

AWS re:INVENT

Machine Learning in Capital Markets

Frank Tarsillo, IHS Markit

Peter Williams, AWS Financial Services

November 28, 2017

AWS
re:Invent

© 2017 Amazon Web Services, Inc. or its Affiliates. All rights reserved.



Financial services companies are using machine learning to reduce fraud, streamline processes, and improve their bottom line. AWS provides tools that help them easily use AI tools like MXNet and Tensor Flow to perform predictive analytics, clustering, and more advanced data analyses. In this session, hear how IHS Markit has used machine learning on AWS to help global banking institutions manage their commodities portfolios. Learn how Amazon Machine Learning can take the hassle out of AI.

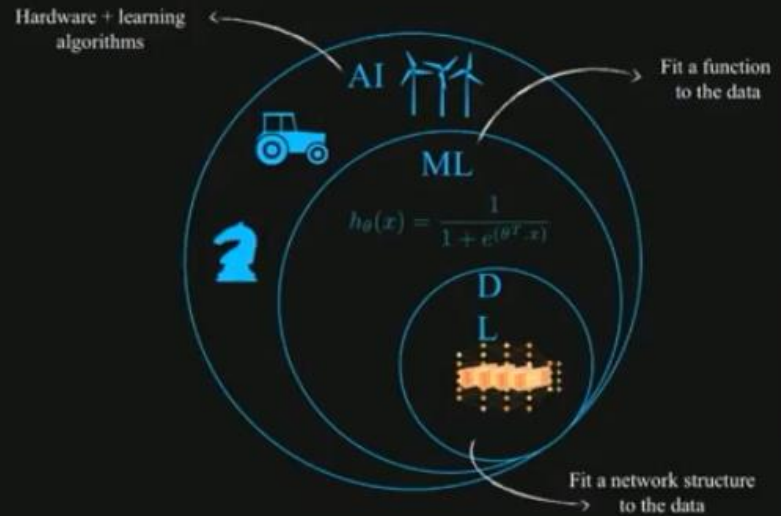
Takeaways from today's session

1. How is machine learning changing financial services?
2. Case study: IHS Markit's commodity inventory solution
3. How do customers get started?
4. Best practices when building machine learning in the cloud

How is machine learning changing financial services?

Artificial intelligence

A system or service that can perform tasks that usually require human intelligence

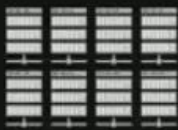


AWS
re:Invent



The challenge:

Scale



Data

PBs and EBs of existing data
Aggressive migration



Training

Tons of GPUs
Elastic capacity
Pre-built images



Prediction

Tons of GPUs and CPUs
Serverless
At the edge,
on IoT devices

Financial services use cases

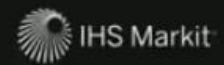
Business group	Use case	ML models	AWS tools
Asset management	Equity ratings Portfolio management Portfolio optimization	Machine learning (clustering, SVM, logistic regression, Gaussian processes)	Amazon Machine Learning Deep Learning MXNet SparkML on Amazon EMR
Banking	Credit card/lending fraud Customer onboarding Digital banking	Deep learning (embedding, matrix factorization, LSTM, conversational UI)	TensorFlow
Compliance	Anti-money laundering, Counter-terrorism financing Due diligence	Graphical models	
Security	Cyber-attack detection Data leak detection Virus detection		

AWS has become the center of gravity for AI



Case study: Commodity Tracker

Commodity inventory challenges



Global warehouses are all different

Diverse flow of paper documents, from warehouse inventory reports to bills of lading

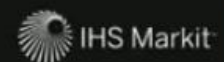
Digitize, classify, and aggregate the document flow

Improve cost and risk management

Localization diversity

Eliminate error-prone manual operations

Commodity inventory solution



Vendors

IHS Markit

Trading firms

Report generation

Vendors continue to generate their reports as usual. Documents are redirected to a Markit server.

Process

Markit scrapes the critical data from the warehouse's unique report using a variety of technologies, including OCR.

Audit

Markit stores a digital copy of the original document for auditing purposes.

Disseminate

The data is standardized for language and terminology, then compiled onto a customer portal.

Data access

Dependent teams at the trading firms connect to the portal to pull a personalized view of the required data.

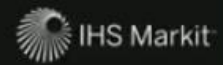
AWS re:Invent



© 2017 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Commodity Tracker app is used to collect, scan, analyze the documents, digitize all the paper data in a structured way for all the different locales, and then provide a service out to our customer in terms of aggregation and commissioning.

IHS Markit—cloud strategy



Cloud-first strategy

Cloud native vs. lift and shift

Time-to-market focus

Investment upfront

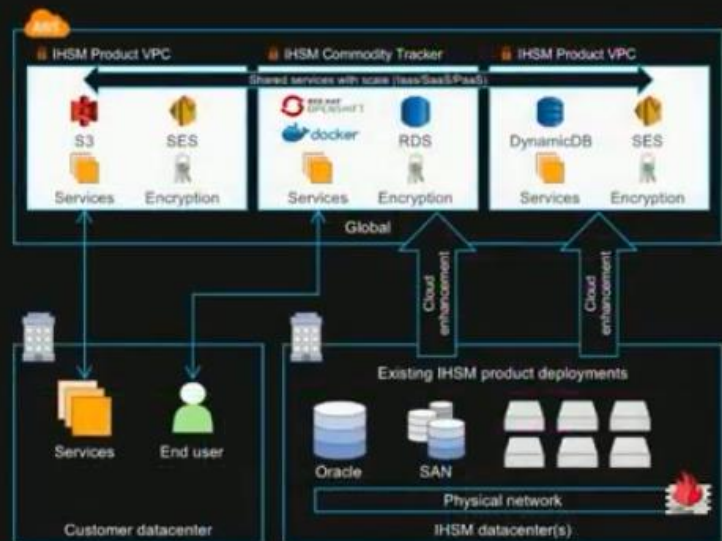
Refreshing the estate

Financial transparency

A DevOps world

Global scale

Cost-savings in the end



AWS re:Invent

© 2017, Amazon Web Services, Inc. or its Affiliates. All rights reserved.



Architecture approach



Microservices design approach

PaaS supporting orchestration and discovery

OCR, classification, and analysis documents

Cloud enablement with some lock-in

API-driven

Modern UI, SSO, preferences, notifications

SDLC-Agile, CI/CD, high velocity of change

Highly available (multi-AZ/global)

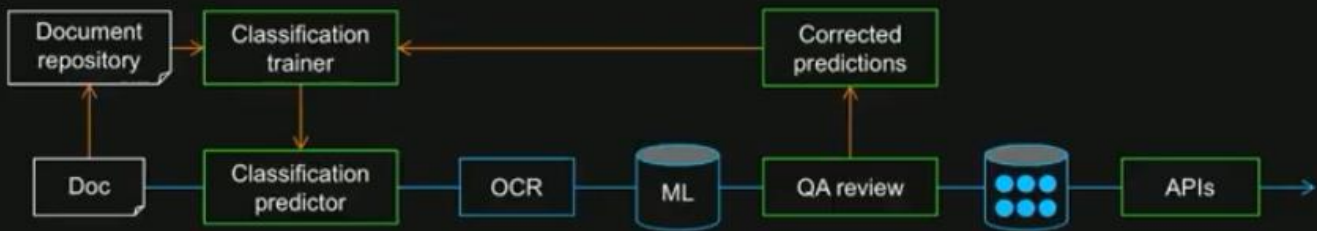
```
public void processData()
{
    do
    {
        int data = getData();
        if(data < 0)
    }
```



GitLab



Machine learning enablement



Document classification and templates to reduce exceptions

Configuration data and models (Keras + TensorFlow)

Use of AML-based images to tool vs. DIY

AWS compute with GPU on-demand for trainer

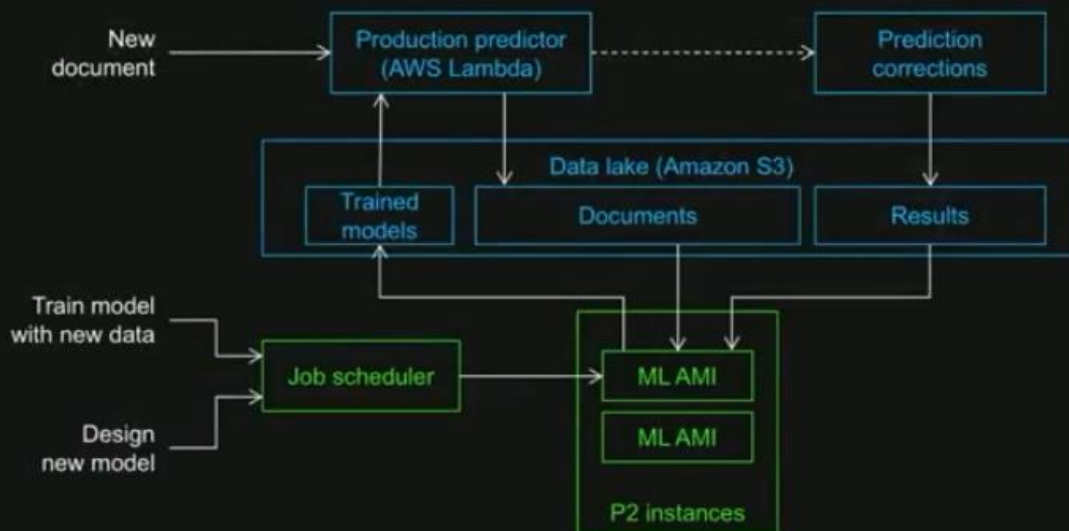
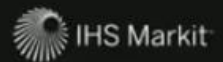
Data lake of document classifications (strategic)

AWS
re:Invent



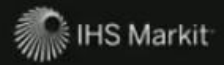
Commodity Tracker takes documents as paper and run them through a machine-enabled document classification system that does a preliminary OCR scan run of the documents like inventory receipts/bills of ladens, commodity description document, etc. We simply classify what a scanned document is in this first step. We simply run this document through a Classification Predictor that we also have to train continuously using a trained model of all the different documents that we can possibly scan. We create a model that we can then use in our predictor system to further classify newly scanned documents. Further downstream, we have another moder and predictor that digitizes an already scanned record document depending on the document's classification like 'Bill of Laden', this will get a 'Bill of Laden' ML model applied to it to extract the details of the document into a suitable format using the correct template.

ML product maintenance



We are feeding the documents and all the trained models into the Data lake.

