



# Machine Learning Workshop I LABS

Elastic Stack | [ML](#) | [Labs](#)

---

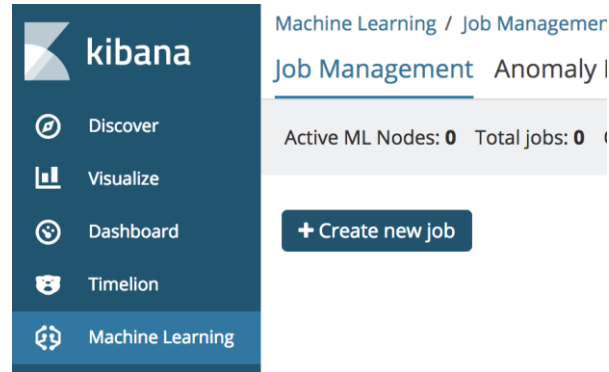
Shawn Hooton | Solution Architect



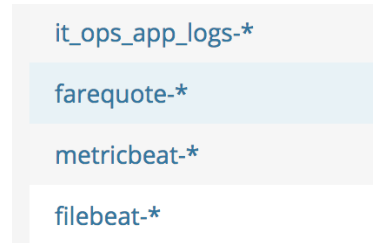
# Lab 1: The Simplest job

# Steps to Complete

1) In Machine Learning,  
*Create new job*

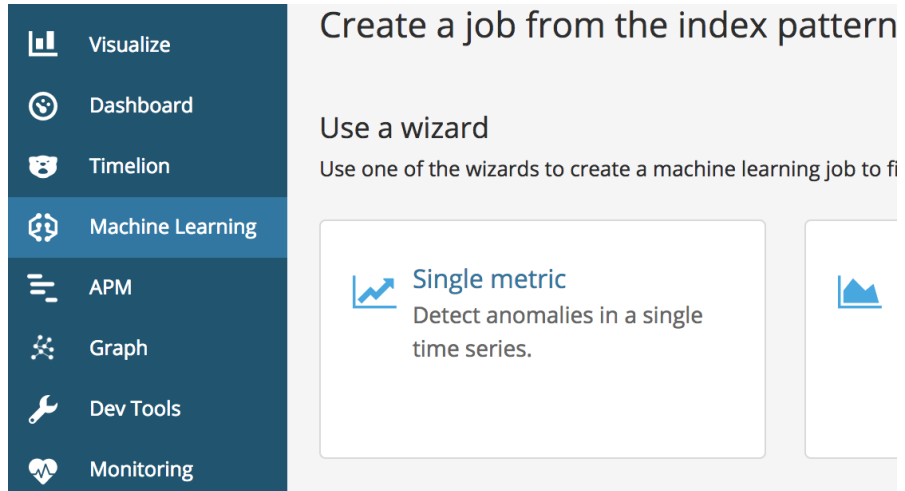


2) Choose the “farequote-\*”  
*Index pattern*



# Steps to Complete

3) pick the *Single Metric* wizard



The screenshot displays the Elastic ML interface. On the left is a dark blue sidebar with a vertical list of icons and labels: Visualize, Dashboard, Timelion, Machine Learning (highlighted with a white border), APM, Graph, Dev Tools, and Monitoring. The main content area has a light gray background. At the top, it says 'Create a job from the index pattern'. Below this, the heading 'Use a wizard' is followed by the text 'Use one of the wizards to create a machine learning job to fi'. Two white rectangular boxes are shown. The left box contains a blue line graph icon, the text 'Single metric' in blue, and the description 'Detect anomalies in a single time series.' The right box contains a blue area chart icon.

