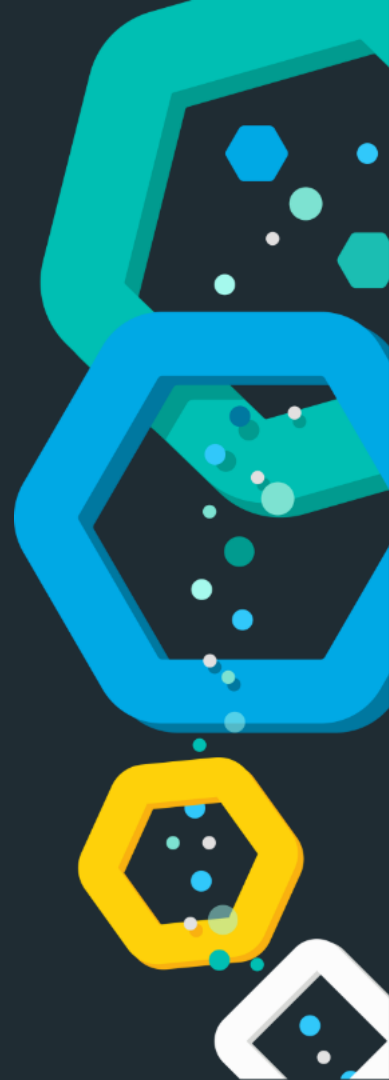




# X-Pack Machine Learning Labs Guide

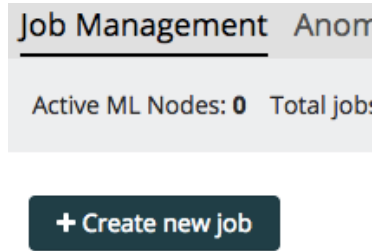
---



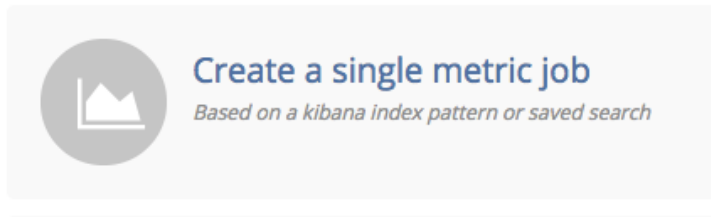
# Lab 1: The Simplest Job

# Steps to Complete

1) In Machine Learning,  
Create new job



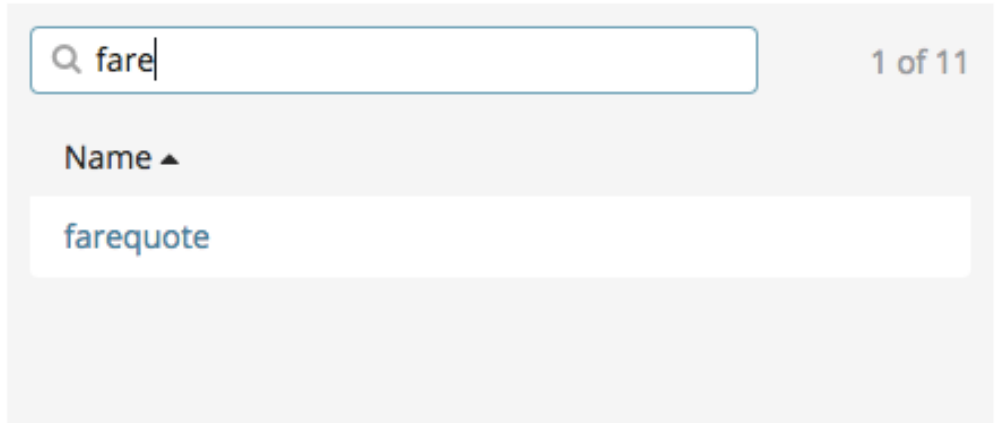
2) Choose single metric



# Steps to Complete

3) pick farequote index

## From a New Search, Select Index



A screenshot of the Elasticsearch index selection interface. At the top, there is a search bar with a magnifying glass icon and the text "fare". To the right of the search bar, it says "1 of 11". Below the search bar, the word "Name" is followed by a small upward-pointing triangle. Below this, the index name "farequote" is displayed in a blue font, indicating it is the selected index.