Commodity tracker benefits



80%

of docs STP

Focus human effort
on the exceptions

60%

improvement
in timeliness

Reduce risk by cutting
the lag between data
collection and use

50%

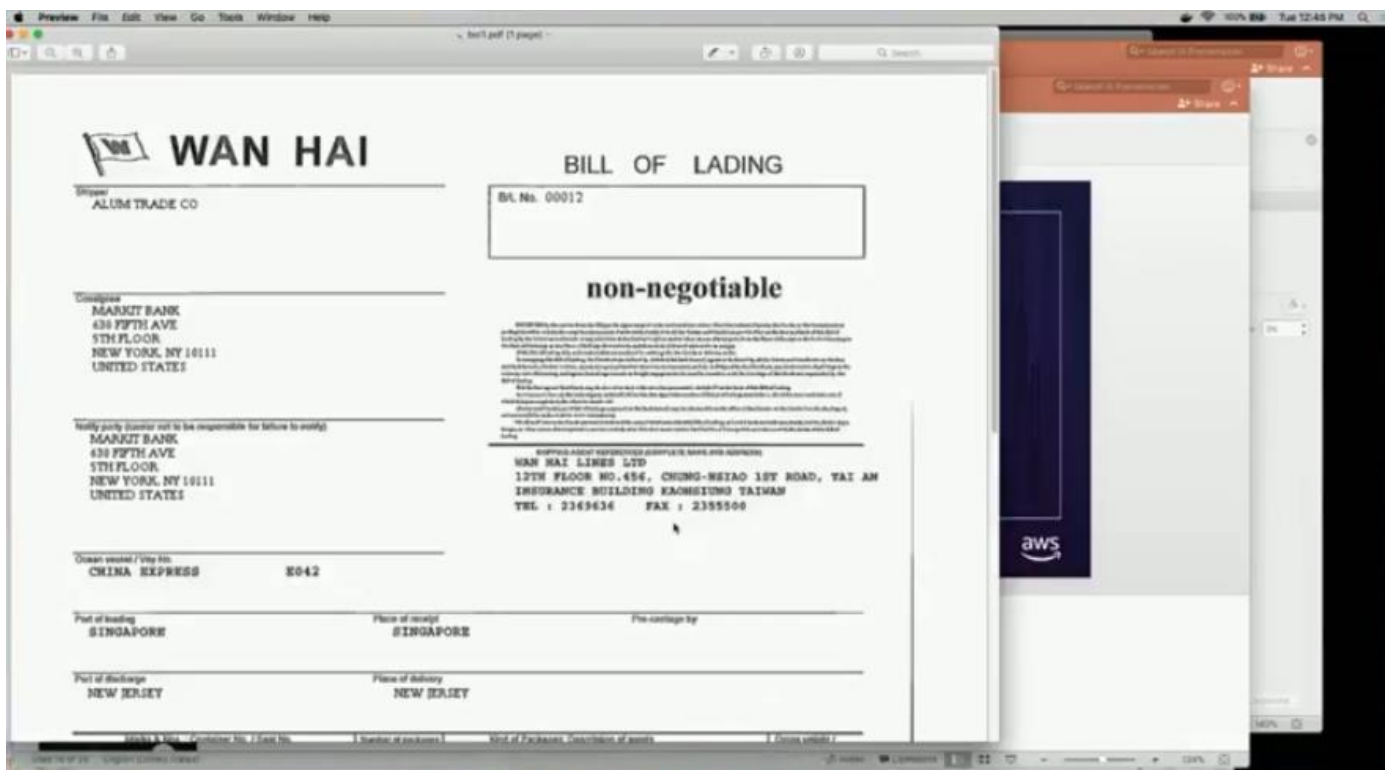
increase in
efficiency

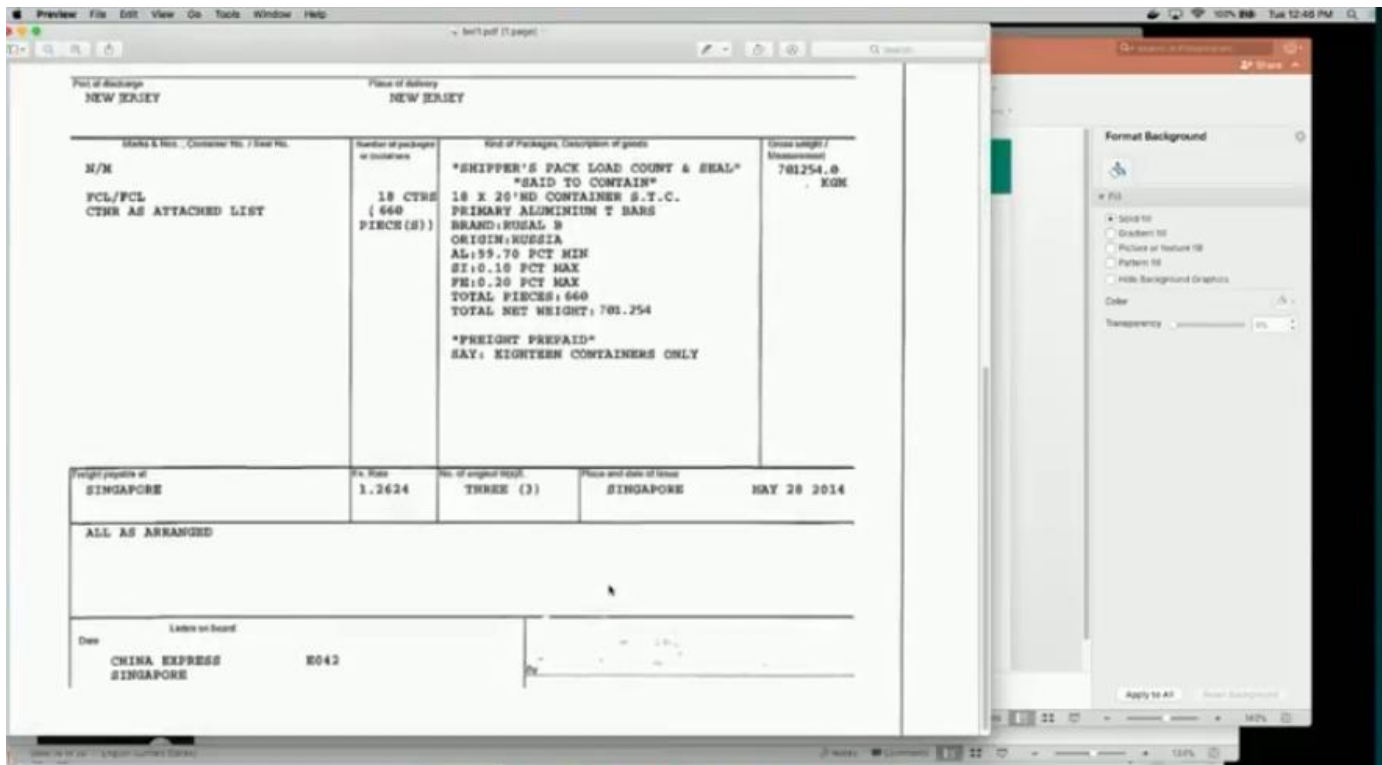
Empower teams
to focus energy on
value-creation activities

AWS
re:Invent

© 2017 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

A Quick Demo





Mark & No. / Seal No.	Number of packages or containers	Kind of Packages; Description of goods	Gross weight / Measurement
ST	18 CTRS (660 PIECE(S))	"SHIPPER'S PACK LOAD COUNT & SEAL" "SAID TO CONTAIN" 18 X 20'HD CONTAINER S.T.C. PRIMARY ALUMINIUM T BARS BRAND:RUSAL B ORIGIN:RUSSIA AL:99.70 PCT MIN SI:0.10 PCT MAX FE:0.20 PCT MAX TOTAL PIECES: 660 TOTAL NET WEIGHT: 701.254 "FREIGHT PREPAID" SAY: EIGHTEEN CONTAINERS ONLY	701254.0 KGM
Ex. Rate 1.2624		No. of original ft(s) THREE (3)	Place and date of issue SINGAPORE MAY 28 2014

The system scans this document and digitizes it as records in the database.

The diagram illustrates a document processing pipeline using AWS services. At the top, 'Unclassified Documents' are represented by icons for PNG, PDF, X (Excel), JPG, and W (Word). These documents enter a central processing block. The pipeline consists of several services and steps:

- Doc workflow service**: The entry point for document processing.
- Doc repository services**: A central hub containing:
 - Upload/download**: For moving documents in and out of storage.
 - Search by tags**: For finding documents based on metadata.
 - Extract text**: For pulling text content from documents.
- Doc utils services**: Contains 'Split files' and 'Convert to pdf'.
- Model services**: Contains 'Get doc type (prediction)' and is associated with the 'OPENSHIFT' logo.
- Lambda services**: Includes 'Doc virus scan' and 'OCR service'.
- Document metadata store**: Stores document metadata, represented by an Amazon S3 icon.
- Document Store**: Stores the document content, represented by a red puzzle piece icon.

Arrows indicate the flow of data and control between these components. The final output is a '<prediction>'.

This is the Document Classification application/sub-system that we run as a set of microservices.

The diagram illustrates the document classification pipeline. It starts with a blue box labeled "Upload Document". A green arrow points to a box titled "Document Metadata" containing the following information:

- Date: 11/10/2017
- DocExt: JPG
- Sub: Classifix
- Usage: General
- IsPublic: No

