How we use machine learning in Project Natural Language Search

August 2018 | Version 1.0

Our Speakers

DIPOCK DAS

Senior Director, Products, Splunk

AUNGON NAG RADON

Data Scientist, Splunk

Forward-Looking Statements

During the course of this presentation, we may make forward-looking statements regarding future events or the expected performance of the company. We caution you that such statements reflect our current expectations and estimates based on factors currently known to us and that actual events or results could differ materially. For important factors that may cause actual results to differ from those contained in our forward-looking statements, please review our filings with the SEC.

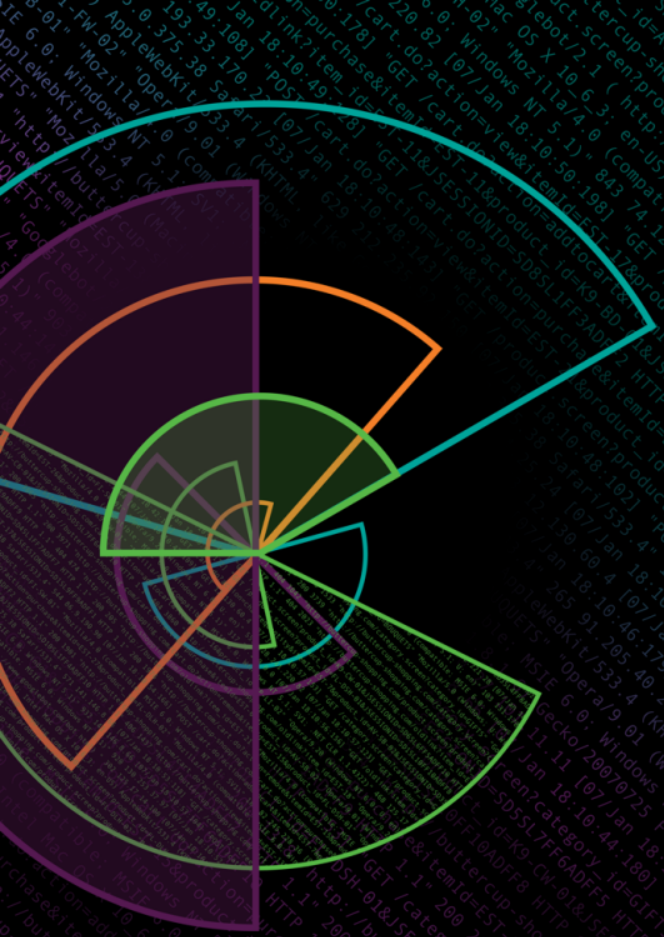
The forward-looking statements made in this presentation are being made as of the time and date of its live presentation. If reviewed after its live presentation, this presentation may not contain current or accurate information. We do not assume any obligation to update any forward-looking statements we may make. In addition, any information about our roadmap outlines our general product direction and is subject to change at any time without notice. It is for informational purposes only and shall not be incorporated into any contract or other commitment. Splunk undertakes no obligation either to develop the features or functionality described or to include any such feature or functionality in a future release.

What you will learn today

Machine learning can be used to solve many different problems.

We will show you examples of applied machine learning in Project Natural Language Search.

We will walk through high level examples before diving into one example in depth.



Introduction

What is Project Natural Language Search?

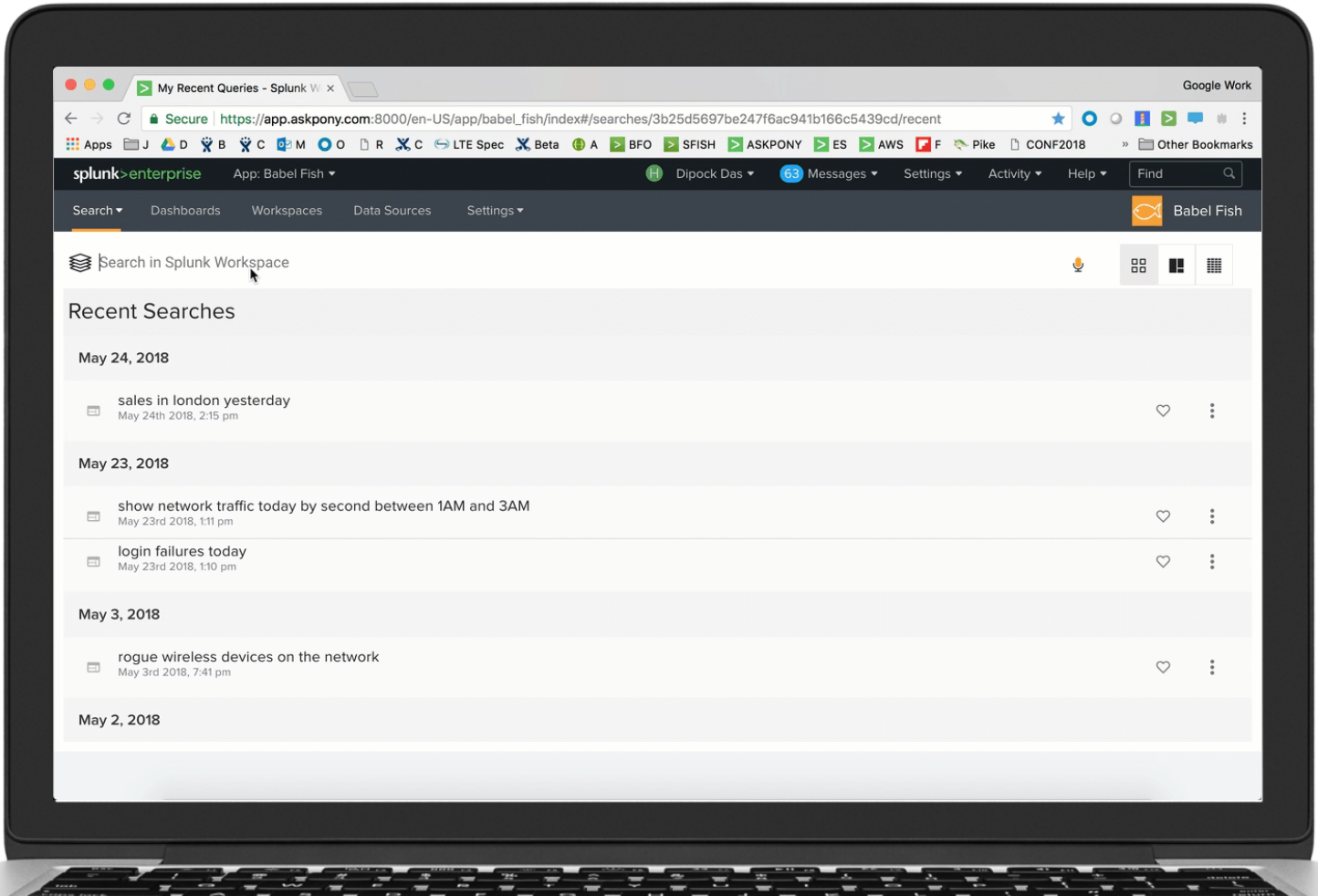
Project Natural Language Search is a natural language platform for machine data that delivers Natural Language Search, Understanding and Generation for Splunk and SQL data.

Natural Language Search

Communicate instantly in charts, text and speech

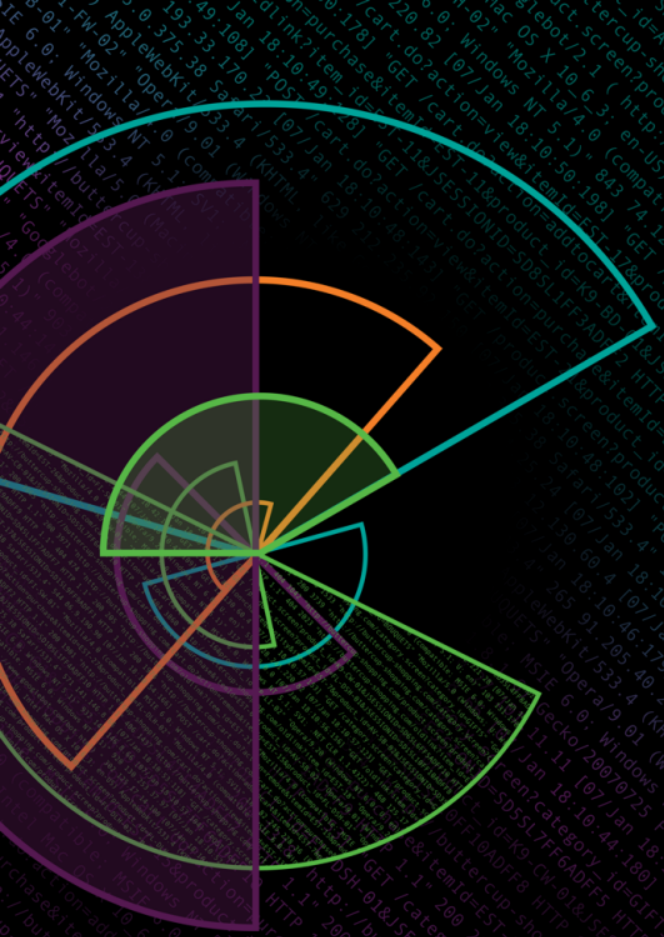
Access anywhere with type, touch, voice

Demo



Overview of Machine Learning in Project Natural Language Search

- Natural Language Understanding
- Visual Interface Determination
- Data Driven Drill Downs
- Handling Ambiguity
- Natural Language Speech Recognition
- Search Query Recommendation



Project NLS: Natural Language Understanding

splunk> .conf18

What is Natural Language Understanding?

