Summary

Feature Engineering

Lak Lakshmanan

You learned how to...

You learned how to...

Convert raw data into features

You learned how to...

Convert raw data into features

Preprocess data in such a way that the preprocessing is also done during serving

You learned how to...

Convert raw data into features

Preprocess data in such a way that the preprocessing is also done during serving

Choose among the various feature columns in TensorFlow

You learned how to...

Memorize large datasets using feature crosses and simple models
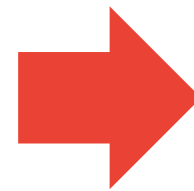
You learned how to...

Memorize large datasets using feature crosses and simple models

Simplify preprocessing pipelines using TensorFlow Transform

Raw data to features

Feature engineering is necessary because raw data doesn't come to us as feature vectors

```
0 : {
  house_info : {
    num_rooms: 6
    num_bedrooms: 3
    street_name: "Main
Street"
    num_basement_rooms: -1
    ...
  }
}
```

```
[
  6.0,
  1.0,
  0.0,
  0.0,
  0.0,
  9.321,
  -2.20,
  1.01,
  0.0,
  ...,
]
```

Process of creating features from raw data is Feature Engineering

Things that are commonly done in preprocessing

Remove examples that you don't want to train on

Compute vocabularies for categorical columns

Compute aggregate statistics for numeric columns

In BigQuery or Beam

Compute time-windowed statistics (e.g. number of products sold in previous hour) for use as input features

In Beam only

Scaling, discretization, etc. of numeric features

Splitting, lower-casing, etc. of textual features

In TensorFlow or Beam

Resizing of input images

Normalizing volume level of input audio

Things that are commonly done in preprocessing

Remove examples that you don't want to train on

Compute vocabularies for categorical columns

Compute aggregate statistics for numeric columns

In BigQuery or **Beam**

Compute time-windowed statistics (e.g. number of products sold in previous hour) for use as input features

In **Beam** only

Scaling, discretization, etc. of numeric features

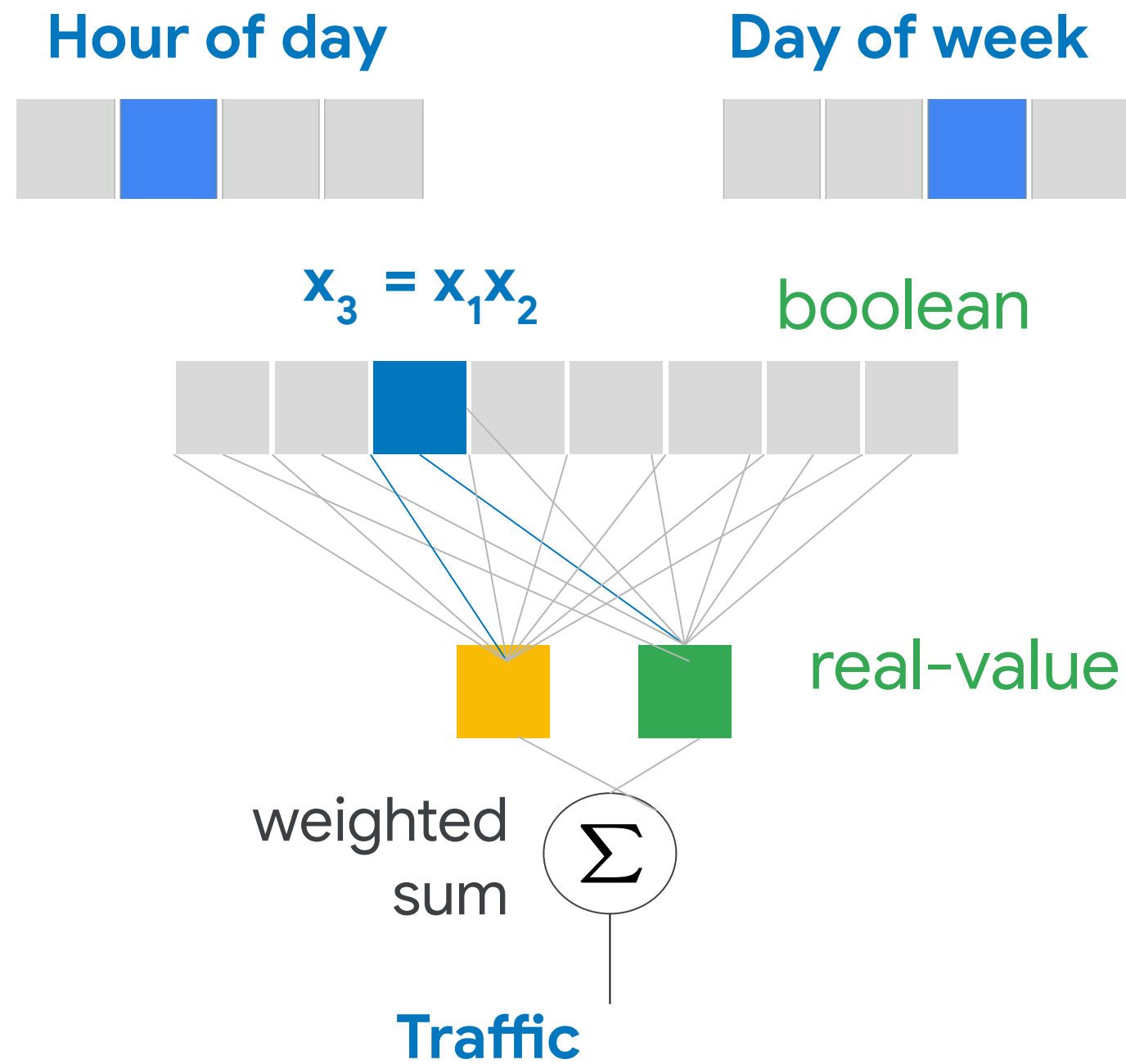Splitting, lower-casing, etc. of textual features