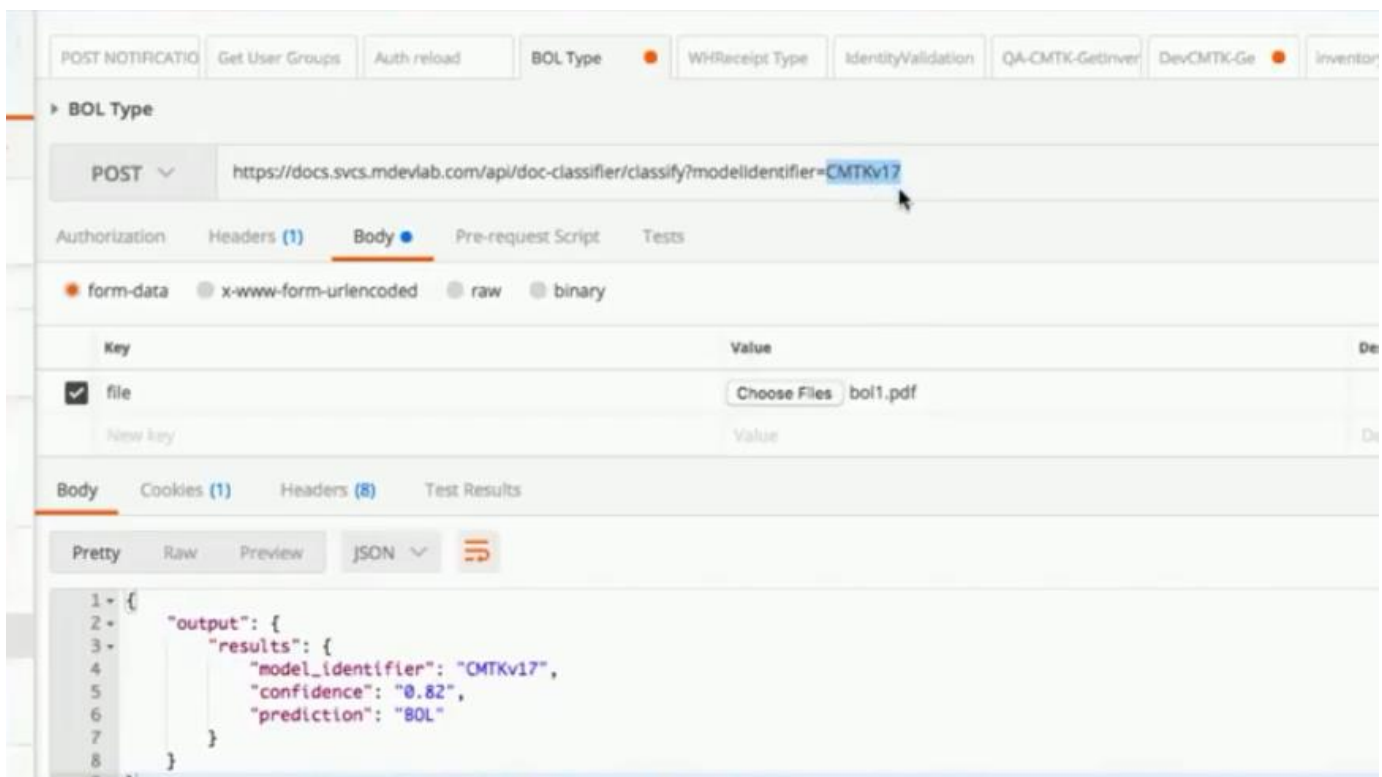
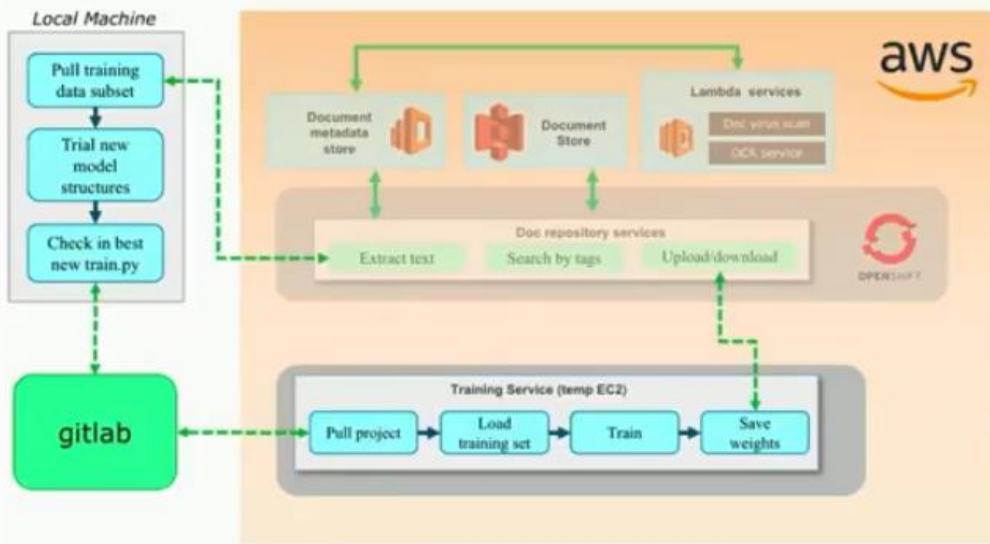
A second green arrow points to a box titled "OCR Results" containing the following text:

Long Form Bill of Lading; Received subject to the classification and tariffs in effect on the date of the issue of the bill of lading the property described below in apparent good order except as noted (contents and condition of contents of packages unknown) marked consigned and destined as indicated below, which said carrier (the word carrier being understood throughout

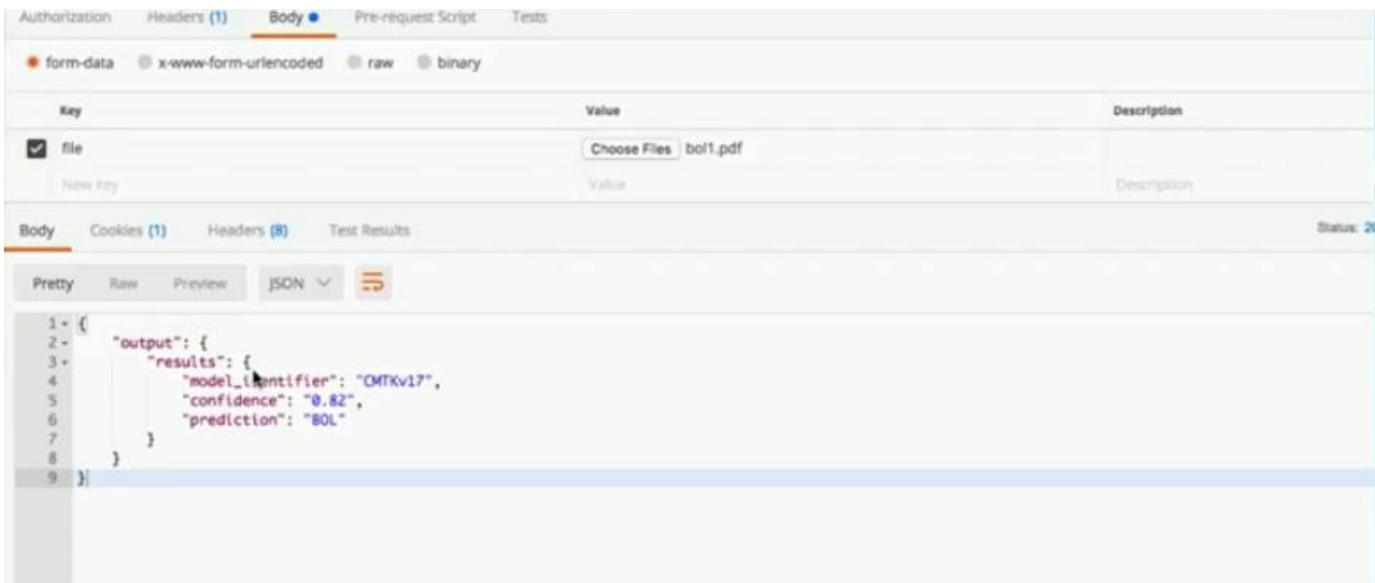
A third green arrow points to a box titled "Prediction results" containing the following list:

- 0.02 Bill of Lading
- 0.07 Receipt
- 0.00 Invoice
- 0.91 Cert of Analysis
- 0.00 Misc

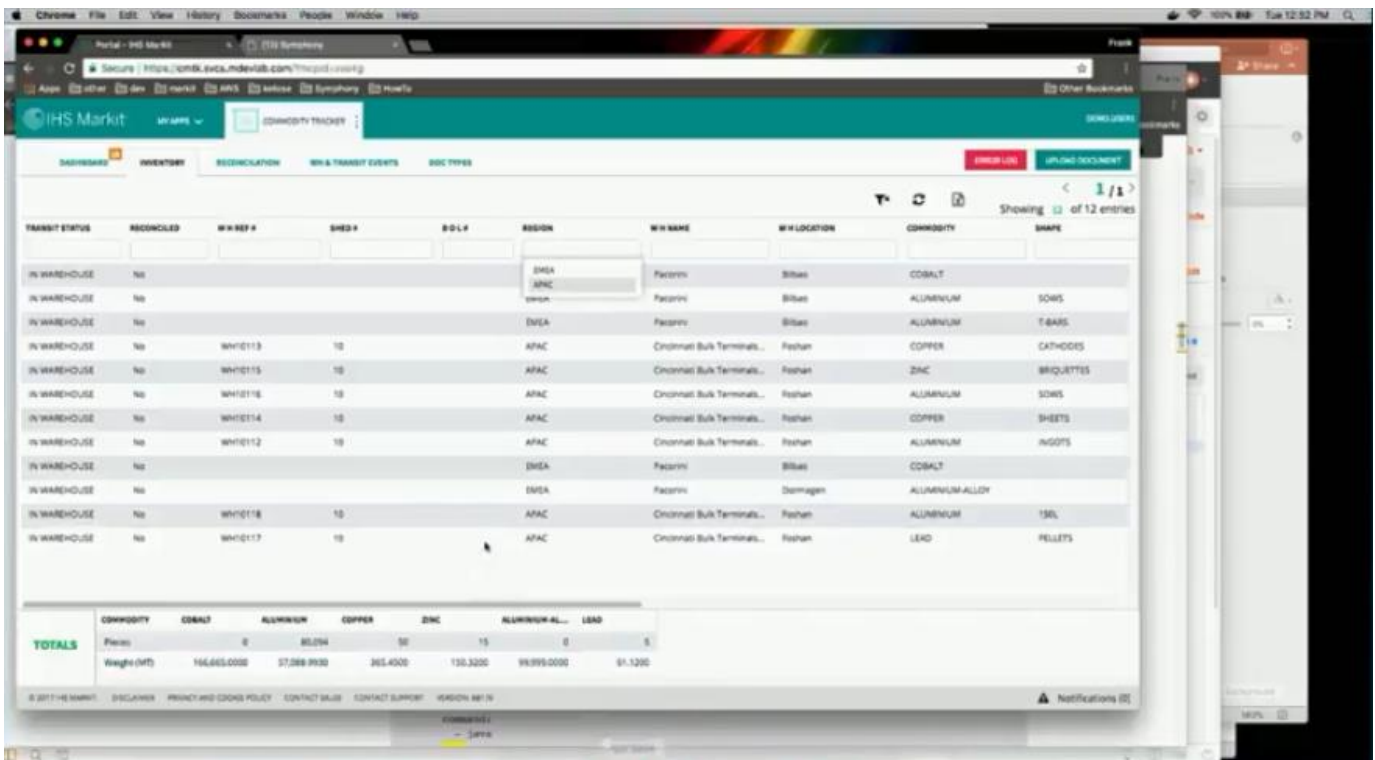
Training and extending models from the pipeline



We are using the CMTKv17 model for this classification task and pass it the bol1.pdf document to scan and digitize with the earlier microservices processes



We then get the classification/prediction process. The CMTKv17 model predictor returns a result is 82% accuracy that the document we just scanned is a bill of lading BOL. We then take this result and send it downstream to be further processed and into the database for storage.