- Understand the intent behind a textual query
- NLU is a hard AI problem!

show	me	daily	sales	of	cappuccino	in	Vancouver	last	10	days
verb	pronoun	adjective	plural	preposition	singular noun	preposition	singular noun		value	plural
	predicate	clause	subject		subject		clause			clause
		time grain	sales data		product data		cities			time range from now

Intent: {sum} {sales} {product} {timegrain} {city} {time range}

Project NLS turns Intents into SPL (or SQL)

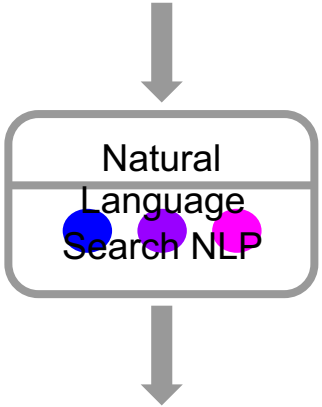
Intent: {sum} {sales} {product} {timegrain} {city} {time range}

SUM sales BY Day
WHERE Product = Cappuccino, AND City IN Vancouver, San Francisco, San Jose,
AND timerange is (today minus 10 days) to today

```
| tstats allow_old_summaries=t summariesonly=t
SUM("All_Sales.grossSales") AS "All_Sales.grossSales"
FROM datamodel=Retail.All_Sales
WHERE ((("All_Sales.productName"="cappuccino")
AND ("All_Sales.city"
IN ("vancouver","san francisco","san jose")))
AND ((earliest=1533427200)
AND(latest=1534291199)))
BY "All_Sales.city" _time span=1d
| eval All_Sales__time_date=strftime(_time, "%Y-%m-%d")
| table All_Sales__time_date "All_Sales.city" "All_Sales.grossSales" |
sort 150 "All_Sales.grossSales" DESC, All_Sales__time_date DESC,
"All_Sales.city" DESC
```

Project NLS machine learning obliterates hard coded rules systems

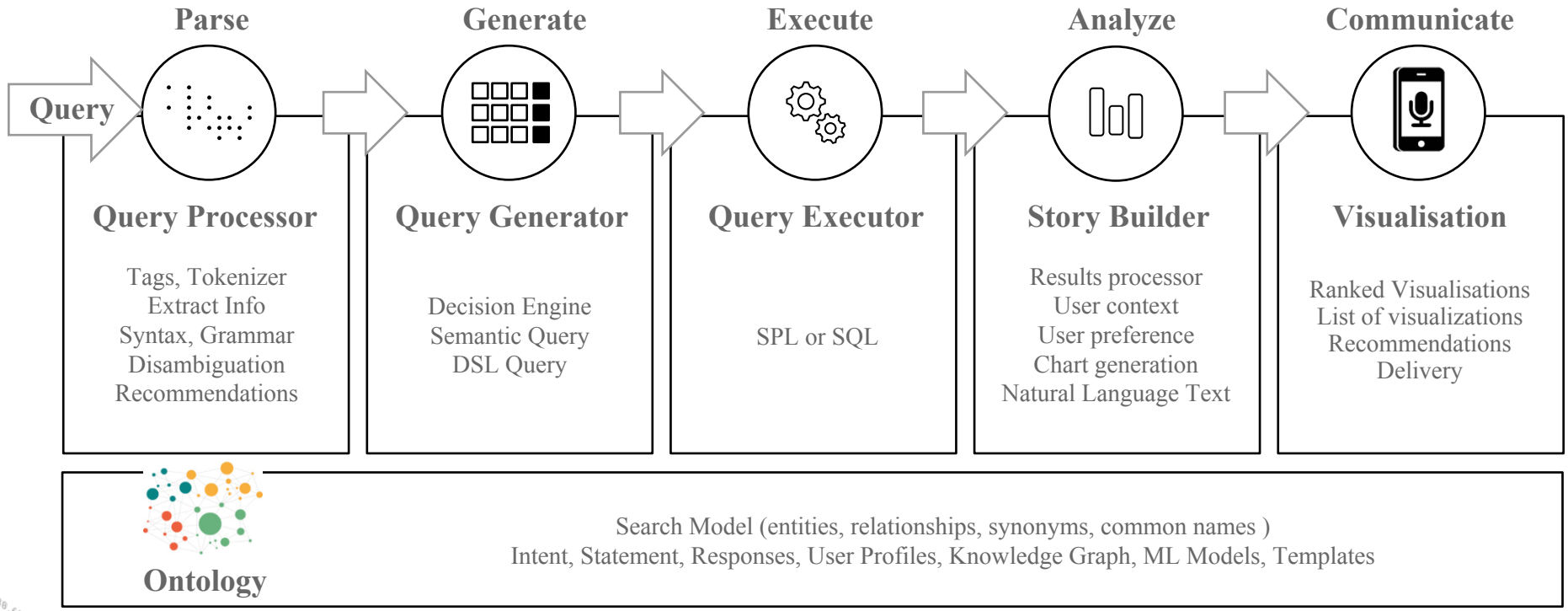
Show me **daily sales** for **San Francisco** in **September**
What were **daily sales** in **SF last month**
Display the **Vancouver** store **gross sales** for the **last 30 days** by **day**
Get me **San Jose takings** for the **last four weeks** by **day**

