



# TF 변환

발표자: 권기호

# 목차

A table of contents

1 전처리

2 TF 변환





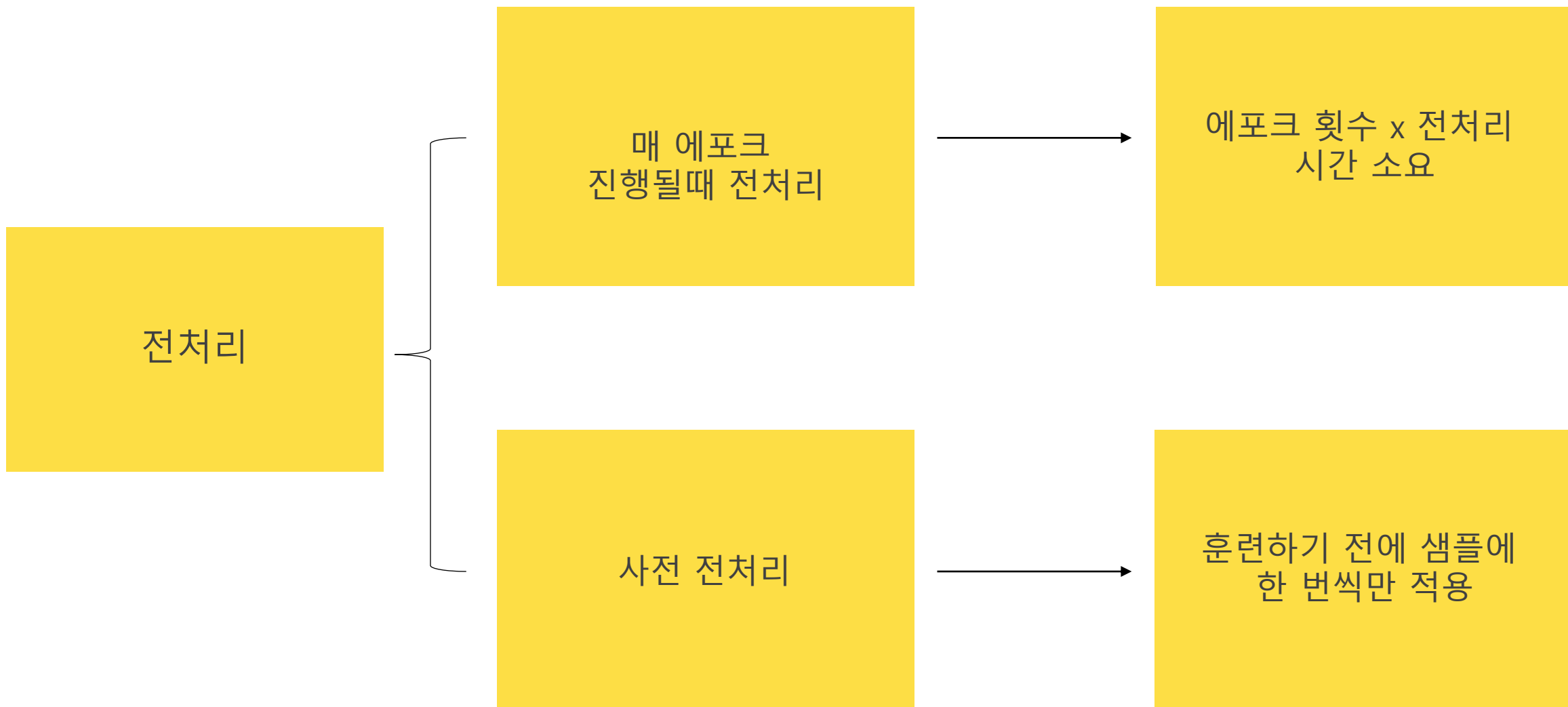
1

# 전처리

전처리  
Dataset: Small, Big  
훈련/서빙 왜곡

# 1 전처리

데이터 사전 전처리



SMALL

VS.

