Lucidworks

LAB: Building a Recommender

Prerequisites

- Register for Kaggle
  - Instructions here: Registering for Kaggle

- LAB: Spark and Data Prep
- LAB: Working with Signals

Getting Started

Lucidworks

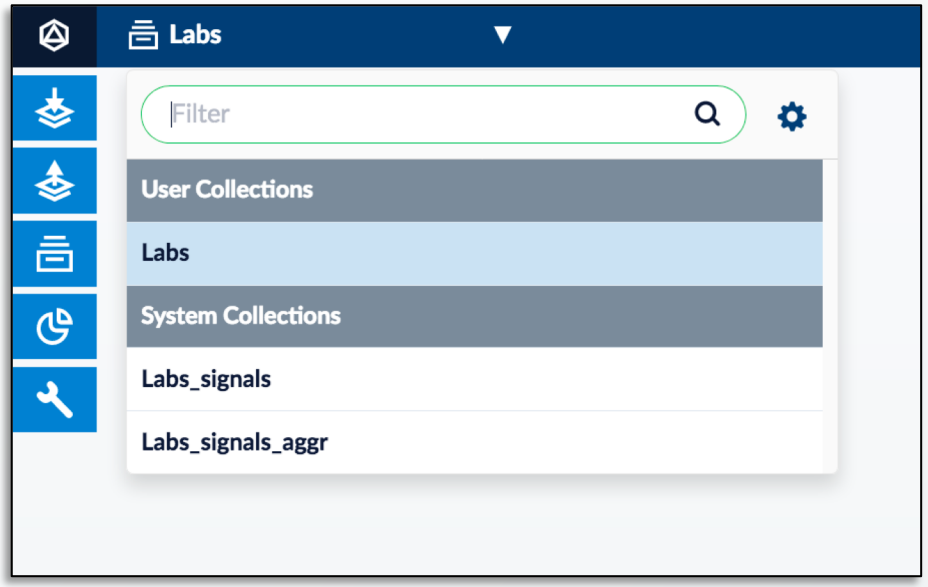
This lab assumes that you are using an AWS virtual machine provided by Lucidworks Training. If this is not the case, your filepaths and IP addresses will vary significantly from those shown.

- In a bash/shell terminal, start Fusion
  `./fusion/4.0.1/bin/fusion start`

- In a web browser, open Fusion Admin
  `<your-vm-ip>:8764`

Lucidworks

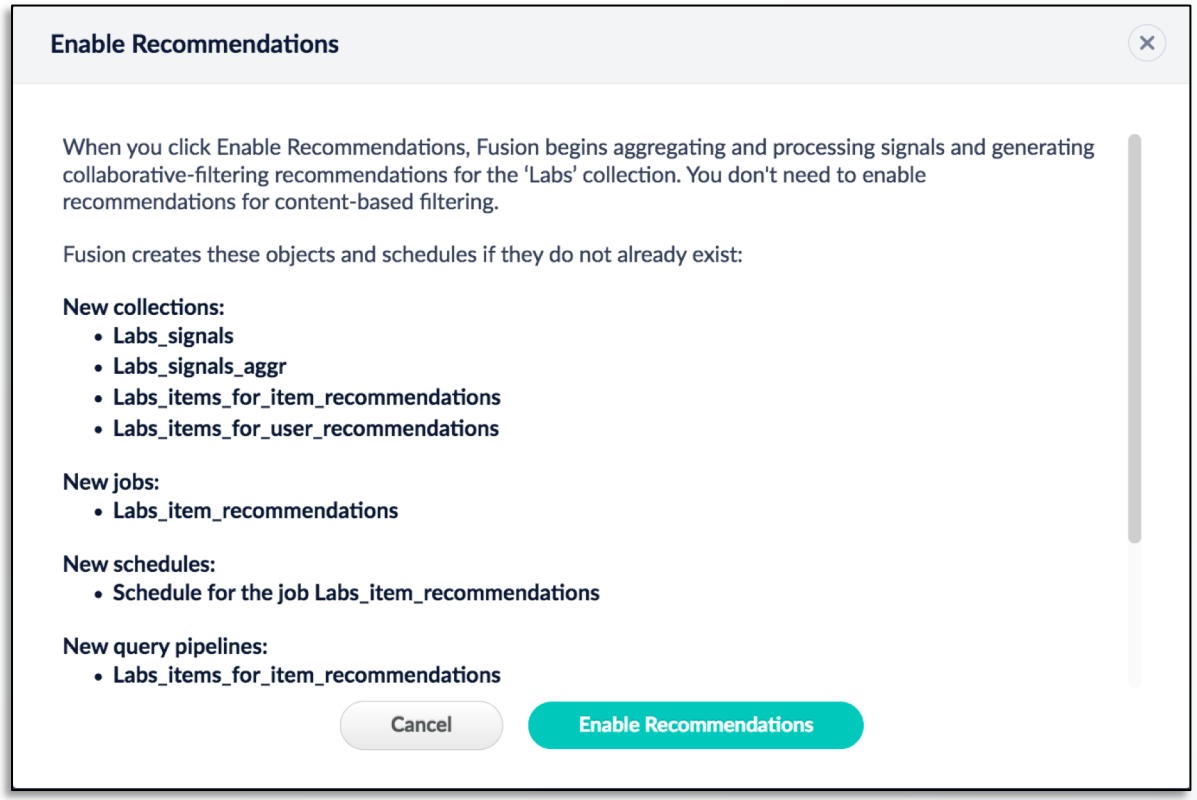
- Enter your username and password
  *The default is **admin** and **Lucidworks1***
- Click into the **Labs** Fusion App
- In the top left dropdown, change to the **Labs** collection

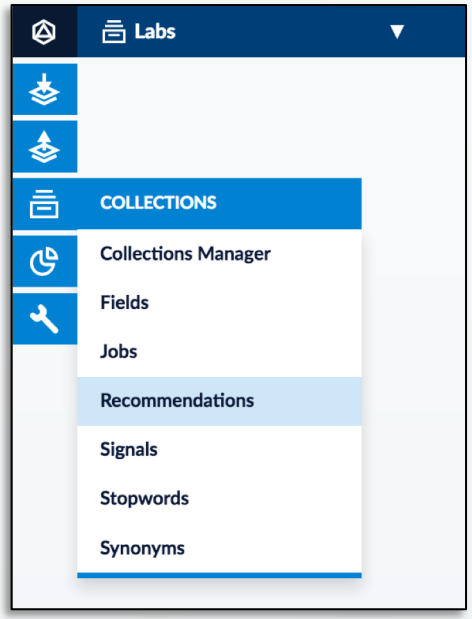


Lucidworks

- In the left side menu, go to **COLLECTIONS > Recommendations**

Be sure to read and understand the dialog box that appears.





- Click **Enable Recommendations**

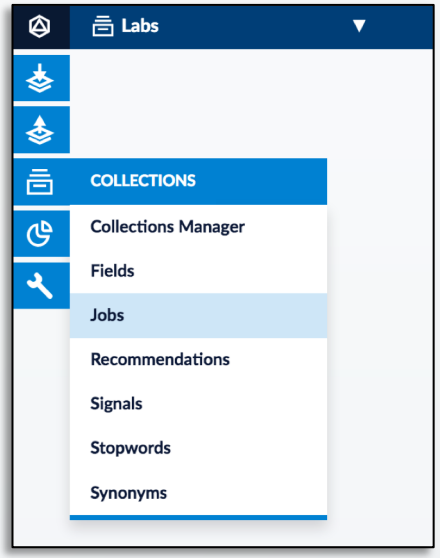
Configuring Recommendations

Configuring Recommender Jobs

Lucidworks

*In this lab, we will set up **items-for-user** recommendations (i.e., "Recommended for you"), as well as **item-for-item** recommendations (i.e., "Customers who bought this also bought XYZ"), on the Best Buy catalog.*

There are two main tasks involved in this. First, we will use existing signals to determine each past user's level of interest in each item. Second, we will use these past user-item preferences to create a user-item matrix, which can be used both to recommend items to new users and recommend other items that tend to collocate with the one currently clicked.