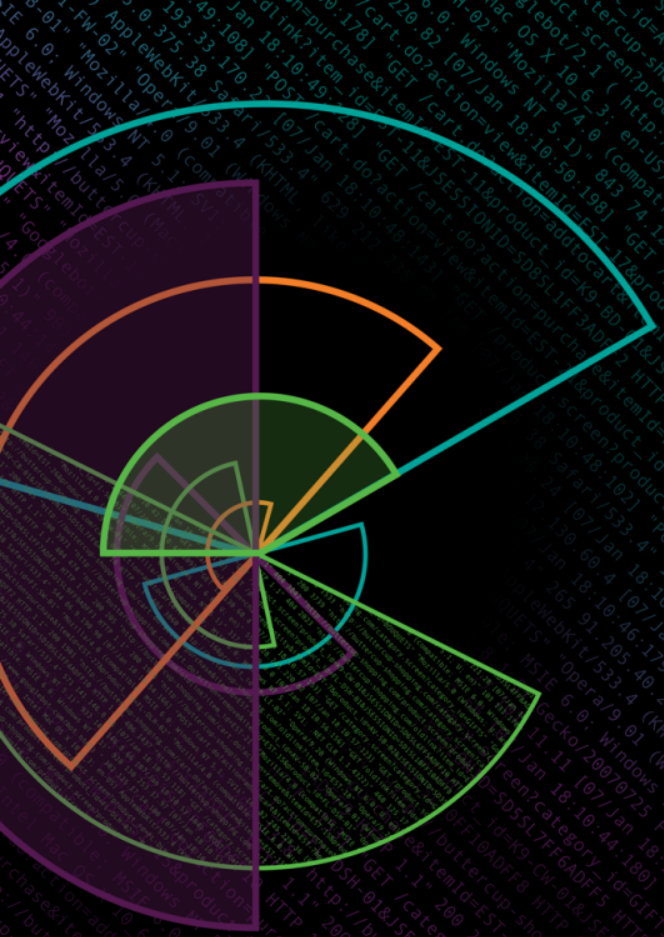


sales, **city**, **time period**, **time grain**

Model Driven Natural Language Understanding

Natural Language Processing Pipeline



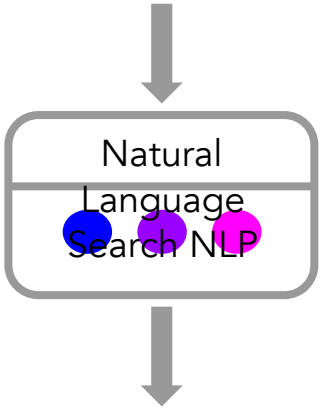


Project NLS: Handling Ambiguity

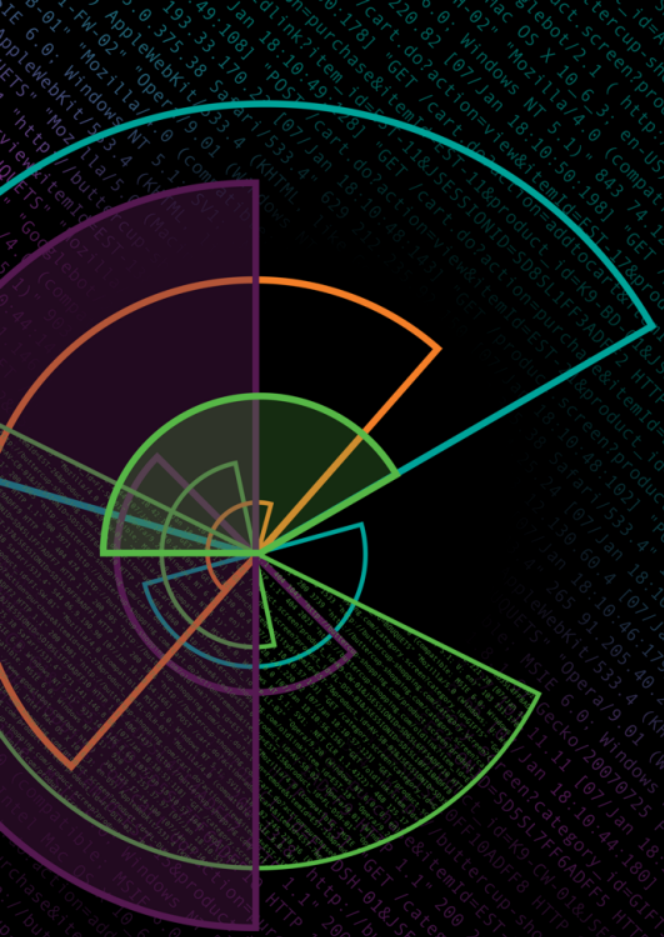
splunk> .conf18

Natural Language search machine learning disambiguates a query

Top 5 products

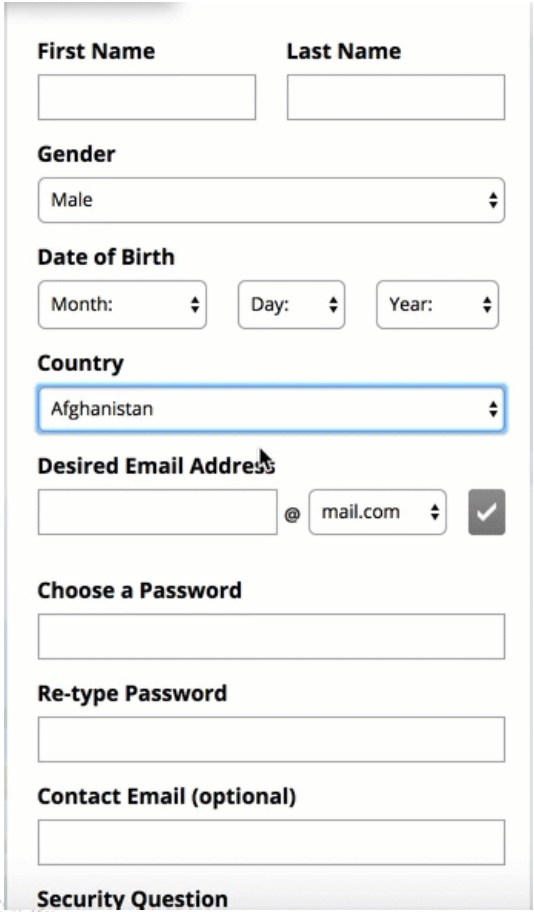


Top 5 products by sales
Top 5 products by location
Top 5 products by quantity
Top 5 products by sugar content



Project NLS: Visual Interface Determination

What's wrong with User Interface Design?



- We can do better than sort alphabetically..
- Afghanistan - 0.48% of World's population
- Default using browser getCurrentPosition()?

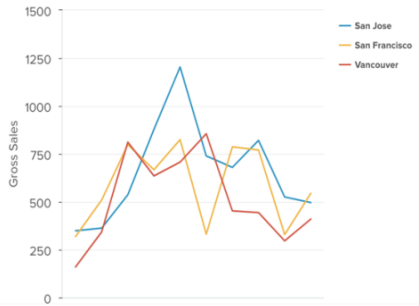
- Why not predict top 5 by signups?

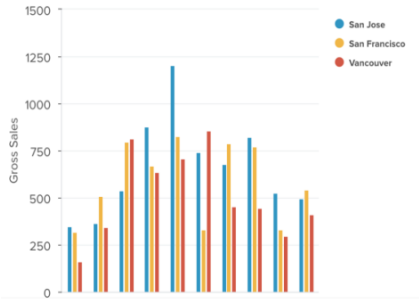
United States	45%
United Kingdom	28%
Canada	16%
Australia	5%
India	2%

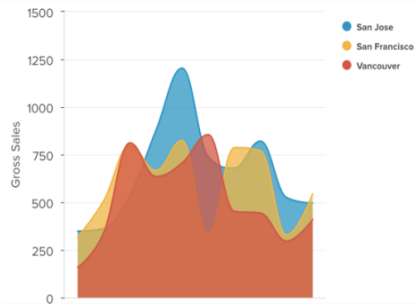
Rest of World listed alphabetically

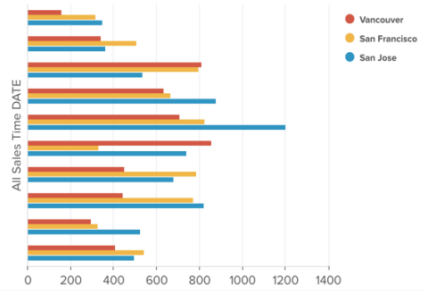
Using ML to display the right chart for the result

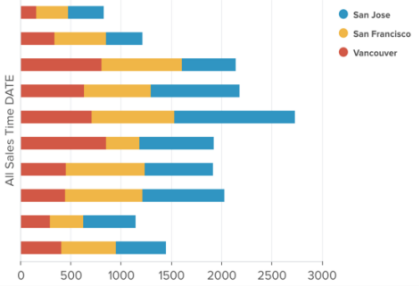
"daily sales of cappuccino in Vancouver, San Francisco and San Jose last 10 days"

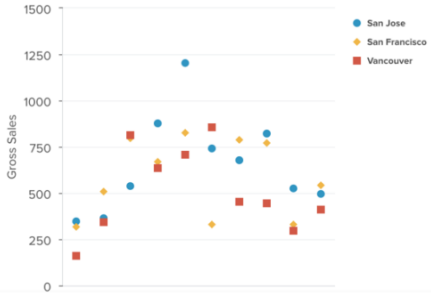








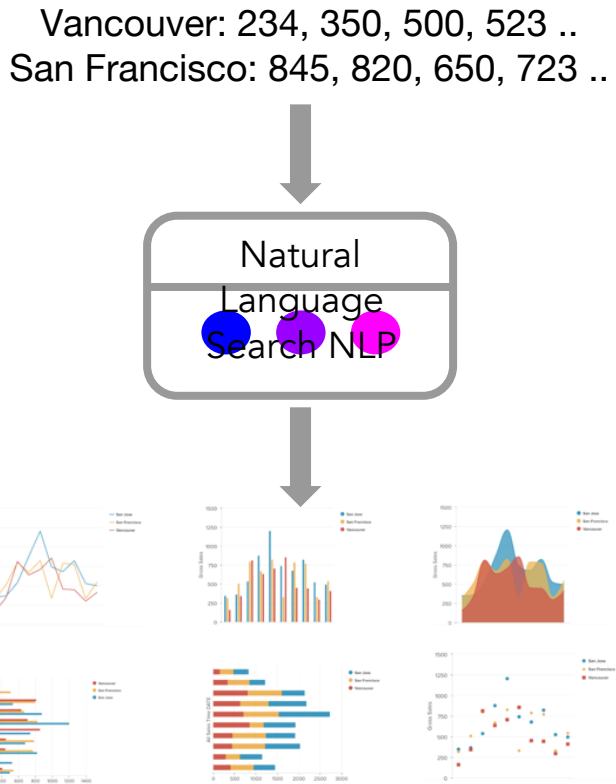




splunk> .conf18

How ML Improves Visual Interface Determination

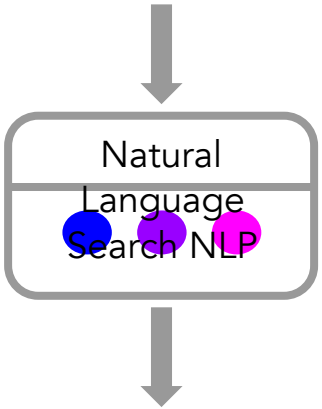
- Learn from past interactions
- Recommend a ranked list of charts to user for a given query
- Incorporate user's preference



What is Data Driven Drill Path?