# Steps to Complete

4) choose the “Count” aggregation

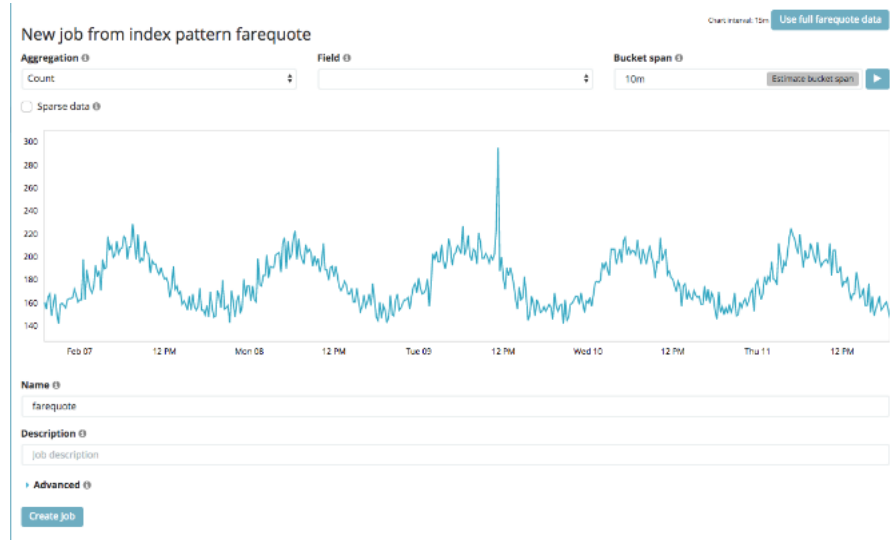
5) select 10m for bucket span

6) leave “field” blank (we don’t count fields, we’re counting documents)

7) click the “use full farequote data” button

8) name job “farequote”

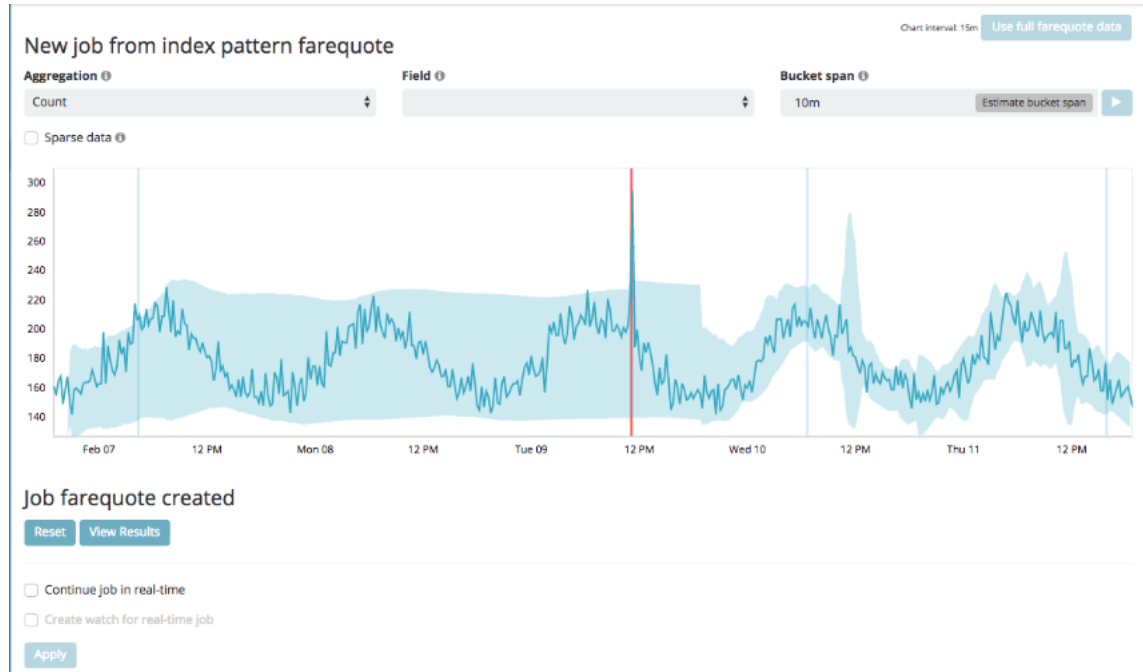
9) click “Create Job”



