

AWS Big Data Crash Course

Noah Gift, UC Davis & Northwestern Lecturer (Cloud, ML, AI), Founder @Pragmatic AI Labs, Author: Pragmatic AI

Day 1 Schedule

- **Part 1: AWS Machine Learning-Specialty (ML-S) Certification Overview, Collection, and Storage (90 min)**
- *QA (15 min)*
- *Break (15 min)*
- **Part 2: Processing for Big Data on AWS (45 min)**
- *QA (10 min)*
- *Break (5 min)*
- **Part 3: Analysis for Big Data on AWS (45 min)**
- *QA (15 min)*

Survey: Experience with AWS

- Novice (No experience)
- Beginner (< 1 Year)
- Intermediate (1-3 Years)
- Advanced (3+ Years)

Survey: Experience with Big Data

- Novice (No experience)
- Beginner (< 1 Year)
- Intermediate (1-3 Years)
- Advanced (3+ Years)

Part 1: AWS Machine Learning-Specialty (ML-S) Certification Overview, Collection, and Storage (90 min)

- Get an overview of the certification
- Use exam study resources
- Review the exam guide
- Learn the exam strategy
- Learn the best practices of Big Data on AWS
- Learn the techniques to accelerate hands-on practice
- Understand important Big Data related services
- Determine the operational characteristics of the collection system
- Determine and optimize the operational characteristics of the storage solution

The screenshot displays the AWS Lambda console interface. The main area shows the code for a function named `index.js`. The code is written in JavaScript and includes several functions: `LaunchRequest`, `GetNewFactIntent`, `GetFact`, and `AMAZON.HelpIntent`. The `GetFact` function is highlighted with a blue circle, and the `AMAZON.HelpIntent` function is highlighted with a red circle. The right sidebar shows the 'Environment Members' section, which lists 'You (online)', 'aaron (online)', and 'rob (online)'. Below this is the 'Group Chat' section, which displays a chat history. The chat history shows messages from 'You', 'aaron', and 'rob'. The chat interface is connected to a circular icon labeled 'ROB' at the top right and a circular icon labeled 'CLAIRE' at the bottom right. The 'ROB' icon is a black circle with a white face and a red circle with a white 'R' next to it. The 'CLAIRE' icon is a black circle with a white face and an orange circle with a white 'C' next to it. The chat history shows a conversation where 'You' asks 'Hey Aaron, can you jump in here quick and look at these variables?', 'aaron' responds 'Sure, looking now', 'aaron' says 'Ok, I've fixed the variables. Let's test it', 'You' says 'thanks, before testing i want to show it to Rob real quick', and 'rob' responds 'Looks ok. I don't see my Star trek facts though 😊'.

```
90 };
91
92 const handlersp = {
93   'LaunchRequest': function () {
94     this.emit('GetFact');
95   },
96   'GetNewFactIntent': function () {
97     this.emit('GetFact');
98   },
99
100 },
101
102 'GetFact': function () {
103   // Get a random space fact from the space facts list
104   // Use this.t() to get corresponding language data
105   const factArr = this.t('FACTS');
106   const factIndex = Math.floor(Math.random() * factArr.length);
107   const randomFact = factArr[factIndex];
108
109   // Create speech output
110   const speechOutput = this.t('GET_FACT_MESSAGE') + randomFact;
111   this.emit(':tellWithCard', speechOutput, this.t('SKILL_NAME'),
112 },
113 'AMAZON.HelpIntent': function () {
114   const speechOutput = this.t('HELP_MESSAGE');
115   const reprompt = this.t('HELP_MESSAGE');
116   this.emit(':ask', speechOutput, reprompt);
117 }
```

Environment Members

- You (online) RW
- aaron (online) RW
- rob (online) RW

Group Chat

Chat history is stored on the environment and can be both read and modified by ReadWrite members.

You 8 minutes ago
Hey Aaron, can you jump in here quick and look at these variables?

aaron 7 minutes ago
Sure, looking now

aaron 5 minutes ago
Ok, I've fixed the variables. Let's test it

You 5 minutes ago
thanks, before testing i want to show it to Rob real quick

rob 2 minutes ago
Looks ok. I don't see my Star trek facts though 😊

Enter your message here

ROB

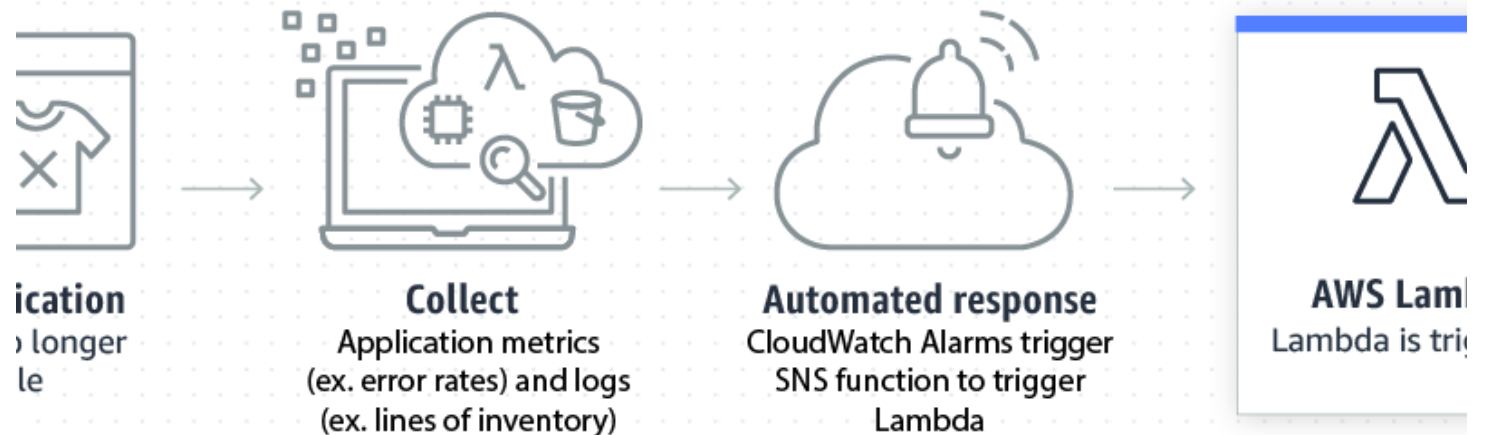
CLAIRE

QA (15 min)
Break (15 min)

QA & Break Part 1

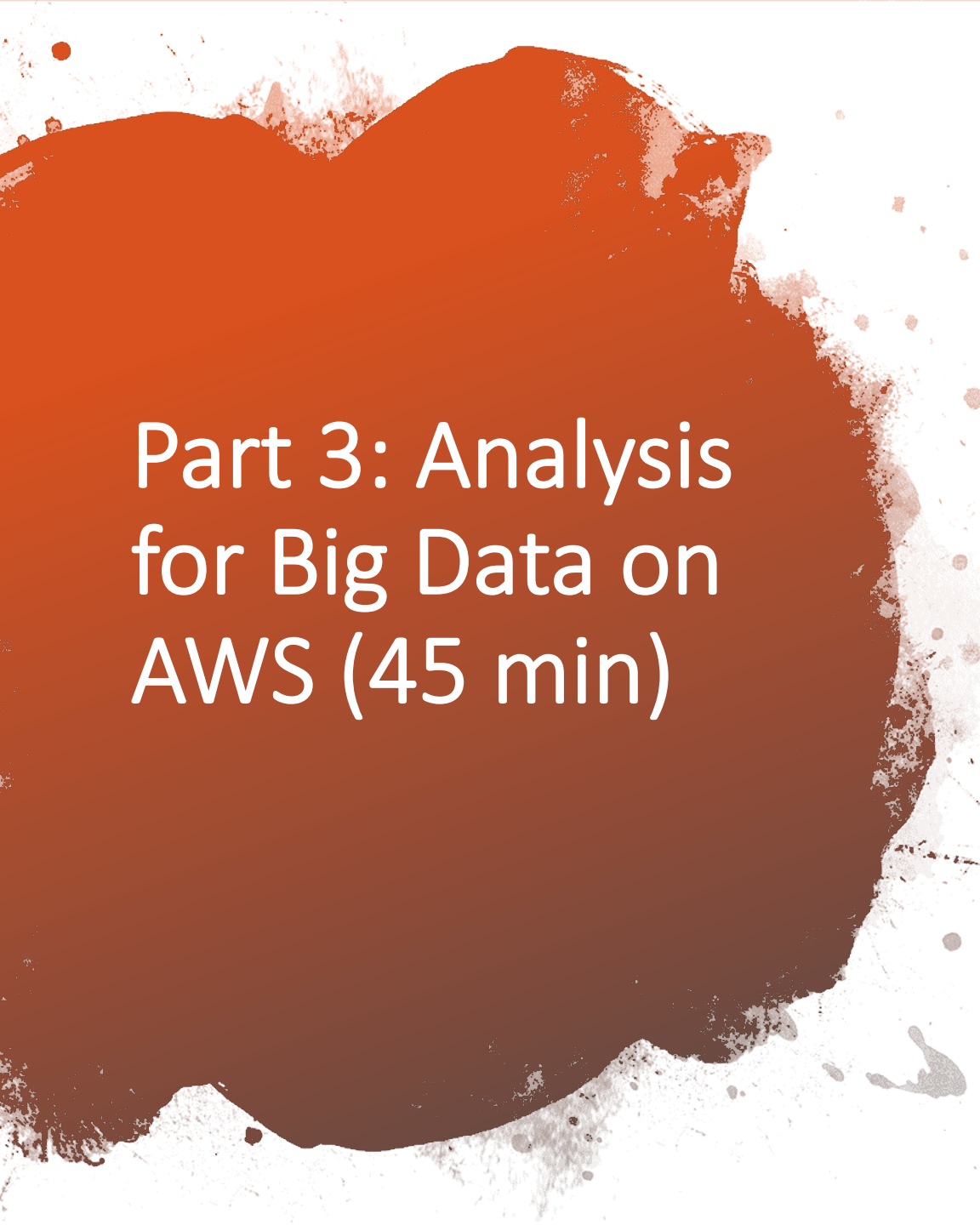
Part 2: Processing for Big Data on AWS (45 min)

- Identify the appropriate data processing technology for a given scenario
- Determine how to design and architect the data processing solution
- Determine the operational characteristics of the solution implemented
- Understand Overview of AWS Processing
- Understand Elastic MapReduce (EMR)
- Learn about Apache Hadoop - Intro
- Apply EMR - Architecture



QA (10 min)
Break (5 min)

QA and Break Part 2

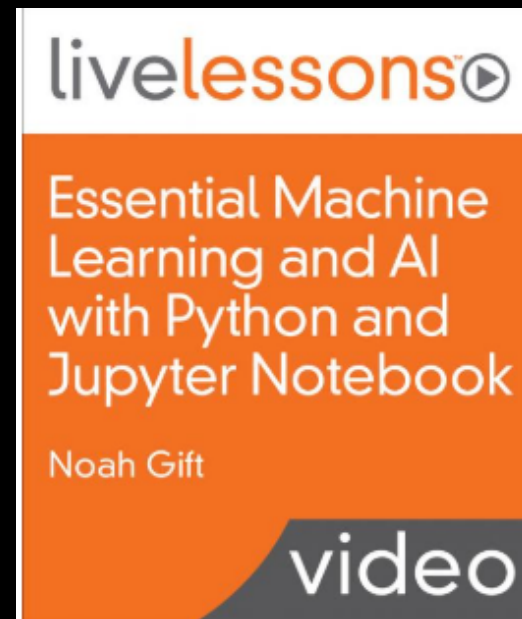
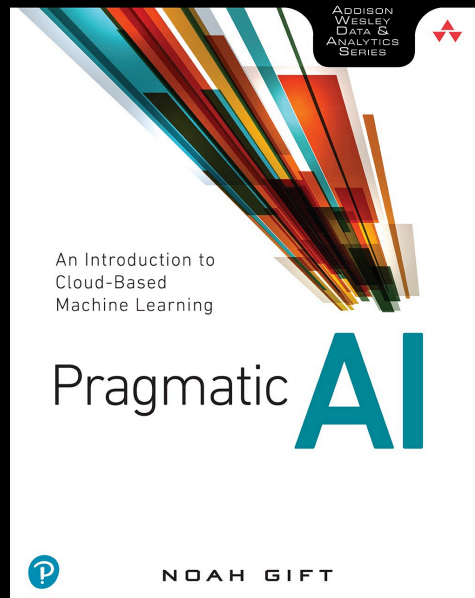
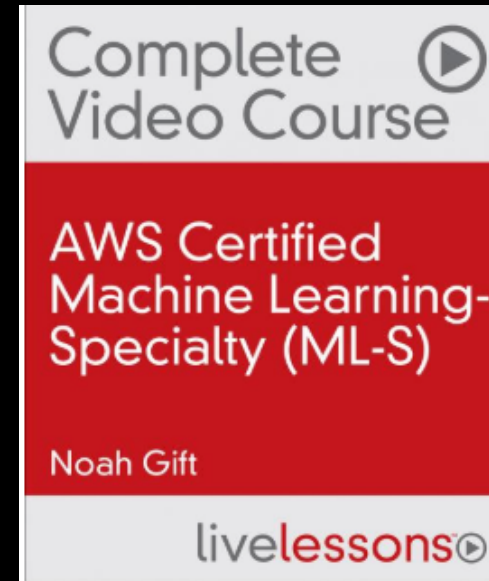
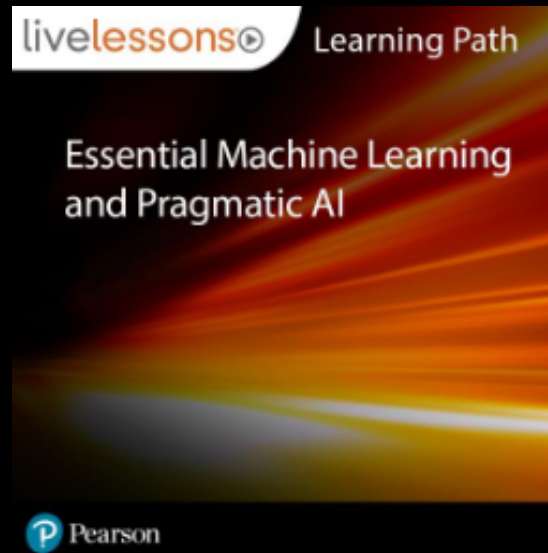
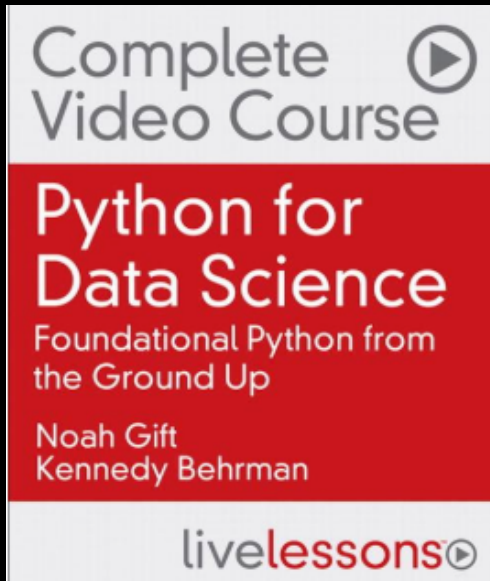
A large, textured orange watercolor splash shape on the left side of the slide, with various shades of orange and some darker spots, creating a dynamic, artistic background element.

Part 3: Analysis for Big Data on AWS (45 min)

- Determine the tools and techniques required for analysis
- Determine how to design and architect the analytical solution
- Determine and optimize the operational characteristics of the Analysis
- Understand Redshift Overview
- Learn Redshift Design
- Use Redshift Data Ingestion
- Apply Redshift Operations
- Use AWS Elasticsearch - operational analytics
- Implement Machine Learning - Clustering & Regression
- Use AWS Athena - interactive analytics

QA (10 min)

QA and Day 1 Wrap up



Related Safari Properties

- [Pragmatic AI \(Book\)](#)
- [Essential Machine Learning and AI \(Video\)](#)
- [AWS Certified Machine Learning-Specialty \(Video\)](#)
- [Essential Machine Learning and Pragmatic AI \(Learning Path\)](#)
- [Python for Data Science \(Video\)](#)
- [AWS Certified Big Data-Specialty \(Video\)](#)

Day 2 Schedule

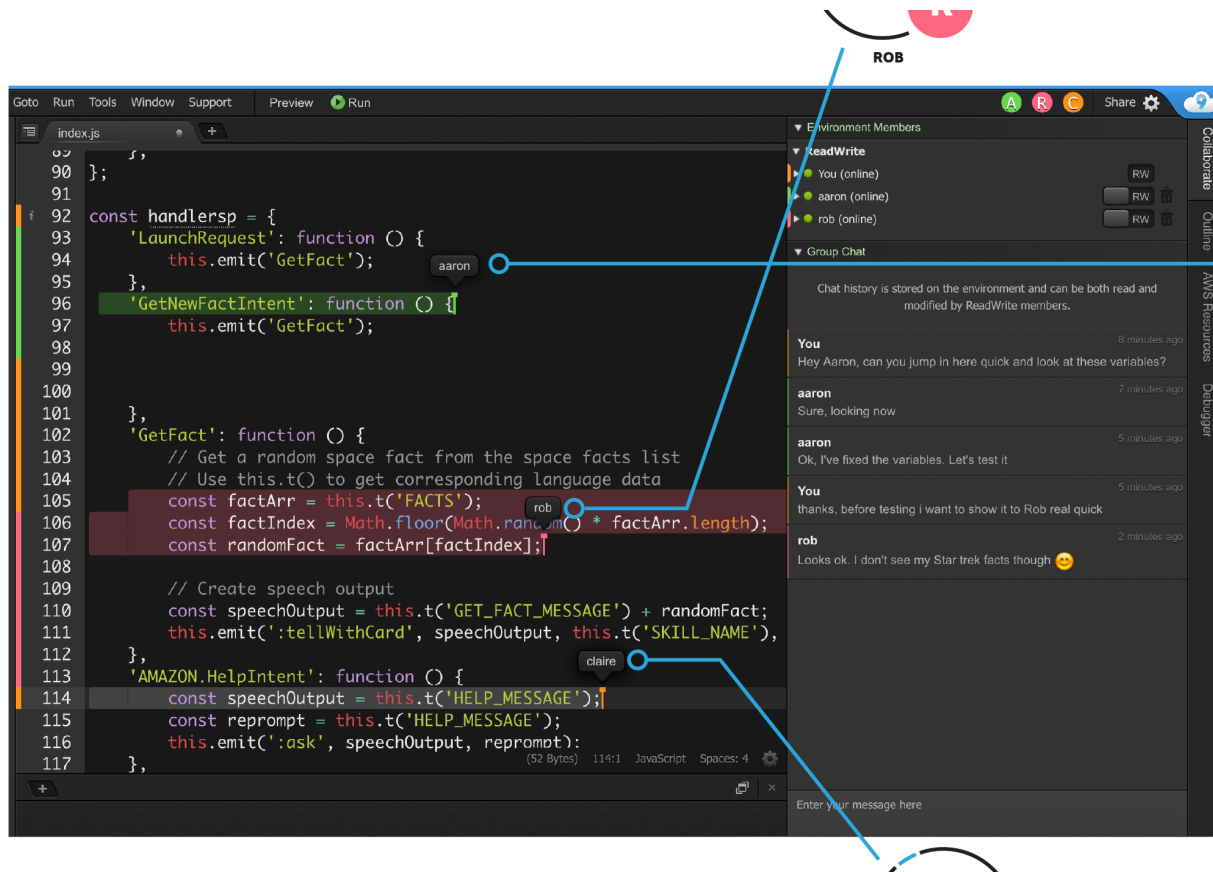
- **Part 4: Visualization & Data Security for Big Data on AWS (90 min)**
- *QA (15 min)*
- *Break (15 min)*
- **Part 5: Case Studies Part 1(45 min)**
- *QA (10 min)*
- *Break (5 min)*
- **Part 6: Case Studies Part 2 and Exam Sample Questions Review (45 min)**
- *QA (15 min)*

Survey: Experience with Visualization

- Novice (No experience)
- Beginner (< 1 Year)
- Intermediate (1-3 Years)
- Advanced (3+ Years)

Survey: Experience with Containers

- Novice (No experience)
- Beginner (< 1 Year)
- Intermediate (1-3 Years)
- Advanced (3+ Years)



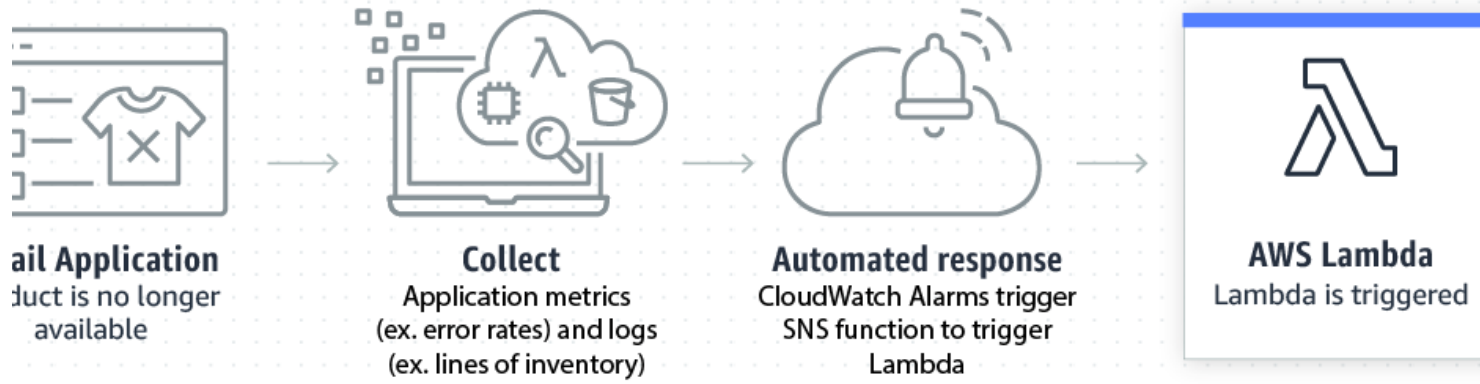
Part 4: Visualization & Data Security for Big Data on AWS

Length
(90 min)

- Determine the appropriate techniques for delivering the results/output
- Determine how to design and create the Visualization platform
- Determine and optimize the operational characteristics of the Visualization system
- Understand AWS Visualization - Overview
- Use AWS Quicksight - dashboards & visualizations
- Determine encryption requirements and/or implementation technologies
- Choose the appropriate technology to enforce data governance
- Identify how to ensure data integrity
- Evaluate regulatory requirements
- Implement AWS IAM
- Implement EMR Security
- Implement Redshift Security

QA (15 min)
Break (15 min)

QA & Break Part 1



- Understand Big Data for Sagemaker
- Learn Sagemaker and EMR Integration
- Learn Serverless Production Big Data Application Development

Part 2: Processing for Big Data on AWS (45 min)

QA (10 min)
Break (5 min)

QA and Break Part 2

Part 3: Case Studies Part 2 and Exam Sample Questions Review (45 min)