Recommend drill path navigation in a selected visualization based on learned preference

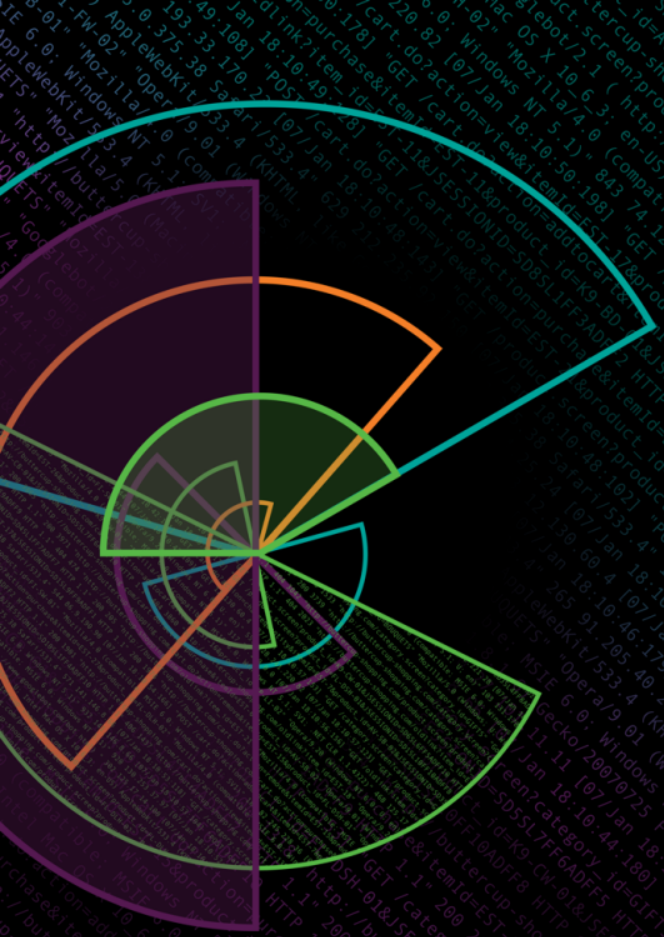
Sales by Country



Learns user's preferred navigation path in a chart type

Capture preference change over time!

Sales by City
Sales by State
Sales by Product



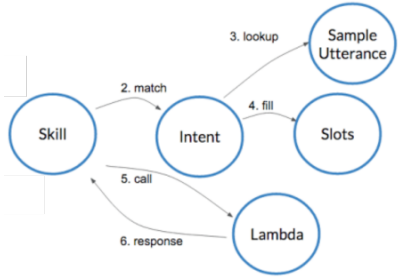
Project NLS: Natural Language Speech Recognition

splunk> .conf18

Natural Language Speech Recognition

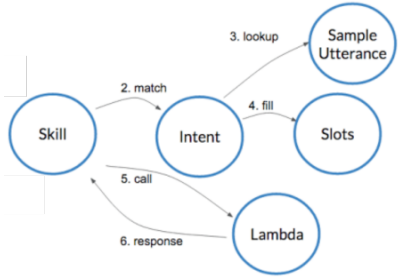
- The User asks a question using a voice enabled device (VTT)
- The User gets an answer from the device (TTS)

NLU reduces the burden of skill building for so many scenarios



Trained skill

Data set

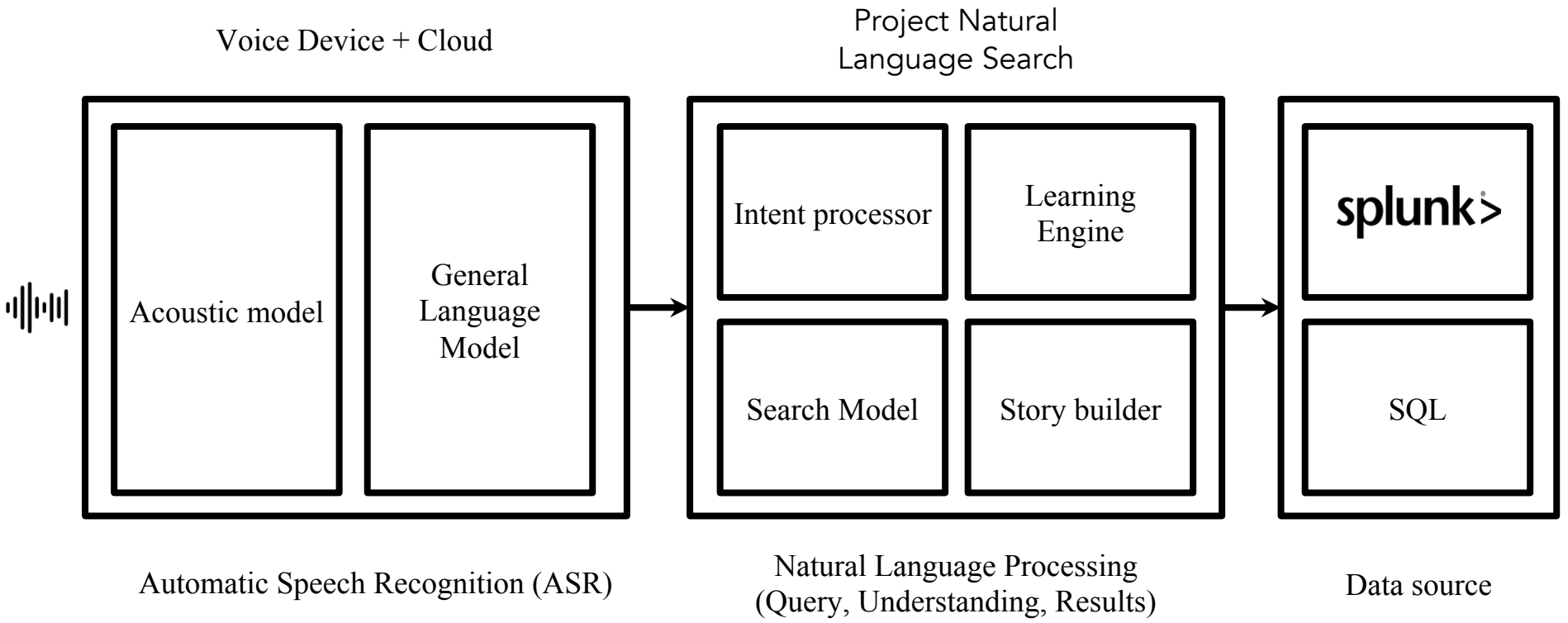


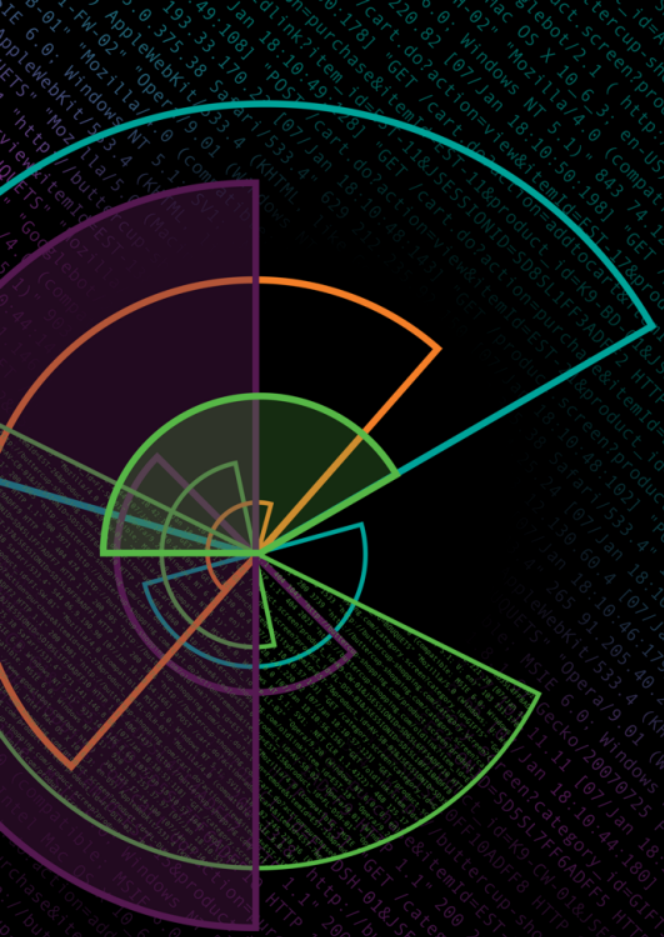
Trained skill

Data set

The current method for building speech (and chatbot) apps is to encode all the possible questions someone might ask.