# Steps to Complete

4) Aggregation: *count*

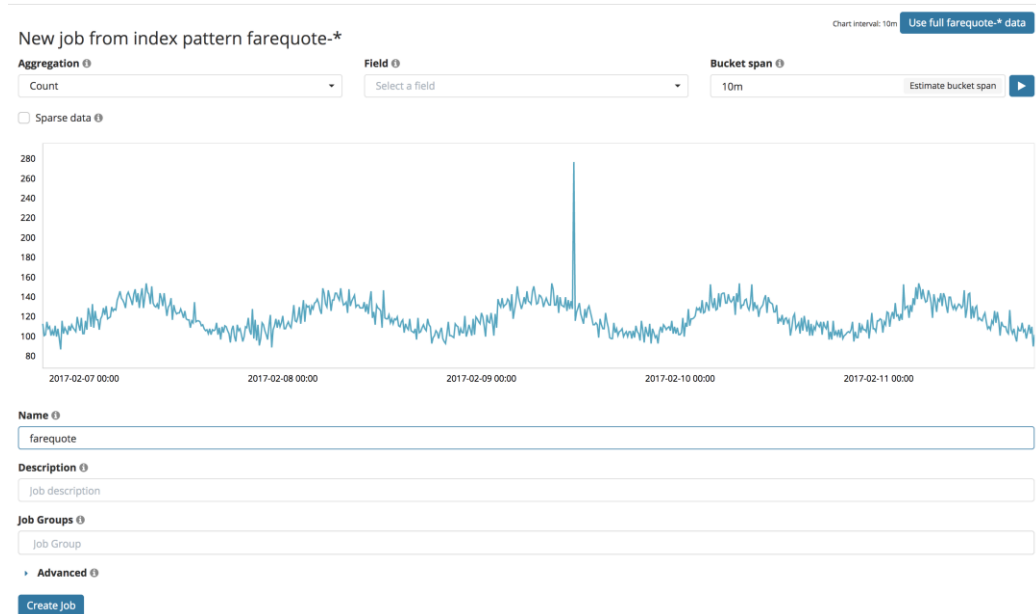
5) Field: *<leave blank>*

6) Bucket span: *10m*

7) Click the *“use full farequote data”*

8) Name: *“farequote”*

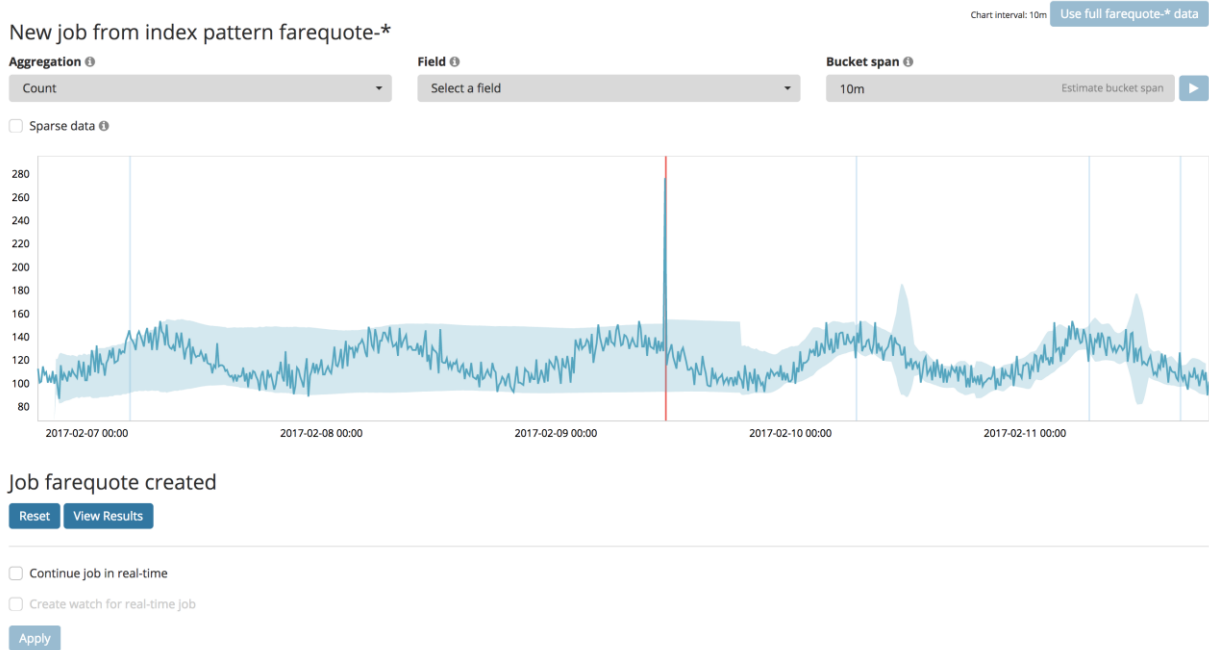
9) click “Create Job”



# Steps to Complete

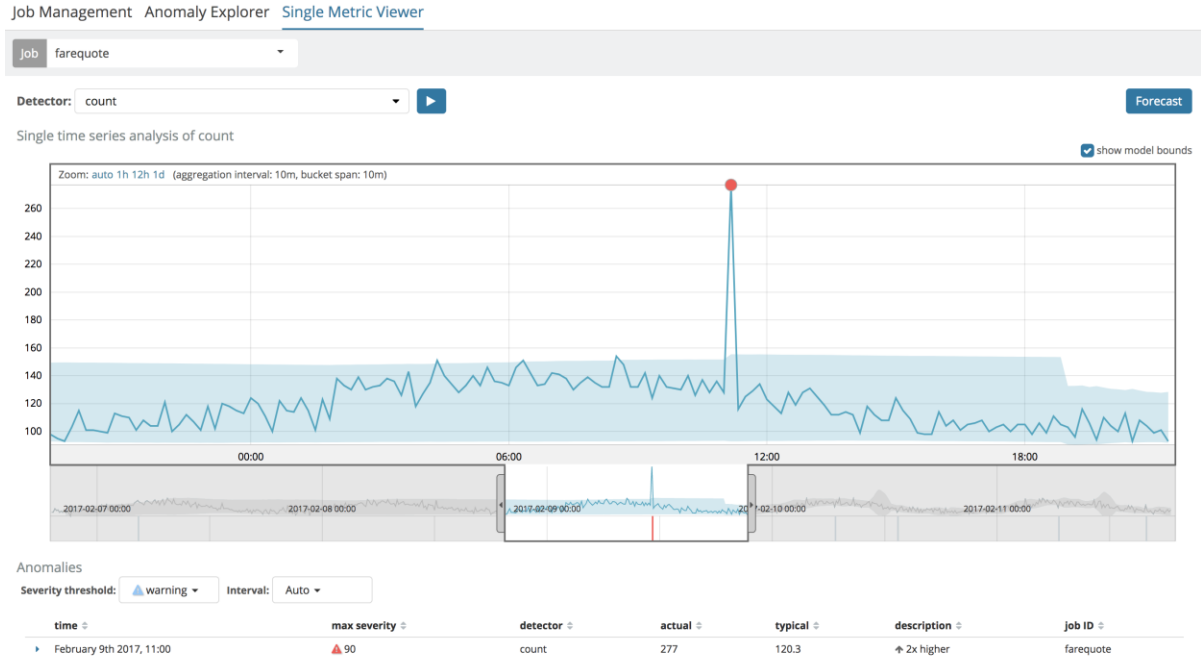
10) See animated learning

11) click “*View Results*”



# Steps to Complete

## 12) Zoom in on anomaly



# Lab 2: Advanced Jobs

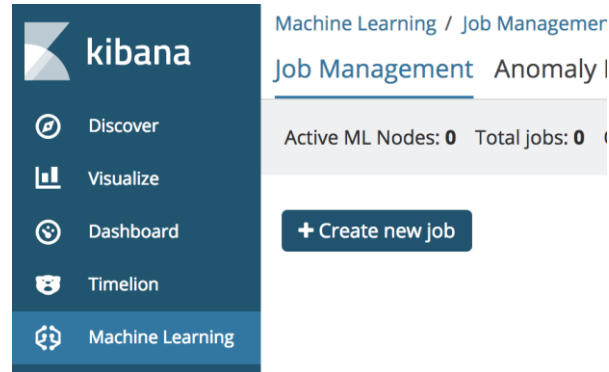


# Steps to Complete

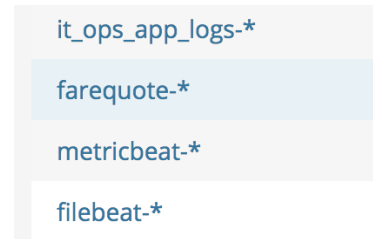
- Index: “farequote”
- ML job type: “advanced”
  - job name: “farequote\_response”
  - bucket\_span: 10m
  - Detector:
    - function: max
    - field\_name: responsetime
    - partition\_field\_name: airline
  - influencer: airline
- Run the job over the entire data set (data is not real-time)