This is the commodity tracker product

INVENTORY

RECONCILIATION

WH & TRANSIT EVENTS

DOC TYPES

ERROR LOG

</

EHOUSE	No			EMEA	Pacorini	Bilbao
EHOUSE	No			EMEA	Pacorini	Bilbao
EHOUSE	No	WH10113	10	APAC	Cincinnati Bulk Terminals...	Foshan
EHOUSE	No	WH10115	10	APAC	Cincinnati Bulk Terminals...	Foshan
EHOUSE	No	WH10116	10	APAC	Cincinnati Bulk Terminals...	Foshan
EHOUSE	No	WH10114	10	APAC	Cincinnati Bulk Terminals...	Foshan
EHOUSE	No	WH10112	10	APAC	Cincinnati Bulk Terminals...	Foshan
EHOUSE	No			EMEA	Pacorini	Bilbao
EHOUSE	No			EMEA	Pacorini	Dormagen
EHOUSE	No	WH10118	10	APAC	Cincinnati Bulk Terminals...	Foshan
EHOUSE	No	WH10117	10	APAC	Cincinnati Bulk Terminals...	Foshan

COMMODITY	COBALT	ALUMINIUM	COPPER	ZINC	ALUMINIUM-AL...	LEAD
Pieces	0	80,094	50	15	0	5
Weight (MT)	166,665.0000	57,088.9930	365.4500	150.3200	99,999.0000	61.1200

IHS MARKET, [DISCLAIMER](#), [PRIVACY AND COOKIE POLICY](#), [CONTACT SALES](#), [CONTACT SUPPORT](#), VERSION: 88176

command 1

Chrome File Edit View History Bookmarks People Window Help

Portal - IHS Market

Secure: https://portal.ihs.com/.../index.html

IHS Market

COMMODITY TRACKER

DEMO USER

DESKTOP INVENTORY RECONCILIATION NEW & TRADING EVENTS DOC TYPES EDITOR LINK UPLOAD DOCUMENT

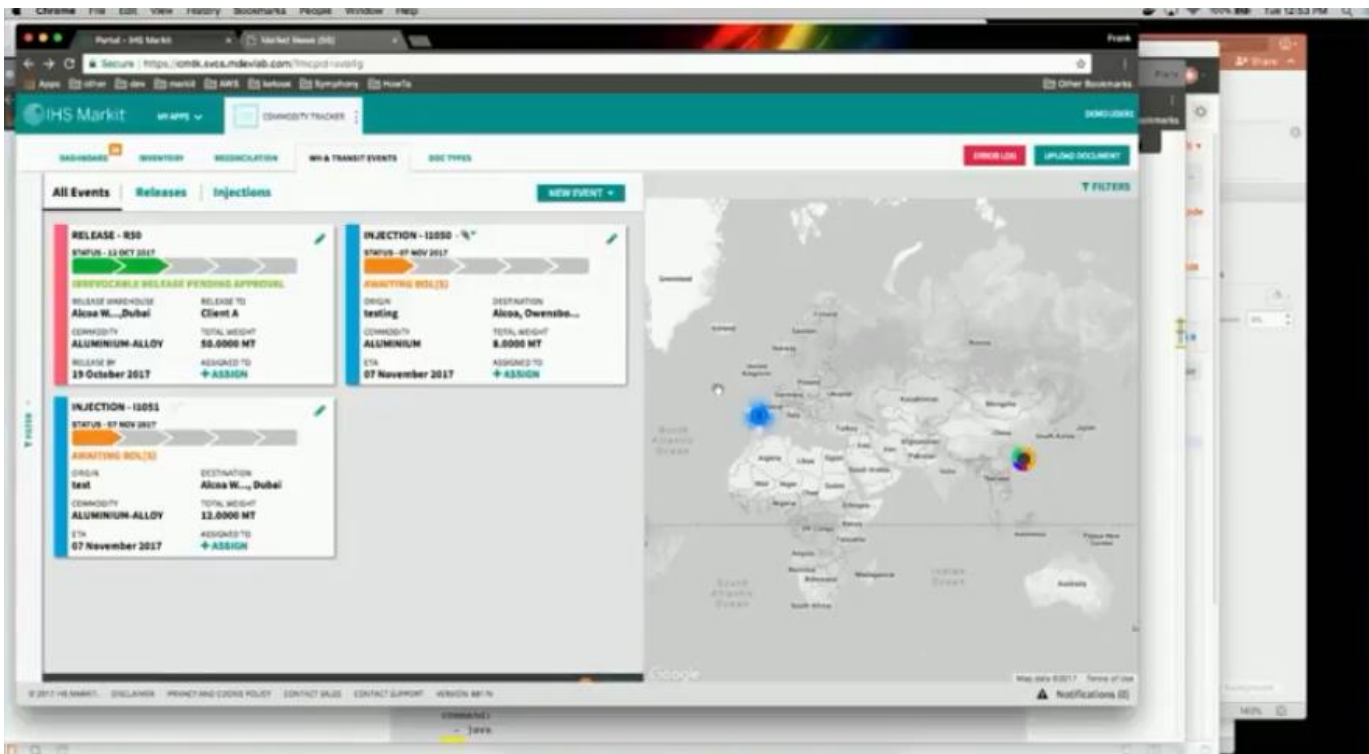
DOCUMENT NAME DATE/TIME RECEIVED FILE TYPE DOCUMENT TYPE

No Document Found.

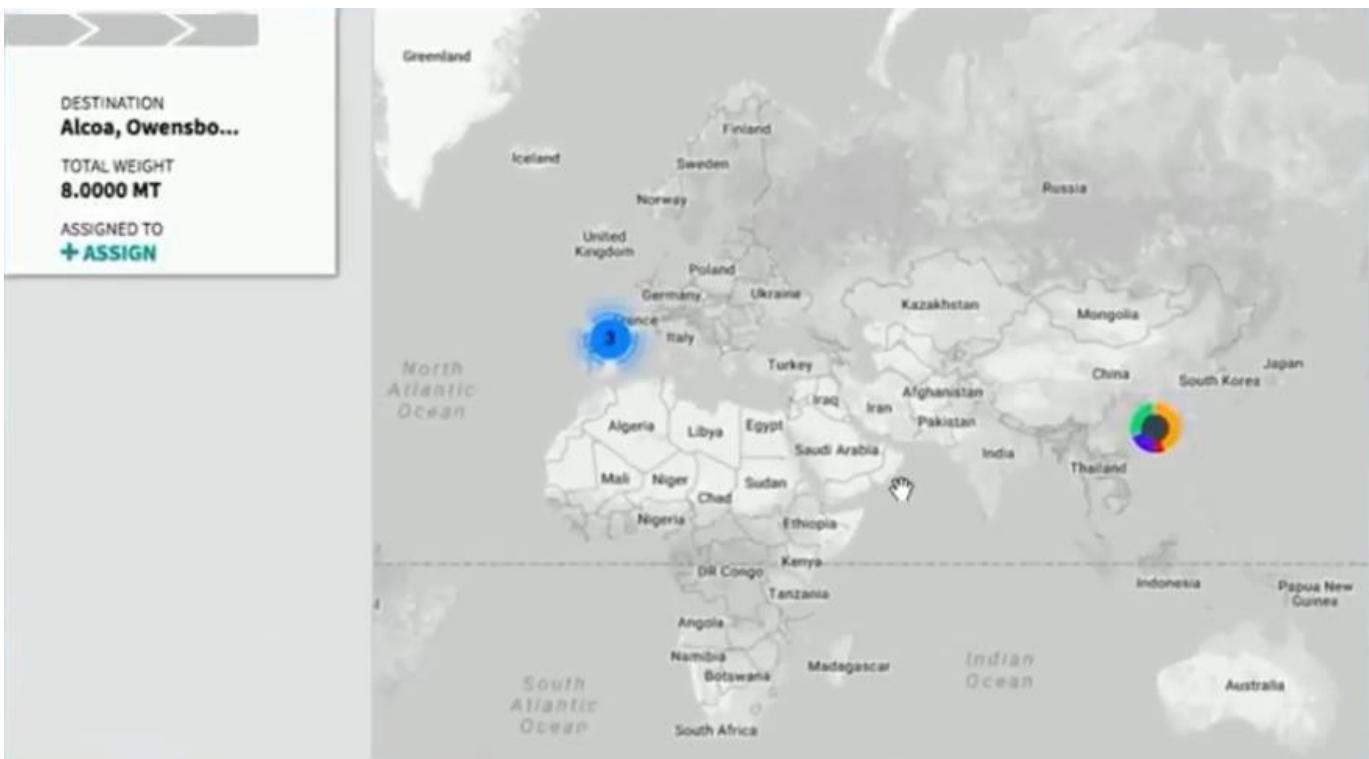
© 2017 IHS MARKIT, [DISCLAIMER](#), [PRIVACY AND COOKIE POLICY](#), [CONTACT SALES](#), [CONTACT SUPPORT](#), VERSION: 88176

command 1

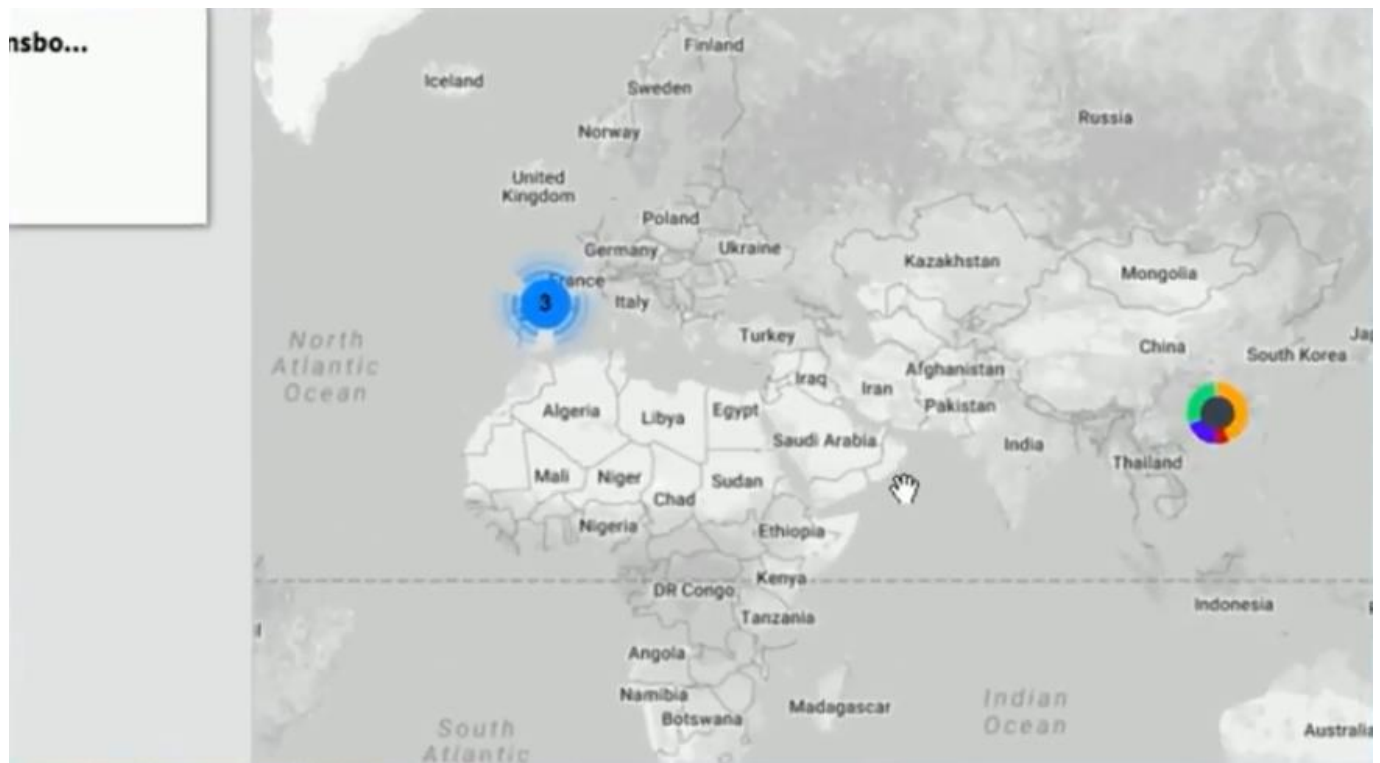
Notifications (0)