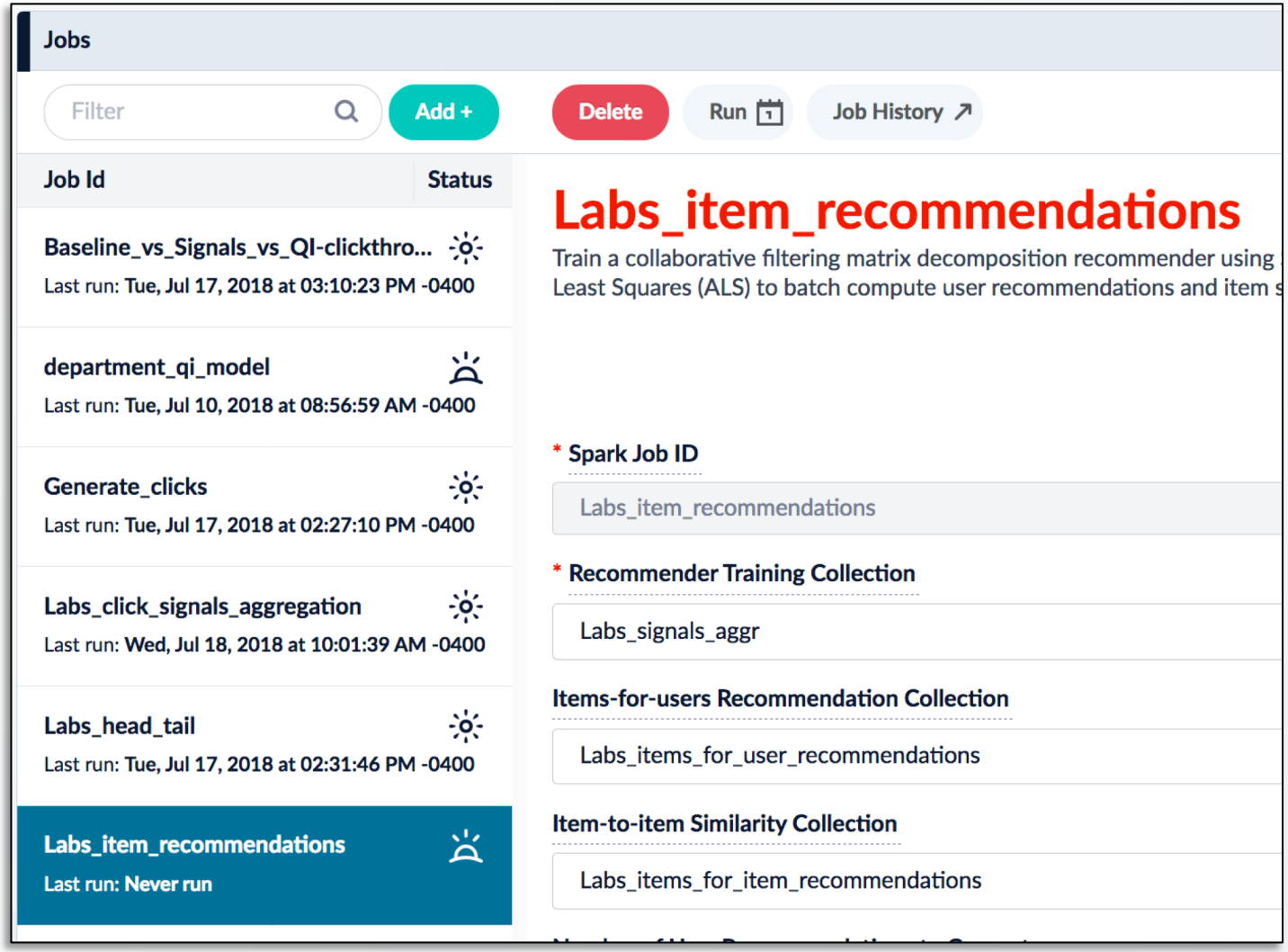
- In the left side menu, go to **COLLECTIONS > Jobs**
- Select **Labs_item_recommendations**

Configuring Recommender Jobs

Lucidworks

This job is the one that creates the user-item matrix.

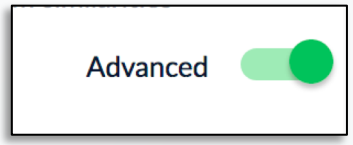
*If you click the **Run** button, you will see that it is scheduled to automatically run after the **Labs_user_item_preferences_aggregation** job, which is the one that calculates past user-item preferences.*



Configuring Recommender Jobs

Lucidworks

- Toggle the **Advanced** switch at the top right
- Fill out the job parameters according to the following table:



Parameter	Value	Explanation
Training Data Sampling Fraction	0.05	*Downsampling makes the job finish faster. You would probably not do this in production, though, as this can degrade output quality.*
Item Metadata Fields	department	*Additional fields to add to the aggregation documents. These are not used in the matrix calculations, but can be used to fine-tune the time usage of recommendations.*
Item Metadata Collection	Labs	*Collection from which to pull the metadata fields*
Item Metadata Join Field	id	*Join key between aggregate document ID (default: **doc_id_s** in the **Labs_signals_aggr** collection) and metadata ID (in this case, from **Labs**)*

- **Save** the **Labs_items_recommendations** job. Do not run it yet!

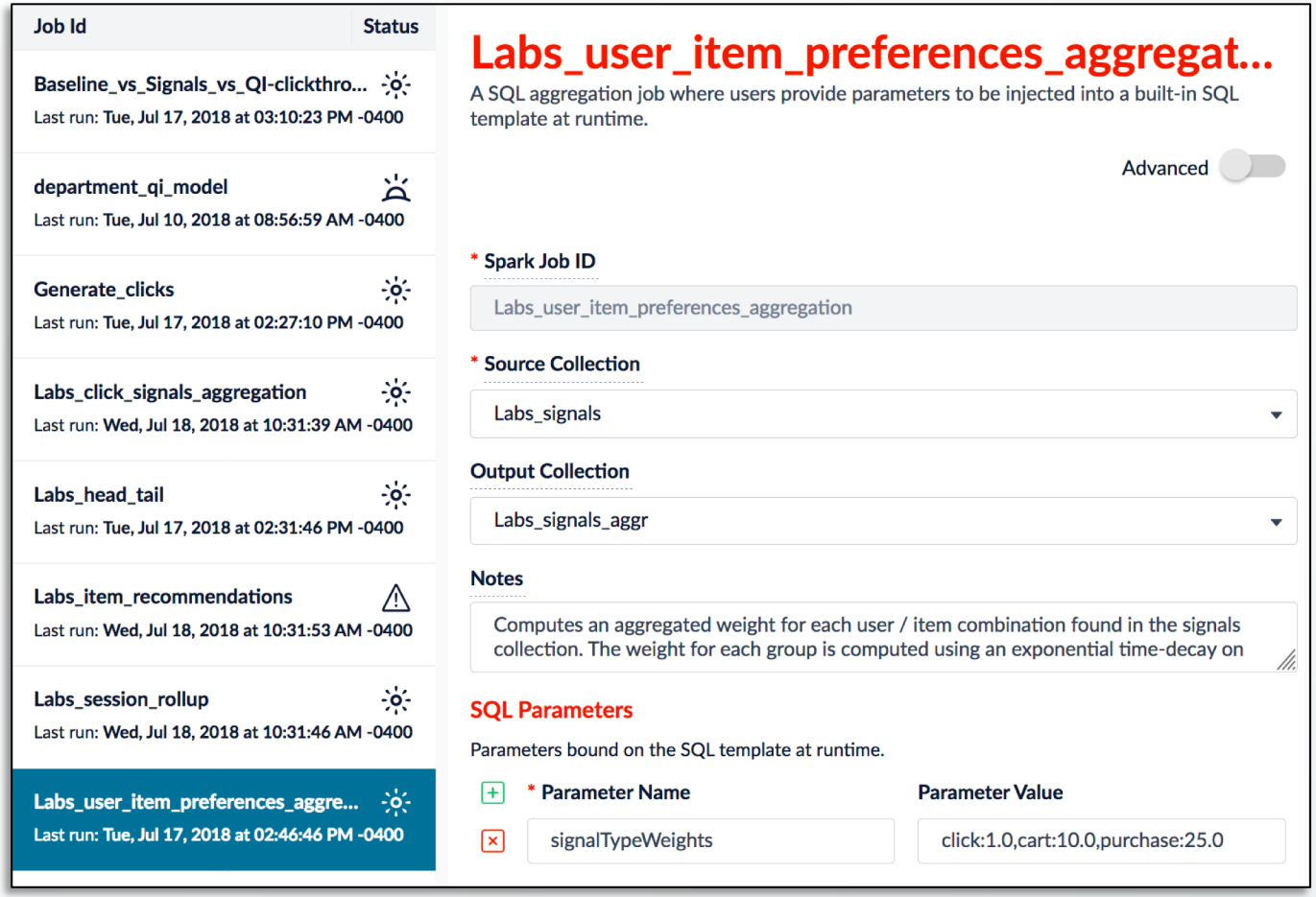
Configuring Recommender Jobs

Lucidworks

- Open the **Labs_user_item_preferences_aggregation** job

As mentioned before, this job is the one that will calculate user-item preferences based on past behavior.