Project NLS support for Voice enabled devices





Project NLS: Search Query Recommendation

splunk> .conf18

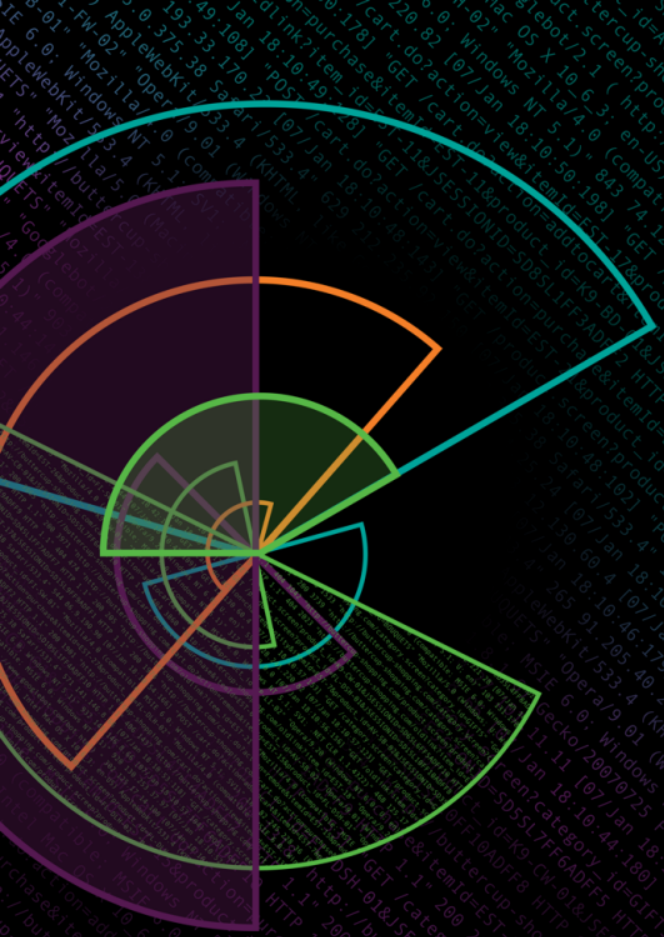
Search Query Recommendation

Trending Search Recommendation:

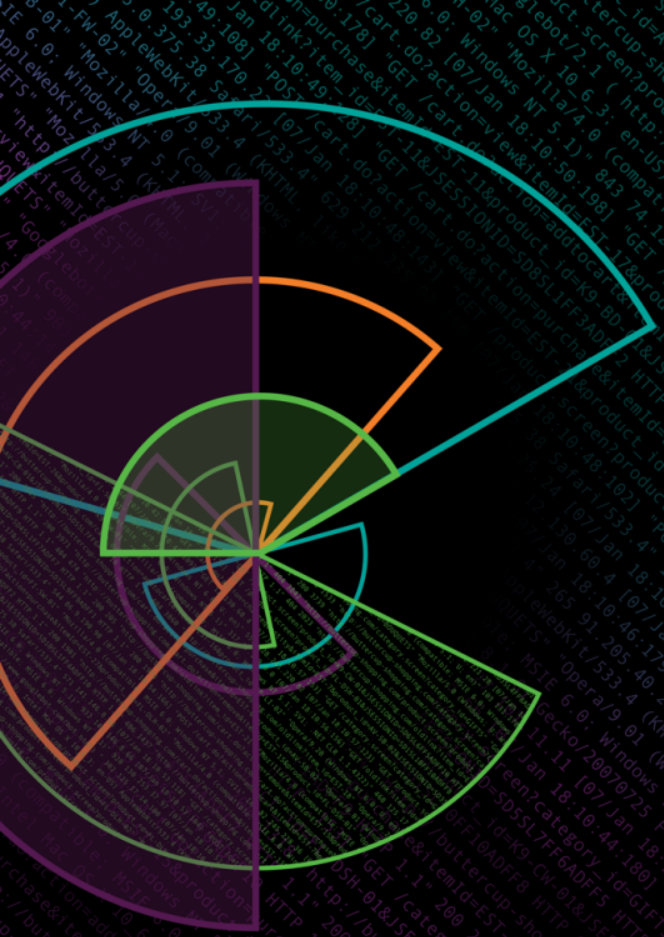
- Recommend top N search queries to user

Relevant Query Recommendation:

- Recommend top N search queries relevant to a selected search query



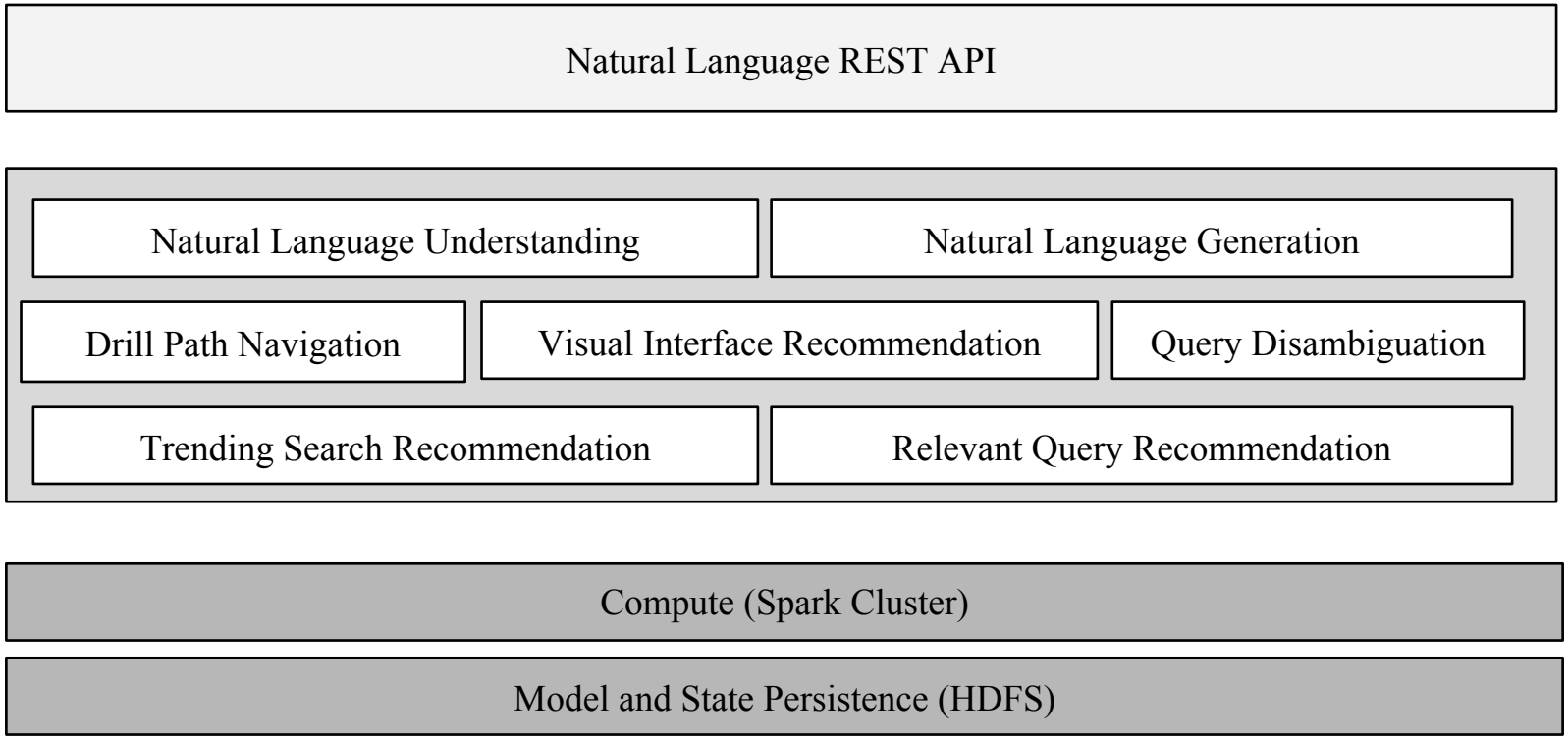
Demo

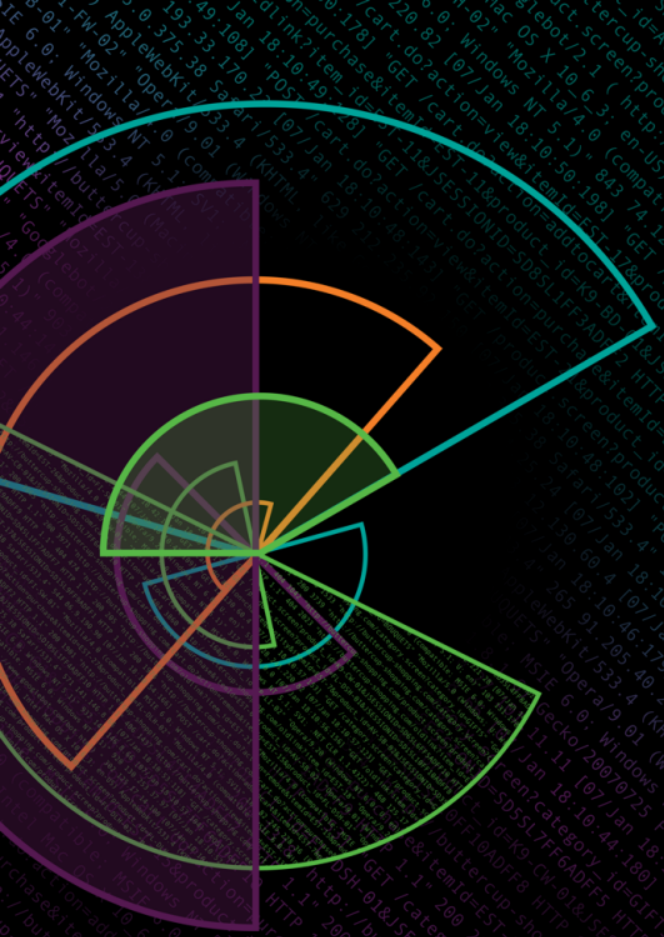


Project NLS: Machine Learning Deep Dive

splunk> .conf18

Natural Language Search Machine Learning Architecture



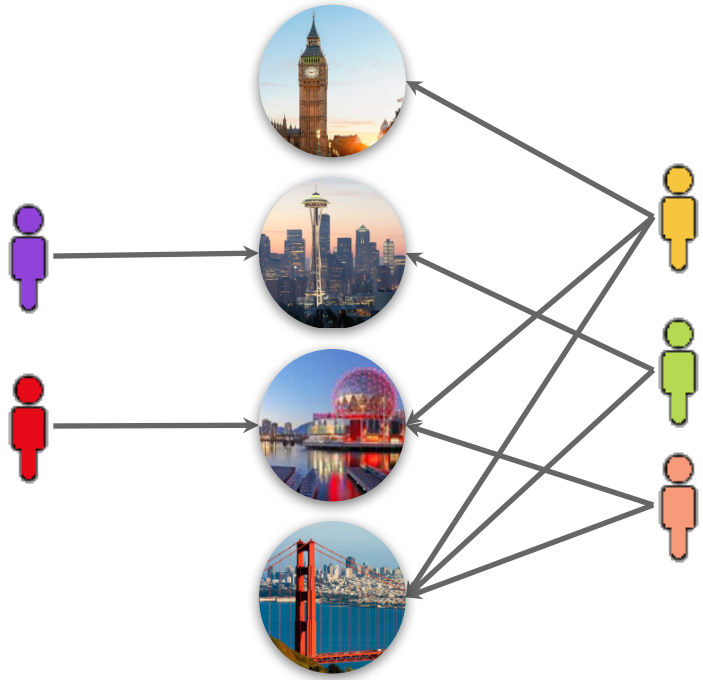


Deep Dive: Trending Search Recommendation

splunk> .conf18

Collaborative Filtering

make automatic predictions (filtering) about the interests of a user by collecting preference information from many users (collaborating)



daily sales in seattle last month?

daily sales in vancouver last week?

daily sales last week in london?
how much espresso sold in vancouver?
daily sales in san francisco last month?

how much espresso sold in seattle?
-?-

how much spicy chai sold in vancouver?
daily sales in san francisco last week?

in this case we observe the questions users ask about stores in specific city locations

Collaborative Filtering

Purple asked questions about **Seattle** and **not London**

Pink asked questions about **San Francisco** and **Seattle**

Green asked questions about **San Francisco**, **Seattle** and **not London**

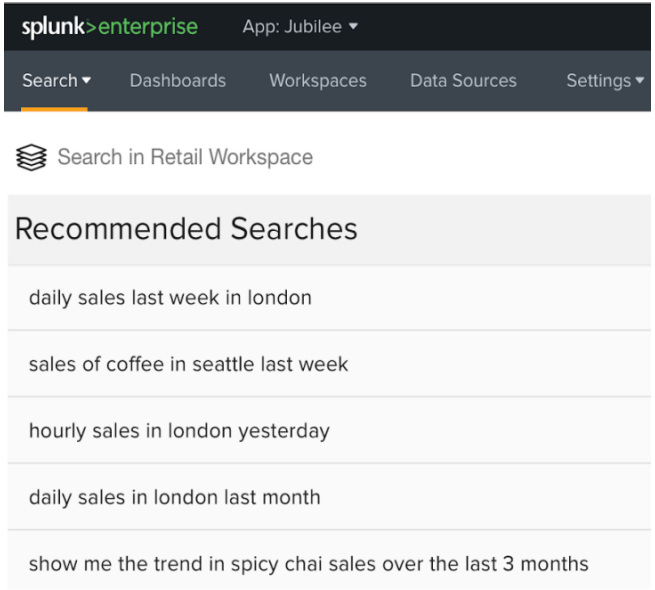
Purple and Pink have asked the same questions as Green for matching location.

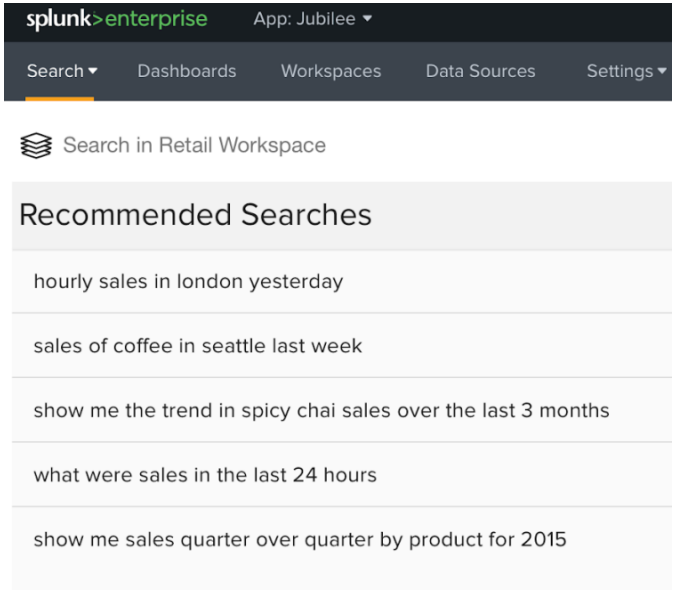
Based on the match, Green is not likely to ask questions about Vancouver

✘

splunk> .conf18

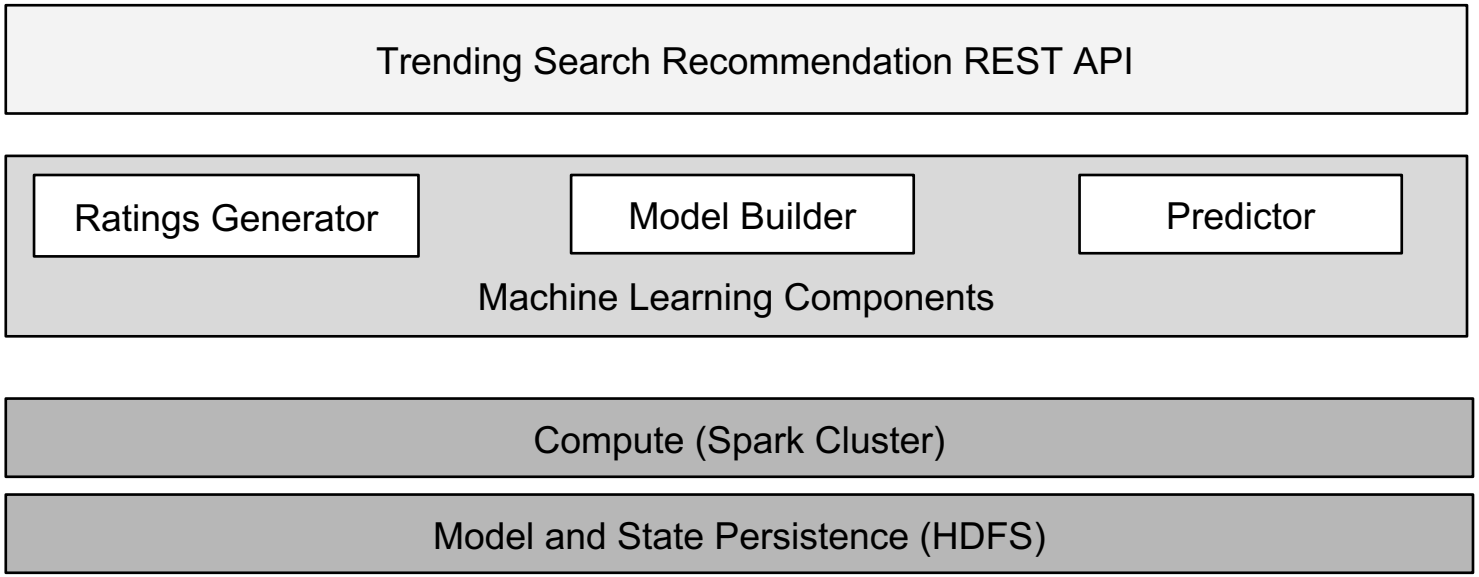
Trending Search Recommendation with Collaborative Filtering