Resizing of input images

Normalizing volume level of input audio

In TensorFlow or **Beam**
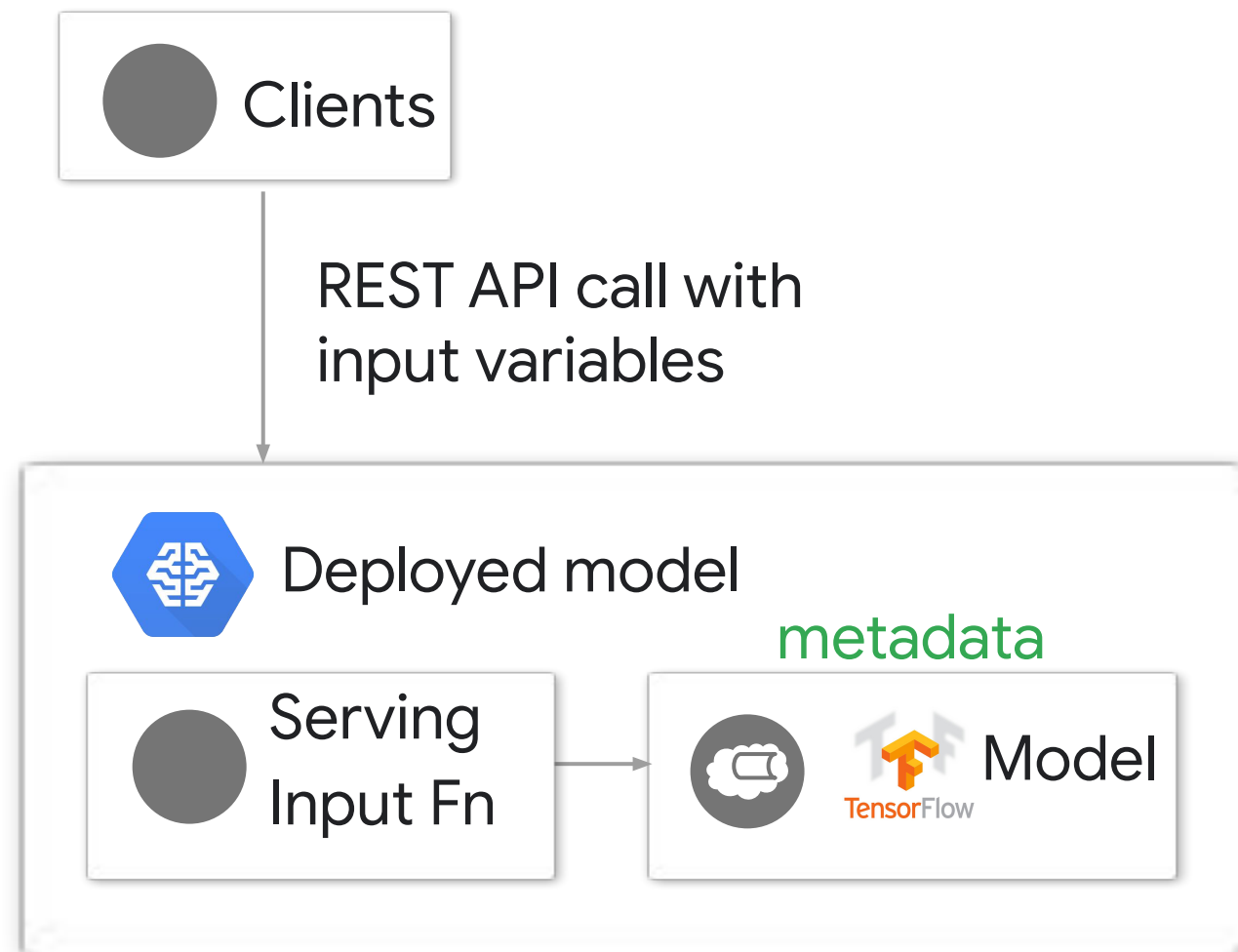
Creating an embedding column from a feature cross

The model graph includes the preprocessing code

Machine Learning on Google Cloud Platform

Machine Learning on Google Cloud Platform

Why ML?

Machine Learning on Google Cloud Platform

Why ML?

ML with TensorFlow

Machine Learning on Google Cloud Platform

Why ML?

ML with TensorFlow

Improving ML Accuracy

Machine Learning on Google Cloud Platform

Why ML?

ML with TensorFlow

Improving ML Accuracy

ML at scale

Specialized ML models

cloud.google.com