비스 & GitHub Actions

Set of Azure
Cloud Services



Python & R
SDKs



GitHub
Actions

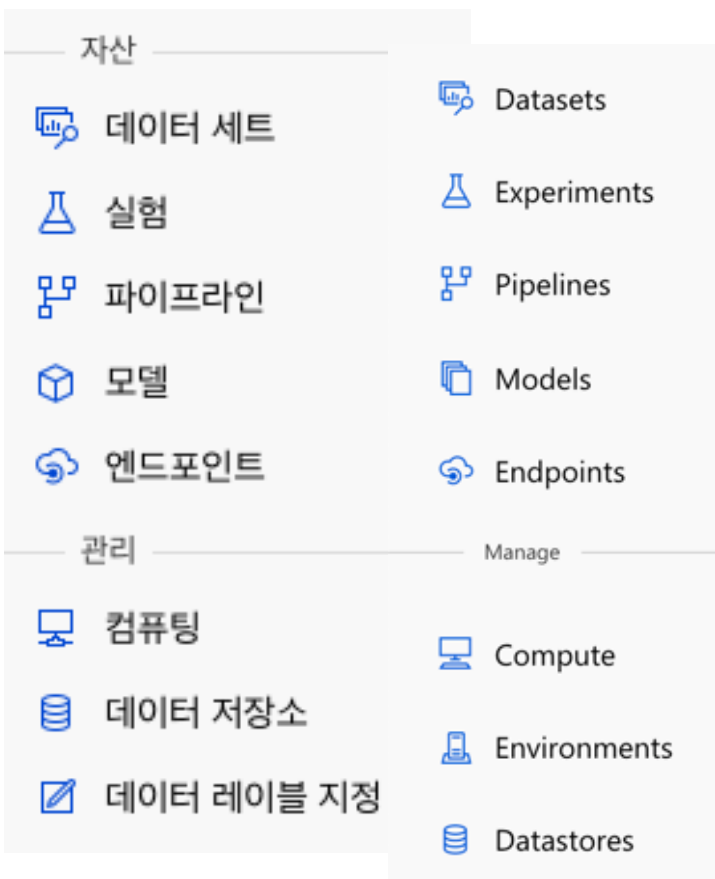
That enables
you to:

- ✓ Prepare Data
- ✓ Build Models
- ✓ Train Models

- ✓ Manage Models
- ✓ Track Experiments
- ✓ Deploy Models

- ✓ Manage Code
- ✓ Collaborate
- ✓ Continuous Integration

Azure Machine Learning



Datasets – registered, known data sets

Experiments – 트레이닝 실행

Pipelines – 트레이닝 워크플로우

Models – 등록된 모델 (버전 관리)

Endpoints – 배포가 이루어진 모델 엔드포인트

Compute – (CPU/GPU) 컴퓨팅 자원 관리

Environments – 트레이닝 및 추론 환경 관리

Datastores – 데이터저장소와 연결

Azure ML service SDK for R

CRAN R과 함께 사용하는 오픈 소스 R 패키지:
azuremlsdk

- R 명령어로 워크스페이스/실험/컴퓨팅/모델 등 생성
- R 기능/패키지 사용 (+ 배포에 대한 요구 사항 추적)
- 트레이닝 클러스터 오토 스케일링 (GPU 클러스터 포함)
- HyperDrive 지원: 병렬 컴퓨팅에서 스마트한 hyperparameter 검색
- 모델을 웹 서비스로 배포 (Azure에, 또는 전용 인프라에)

aka.ms/ian-mlops-r

azuremlsdk 0.10.0

Azure Machine Learning SDK for R (preview)

Data scientists and AI developers use the Azure Machine Learning SDK for R to build and run machine learning workflows with Azure Machine Learning.

cran.r-project.org

azuremlsdk: Interface to the 'Azure Machine Learning' 'SDK'

Interface to the 'Azure Machine Learning' Software Development Kit ('SDK'). Data scientists can use the 'SDK' to train, deploy, automate, and manage machine learning models on the 'Azure Machine Learning' service. To learn more about 'Azure Machine Learning' visit the website: <https://docs.microsoft.com/en-us/azure/machine-learning/service/overview-what-is-azure-ml>.