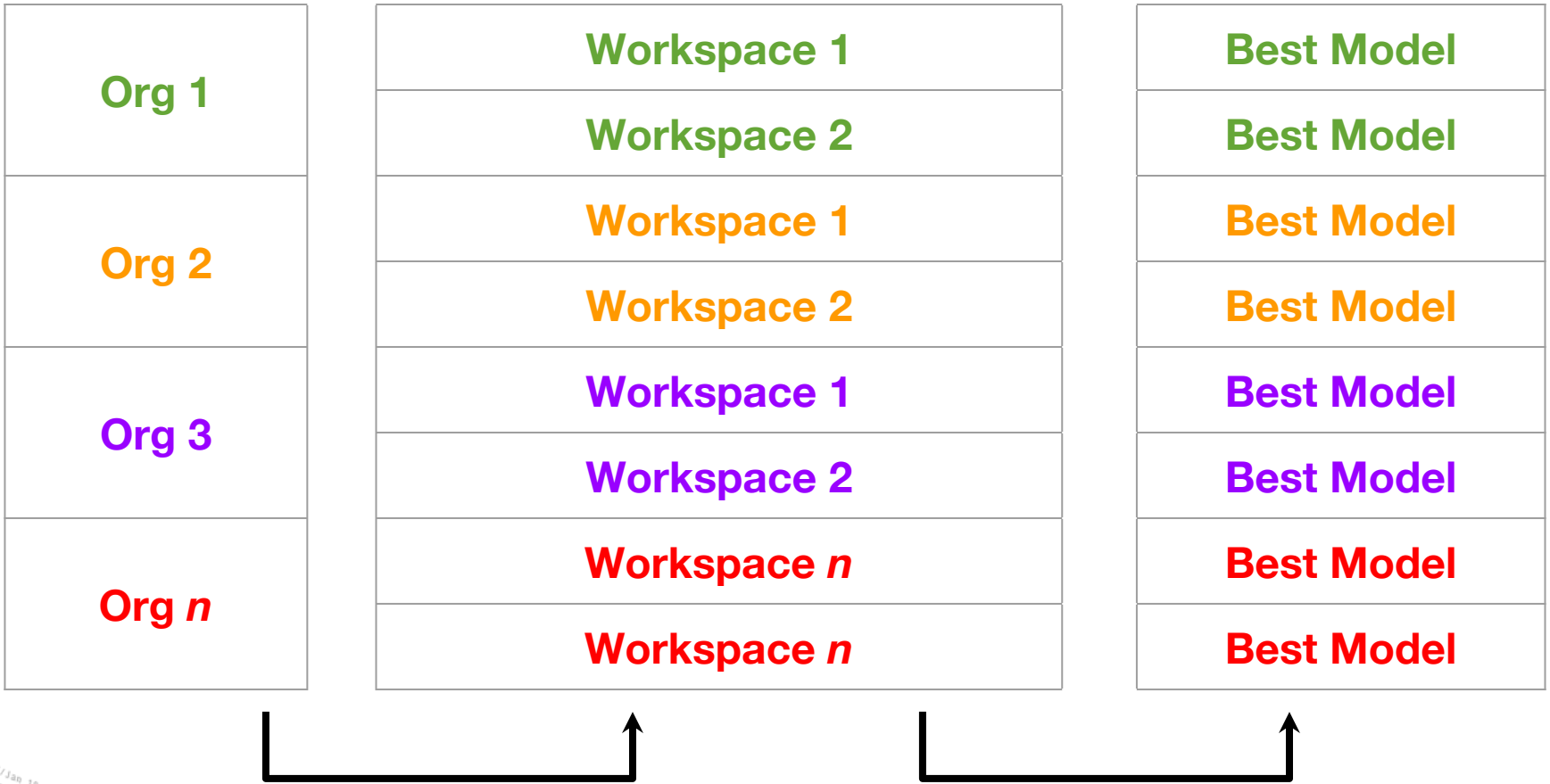


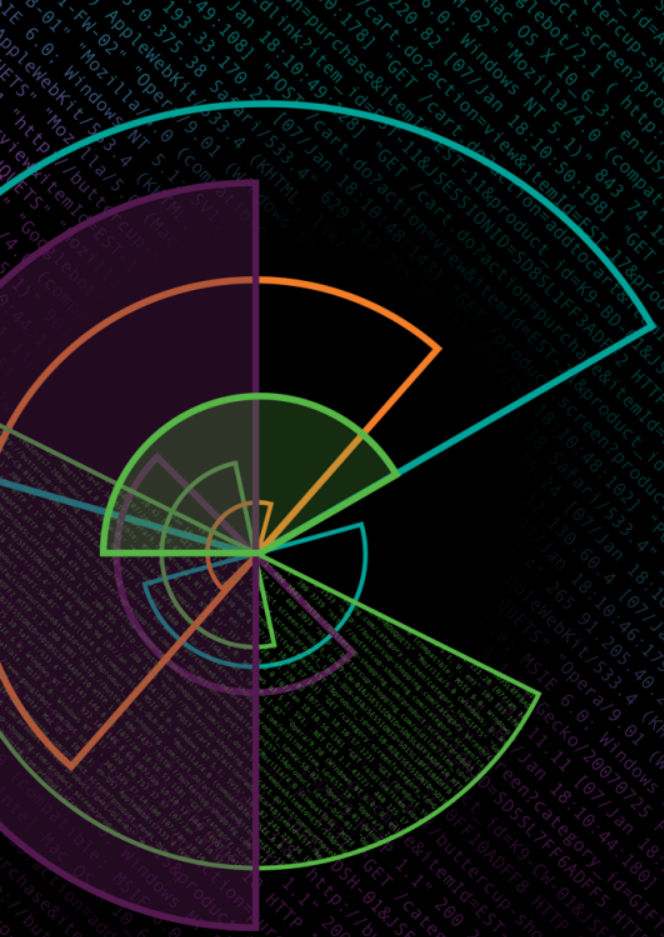
Why was the Recommendations List updated?