# Steps to Complete

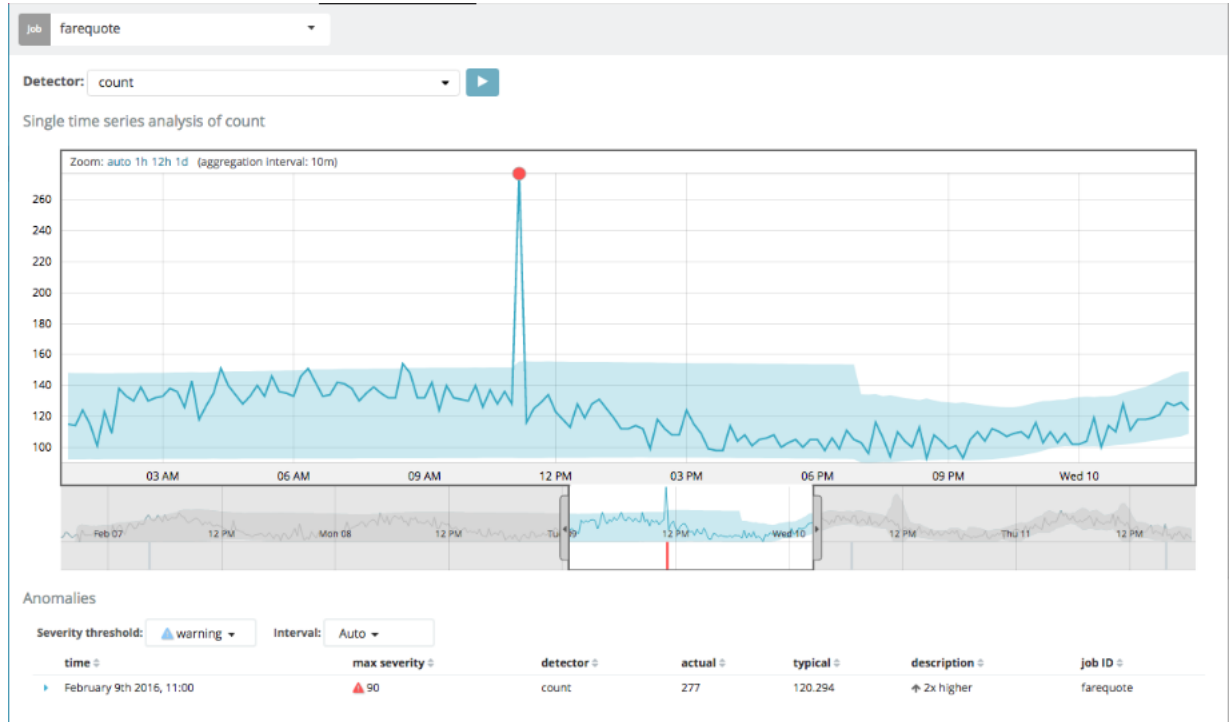
10) See animated learning

11) click “View Results”



# Steps to Complete

## 12) Zoom in on anomaly



# Lab 2: Advanced Jobs

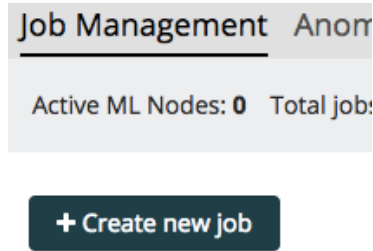


# Steps to Complete

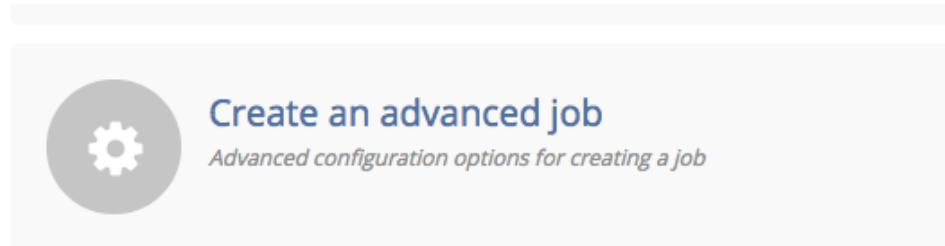
- Using the “farequote” data set:
- Get familiar with how the raw data looks in Kibana Discover
- Load the ML home screen and create a new “advanced” job to analyze unusually slow response times per airline
  - choose a bucket\_span of 10m
  - pick an appropriate way to split data (hint: partition using airline)
  - select airline as influencer
- Run the job over the entire data set (data is not real-time)