Version:

0.6.85

Depends:

R (≥ 3.5.0)

Imports:

[ggplot2](#), [reticulate](#) (≥ 1.12), [plyr](#) (≥ 1.8), [DT](#), [rstudioapi](#) (≥ 0.7), [htmltools](#), [servr](#), [shiny](#), [shinycssloaders](#)

Suggests:

[rmarkdown](#), [knitr](#), [testthat](#), [dplyr](#), [jsonlite](#)

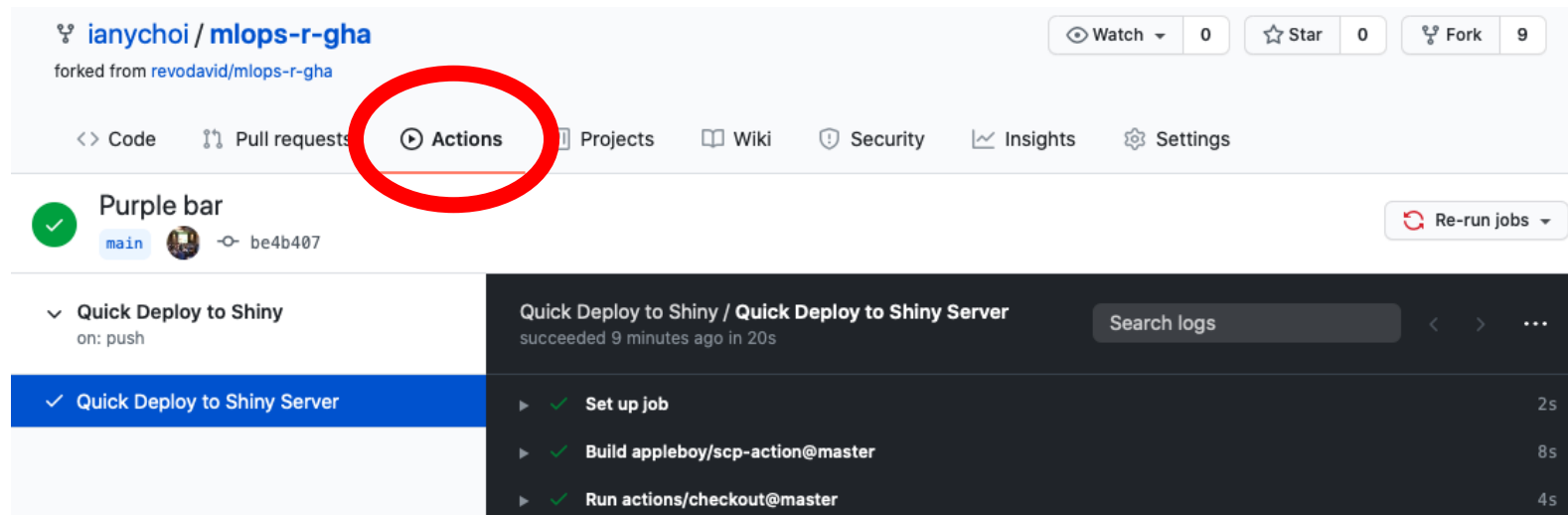
Published:

2020-02-05

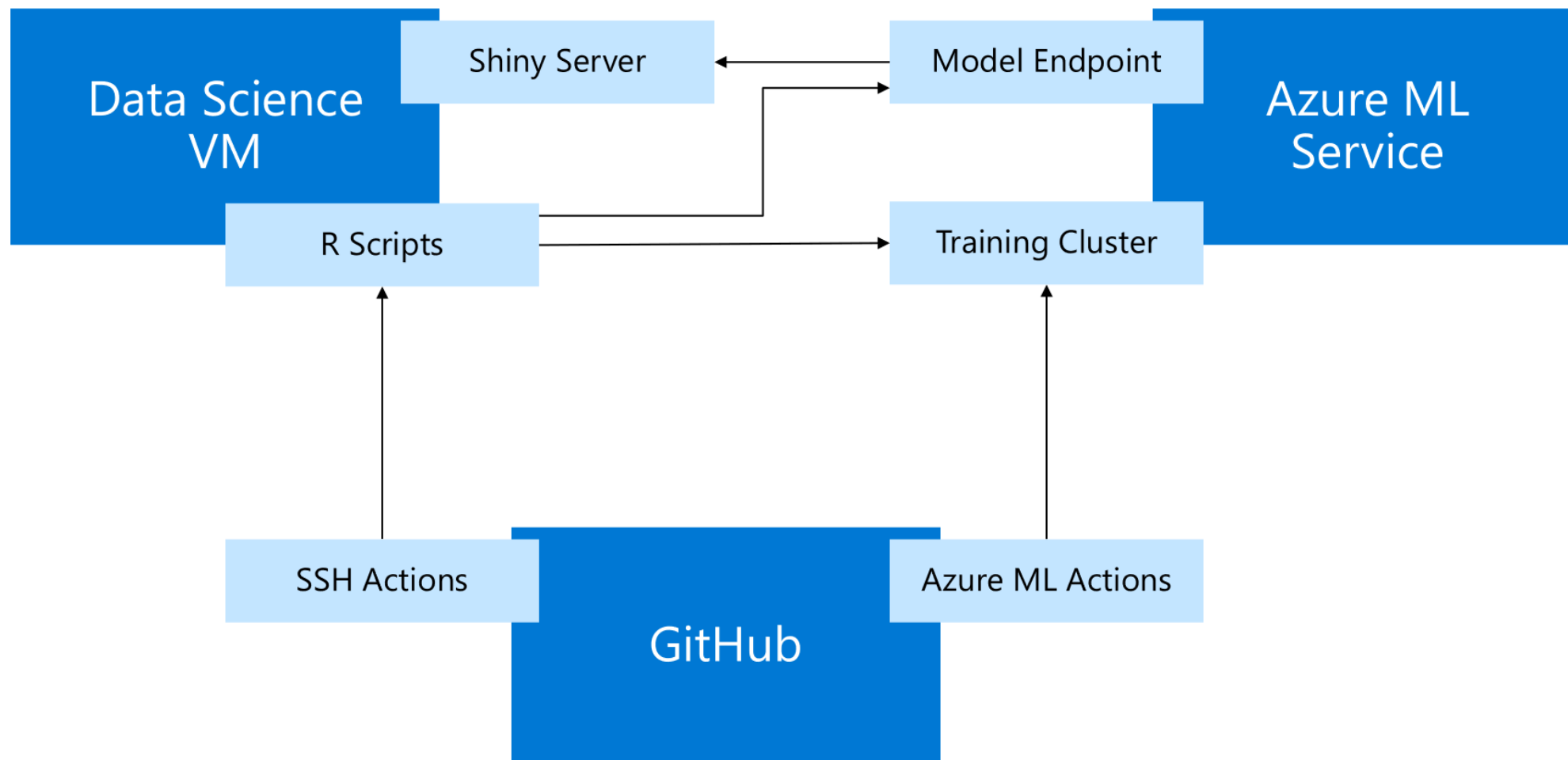
@ianychoi at #SeoulRMeetup

GitHub Actions (짧은 소개)

- YAML파일을 .github/workflows 경로에 생성하여 작업을 정의
- GitHub Actions 마켓플레이스에서 미리 정의된 템플릿 검색/사용
- Actions 탭에서 워크플로우를 모니터링