# Steps to Complete

1) In Machine Learning,  
*Create new job*



2) Choose the “farequote-\*”  
*Index pattern*



# Steps to Complete

## 3) Choose Advanced Job Wizard



### Advanced

Use the full range of options to create a job for more advanced use cases.

# Steps to Complete

## 5) Name job

*“farequote\_response”*

### Create a new job

Job Details

Analysis Configuration

Datafeed

Edit JSON

Data Preview

#### Name ⓘ

farequote\_response

#### Description ⓘ

Job description

#### Job Groups ⓘ

Job Group

#### Custom URLs ⓘ

+ Add Custom URL

☐ Use dedicated index ⓘ

#### Model memory limit ⓘ

1024MB

Save

Cancel

# Steps to Complete

6) Set bucket\_span= *10m*

bucket\_span ⓘ

10m

7) Add a Detector:

function: *max*

field\_name: *responsetime*

partition\_field\_name: *airline*

Add new detector

Description ⓘ

max(responsetime) partition\_field\_name=airline

function ⓘ

max

x

field\_name ⓘ

responsetime

by\_field\_name ⓘ

Select...

over\_field\_name ⓘ

Select...

partition\_field\_name ⓘ

airline

exclude\_frequent ⓘ

Select...

[Help for max](#)

Add

Cancel



# Steps to Complete

7) Select *"airline"* as  
Influencer

8) Save job

9) Start datafeed

## Detectors

*max(responsetime) partition\_field\_name=airline*  

+ Add Detector

## Influencers

☒ airline

Custom influencer

+ Add

Save

Cancel

# Steps to Complete

8) Start at beginning of data  
leave “now” as end time

New Job 'farequote\_response' added x

Start datafeed for farequote\_response

Search start time

Start at beginning of data

Start now

Specify start time

Search end time

No end time (Real-time search)

Specify end time

2017-12-21 14:19:08.231

YYYY-MM-DD HH:mm:ss.SSS

< December 2017 >

Sun	Mon	Tue	Wed	Thu	Fri	Sat
					01	02
03	04	05	06	07	08	09
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30
31						

Start

Cancel

# Steps to Complete

9) wait until all 86274 events are processed













[Job Management](#) [Anomaly Explorer](#) [Single Metric Viewer](#)

Active ML Nodes: 1 Total jobs: 2 Open jobs: 0 Closed jobs: 1 Active datafeeds: 0

New Job 'farequote\_response' added

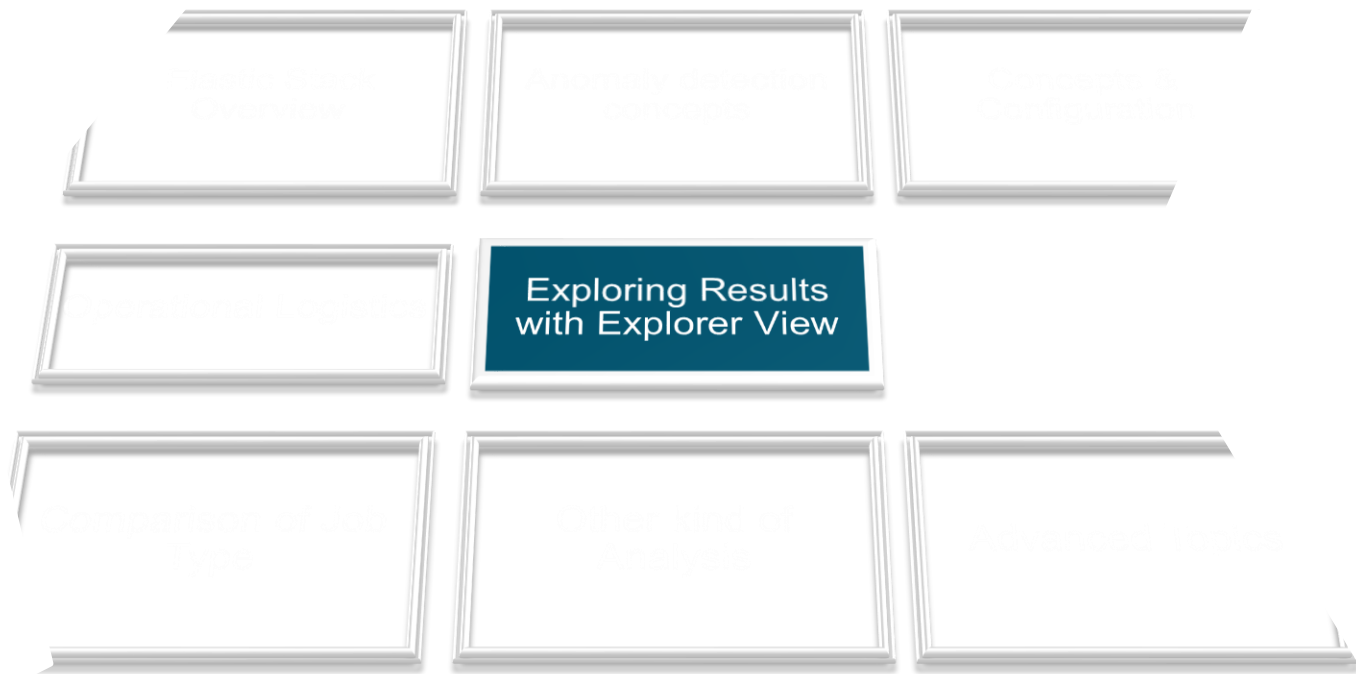
+ Create new job

Job filter

Job ID ↕	Description ↕	Processed records ↕	Memory status ↕	Job state ↕	Datafeed state ↕	Latest timestamp ↕	Actions
▶ farequote		719	ok	closed	stopped	2017-02-11 18:49:56	     
▶ farequote_response		86,274	ok	closing	stopped	2017-02-11 18:59:54	     