Machine Learning Module Components



Model Persistence per customer per workspace



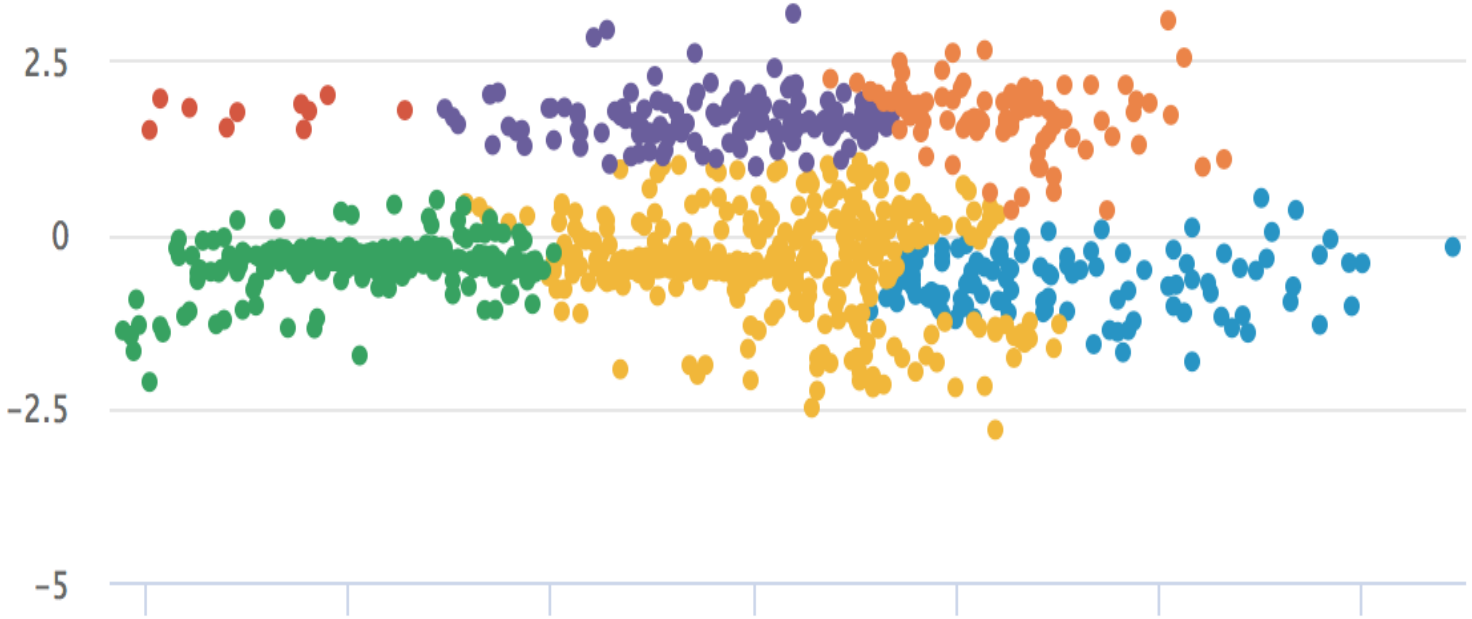


Deep Dive: Relevant Query Recommendation

splunk> .conf18

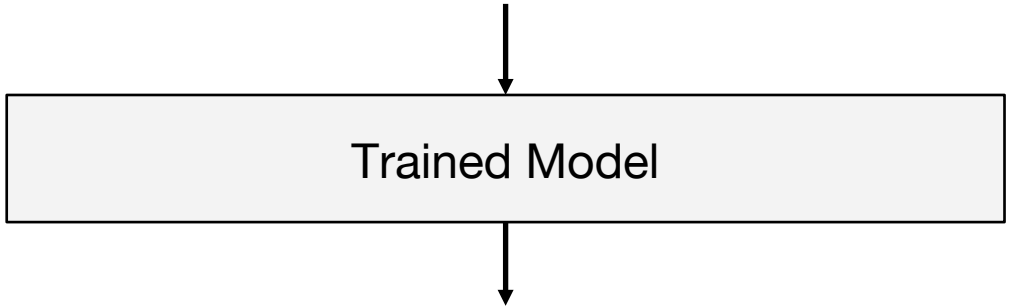
How Do We Recommend Relevant Queries?

- Clustering (Unsupervised ML Technique)



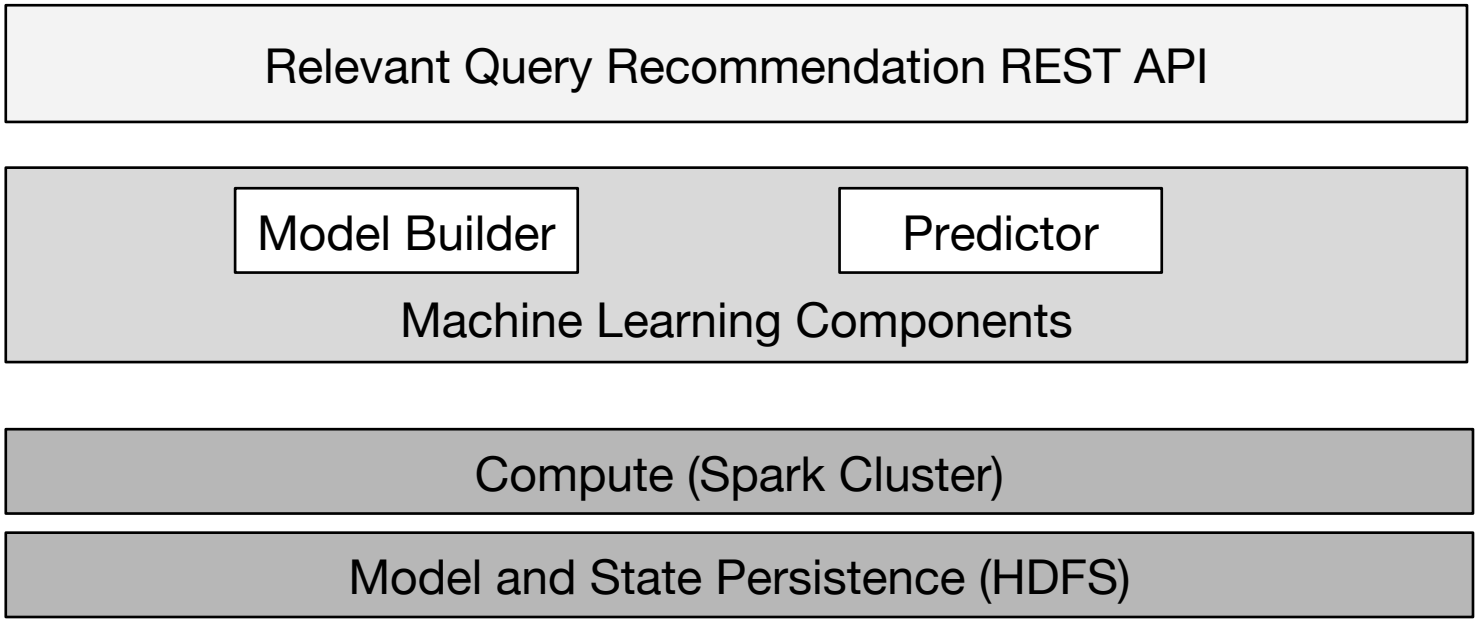
Example of Clustered Queries

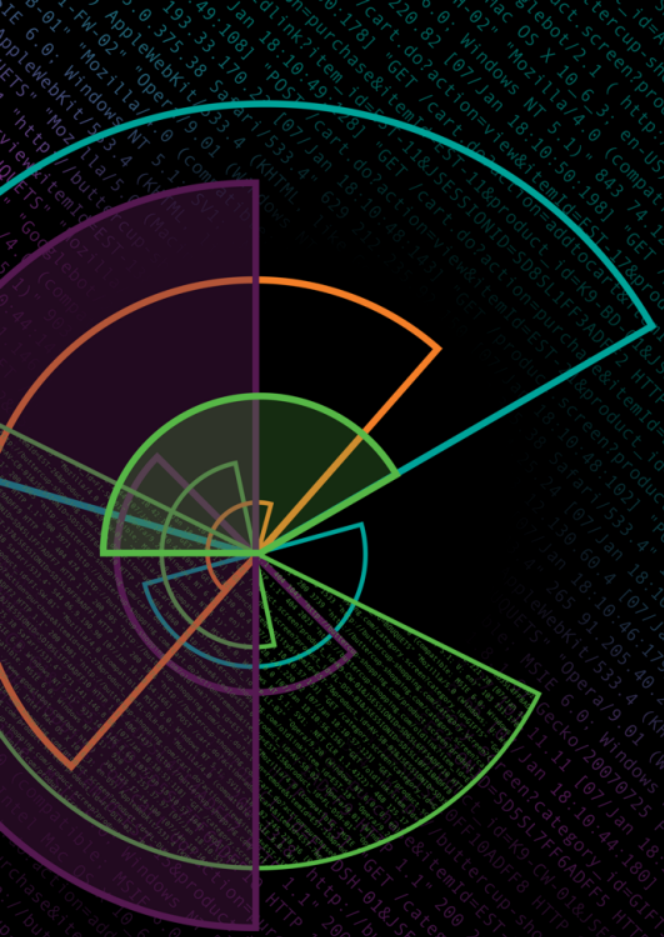
Show me sales last week



Show me sales in SF last week
Show me sales in California last year
Show me sales in 2017 vs 2016 in California

Architecture





Demo

Where next?

1. More user interface driven ML!
2. Sequence to sequence prediction - predict Next Query for Dialog based interaction
3. Federated learning - sharing trained models (not data) to help everyone

Key Takeaway

Machine learning can be used to solve a myriad of problems.

We have shown you production examples of machine learning used to improve the usability of an application.

You can use the same techniques in your daily work to accomplish similar results.

Next steps

- Get Project Natural Language Search
  - Dipock Das - dipock@splunk.com
  - Melissa Gannes - mgannes@splunk.com
- Get the MLTK
- Get on the ML Advisory program
  - Work with your Rep
  - https://www.splunk.com/en_us/software/splunk-enterprise/machine-learning.html#MLTK-FAQ

Other sessions

Machine Learning & Natural Language Processing at BMW (FN1199)	**11:30 yesterday**
Spreading the Word: How Chat and Voice Is Transforming Splunk in Retail AI Ops (FN1572)	**4:30 yesterday**
Ask Splunk! Using natural language, voice and chat with Project Natural Language Search (FN1615)	**12:15 today**

Thank You

Don't forget to rate this session in the .conf18 mobile app