TECHNICAL GUDE

Splongo App Installation and Configuration

Requirements

High speed, low latency connection to the MongoDB Instance.
MongoDB instances and Splunk need to have synchronized time.

Depending on the amount of data:

Below 10gb per day:

- 2-4 CPU cores
- 8GB RAM

Between 10-100GB:

- 4 - 8 cores
- 16GB RAM

Between 100-250GB data per day:

- 12 CPU Cores
- 32GB RAM

Above 250GB per day:

- 16 CPU cores
- 40GB RAM

General Set-up - Interval Query

You can setup an unlimited number of jobs. Each job can handle only one collection. In cases where you want to import from multiple collections you have to use multiple jobs.

Additionally you can also use a multi-job setup for a single collection if you want to use specific queries for that collection.

Click on the ***"Create new input"*** button and select ***"Interval Query"***. Fill out the parameters according to your specifications.

Required fields:

- Name - Name the data ingestion job;
- Interval - Interval on which the job will be executed;

BRIGHT

- Index - Which Splunk index should the data be in;
- Collection - MongoDB Collection to be queried;
- Time Field - The job requires the Collection to have a timestamp field, based on which queries will be prepared and executed. If the field is nested use JSON pathing*;
- Cores - Number of parallel processes to be spawned to collect the data from MongoDB, see Performance section for further details.

Optional Fields:

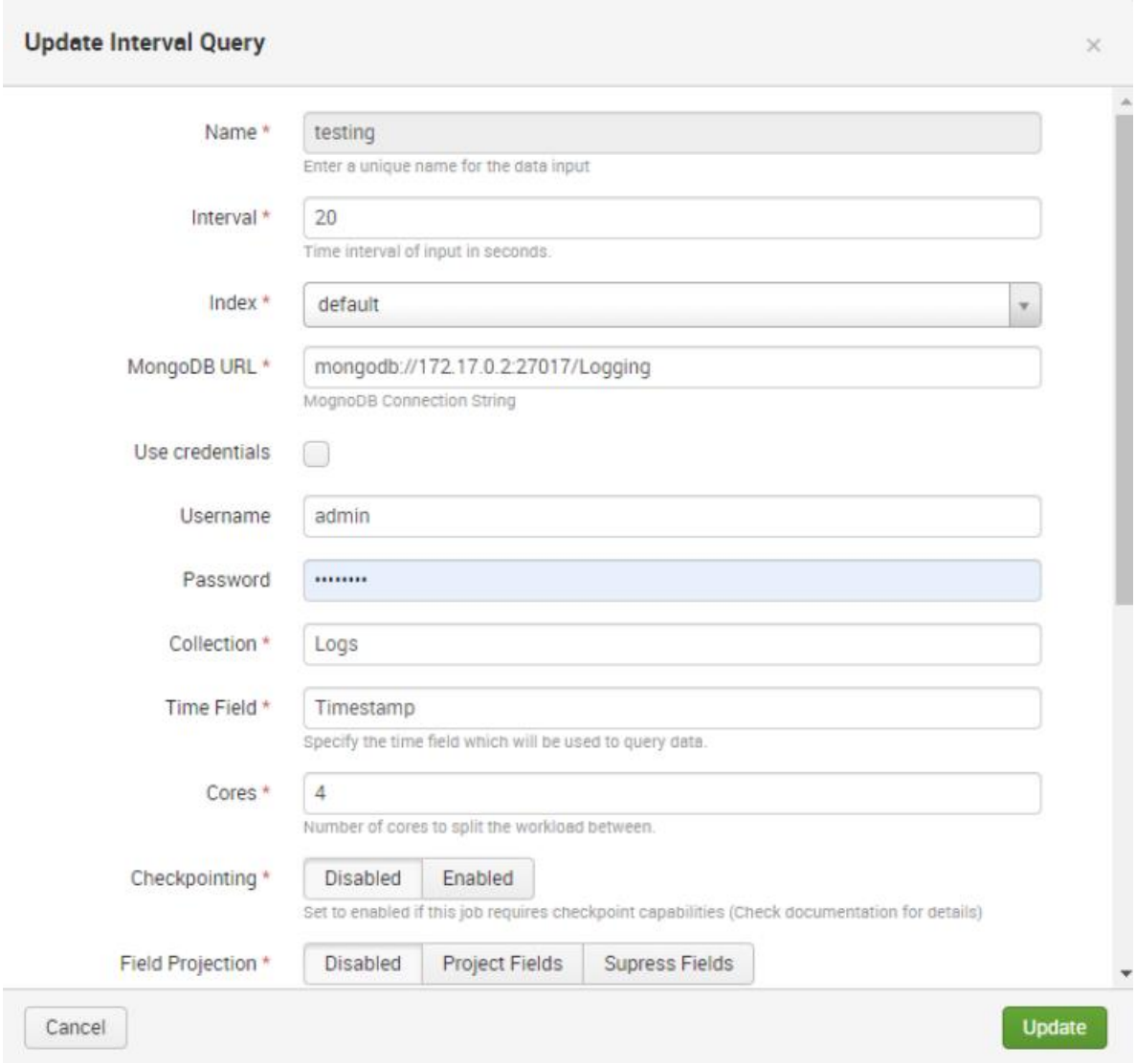
- Field Projection - If enabled MongoDB will be queried with field projection;
- Supress fields means that the fields specified in ProjectedFields will not be ignored by MongoDB;
- Project fields means that only the fields specified in ProjectedFields will be returned;
- Projected Fields - A comma-delimited list of fields to be projected;
- Keyword Filtering - If enabled, the script will scan your specified "Payload" field and extract only the relevant data. For further information see Payload Filtering;
- Keywords - A comma delimited list of keywords to look for. Not case sansitive;
- Payload Field - Specify the field that contains the payload. (All child nodes will be scanned);
- Source field Path - Specify a field from the ingested document to be used as "Source" field in Splunk;
- Host field Path - Specify a field from the ingested document to be used as "Host" field in Splunk.