Page Size 10





# Explorer View of a Job










[Job Management](#) [Anomaly Explorer](#) [Single Metric Viewer](#)

Active ML Nodes: 1 Total jobs: 2 Open jobs: 0 Closed jobs: 1 Active datafeeds: 0

New Job 'farequote\_response' added

+ Create new job

Job filter

Job ID	Description	Processed records	Memory status	Job state	Datafeed state	Latest timestamp	Actions
farequote		719	ok	closed	stopped	2017-02-11 18:49:56	     
farequote_response		86,274	ok	closing	stopped	2017-02-11 18:59:54	     

Page Size 10

Or here

click here

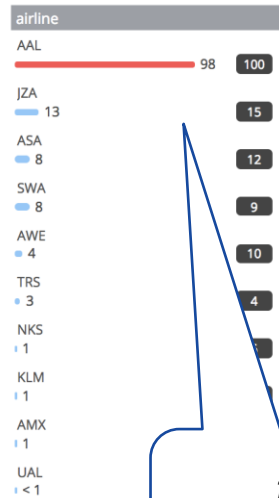
# Anomaly Explorer

Machine Learning / Anomaly Explorer

Job Management Anomaly Explorer Single Metric Viewer

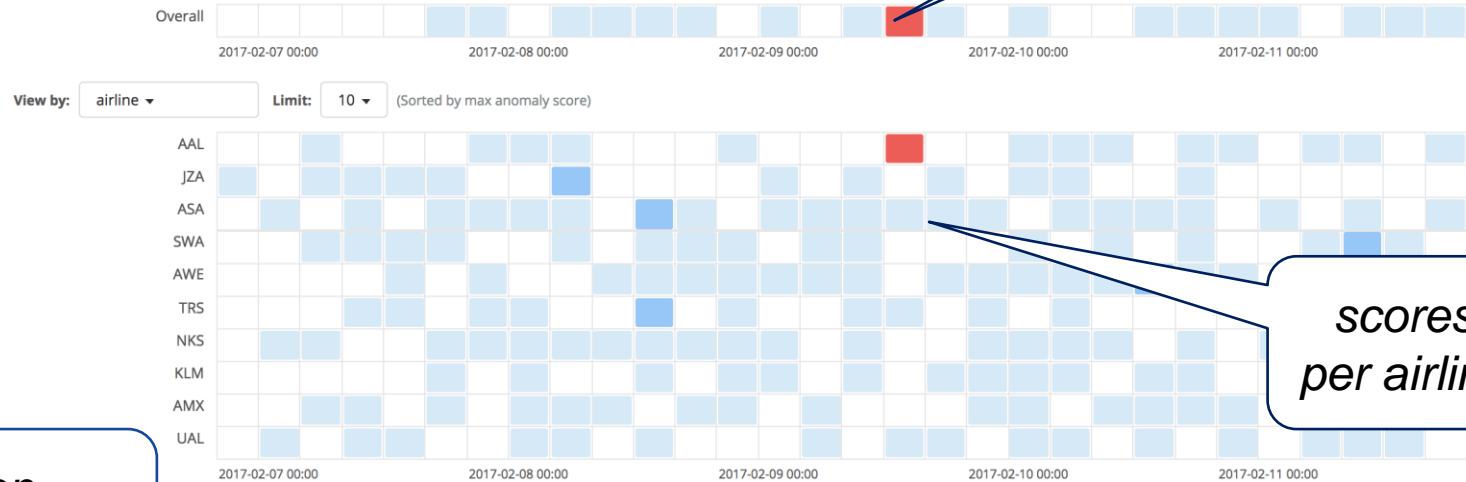
Job: farequote\_response

## Top Influencers



*top influencers*

## Anomaly timeline



*scores per airline*

*job level score*

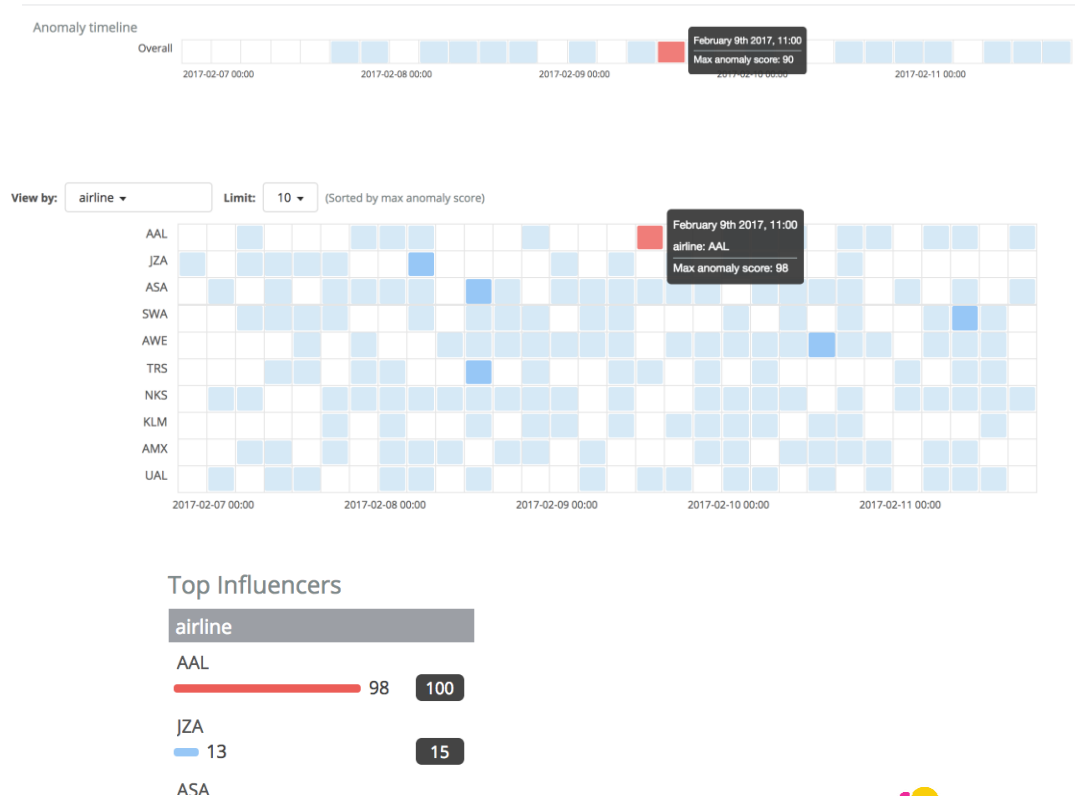
	max severity	detector	found for	influenced by	actual	typical	description	job ID	links
February 9th 2017	97	max(responsetime) partitionfield=airline	AAL	airline: AAL	378.2	137.9	3x higher	farequote_response	<a href="#">Open link</a>
February 8th 2017	13	max(responsetime) partitionfield=airline	JZA	airline: JZA	1042.1	1011.4	Unusually high	farequote_response	<a href="#">Open link</a>

# Concept: What is an Influencer?

- An Influencer is a field, selected at configuration time, that would be a logical **entity "to blame"** if an anomaly were to exist
- Doesn't have to be a field in the actual detector, but **fields used to split the data are often good candidates**
- Will get its own score based upon how influential that entity is on the anomaly

# Scoring

- Overall Job score is 90
  - How unusual is that bucket, given all airlines?