“Accident” 치사율에 대한 Shiny App 아키텍처



4개 노드 트레이닝 클러스터

10-azureml.yml

Connect or Create a Compute Target in Azure Machine Learning

- name: Connect/Create Azure Machine Learning Compute Target

id: aml_compute_training

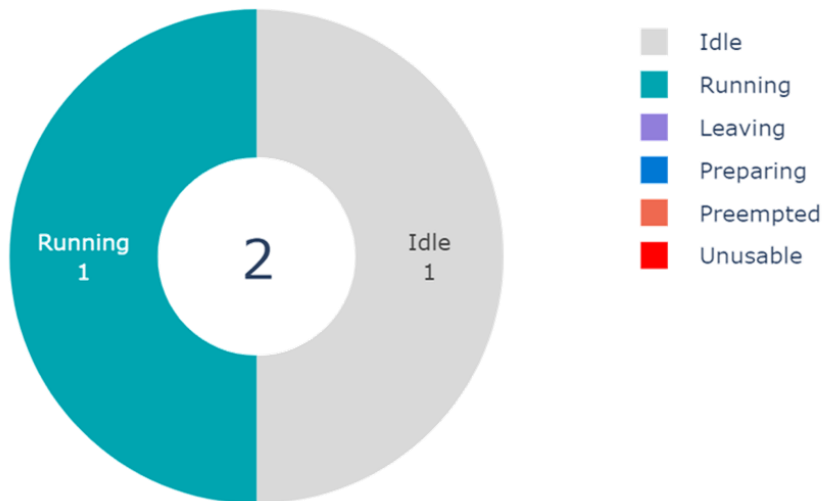
uses: Azure/aml-compute@v1

with:

azure_credentials: \${ secrets.AZURE_CREDENTIALS }

compute.json

```
{  
  "name": "rcluster",  
  "compute_type": "amlcluster",  
  "min_nodes": 0,  
  "max_nodes": 4,  
  "idle_seconds_before_scaledown": 600  
}
```



aka.ms/ian-mlops-r

- 사용하지 않는 노드는 자동으로 할당 해제
- 개발 중일 때는 idle 타임아웃을 늘리자

@ianychoi at #SeoulRMeetup

모델 트레이닝

10-azureml.yml

```
# Connect to the Shiny VM to train the model
- name: Train model
uses: JimCronqvist/action-ssh@master
env:
  AZURE_CREDENTIALS: '${{ secrets.AZURE_CREDENTIALS}}'
with:
  hosts: '${{ secrets.SHINYUSERNAME }}@${{ secrets.SHINYHOST }}'
  privateKey: '${{ secrets.SHINYKEY }}'
command: |
  cd mlops-r-gha/model
  export AZURE_CREDENTIALS
  Rscript train-model.R
```

train-model.R

```
est <- estimator(source_directory=".",
  entry_script = "accident-glm.R",
  script_params = list("--data_folder" =
    ds$path(target_path)),
  compute_target = compute_target)
cat("Created Estimator\\n")

run <- submit_experiment(exp, est)
cat("Submitted Run\\n")
```

- 실험: 소스 스크립트 및 기록된 지표에 의해 추적
- 명령 프롬프트 매개변수에 의해 실행 제어
- 대부분 패키지들이 미리 로드되어 있으며, 사용자 정의 패키지 또한 지원함

accident

테이블 편집 새로 고침 보기 다시 설정 차트 추가

이 페이지에 대한 사용자 지정은 이 브라우저에서 계속 유지되며 다른 사람이 동일한 페이지를 경험하는 방식에는 영향을 주지 않습니다.

+ 필터 추가

자식 실행 포함

실행 상태

실행 중
0

완료됨
2

실패
3

기타
0

선택한 행만 표시 (5개 선택됨)

Accuracy



✓ 실행	실행 ID	상태	제출된 시간	기간
✓ 실행 5	accident_1599403704068	완료됨	Sep 6, 2020 11:48 PM	2분 22초
✓ 실행 4	accident_1599399149_a4acbd9f	완료됨	Sep 6, 2020 10:32 PM	1분 58초

Shiny 앱과 모델 Integration

```
app.R
```

```
library(httr)
```

```
v <- POST(accident.endpoint,  
  body=input,  
  encode="json")
```

```
pred() <- content(v)[[1]]
```

```
20-shiny.yml
```

```
- name: copy files via ssh key  
uses: appleboy/scp-action@master  
with:  
  host: ${ secrets.SHINYHOST }  
  username: ${ secrets.SHINYUSERNAME }  
  key: ${ secrets.SHINYKEY }  
  source: "accident-app/app.R"  
  target: "~"
```

aka.ms/ian-mlops-r

Accident Fatality Probability Estimator

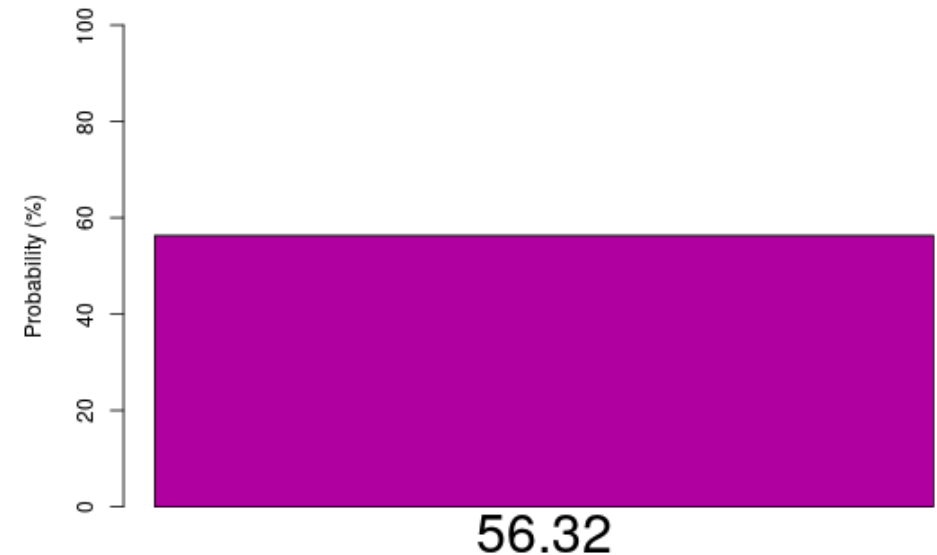
Occupant Age: 16 24 32 40 48 56 64 72 80 88 95

Occupant gender:

Occupant role:

Vehicle Year: 1,955 1,965 1,975 1,985 1,995 2,005

Seatbelt:



@ianychoi at #SeoulRMeetup

Demo: R 파일 업데이트에 따른 GitHub Actions

예상 일일 비용*

(*\$USD 기준, 비용은 리전 및 서비스에 따라 다를 수 있음. 아래는 설명을 위한 예시)

GitHub Actions	\$0.00
Azure ML Service	\$0.00 (스튜디오, 로깅, 오케스트레이션)
트레이닝 클러스터	\$0.10 (필요시만 사용)
Scoring을 위한 엔드포인트	\$1.67 (Azure 컨테이너 인스턴스)
Shiny Server	\$3.59 (Data Science VM, 4CPU 14Gb)
합계	\$5.36

Azure 체험 계정을 통해 약 200달러 무료 크레딧 이용 가능: **aka.ms/AML-SEOULR**

Thank you!

Ian Choi
Field Developer Relations,
Microsoft APAC

Slides and links:
aka.ms/ian-mlops-r

Note: This is a Korean localized version of <https://youtu.be/hCCZZyHz-ko> by David Smith, Cloud Advocate
- thanks a lot David for your love to R communities!