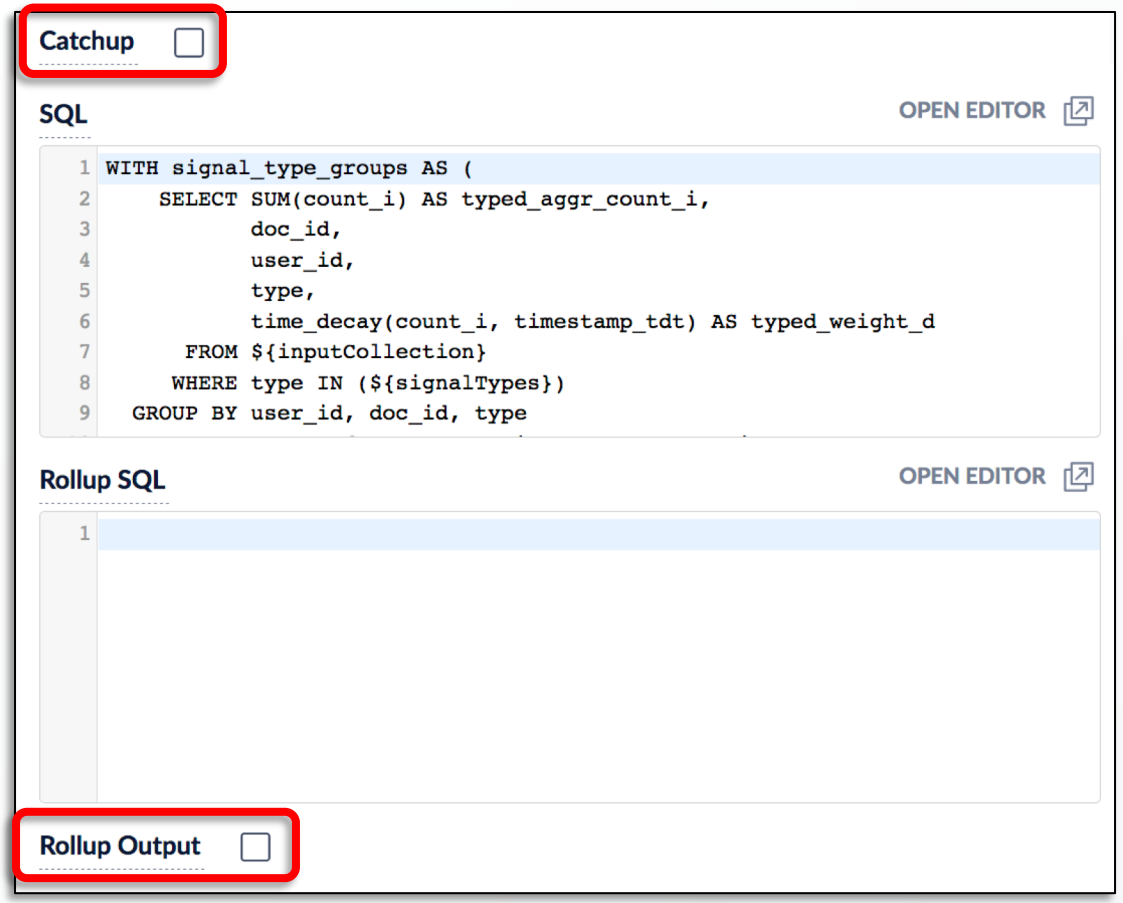
*The output of this will be used by **Labs_item_recommendations** to create a user-item matrix*



Configuring Recommender Jobs

Lucidworks

- Toggle the **Advanced** switch in the top right



- Uncheck the boxes **Catchup** and **Rollup Output**

These parameters are only functional for subsequent runs of the job.

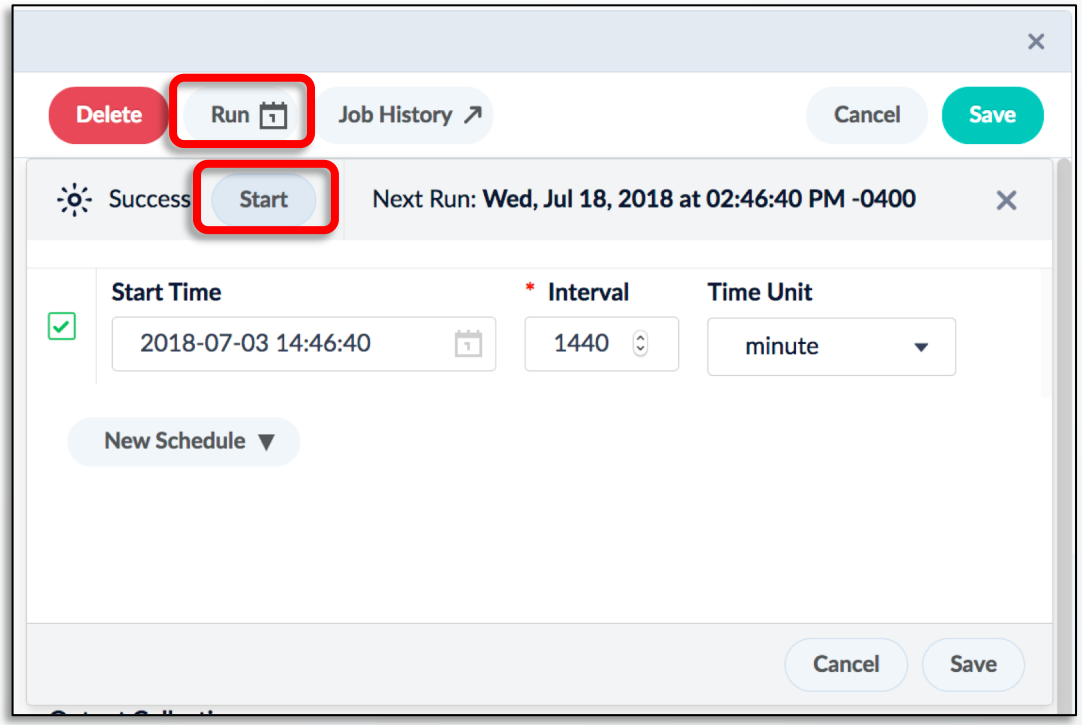
- **Save** the modified job

Running the Recommender Jobs

Lucidworks

- Click **Run > Start**

*The job should take ~30 seconds to finish. Once it does, the **Labs_item_recommendations** job will kick off. That job will take a bit longer; ~2 minutes.*

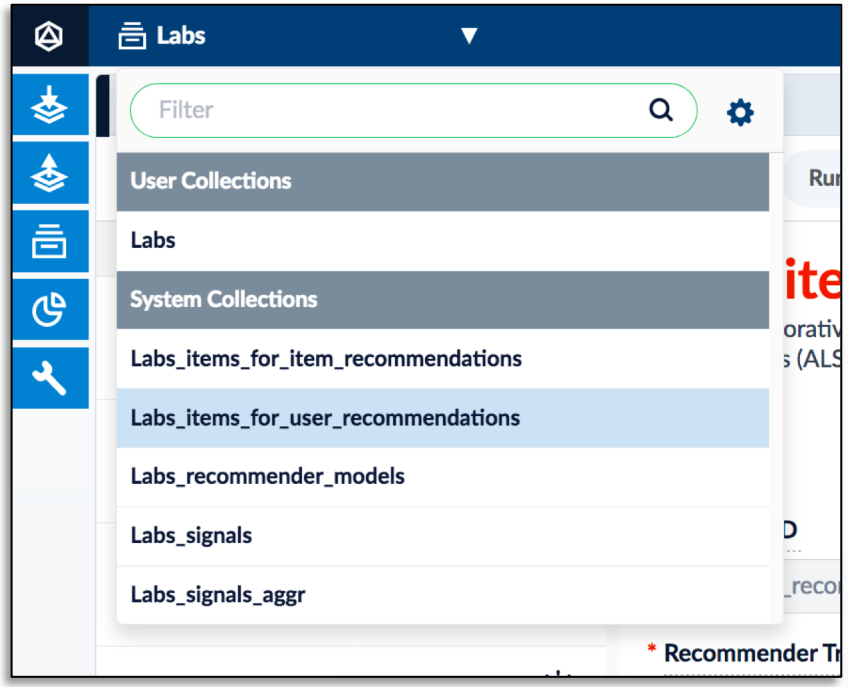


Verifying Successful Model Build

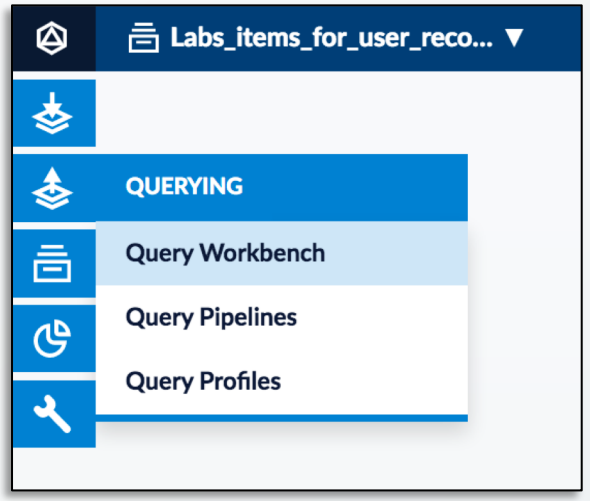
Verifying Items-for-User Data

Lucidworks

- In the top left dropdown of Fusion Admin, change to the **Labs_items_for_user_recommendations** collection



- In the left side menu, go to **QUERYING > Query Workbench**



Verifying Items-for-User Data

Lucidworks

- Click **show fields** under any document

c2cb85bd-7679-4563-87c2-66aed9a64de6

1251309

Score: 1 hide fields

department	COMPUTERS
id	c2cb85bd-7679-4563-87c2-66aed9a64de6
itemId	1251309
jobId	164aefc7dddT5c052642
modelId	Labs_recommender
numInteractionsForItem	0
numInteractionsForUser	0
score	1
type	items-for-user
userId	d3b26eb7245cbc325ad8aaaf24d1c8214329411b
weight	0.09701470803776147
version	1606359972646486000