- Detector score is 98
  - How unusual is the response time of airline=AAL?
- airline=AAL is the top influencer in this time range
  - 98 is the max anomaly score
  - 100 is the sum of anomaly scores in this time range

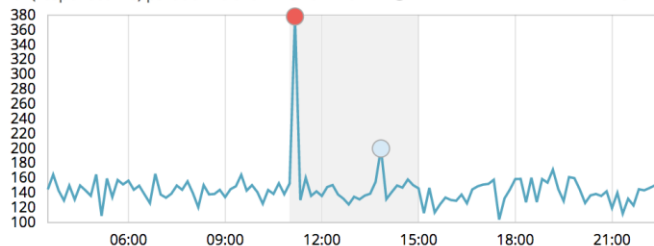


# Anomaly Details

## Anomalies

Severity threshold: ▲ warning ▼ Interval: Auto ▼ ☒ Show Charts

max(responsetime) partitionfield=airline - airline AAL ⓘ [View](#)



time	max severity	detector	found for	influence	actual	typical	description	job ID	links
February 9th 2017, 11:00	▲ 97		max(responsetime) partitionfield=airline	AAL	378.2	137.9	▲ 3x higher	farequote_res	

### Description:

critical anomaly in max(responsetime) partitionfield=airline found for airline AAL

### Details on highest severity anomaly:

airline: AAL  
time: February 9th 2017, 11:10:00 - February 9th 2017, 11:20:00  
function: max  
fieldName: responsetime  
actual: 378.2  
typical: 137.9  
job ID: farequote\_response  
probability: 3.95474803120103e-22

### Influenced by:

airline AAL

*view of  
response  
time for AAL*

*actual vs.  
"typical"*

*raw  
probability*

# Lab 3: Multi-Metric Jobs

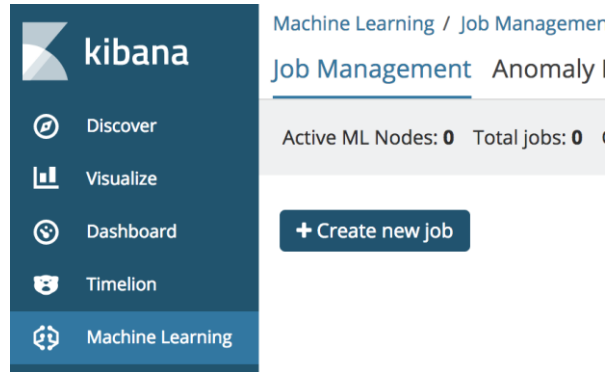
# Steps to Complete

- Again, use index: *“farequote”*:
- Job type: *“multi-metric”*
  - Re-create the *“max(responsetime) per airline”* job
- Also add *“count per airline”* in the same job

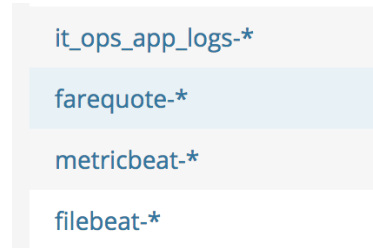


# Steps to Complete

1) In Machine Learning,  
*Create new job*



2) Choose the “farequote-\*”  
*Index pattern*



# Steps to Complete

3) pick Multi-Metric  
Job Wizard



## Multi metric

Detect anomalies in multiple metrics by splitting a time series by a categorical field.