We can now search for the data of how much physical commodities we have in various warehouse locations worldwide based of real time information including the result of the new document we just scanned in.



nsbo...



SECTION - 11050 -

US - 07 NOV 2017

LOADING BOL(S)

IN **ing**

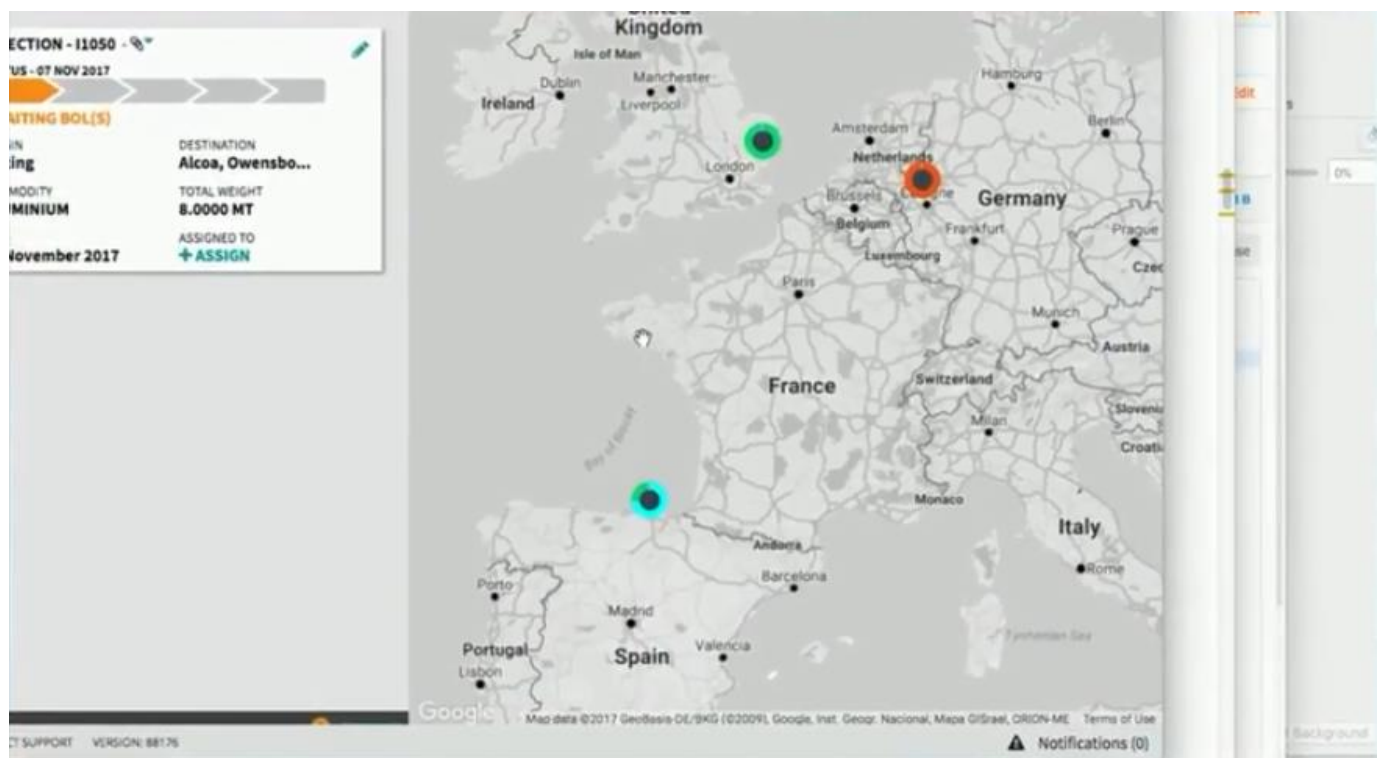
MODITY **MINIUM**

November 2017

DESTINATION **Alcoa, Owensbo...**

TOTAL WEIGHT **8.0000 MT**

ASSIGNED TO **+ASSIGN**



CT SUPPORT VERSION: 88176

Notifications (0)

INJECTION - 11050 -

US - 07 NOV 2017

AWAITING BOL(S)

ORIGIN
ing

MODIFY
MINIMUM

07 November 2017

DESTINATION
Alcoa, Owensbo...

TOTAL WEIGHT
8.0000 MT

ASSIGNED TO
[+ ASSIGN](#)

Ports America - Totals

COMMODITY	SHAPE	BRAND	PIECES	WEIGHT (M T)
ALUMINIUM			1,156	1,138.7872

NOTIFICATIONS (0)

Window Help

Frank

100% 88% Tue 12:51

Frank

Other Bookmarks

DEMO USER

ERROR LOG

UPLOAD DOCUMENT

TRANSIT EVENTS

DOC TYPES

NEW EVENT +

INJECTION - 11050 -

STATUS - 07 NOV 2017

AWAITING BOL(S)

ORIGIN
testing

COMMODITY
ALUMINIUM

ETA
07 November 2017

DESTINATION
Alcoa, Owensbo...

TOTAL WEIGHT
8.0000 MT

ASSIGNED TO
[+ ASSIGN](#)

Ports America - Totals

story Bookmarks People Window Help

k.svcs.mdevlab.com/mcpldview4g

COMMODITY TRACKER

RECONCILIATION WH & TRANSIT EVENTS DOC TYPES

INJECTIONS

NEW EVENT

INJECTION - 11050

STATUS - 07 NOV 2017

AWAITING BOL(S)

ORIGIN testing DESTINATION Alcoa, Owensbo...

COMMODITY ALUMINIUM TOTAL WEIGHT 8.0000 MT

ETA 07 November 2017 ASSIGNED TO

RELEASE TO Client A

TOTAL WEIGHT 50.0000 MT

ASSIGNED TO

DESTINATION Alcoa W..., Dubai

TOTAL WEIGHT 12.0000 MT

ASSIGNED TO

Google Maps

Ports America - Totals

COMMODITY	SHAPE	BRAND	PIECES	WEIGHT (MT)
ALUMINIUM			1,156	1,136.7672

story Bookmarks People Window Help

k.svcs.mdevlab.com/mcpldview4g

COMMODITY TRACKER

RECONCILIATION WH & TRANSIT EVENTS DOC TYPES

INJECTIONS

NEW EVENT

INJECTION - 11050

STATUS - 07 NOV 2017

AWAITING BOL(S)

ORIGIN testing DESTINATION Alcoa, Owensbo...