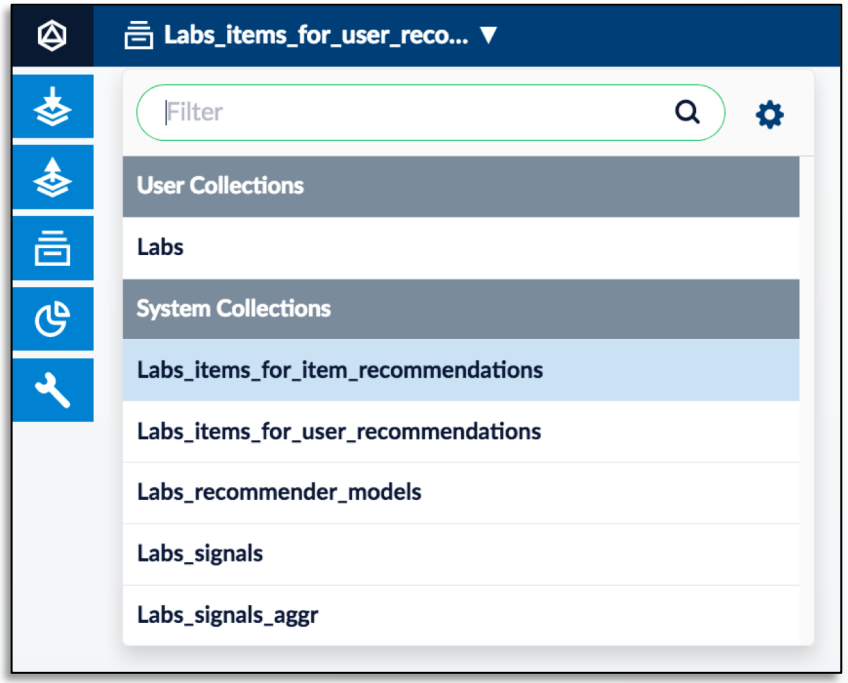
*Key among the fields here are **itemID** and **userID**, representing a single, unique user-item pairing. The **weight** field is a prediction of how well this item matches this user, based on the behavior of other users with similar activity.*

*Also note the **department** field. This was pulled in by the Metadata Join we requested in the **Labs_item_recommendations** job. This field can be used at query time; for example, to only recommend items from certain departments.*

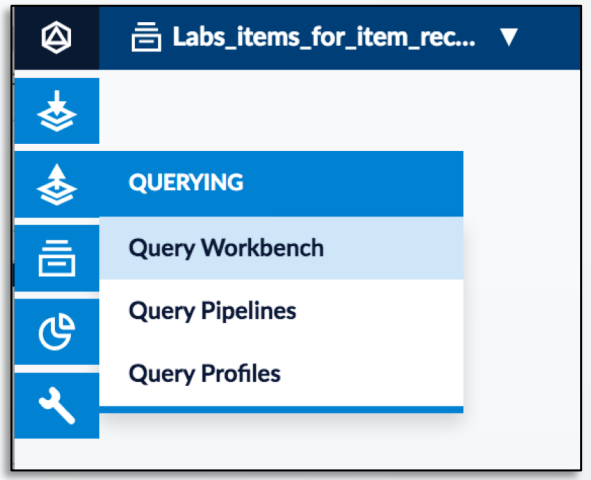
Verifying Items-for-Item Data

Lucidworks

- In the top left dropdown of Fusion Admin, change to the **Labs_items_for_item_recommendations** collection



- In the left side menu, go to **QUERYING > Query Workbench**



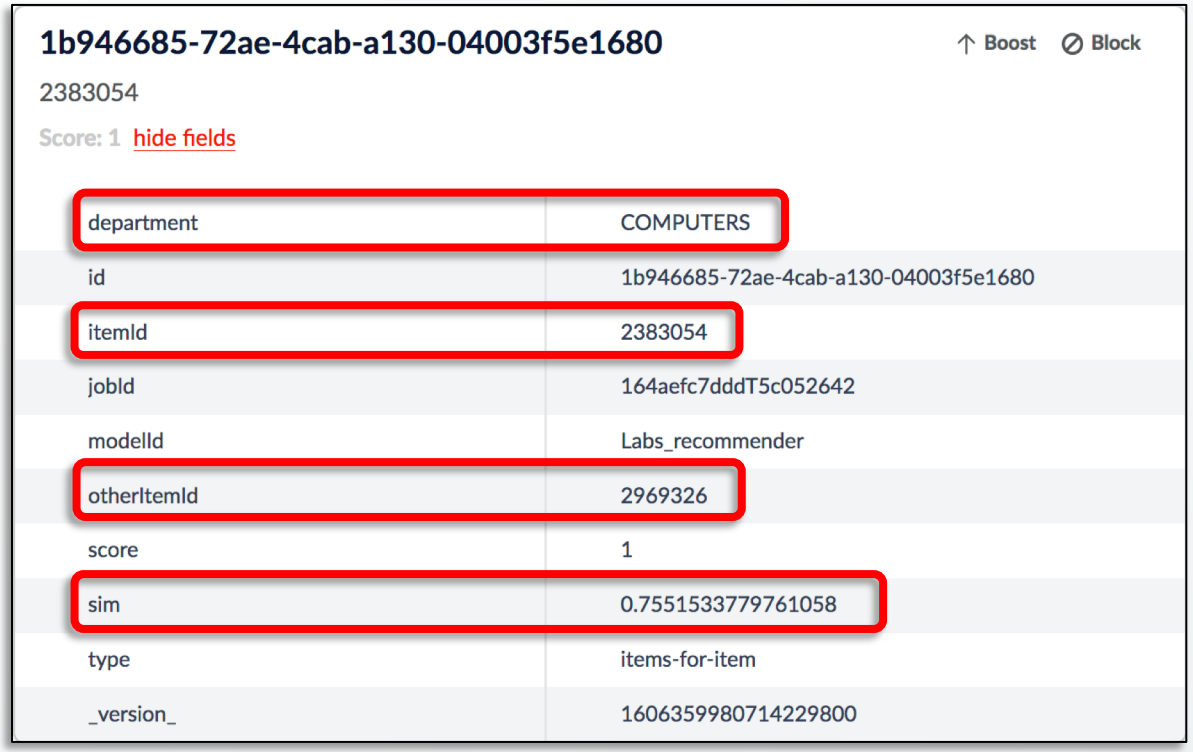
Verifying Items-for-Item Data

Lucidworks

- Click **show fields** under any document

*Notable fields here are **itemId** and **otherItemId**, representing a single unique pairing of items. The **sim** field is a measurement of how "similar" these two items are, based on collocation in user activities.*