# Steps to Complete

## 4) Choose

- *event rate, count*
- *responsetime, max*

## 5) Bucket span: *10m*

## 6) Split Data: *airline*

## 7) Click “*use full farequote data*”

## 8) Name: “*farequote\_multi*”

## 9) Click “*Create Job*”

New job from index pattern farequote-\*

Chart interval: 15m Use full farequote-\* data

**Job settings**

Fields

☒ event rate

Count

☒ responsetime

Max

☐ airline

Distinct count

☐ Sparse data ⓘ

Split Data

Remove split

t airline

Key Fields (Influencers)

t airline ✕

Bucket span ⓘ

10m

Estimate bucket span

Job Details

Name ⓘ

farequote\_multi

Description ⓘ

job description

Job Groups ⓘ

Job Group

Advanced ⓘ

Create job

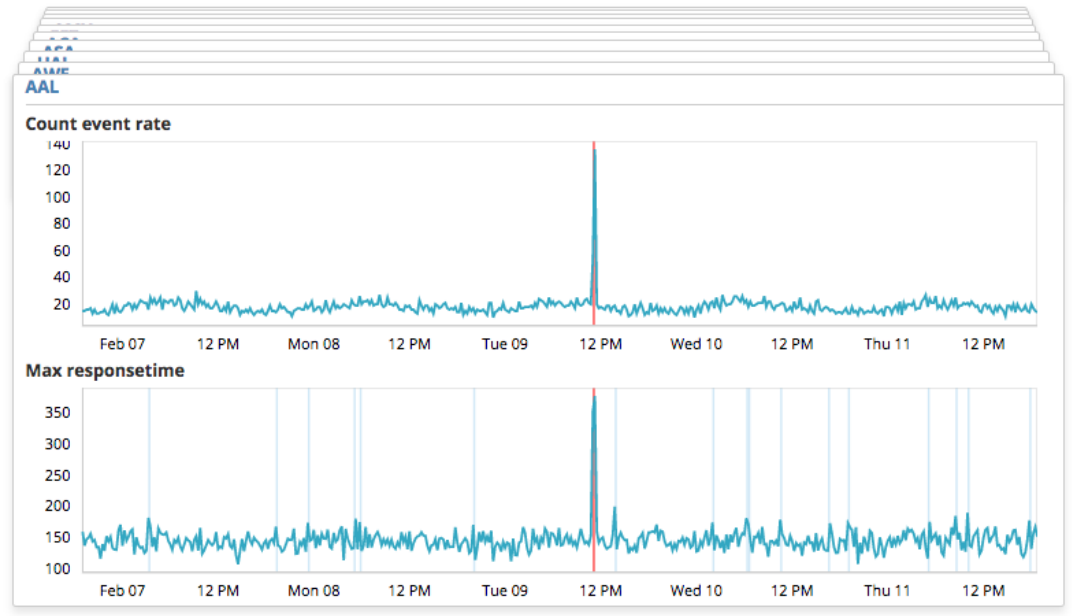


# Steps to Complete

10) See animated learning

11) Click “View Results”

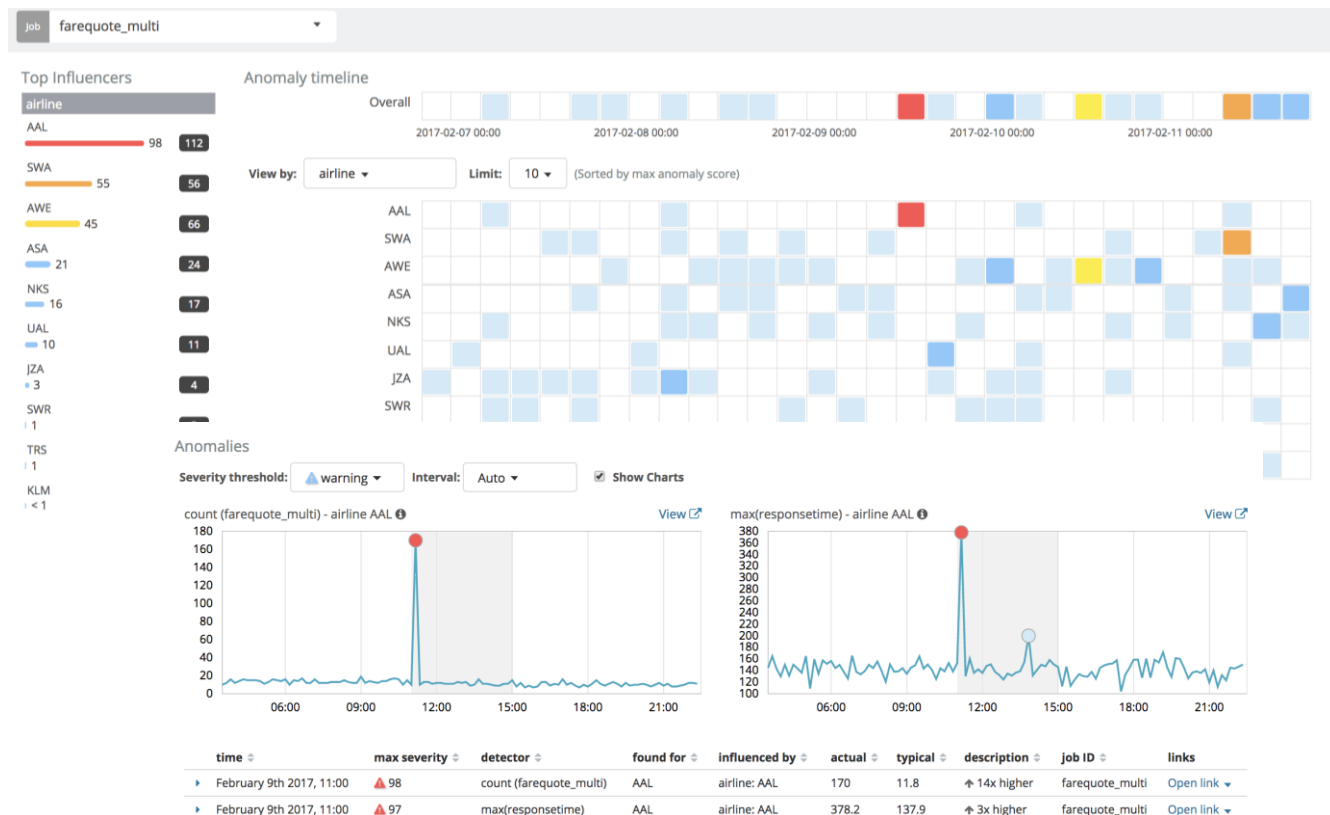
Data split by airline



# Steps to Complete

## 12) Result:

*anomalies for AAL  
in both count  
and response time*



# Lab 4: Multi-Job Analysis

# Steps to Complete

- Create and [Advanced job](#) for:
  - “it\_ops\_app\_logs-”
  - Create a “count by mlcategory” job for the log events
    - use “message” as the categorization\_field\_name
- Create [Multi-Metric jobs](#) for the following indices:
  - it\_ops\_sql-\*
  - it\_ops\_network-\*
- Create [Single-Metric job](#) for the following :
  - it\_ops\_kpi-\*
- Group all jobs into the [same Job Group](#) name
- View all jobs overlaid in the Explorer View