COMMODITY ALUMINIUM TOTAL WEIGHT 8.0000 MT

ETA 07 November 2017 ASSIGNED TO

RELEASE TO Client A

TOTAL WEIGHT 50.0000 MT

ASSIGNED TO

DESTINATION Alcoa W..., Dubai

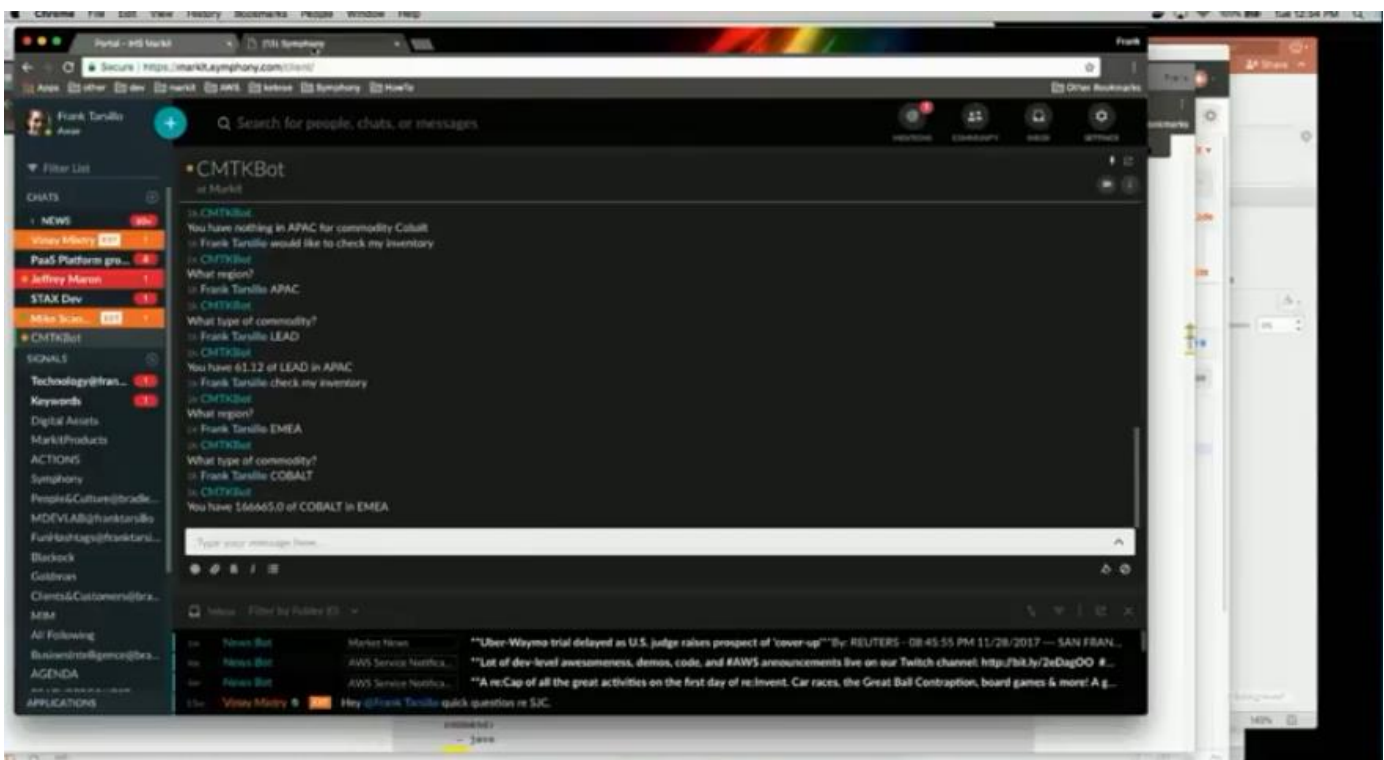
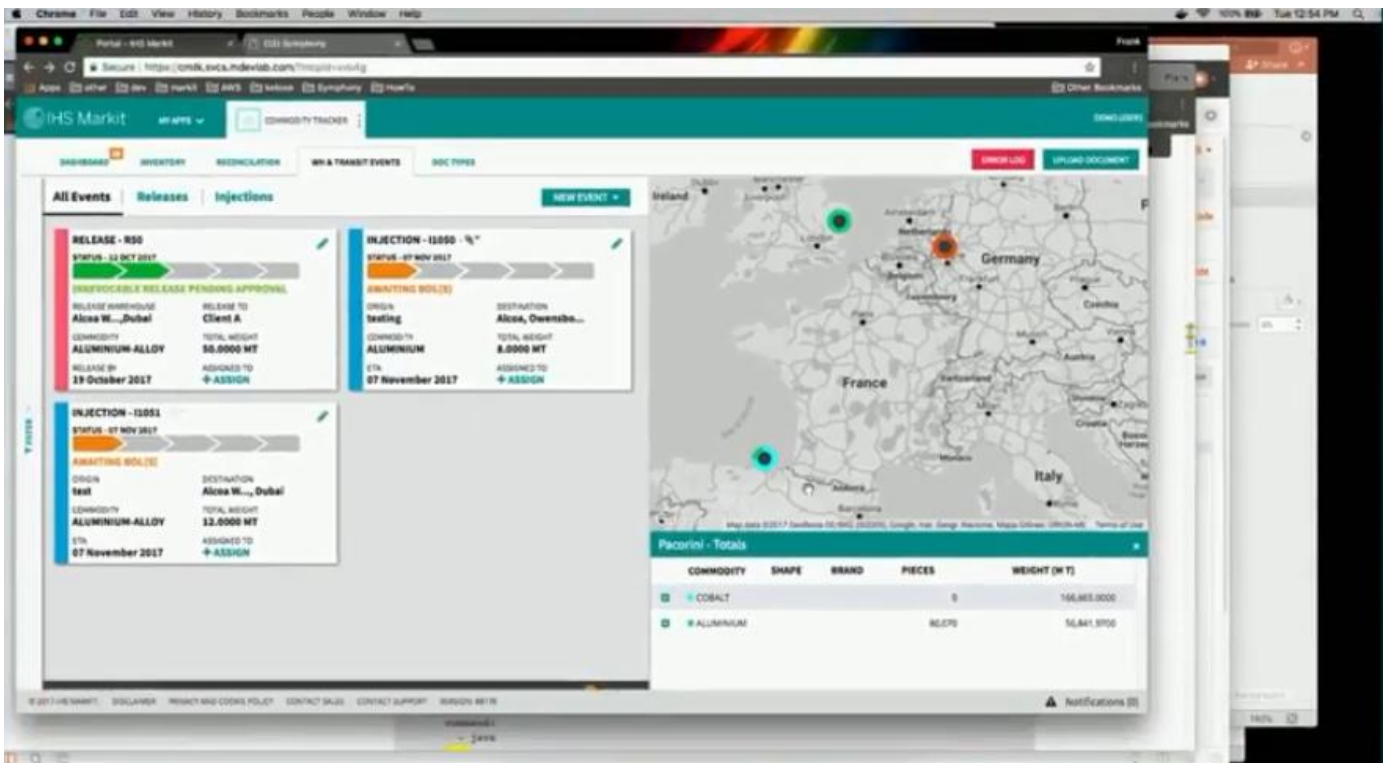
TOTAL WEIGHT 12.0000 MT

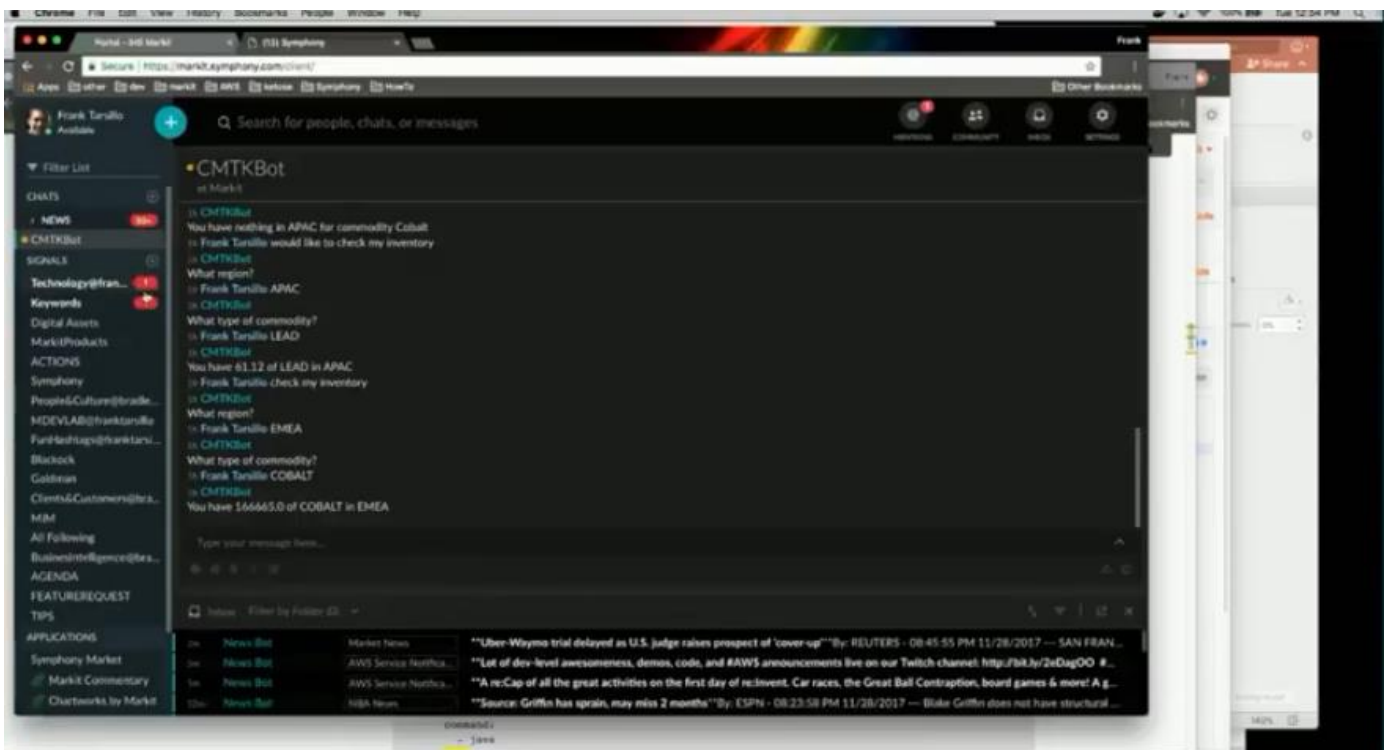
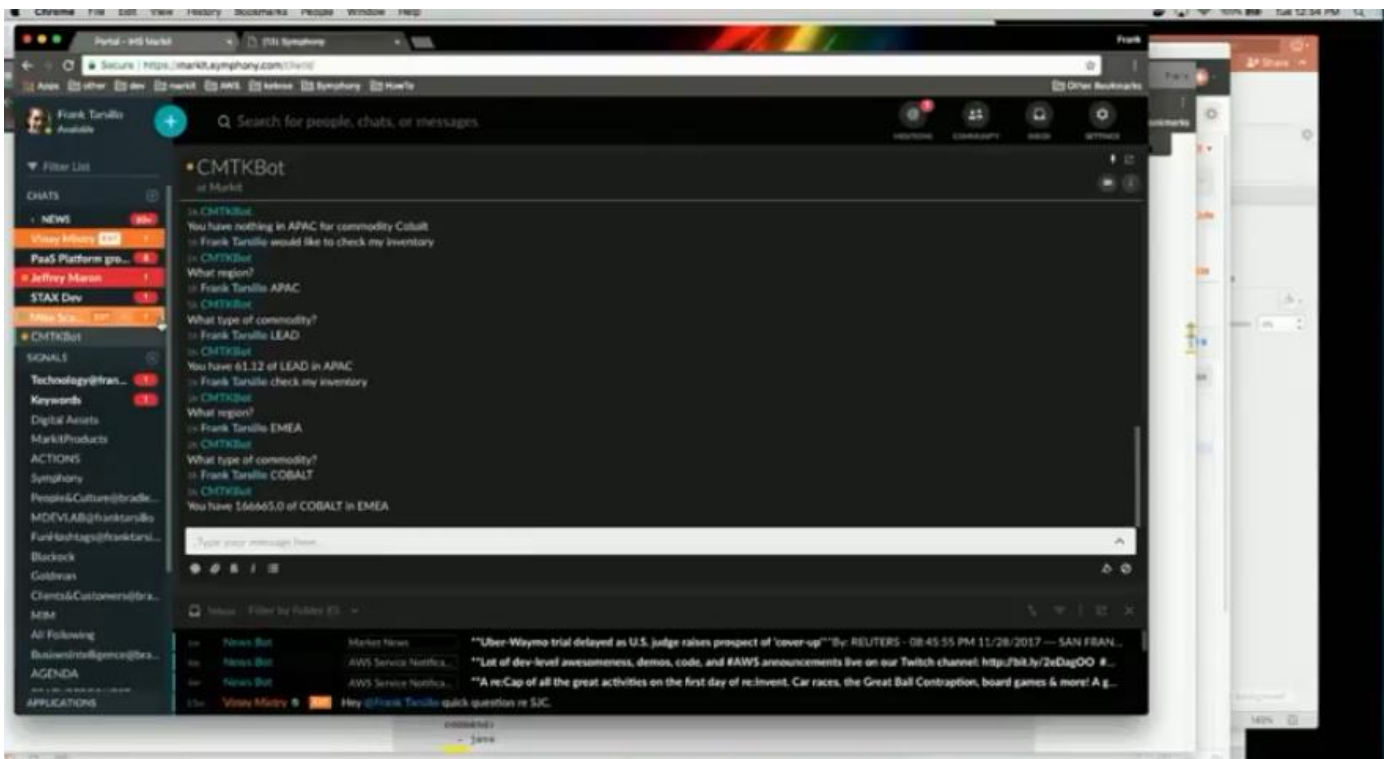
ASSIGNED TO