*Again, also note the **department** field from the Metadata Join.*

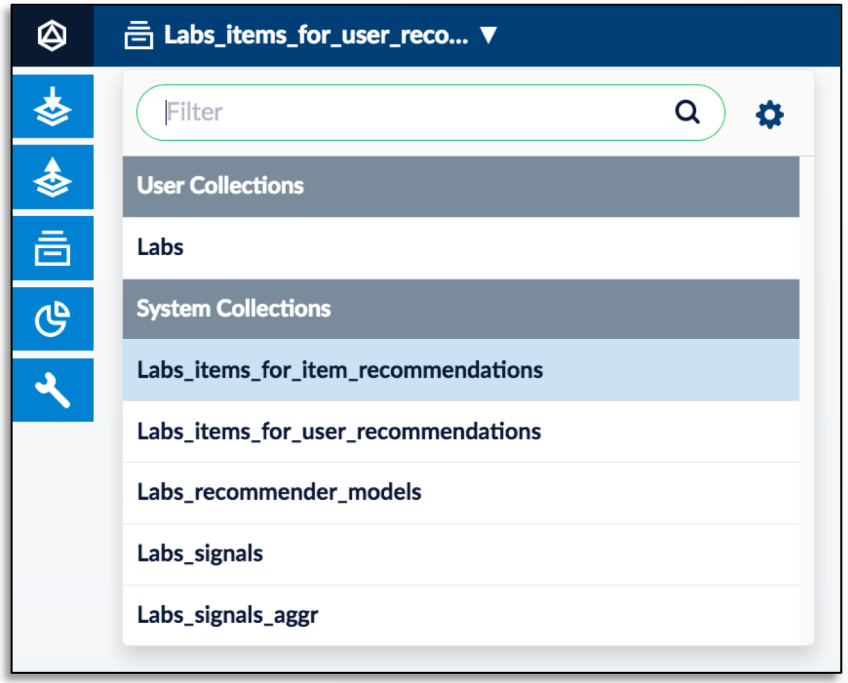


Testing Item-Item Recommendations

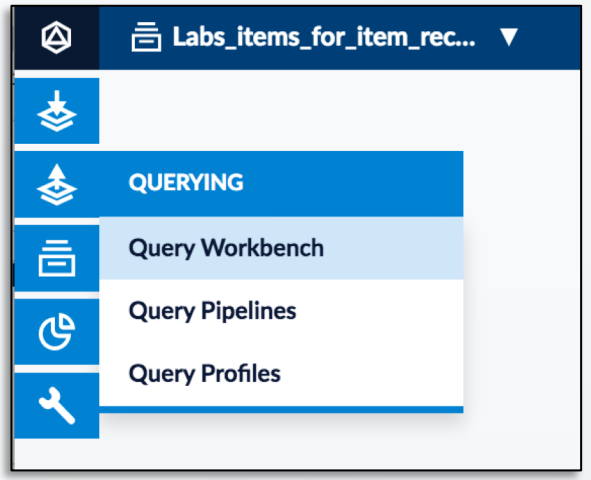
Testing Item-Item Recommendations

Lucidworks

- In the top left dropdown of Fusion Admin, change to the **Labs** collection



- In the left side menu, go to **QUERYING > Query Workbench**

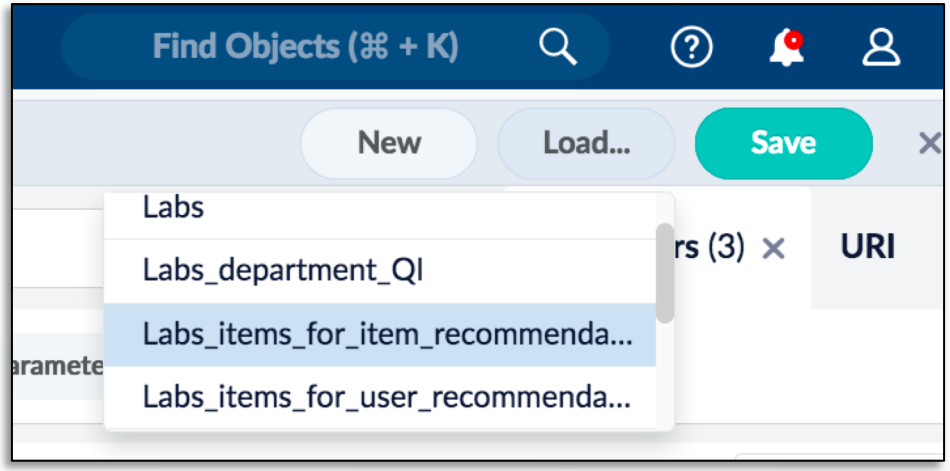


Testing Item-Item Recommendations

Lucidworks

- At the top right, click **Load**

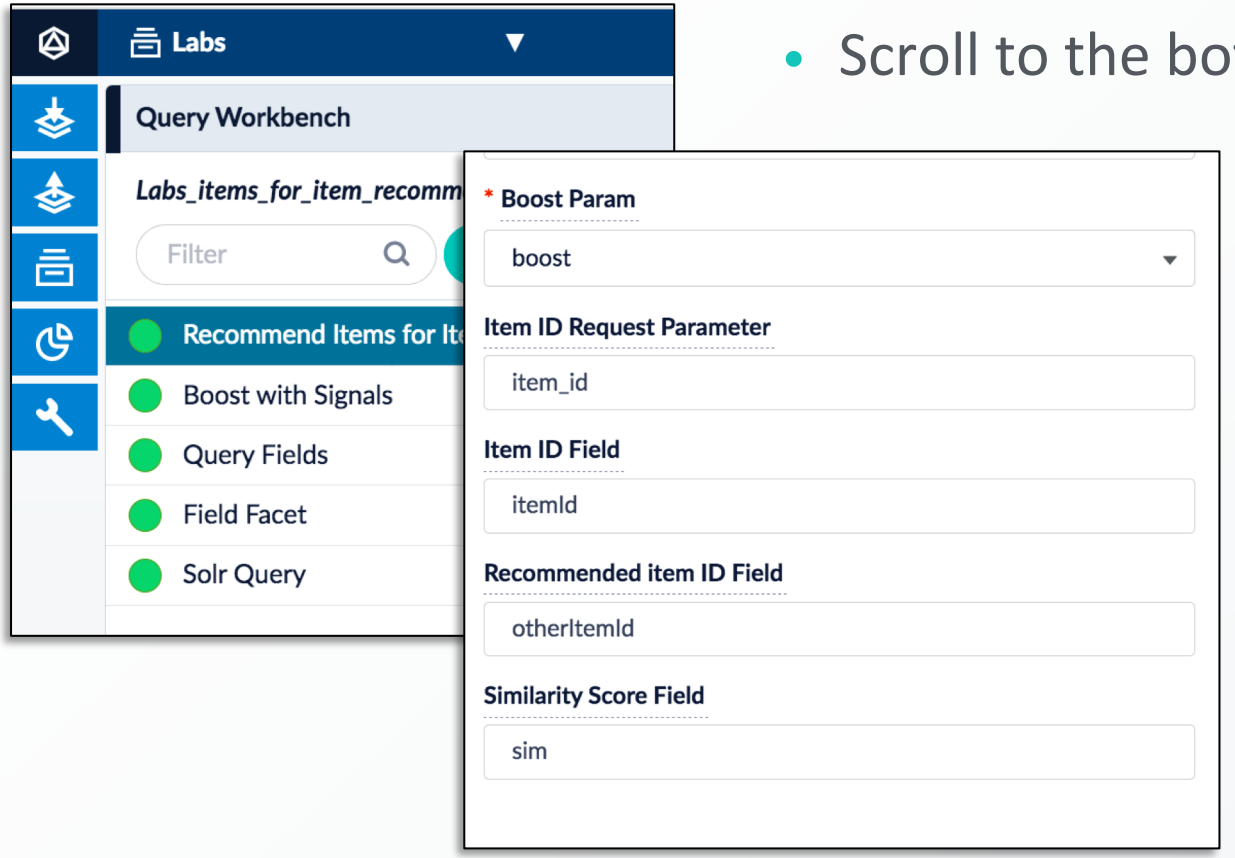
- Select the pipeline **Labs_items_for_item_recommendations**



Testing Item-Item Recommendations

Lucidworks

- In the Stages list, select **Recommend Items for Item**



- Scroll to the bottom of the Stage Editor pane

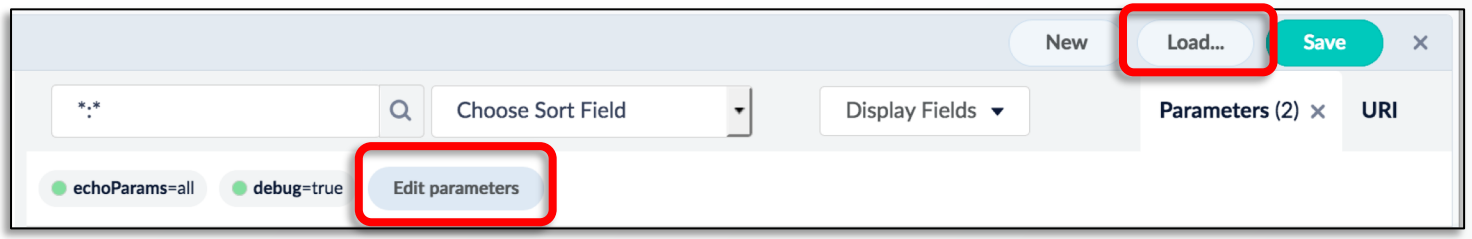
*This stage implements a relevance boost for similar items, as long as a single item is "selected" via an **item_id** parameter attached to the query. This parameter will be populated, for example, by going to the details page for a given item.*

The last three parameters relate to the structure of the items-for-item documents we examined a moment ago.

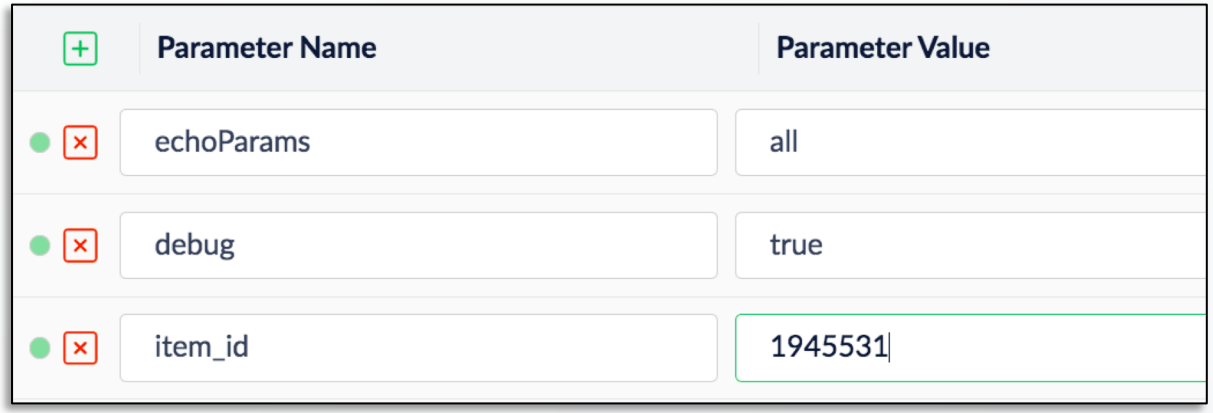
Testing Item-Item Recommendations

Lucidworks

- At the top of the page, click **Parameters > Edit Parameters**



- Add a parameter named **item_id** with value **2460117**



*As mentioned before, this parameter prompts the **Recommend Items for Item** stage to boost items similar to the one with doc id* **2460117**

Testing Item-Item Recommendations

Lucidworks

- Execute the query **id:2460117**

As this document is a SSD, items boosted by the recommender should be related to hard drives.