# Steps to Complete

- For index:it\_ops\_app\_logs
  - create an advanced job
  - put job in a group
- make sure you choose “message” for categorization\_field\_name
- detector is: count with by\_field\_name of “mlcategory”

**New**

**Name** ⓘ

app\_logs

**Description** ⓘ

job description

**Job Groups** ⓘ

my\_group ✕

Machine Learning / Job Management / Create New Job / Advanced Job Configuration

Create a new job

Job Details

bucket\_span ⓘ

15m

summary\_count ⓘ

Select...

categorization\_field\_name ⓘ

message

Categorization Fields

+ Add Categorization Field

**Detectors** ⓘ

+ Add Detector

**Add new detector**

**Description** ⓘ

count by mlcategory

function ⓘ	field_name ⓘ	by_field_name ⓘ
count ✕	Select...	mlcategory ✕

over_field_name ⓘ	partition_field_name ⓘ	exclude_frequent ⓘ
Select...	Select...	Select...

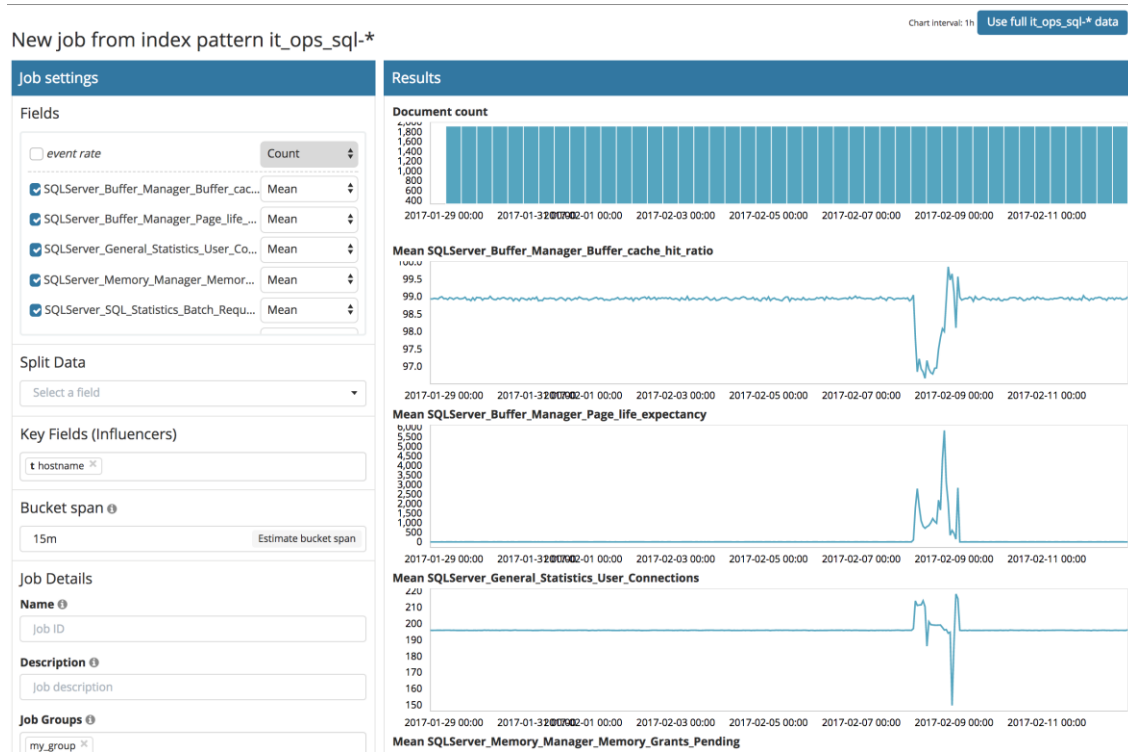
[Help for count ⓘ](#)

**Add** **Cancel**



# Steps to Complete

- For index:it\_ops\_sql-\*
  - create multi-metric job
  - “mean” for all SQL metrics
- Also put job in same Job Group as the app logs job



# Steps to Complete

- For index:it\_ops\_network-\*
  - create multi-metric job
  - “mean” for all metrics
- Also put job in same Job Group as the app logs job