Google Maps

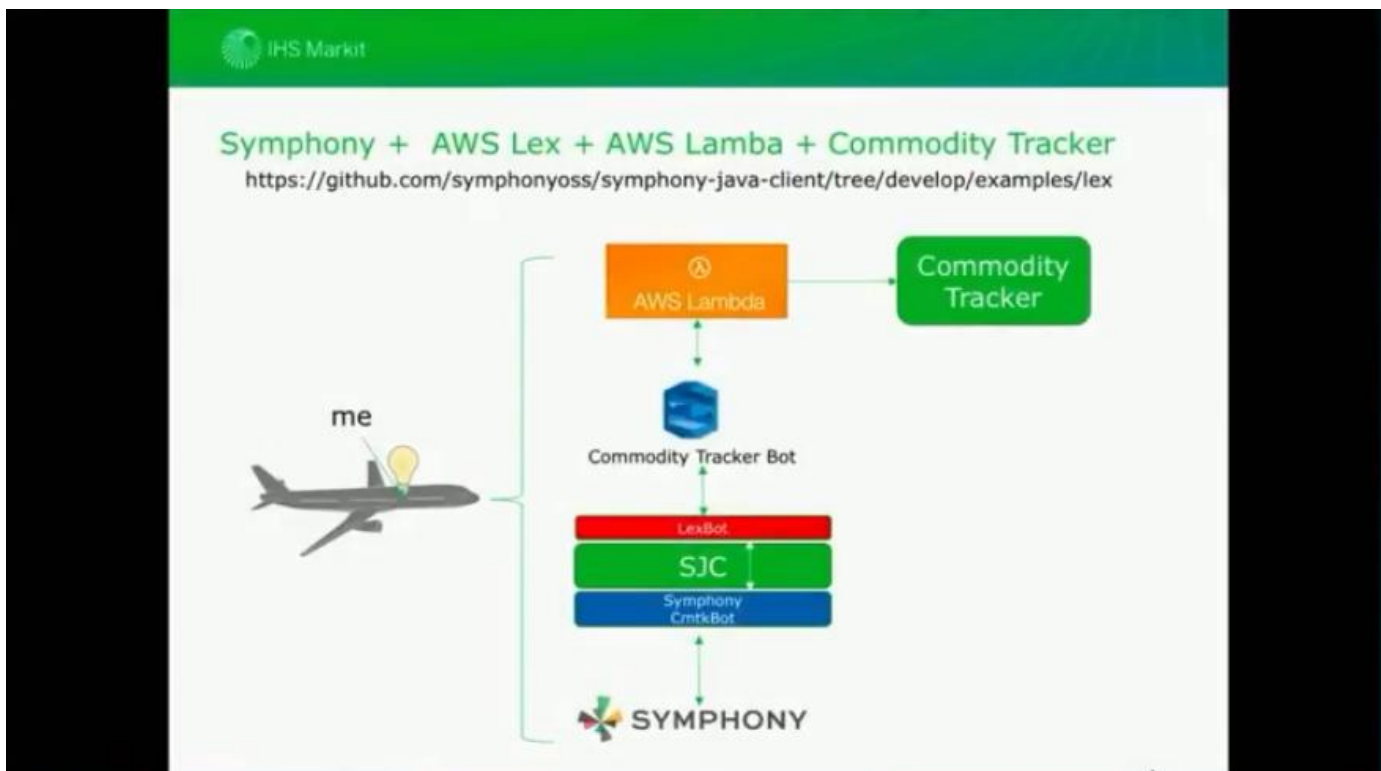
Ports America - Totals

COMMODITY	SHAPE	BRAND	PIECES	WEIGHT (MT)
-----------	-------	-------	--------	-------------

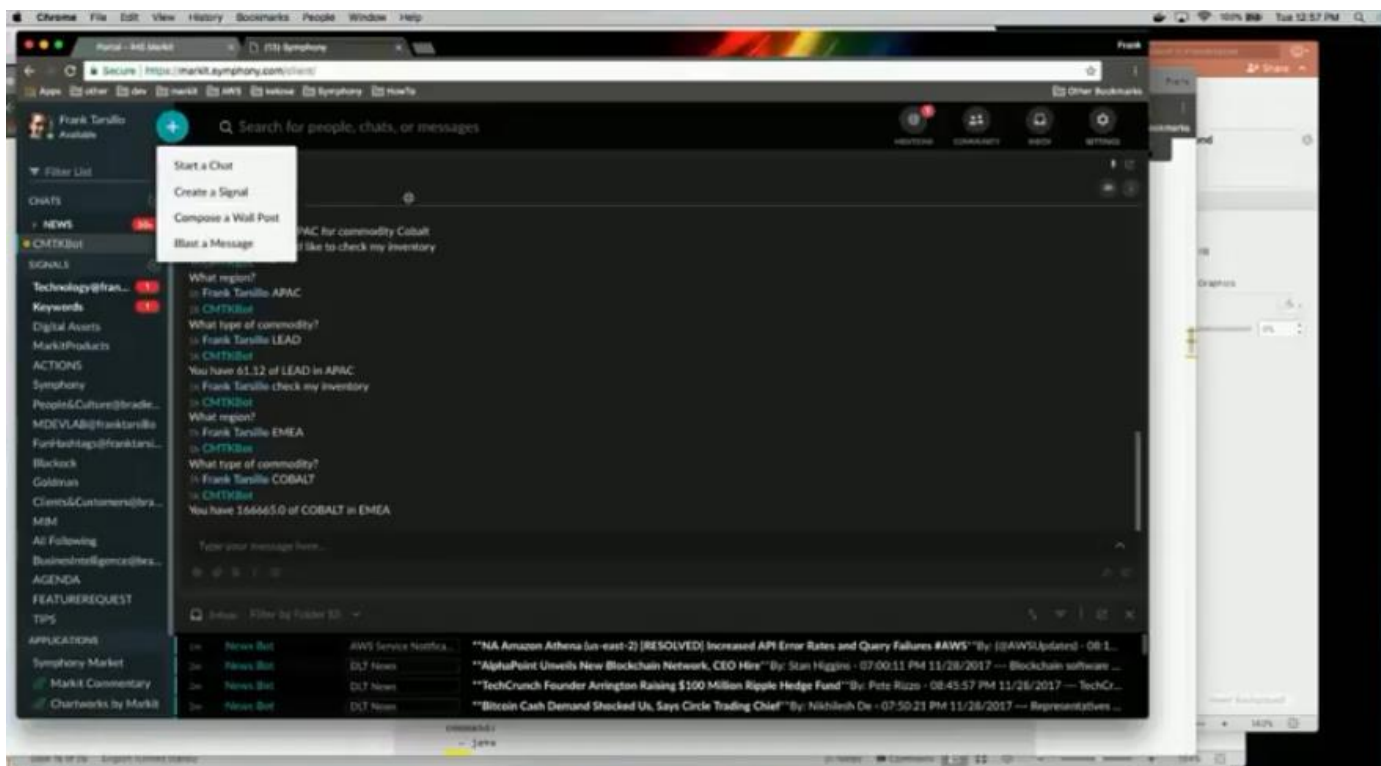




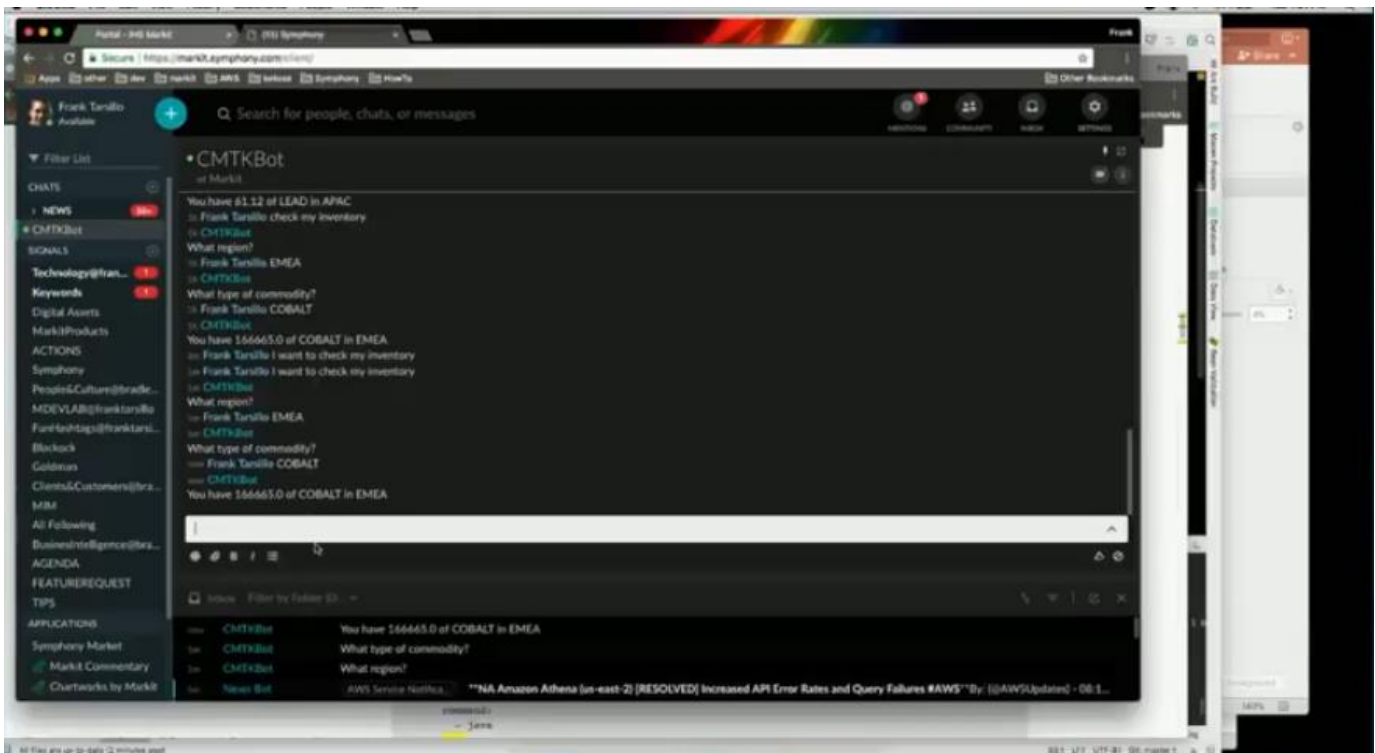
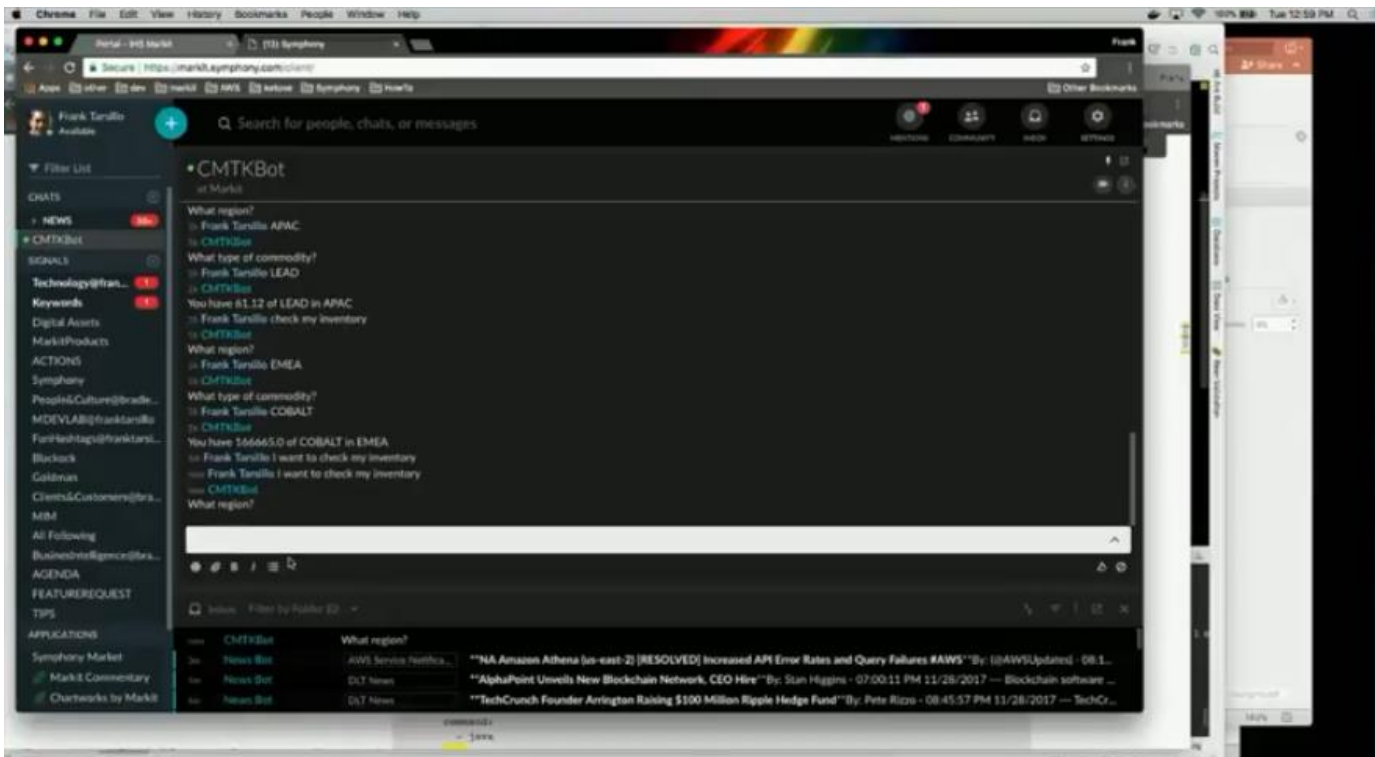
Symphony is a new collaboration network being assembled by all the major financial institutions for providing instant messaging between different parties.

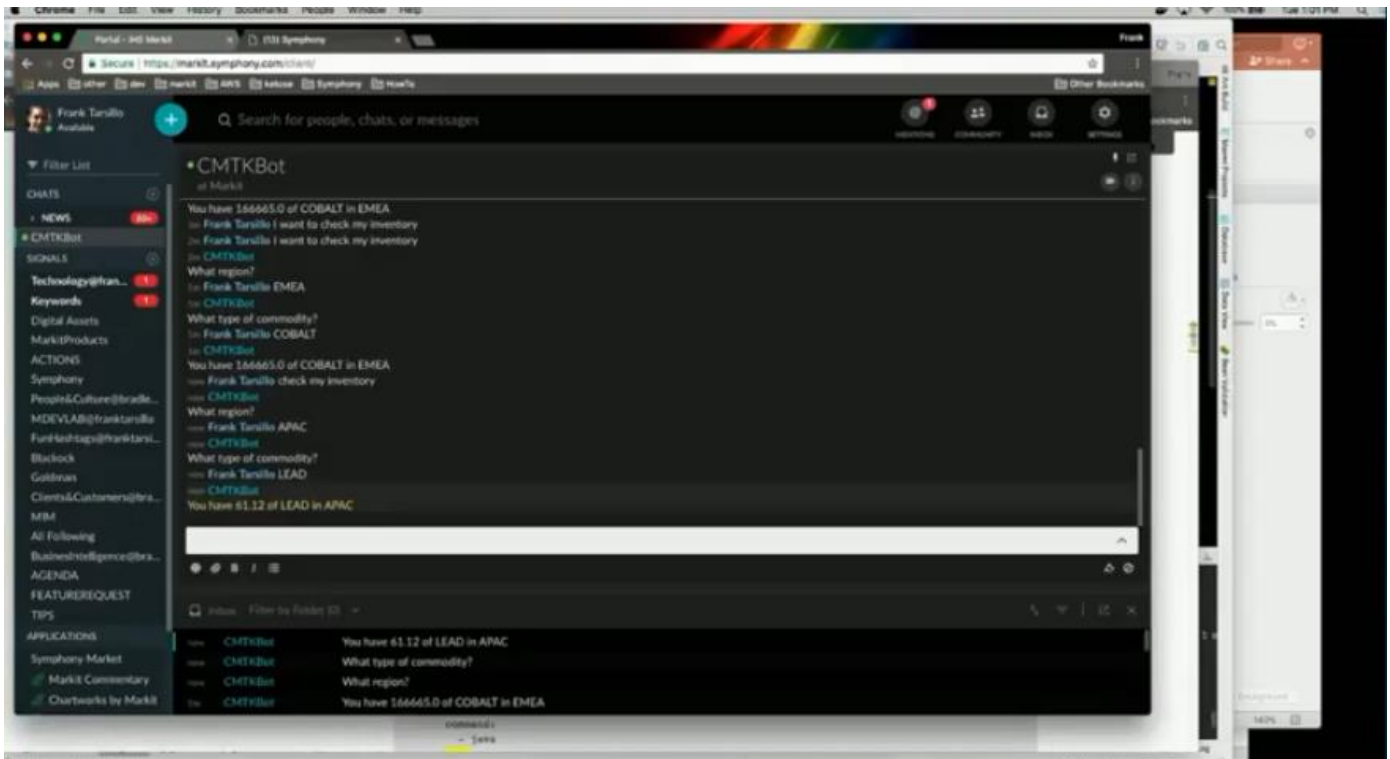


I started to create a bot that makes a bot that can serve real time data using lambda functions and Amazon Lex with its NLP features. The lambda functions can go into Commodity Tracker and pull out needed information that can then be relayed to Lex for the bot.



A trader can simply interact and talk to the bot for real time information when trying to make a commodities deal. These are traders just in Symphony all day long and the ability to chat with a Bot using Lex and Lambda is critical.





Getting started with ML

AWS AI:

Democratized artificial intelligence

Services



Amazon Lex (language)



Amazon Polly (speech)



Amazon Rekognition (vision)

Platforms

Amazon ML

Spark & EMR

Kinesis

Batch

ECS

Frameworks

Apache
MXNet

Tensor
Flow

Gluon

Caffe2
& Caffe

Theano

Keras

Cognitive
toolkit

Torch

AWS Deep Learning AMI

Infrastructure

GPU

CPU

IoT

Mobile

AWS
re:Invent

© 2017 Amazon Web Services, Inc. or its Affiliates. All rights reserved.



Amazon Machine Learning (Amazon ML)

Easy-to-use, managed machine learning service built for developers

Robust, powerful machine-learning technology based on Amazon's internal systems

Create models using your data that is already stored in the AWS cloud