# Steps to Complete

1) Create new job



2) Choose advanced



# Steps to Complete

3) Pick farequote index

☒ Input index

☐ Choose index from list

Index

farequote

Types

☒ responsetime

Time-field name

@timestamp

Next

# Steps to Complete

5) Name job  
“farequote\_response”

## Create a new job

[Job Details](#)[Analysis Configuration](#)[Datafeed](#)[Edit JSON](#)[Data Preview](#)

**Name** ⓘ

**Description** ⓘ

**Custom URLs** ⓘ  
[+ Add Custom URL](#)

☐ **Use dedicated index** ⓘ

[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

field\_name ⓘ

responsetime

by\_field\_name ⓘ

over\_field\_name ⓘ

partition\_field\_name ⓘ

airline

exclude\_frequent ⓘ

[Help for max](#)



Add

Cancel

# Steps to Complete

- 7) select Influencer,  
save job  
start datafeed

## Detectors

`max(responsetime) partition_field_name=airline`  

+ Add Detector

## Influencers

- ☐ @version.keyword
- ☒ airline
- ☐ host.keyword
- ☐ path.keyword
- ☐ type

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

2017-07-10 07:07:27.683

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

< July 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 86275 events are processed

+ Create new job

job filter



Job ID	Description	Processed records	Memory status	Job state	Datafeed state	Latest timestamp	Actions
farequote		720	ok	closed	stopped	2016-02-11 18:59:54	
farequote_response		86,274	ok	closed	stopped	2016-02-11 18:59:54	

Page Size 10



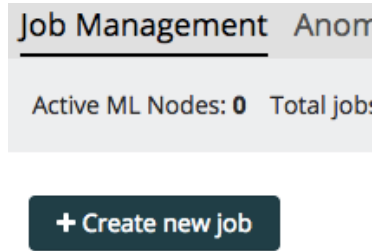
# Lab 3: Multi-Metric Jobs

# Steps to Complete

- Again, using the “farequote” data set:
- Re-create the “max(responsetime) per airline” job using a “multi-metric” job
- Also add “count per airline” in the same job

# Steps to Complete

1) Create new job



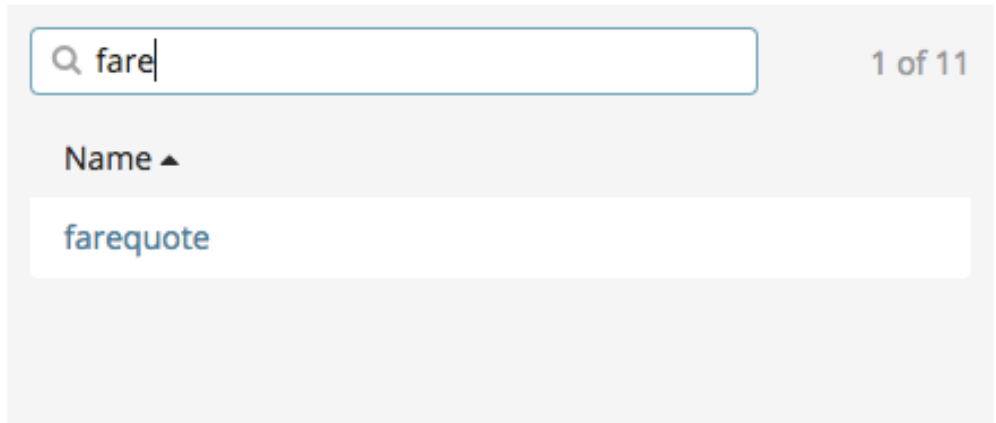
2) Choose multi metric



# Steps to Complete

3) pick farequote index

## From a New Search, Select Index



A screenshot of the Elasticsearch index selection interface. At the top, there is a search bar with a magnifying glass icon and the text "fare". To the right of the search bar, it says "1 of 11". Below the search bar, the word "Name" is followed by a small upward-pointing triangle. A list of index names is displayed below, with "farequote" highlighted in blue.

