Google Cloud

# Production Machine Learning Systems

Here are the assembled readings provided in this course.

## Module 1: Architecting Production ML Systems

- Architecture of a real-world Machine Learning system
- ML Reference Architecture
- Machine Learning Pipeline: Architecture of ML Platform in Production
- Design Decisions for Architecting Production Machine Learning Systems
- Production ML Systems
- 3 Building Blocks of Machine Learning you Should Know as a Data Scientist
- MLOps: Continuous delivery and automation pipelines in machine learning
- Smart Decisions Game: Machine Learning for Architects
- Training and Serving CARET models using AI Platform Custom Containers and Cloud Run
- Using TensorFlow to predict product weight and dimensions
- Getting batch predictions
- How to extend a canned TensorFlow Estimator
- Introduction to loading data
- Google Cloud  Vertex AI
- AI Simplified
- Cloud Blog AI & Machine Learning
- GitHub - Google Cloud Pipeline Components
- Vertex AI: Building a fraud detection model with AutoML

## Module 2: Designing Adaptable ML Systems

- Deep Learning AI Needs Tools To Adapt To Changes In The Data Environment
- MACHINE LEARNING FOR FUTURE SYSTEM DESIGNS
- Three Risks in Building Machine Learning Systems
- Advantages of Adaptive AI Over Traditional Machine Learning Models
- ML Opening New Doors For FPGAs
- Rules of Machine Learning: Best Practices for ML Engineering
- Best Practices for creating training data
- Productionizing Behavioural Features for Machine Learning with Apache Spark Streaming
- TensorFlow Data Validation: Checking and analyzing your data

## Module 3: Designing High-performance ML Systems

- How to Evaluate the Performance of Your Machine Learning Model
- Best practices for performance and cost optimization for machine learning
- How To Improve Machine Learning Model Performance: Five Ways
- Distributed TensorFlow model training on Cloud AI Platform (TF Dev Summit '20)
- Distributed training with TensorFlow
- Speeding Up Neural Network Training with Data Echoing
- Machine Learning Performance Improvement Cheat Sheet
- Building a High-Performance Data Pipeline with Tensorflow 2.x



- Distributed training with TensorFlow
- AutoML Tables

## Module 4: Hybrid ML Systems

- Kubeflow
- Introduction to Kubeflow
- Orchestrating TFX Pipelines
- Introduction to Machine Learning Pipelines with Kubeflow
- Kubeflow — a machine learning toolkit for Kubernetes
- ML for Mobile and Edge Devices - TensorFlow Lite
- TensorFlow Lite Examples | Machine Learning Mobile Apps
- Optimize TensorFlow models for mobile and embedded devices
- The Essential Guide To Learn TensorFlow Mobile and Tensorflow Lite