

Week 1

Week 2

Week 3

Week 4

Grades

Notes

Discussion Forums

Messages

Course Info

# Week 1

Deploying Machine Learning Models in Production

### Week 1 Discuss the topic here. Go to forum 3 threads · Last post 22 days ago

# Week 1: Model Serving: Introduction





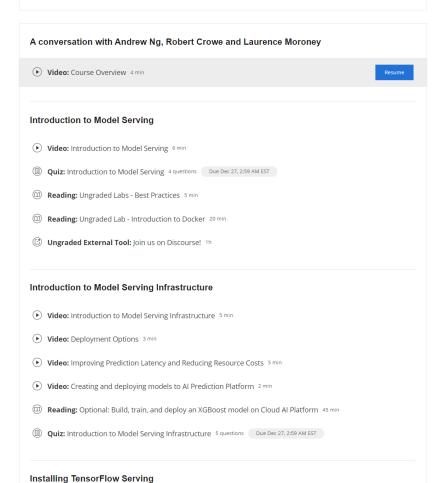
Learn how to make your ML model available to end-users and optimize the inference process

# **Learning Objectives**

- Identify and contrast the challenges for serving inference requests
  Compare cost, latency and throughput metrics to optimize serving inference requests
- Judge the hardware resources and requirements for your serving models so that your system is reliable and can scale based on
- Install and Use TensorFlow Serving to serve inference requests on a simple image classification model









▶ Video: Installing TensorFlow Serving 6 min

Quiz: TensorFlow Serving 2 questions Due Dec 27, 2:59 AM EST Reading: Ungraded Lab - Tensorflow Serving with Docker 20 min © Reading: Ungraded Lab - Serve a model with TensorFlow Serving 30 min