Parameter Name	Parameter Value
echoParams	all
debug	true
item_id	2460117

Corsair - Force Series 3 60GB Internal Serial ATA III Solid State Drive for Laptops

This solid state drive features a 60GB capacity for storing important documents, special photos and more. The Serial ATA III interface allows for a s imple connection to your computer.

Score: 15 show fields + Added to page

Corsair - Professional Gold Series 1200-Watt Power Supply

This power supply features reliable delivery of power for your work and gaming needs. The modular cabling system makes installation and customi zation simple, while short-circuit protection helps prevent damage to your computer and components.

Score: 14.8681 show fields + Added to page

Corsair - Graphite Series 600T Mid-Tower Gaming Case and AX850 Power Supply Bundle

This mid-tower gaming case features dual 200mm fans and a 120mm rear fan to keep your installed components cool. The Professional Series Gol d AX850 power supply delivers plenty of power for your system.

Score: 14.8656 show fields + Added to page

Corsair - Gaming Series 700-Watt ATX CPU Power Supply

Power your gaming computer with this power supply that features 700 watts of output power for enhanced performance. The 0.99 active power f actor correction delivers reliable power.

Score: 12.4673 show fields + Added to page

1 2 3 4 5 6 7 8 ...4202 Next

1-10 of 42,012 docs (1 ms, max-score 15)

- Execute the query ***:***

These results are pretty good. Yours, however may differ, and may not be so good. This is due to the downsampling we did when building the model. For better results, use a larger sample!

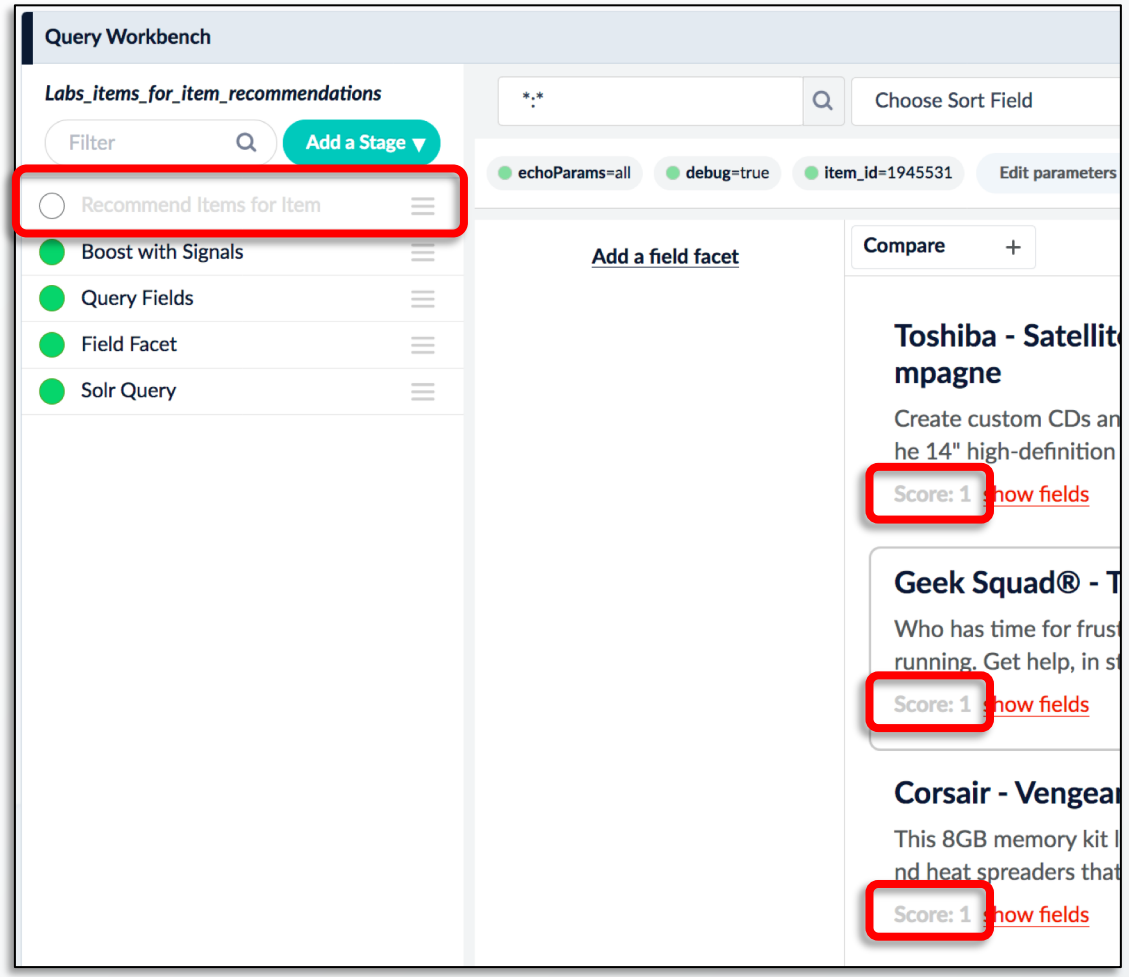
Testing Item-Item Recommendations

Lucidworks

- In the Pipeline pane, disable the **Recommend Items for Item** stage by clicking the green circle

*Now we are back to a pseudo-random document ordering, with all results having an identical score of 1 despite our **item_id** parameter*

- Re-enable **Recommend Items for Item** stage by clicking the circle again



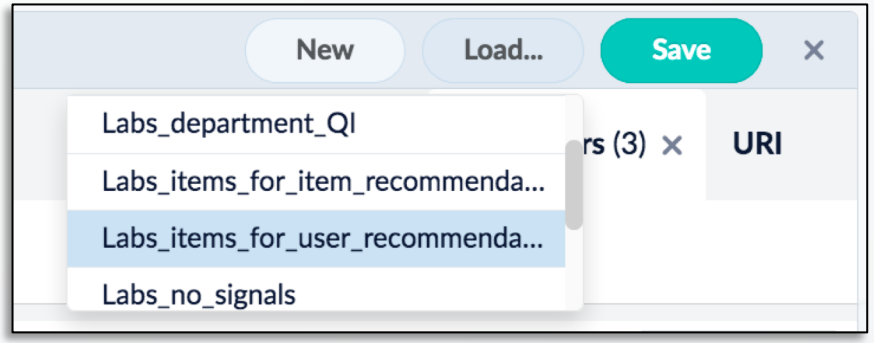
Testing User-Item Recommendations

Testing User-Item Recommendations

Lucidworks

There are two ways to implement a user-item recommendation. One is Personalized Recommendations, where we offer users a list of interesting items before they ask for anything. Another is Personalized Search, where we apply boost factors to items based on the user-item similarity of the person looking. We will start by testing Personalized Recommendations.

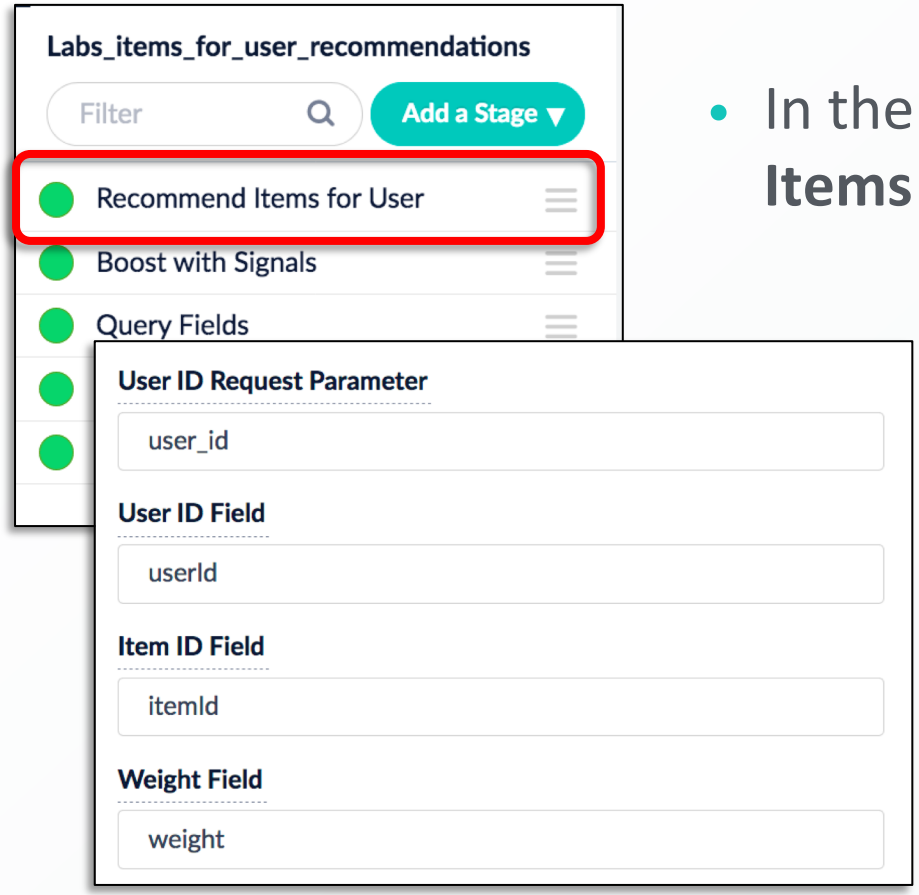
- Go to the **Query Workbench** of the **Labs** collection
- At the top right, click **Load**
- Select the pipeline **Labs_items_for_user_recommendations**



Testing User-Item Recommendations

Lucidworks

Similarly as before, this pipeline has a special stage that implements recommendations.