Examples

1. **Basic single-collection reading:**

    - Name – IntervalQuery
    - Interval – 30
    - MongoDB URL- mongodb://mongodb:27017/Logging
    - Index – mongo_logs
    - Collection – Logs
    - Cores – 4
    - Time Field – Timestamp
    - Checkpointing – Disabled
    - Field Projection – Disabled

The job described above will execute every 30 seconds and ingest the latest 30 seconds of data within your MongoDB Collection determined by the time field specified.



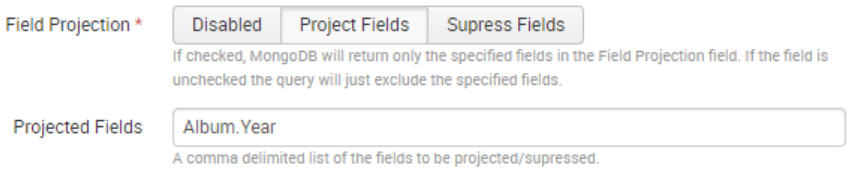
2. Field projection

These are default MongoDB Options for which you can read further here.

If we have the following document in MongoDB:

```
{
"_id" : 276286,
"Timestamp" : ISODate("2020-06-02T18:25:25.681Z"),
"Album" : { "Artist" : "Metallica",
            "Album" : "Metallica",
            "Year" : "1991" }
}
```

and select for field projection the **Project Fields** option and we specify the Album field to be projected, this will return only the fields specified in the Project Field.



This is the result:

You can add multiple fields separated with a comma. You can also specify nested fields here, but you need to specify the full JSON path. We will see this example with the Suppress fields option.

If we select the suppress fields option and use the following in the **Project Fields** parameter.

Result is:

As you can see the Album and Year fields in the sub-element Album are missing.

3. Keyword filtering

What the Keyword filtering option does is it searches the whole document for specific keys, which are specified in the **Keyword Filter** option.

Consider the follwoing Document:

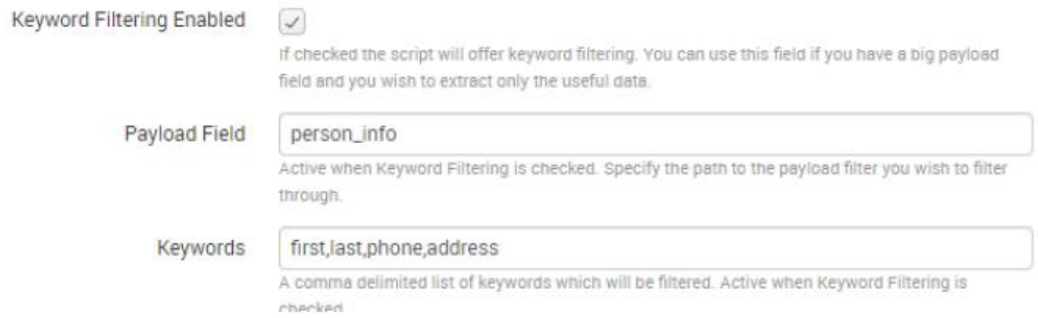

```
{
   "_id":293338,
   "Timestamp":{
      "$date":1591169427101
   },
   "person_info":{
      "_id":"5e8dd3a129bb614e1a38e302",
      "index":0,
      "guid":"17cd57b3-2832-4ee7-88c4-1c1105d921eb",
      "isActive":true,
      "person":{
         "company":"ZENTIME",
         "first":"Robbie",
         "last":"Morton",
         "phone":"+1 (849) 438-3378",
         "address":"579 Bergen Avenue, Bison, Georgia, 9846",
         "email":"robbie.morton@zentime.com",
         "about":"Incididunt Lorem enim est laboris in proident irure sunt.",
         "registered":"Monday, February 17, 2014 2:08 PM",
         "latitude":"-31.948798",
         "longitude":"14.262519"
      }
   }
}
```


Lets say we care only about the person's first, last name, email and phone. Keyword filtering can help with reducing unwanted data.

This can be achieved using the following setup:

- Keyword Filtering - Enabled
- Payload Field - person_info

- Keywords - first,last,phone,address



And this is the result:

The Payload field is removed and in its place now there is a "filtered_payload" field, which holds the matched keywords.

Non-default Timestamp/Timezone

If the Time field in your Collection insn't using the default MognoDB Timestamp or date-time format you will have to manually edit the props.conf file in the app's defalt folder.

Usually located in:

<SPLUNK_HOME>/etc/apps/Splongo/local/props.conf

You need to edit the following fields under the [IntervalQuery] stanza:

TZ = UTC
TIME_FORMAT = \d{4}-\d{2}-\d{2}\s\d{2}:\d{2}:\d{2}.\d{6}

If you need further information or have any questions, do not hesitate to contact our Bright team at support@bright.consulting