Deploy models to production in seconds



AWS
re:Invent

aws

Fully managed model and prediction services

End-to-end service with no servers to provision and manage

One-click production model deployment

Programmatically query model metadata to enable automatic retraining workflows

Monitor prediction usage patterns with Amazon CloudWatch metrics



AWS
re:Invent

aws

Easy to use and developer-friendly

Use the intuitive, powerful service console to build and explore your initial models

Data retrieval

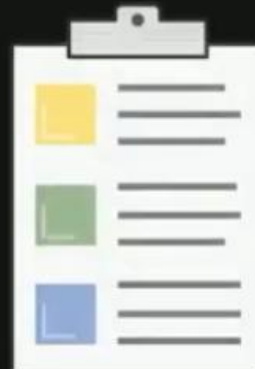
Model training, quality evaluation, fine-tuning

Deployment and management

Automate model lifecycle with fully featured APIs and SDKs

Java, Python, .NET, JavaScript, Ruby, PHP

Easily create smart iOS and Android applications with AWS Mobile SDK



AWS
re:Invent

© 2017 Amazon Web Services, Inc. or its Affiliates. All rights reserved.



Integrated with the AWS data ecosystem

Access data that is stored in Amazon S3, Amazon Redshift, or MySQL databases in Amazon RDS

Output predictions to Amazon S3 for easy integration with your data flows

Use AWS Identity and Access Management (IAM) for fine-grained data access permission policies



Key takeaways

Key takeaways

- ✓ AWS machine-learning services can help transform your applications and improve your bottom line
- ✓ Get started today:
 - Try the AML tutorial at <https://aws.amazon.com/aml/getting-started/>

<https://aws.amazon.com/aml/getting-started/>

Please remember to rate this session in the mobile app



AWS re:Invent

Thank you!

Frank Tarsillo, frank.tarsillo@ihsmarkit.com

Peter Williams, willpe@amazon.com

Please fill out a survey of this session: GPSTEC305

AWS re:Invent

© 2017, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