- In the Pipeline pane, select the **Recommend Items for User** stage

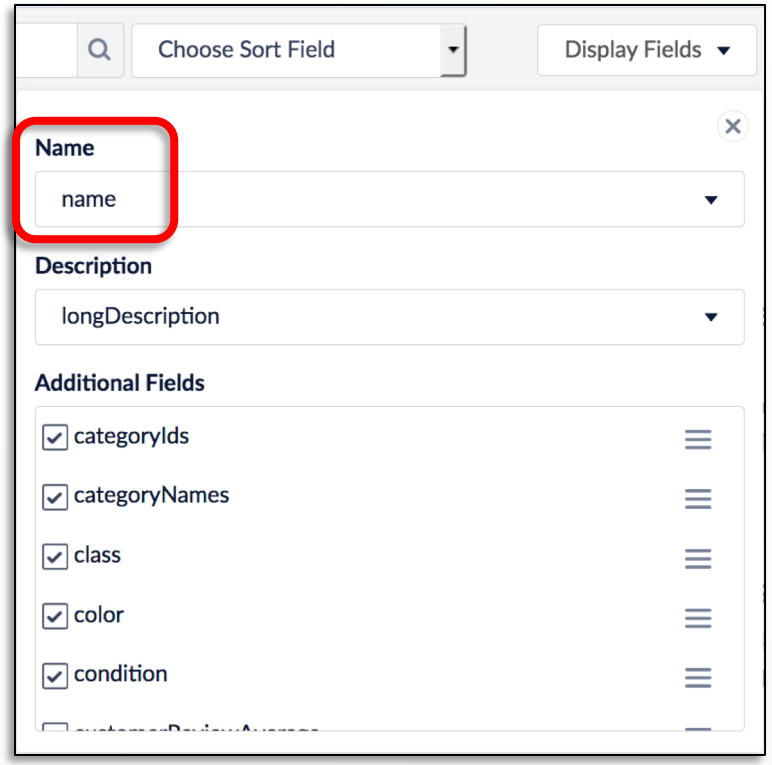
*Near the bottom of the stage editor you will see some familiar parameters. In this case, the key query parameter is **user_id**, which will generally be populated by a login action.*

*The user-item recommendations collections will be searched for items that have a **userId** field equal to user_id, and will provide a relevance boost on the associated **itemId**. The **weight** determines the magnitude of the boost*

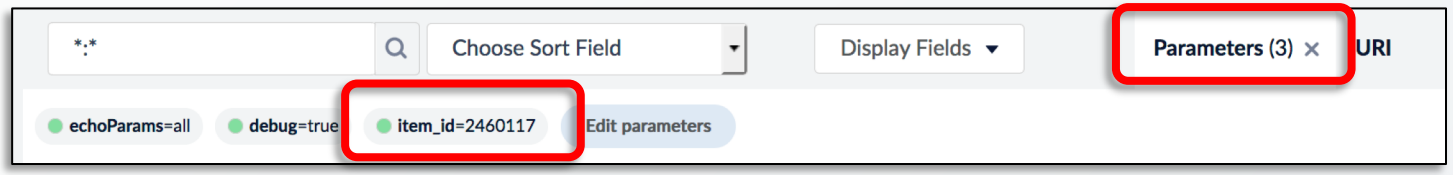
Testing User-Item Recommendations

Lucidworks

- In the **Display Fields** dropdown, set the **Name** field to **name**



- At the top right, click **Parameters > Edit Parameters**

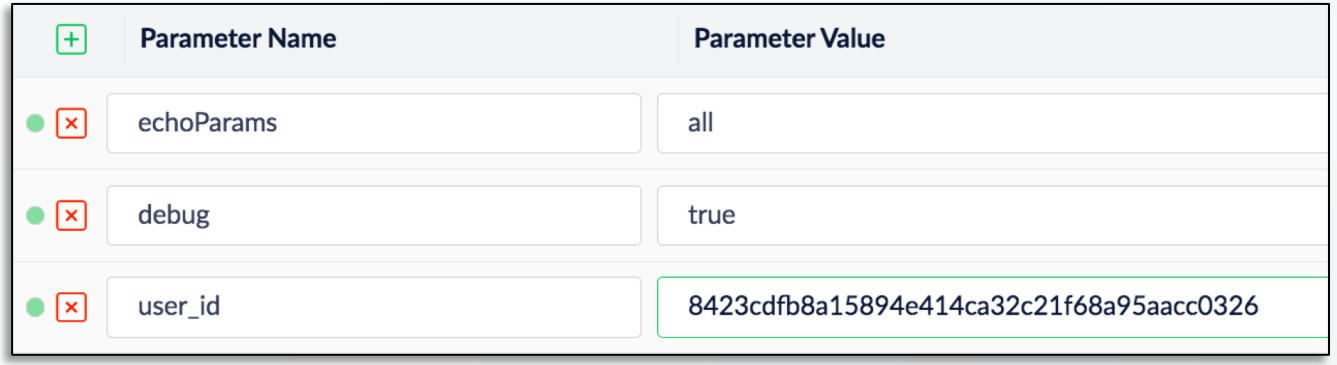


Testing User-Item Recommendations

Lucidworks

- Delete the **item_id** parameter

- Add a parameter named **user_id** with value **8423cdfb8a15894e414ca32c21f68a95aacc0326**



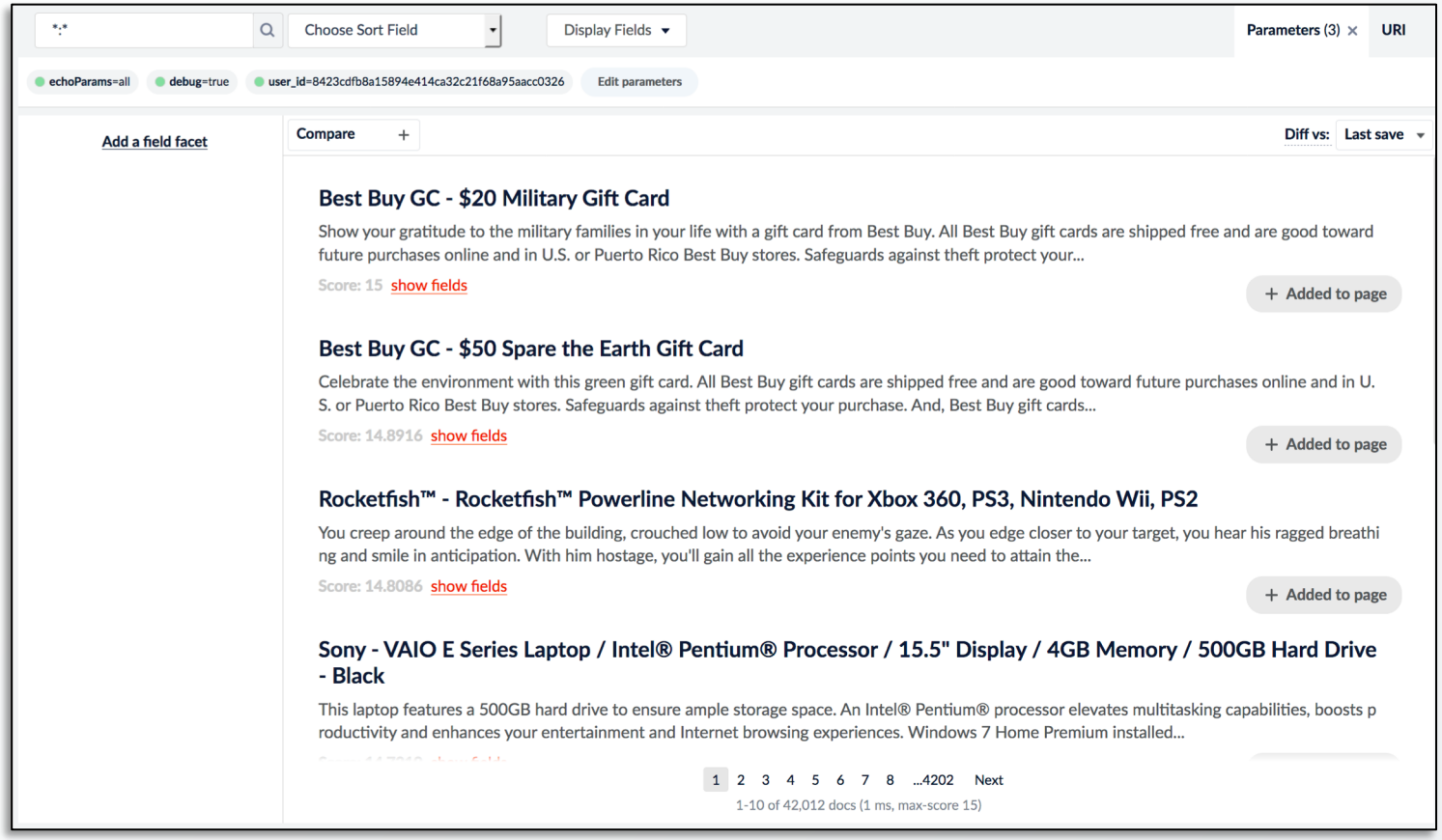
Testing User-Item Recommendations

Lucidworks

*The default *:* query, combined with the user-item recommender, gives us Personalized Recommendations. These are the items that this user would see when they first log on, before executing any queries.*

Again, your results may differ from these. The severe downsampling we did in the model building job yields inconsistency.