New job from index pattern it\_ops\_network-\*

**Job settings**

**Fields**

☐ event rate

Count

☒ In\_Broadcast\_Pkts

Mean

☒ In\_Discards

Mean

☒ In\_Errors

Mean

☒ In\_Octets

Mean

☒ Out\_Broadcast\_Pkts

Mean

**Split Data** Remove split

physicalhost

**Key Fields (Influencers)**

physicalhost

**Bucket span**

15m

Estimate bucket span

**Job Details**

**Name**

network\_metrics

**Description**

job description

**Job Groups**

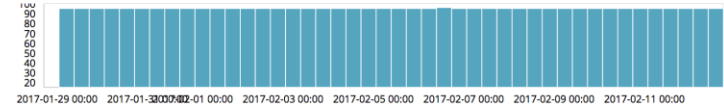
my\_group

**Advanced**

Create Job

## Results

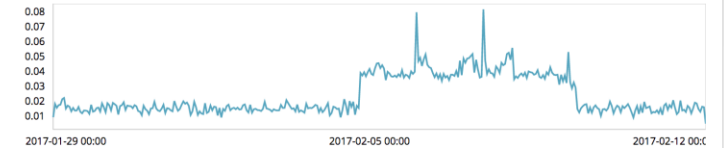
### Document count



### Data split by physicalhost

esxsxserver1.acme.com

#### Mean In\_Broadcast\_Pkts



#### Mean In\_Discards



#### Mean In\_Errors

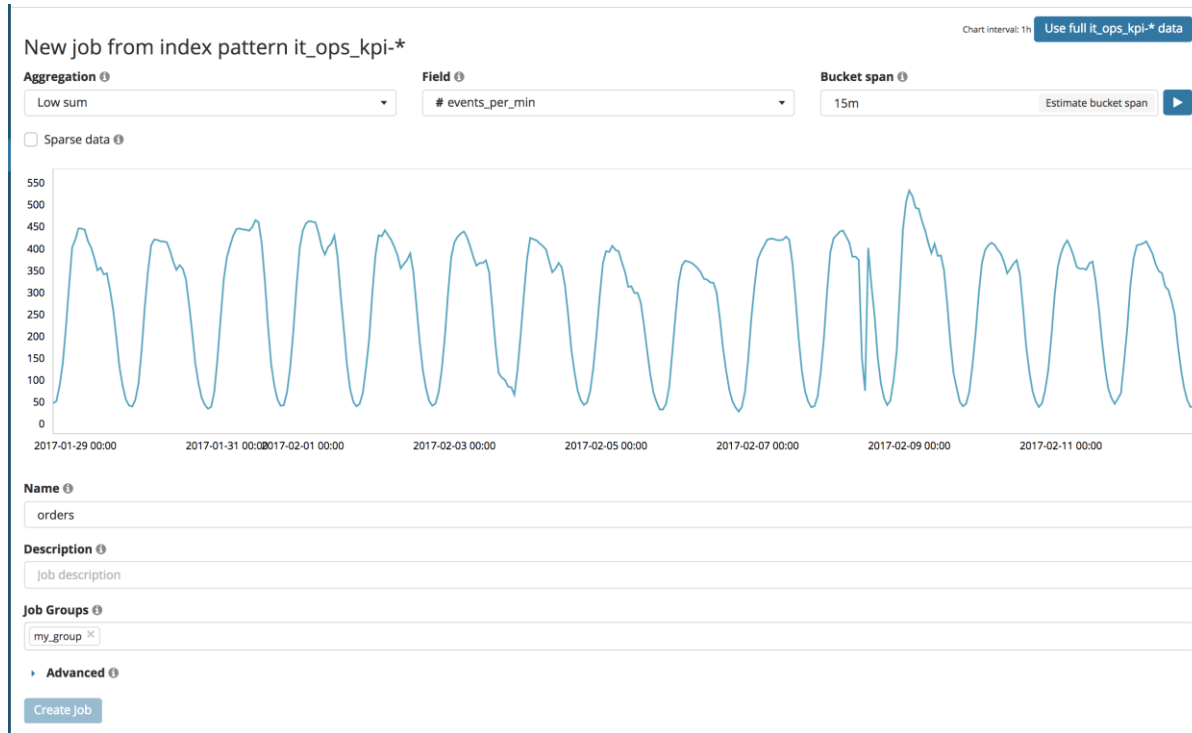


#### Mean In\_Octets



# Steps to Complete

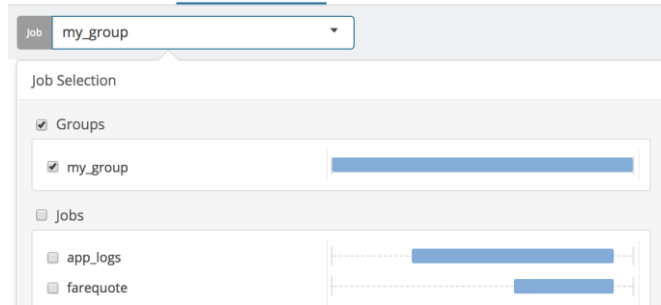
- For index:it\_ops\_kpi-\*
  - create single-metric job
  - “low\_sum” for field “events\_per\_min”
- Also put job in same Job Group as the app logs job



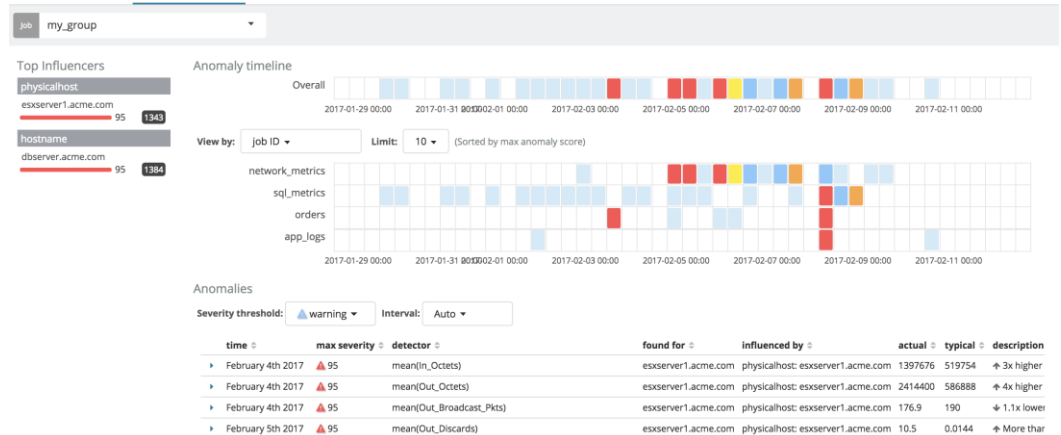
# Steps to Complete

- View all jobs in Group
- See correlated anomalies

Job Management **Anomaly Explorer** Single Metric Viewer



Job Management **Anomaly Explorer** Single Metric Viewer





# THANK YOU