Implement Containerization for Big Data



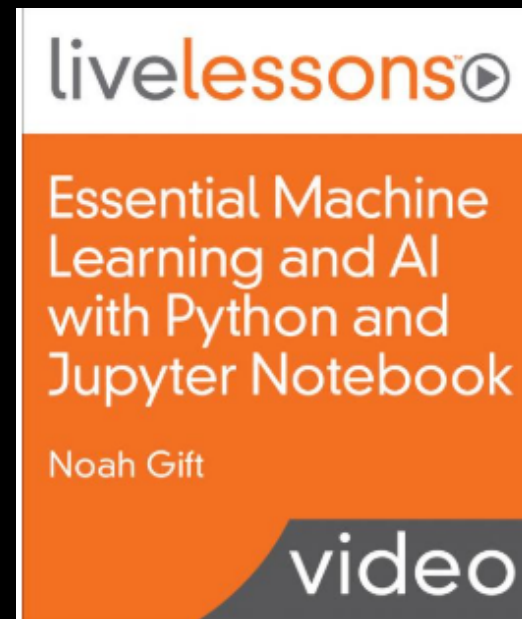
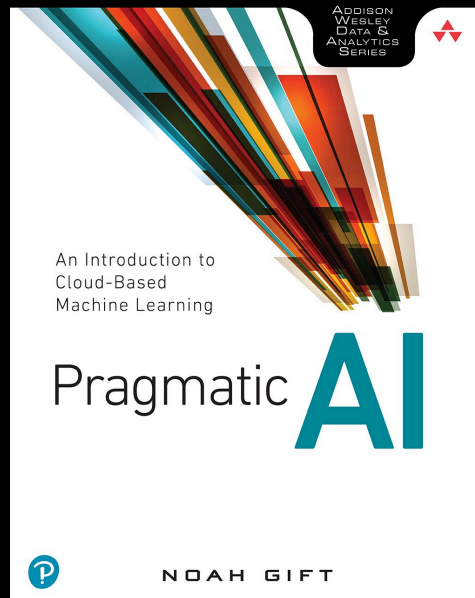
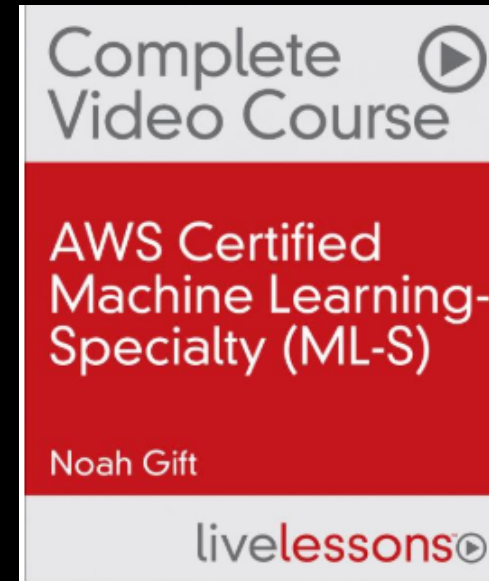
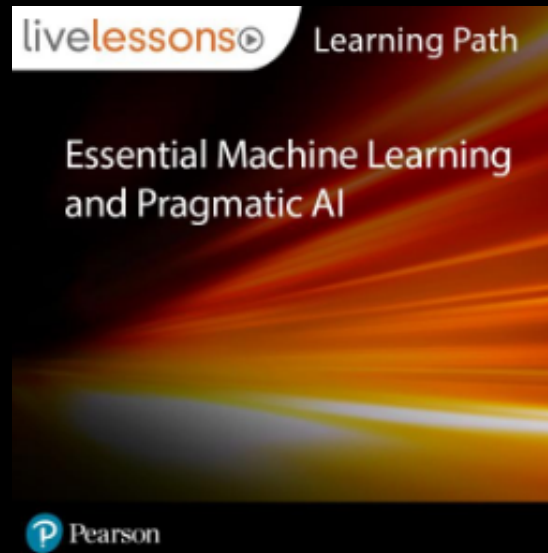
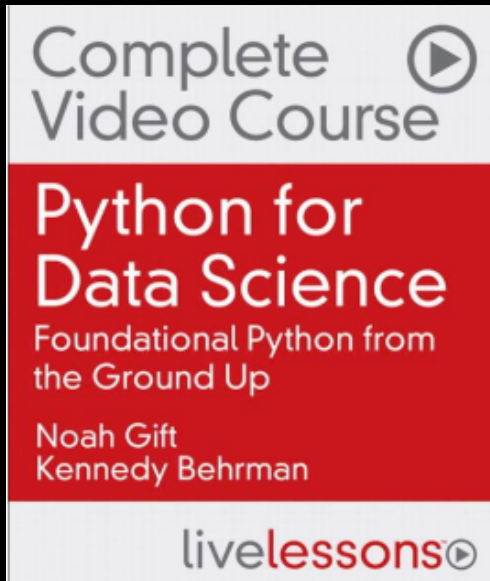
Implement Spot Instances for Big Data Pipeline



Exam Review

QA (10 min)

QA and Day 2 Wrap up



Related Safari Properties

- [Pragmatic AI \(Book\)](#)
- [Essential Machine Learning and AI \(Video\)](#)
- [AWS Certified Machine Learning-Specialty \(Video\)](#)
- [Essential Machine Learning and Pragmatic AI \(Learning Path\)](#)
- [Python for Data Science \(Video\)](#)
- [AWS Certified Big Data-Specialty \(Video\)](#)