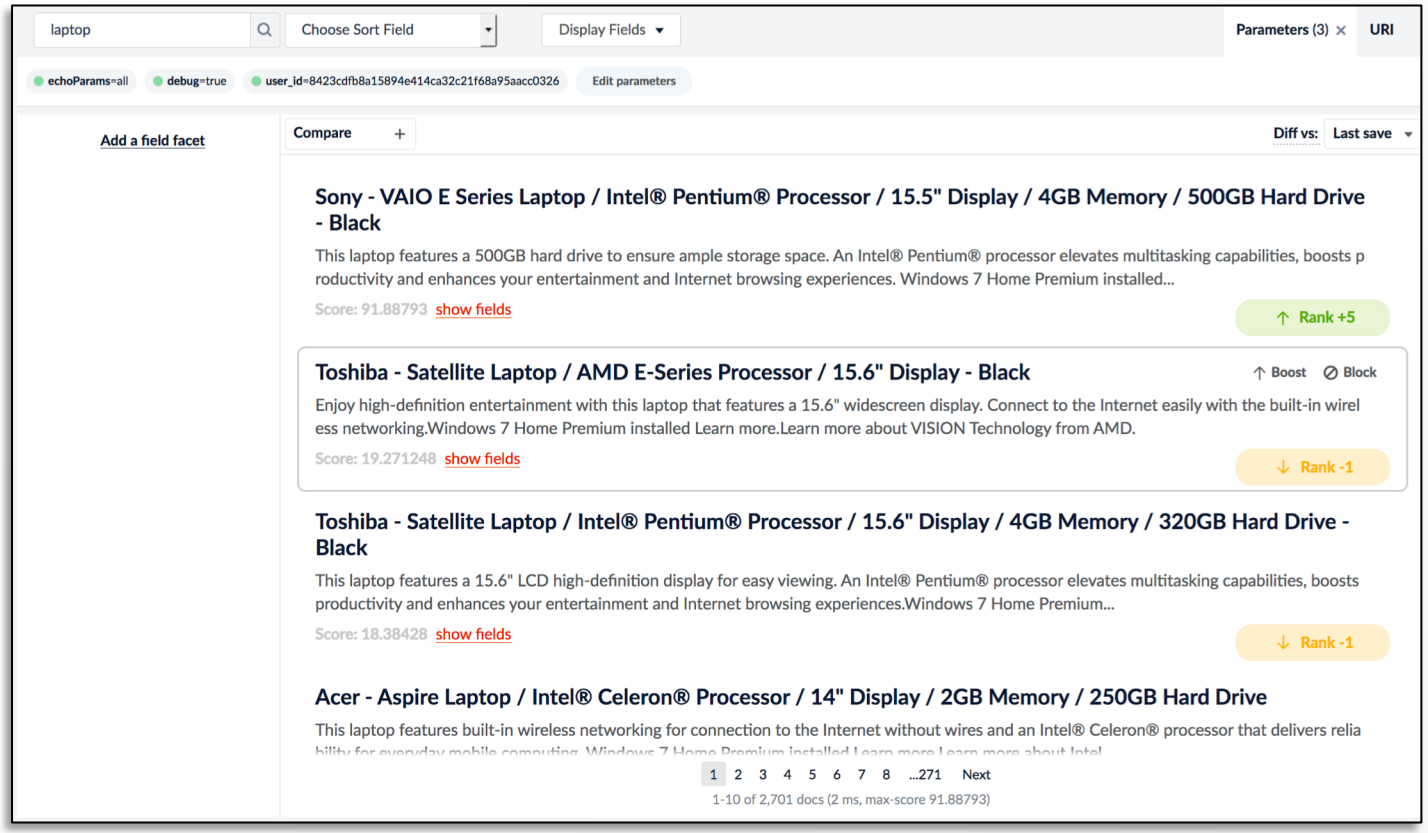
*You can observe the difference in results when the **Recommend Items for User stage** is disabled by clicking the green circle, as before*



Testing User-Item Recommendations

Lucidworks

- Execute the query **laptop**



The top result for this user was originally 5 ranks lower, without the recommender active. This is an example of Personalized Search.

Again, your results may differ from these. For better results, use a larger sample when building the recommender model.

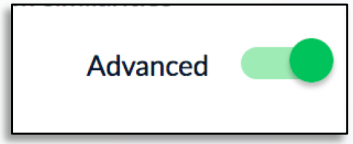
Building a Better Recommender (optional)

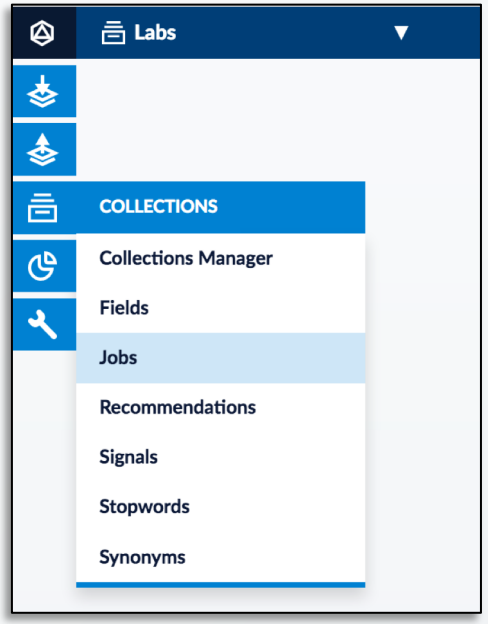
Testing User-Item Recommendations

Lucidworks

If you can spare the time, try the Testing sections of this lab with a non-downsampled model. This section will walk you through the steps for doing so.

- In the left side menu of Fusion Admin, go to **COLLECTIONS > JOBS**
- Select **Labs_items_recommendations**
- Toggle the **Advanced** switch at the top right

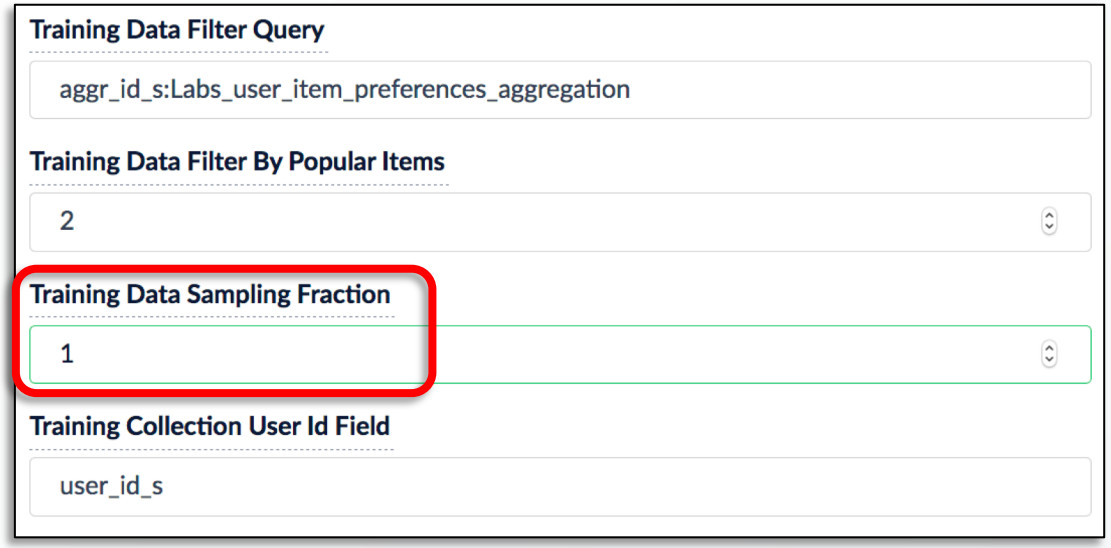




Testing User-Item Recommendations

Lucidworks

- Change the **Training Data Sampling Fraction** parameter to **1**



- **Save** and **Run** the updated job

With full sampling, the job should take ~10 minutes to complete. Once finished, try the Testing steps again, and see the improved results!

End of Lab