BIG

## 2 SMALL

데이터 크기가 작은 경우

### 001 >> Cache?

값을 미리 저장한 임시 데이터 -> RAM에 저장 가능할 정도로 작은 데이터

---

### 002 >> Use Cache Method

처음 불러온 데이터를 저장 해놓고, 이후 필요할 경우 저장된 데이터를 호출하여 사용

---

### 003 >> How?

```
train_univariate = tf.data.Dataset.from_tensor_slices((x_train_uni, y_train_uni))
# cache, shuffle, repeat
train_univariate = train_univariate.cache().shuffle(BUFFER_SIZE).batch(BATCH_SIZE).repeat()

val_univariate = tf.data.Dataset.from_tensor_slices((x_val_uni, y_val_uni))
val_univariate = val_univariate.batch(BATCH_SIZE).repeat()
```

### 001 >> Use Apache Beam or Spark

대용량 데이터 처리를 위한 클러스터 컴퓨팅 엔진을 활용하여 Pipeline 구축

---

### 002 >> 유지 보수의 어려움 (훈련/서빙 왜곡)

훈련 전에 수행한 전처리 연산과 앱이나 브라우저에서 수행하는 전처리 차이 발생

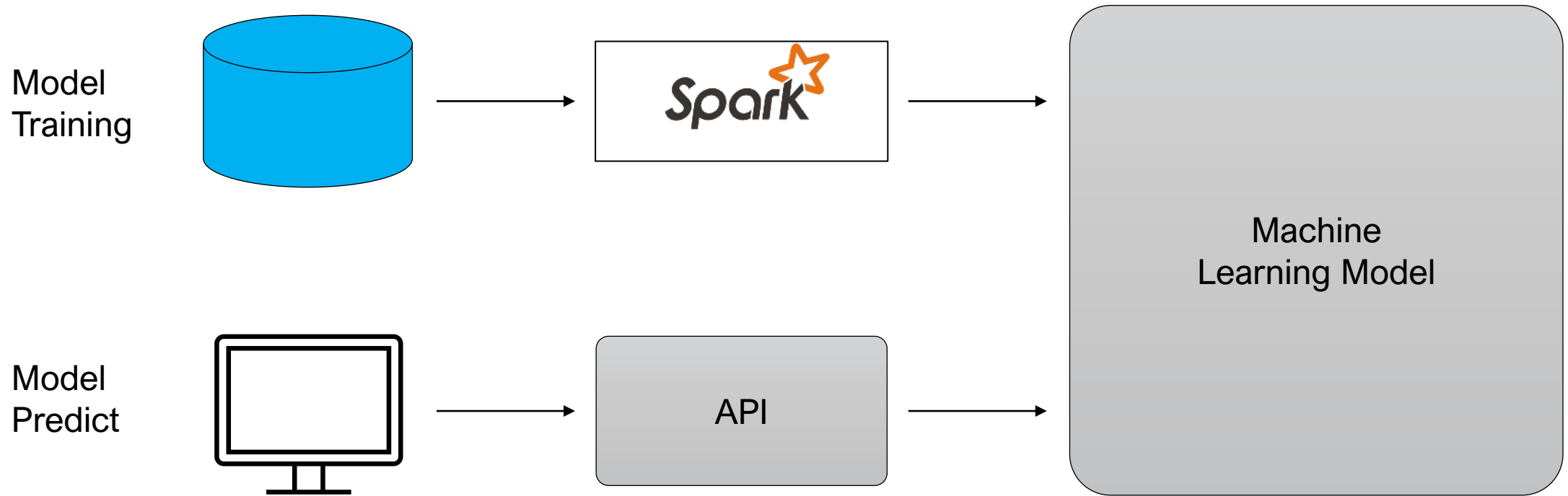
---

### 003 >> Solution?

TF 변환

# 4 훈련 서버 왜곡

Architecture





# TF변환

TF 변환  
TF 변환 적용



2

# 1 TF 변환

## 1. 훈련/서빙 왜곡 방지

Training과 Serving의 데이터  
전처리 괴리 방지

## 2. 유지 보수 향상

한번의 처리로 전처리 과정을  
실행하여 이상적인 MLOps 기여

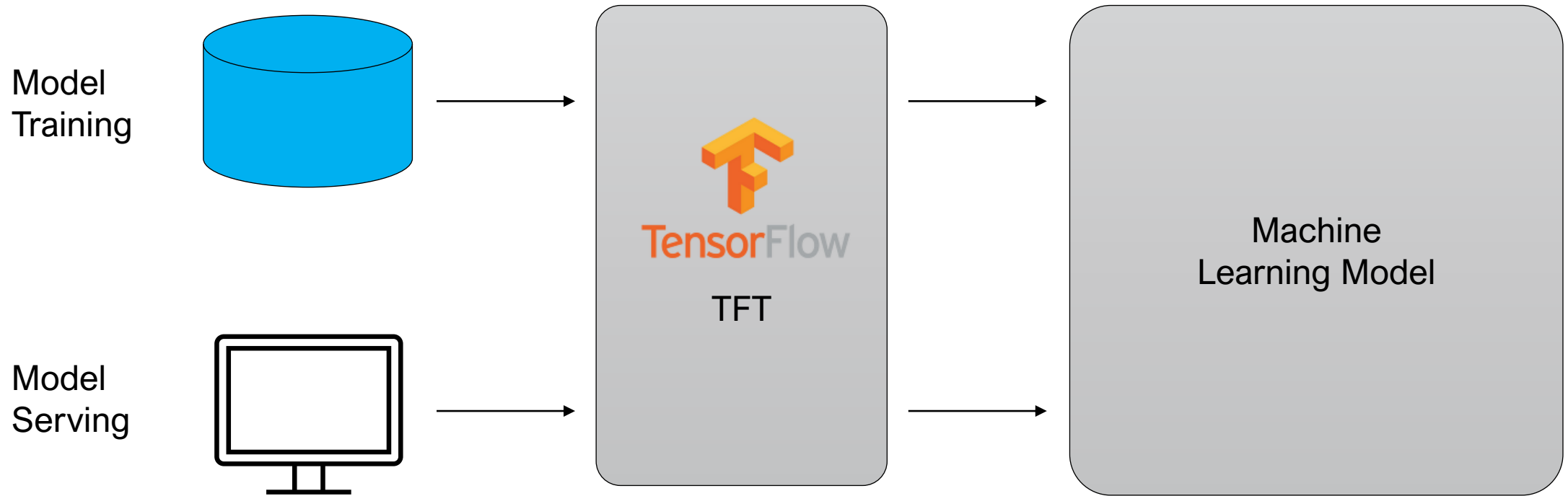
## 3. 확장성

GCP의 Dataflow, Apache spark,  
Apache beam 등 활용 가능

```
import tensorflow_transform as tft
def preprocess(inputs):
    median_age = inputs["housing_median_age"]
    ocean_proximity = inputs["ocean_proximity"]
    standardized_age = tft.scale_to_z_score(median_age)
    ocean_proximity_id = tft.compute_and_apply_vocabulary(ocean_proximity)
    return {
        "standardized_median_age": standardized_age,
        "ocean_proximity_id": ocean_proximity_id
    }
```

## 2 TF 변환 적용

Architecture





The background of the image is a close-up, top-down view of a cobblestone path. The cobblestones are dark grey or black, with a rough, textured surface. A section of the path, running diagonally from the top right towards the bottom right, is painted a bright yellow. The yellow paint is applied in a way that covers the surface of the cobblestones in that area, though some of the original stone texture is still visible through the paint. The lines between the cobblestones are dark and recessed. In the center of the image, there is a horizontal grey rectangular box containing the Korean text '질문 타임' in yellow.

질문 타임





**Thanks!**