Name ▲
farequote

# Steps to Complete

4) choose

- event rate, count
- responsetime, max

5) select 10m for bucket span

6) Split Data by airline  
(influencer for airline is chosen for you)

7) click “use full farequote data”

8) name job “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

☐ Sparse data

### Split Data

airline

### Key Fields

☐ \*\_ip  
☐ \*\_port  
☐ @version.keyword  
☐ \_index  
☒ airline  
☐ host.keyword  
☐ path.keyword

### Bucket span

10m Estimate bucket span

### Job Details

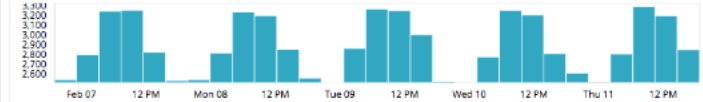
#### Name

farequote\_multi

#### Description

## Results

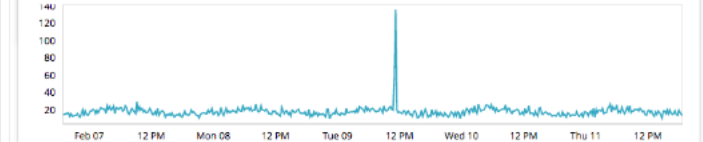
### Document count



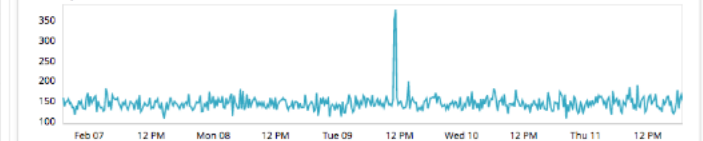
### Data split by airline



### Count event rate



### Max responsetime

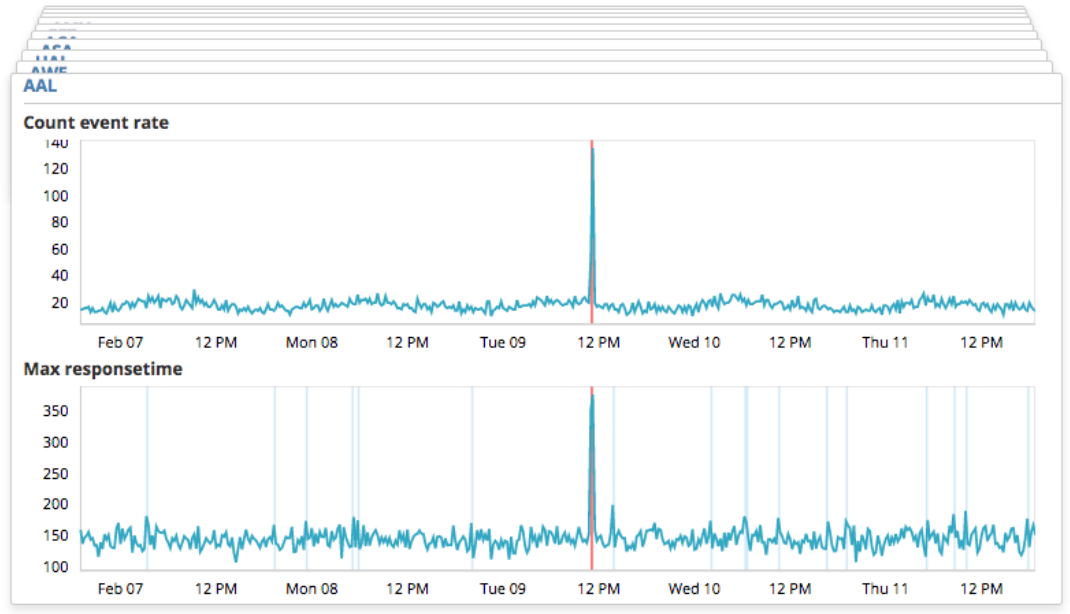


# Steps to Complete

10) See animated learning

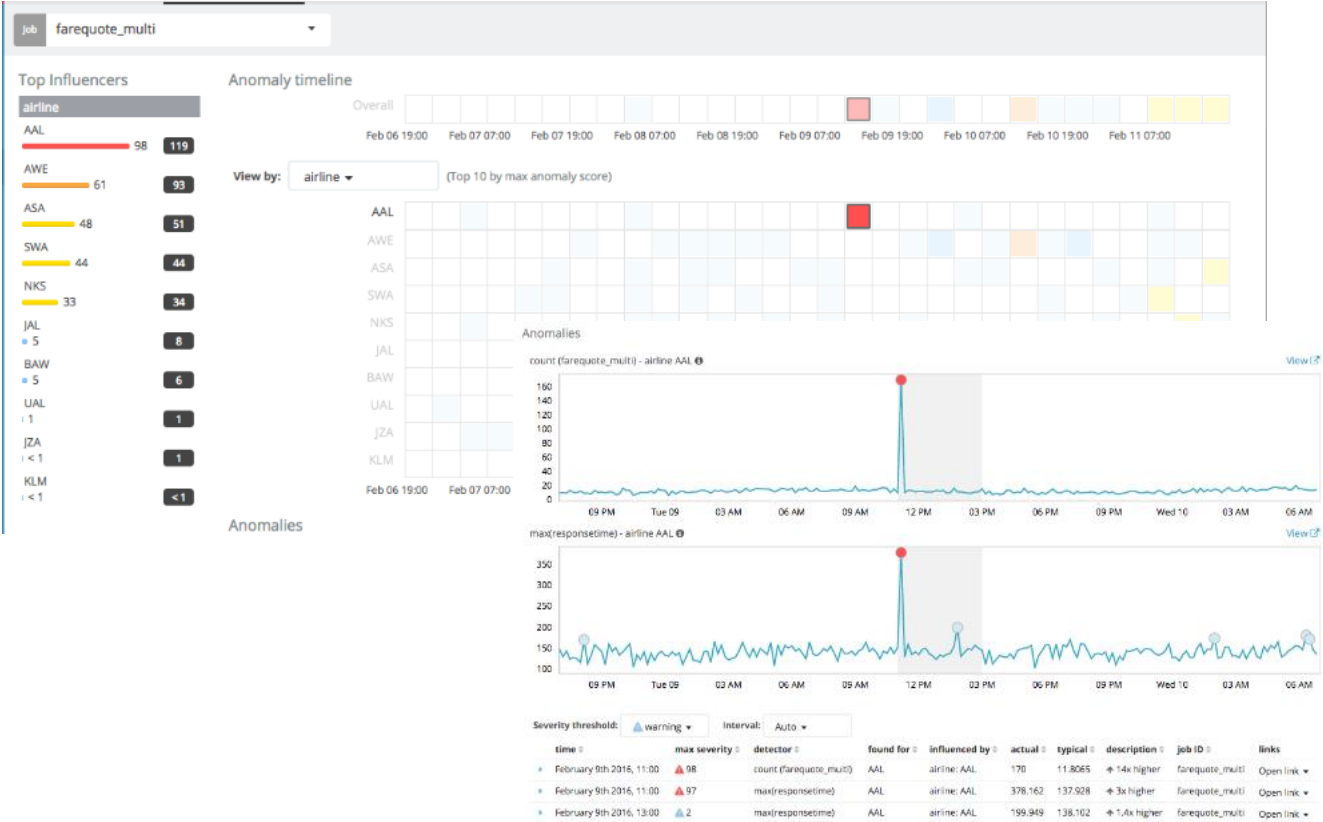
11) click “View Results”

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

- Using the “it\_ops\_logs” data set:
  - Create a “count by mlcategory” job for the log events
    - use “message” as the categorization\_field\_name
- Using the “it\_ops\_metrics” data set:
  - Create a “mean(metricvalue) by metricname” job for the metrics
- View both jobs overlaid in the Explorer View

# Steps to Complete

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

☒ Input index  
☐ Choose index from list

Index

Types  
☒ logs

Time-field name

Next

Create a new job

job Details Analysis

bucket\_span @  
10m

summary.count\_field\_name @  
message.keyword

categorization\_field\_name @  
message.keyword

Categorization Filters @  
Add Categorization Filter

Detectors @  
Add Detector

Add new detector

Description @  
count by mlcategory

function @	field_name @	by_field_name @
count	-	mlcategory
over_field_name @	partition_field_name @	exclude_frequent @

Help for count ?

Add Cancel

# Steps to Complete

- Answer
  - For index:it\_ops\_metrics
    - create multi-metric job
    - mean of metricvalue split on metricname.keyword

New job from index pattern it\_ops\_metrics

Chart interval: 50m Use full it\_ops\_metrics data



# Steps to Complete

- Your goal is to get this View:

