# PostgreSQL FTS Relay Application

PostgreSQL FTS Relay web application is built on Spring MVC framework module. The application gets compiled on JDK 1.8 through Maven, and runs on any environment where the JDK is available. The only external dependency for the application – is external PostgreSQL database.

## Source Repository

The sources for the application are in GitHub <https://github.com/syemialy/postgres-jbi> To work with the project you may fork the repository or request private access to it.

## Configuration

Application has only single configuration file – *application.properties,* where postgre credentials are supplied to the application. When running application on various environment it is suggested to initialize postgre settings not via properties, but through the environment variables.

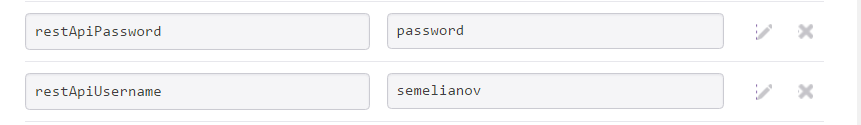
When compiled, the application may run via command:

java -D**database.user**=tnkxkfefnfysiw -D**database.pwd**=eoyUq2ZLcYcPY7U0bvxtRhLAxx -D**database.url**=jdbc:postgresql://ec2-54-225-255-208.compute-1.amazonaws.com:5432/d85lilb3bg1tg0 $JAVA\_OPTS -jar target/dependency/jetty-runner.jar --port $PORT target/\*.war

## Authentication

Basic http authentication is used to prevent unauthorized usage of the REST services. By default the application is protected by the realm where user *test* identified by password *test* is registered. To change credentials, it is suggested to use two additional environment variables: **restApiUsername** and **restApiPassword**.

When running the application on Heroku, one must change the authentication credentials via configuration variables as shown on the screenshot below, the variable values should be kept private and shared with application clients only.



## SQL Injection Protection

The application dynamically assembles three types of statements – SELECT, CREATE INDEX, DROP INDEX. The process of creating SQL statement depends on the content of incoming messages. It is very likely that intruder may take an advantage of SQL Injection attack by sending a command of the following content to the search service

$ curl -X POST -H 'Content-type:application/json' -d'{"table":{"name":"products","columns":[{"name":"description","selectable":false,"tsinclude":true},{"name":"count(id)","selectable":true,"tsinclude":false},{"name":"product\_name","selectable":false,"tsinclude":true}]},"query":"need to connect","**orderby":"product\_name;dElete \* FROM PRODUCTS**"}' http://localhost:8081/srv/search

The application will catch such an attempt and the following error message will be thrown

{"error\_message":"SQL Injection","error":true}

The error will be accompanied by the following log statement:

possible SQL injection attack with statement count(id) FROM products WHERE to\_tsvector('english',coalesce(description,'') || ' ' || coalesce(product\_name,'')) @@ plainto\_tsquery('need to connect') ORDER BY product\_name;dElete \* FROM PRODUCTS

The SQL Injection protection is made using well known principles:

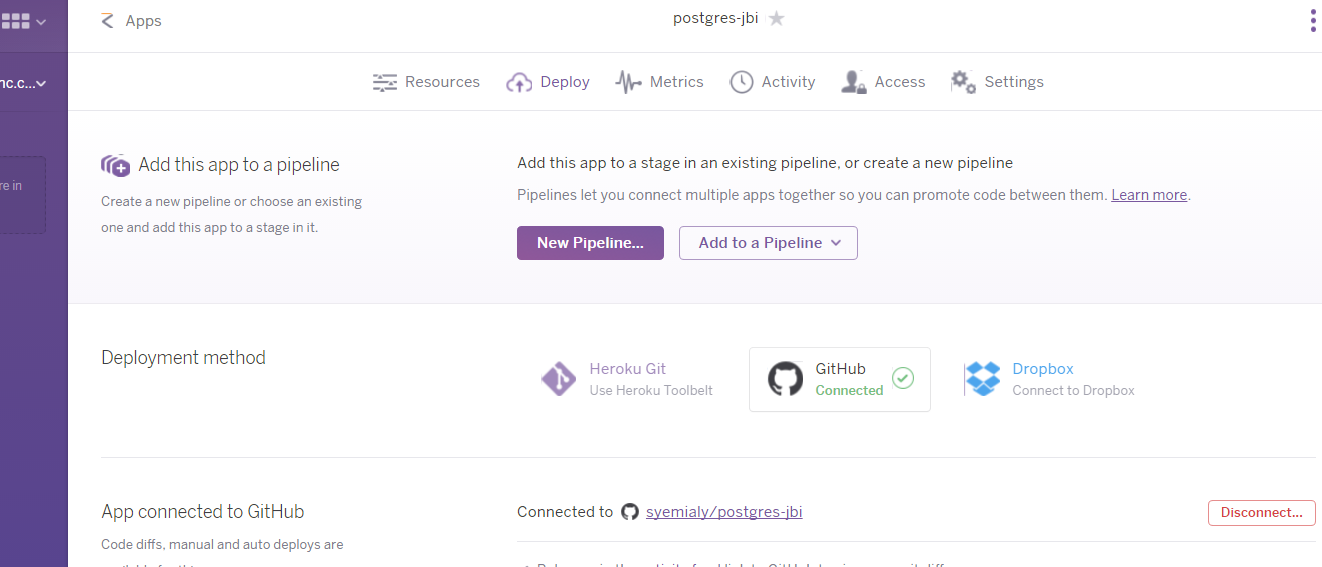
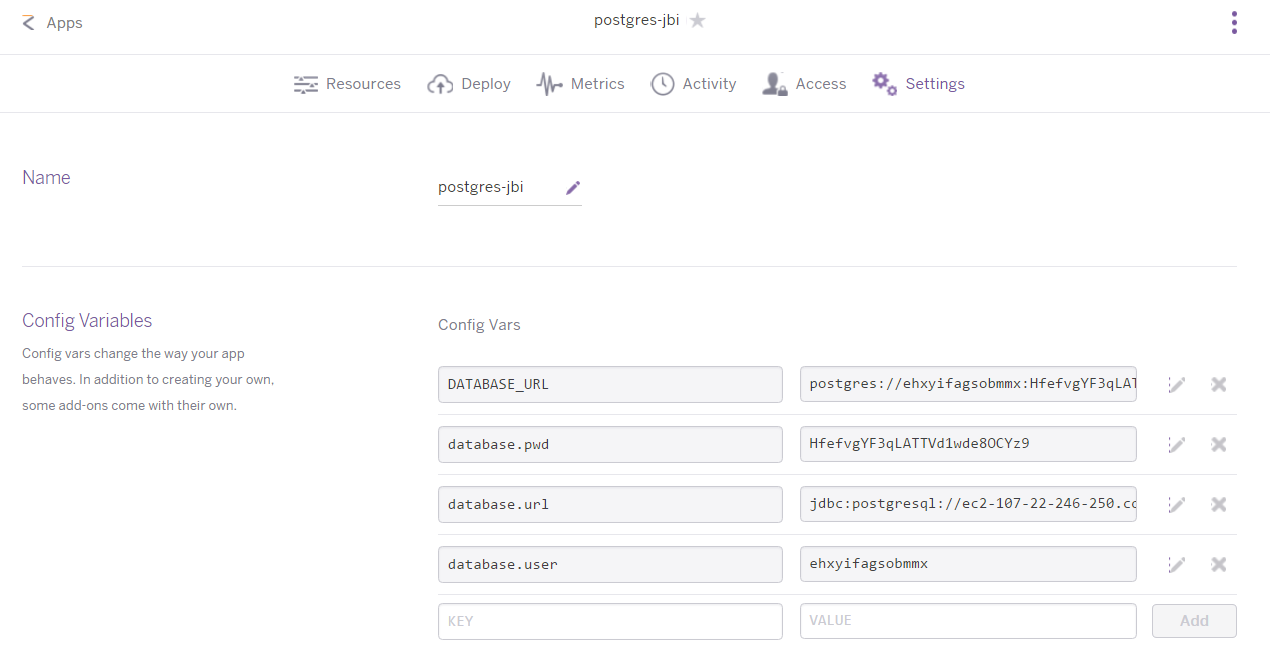
1. A statement is checked for having statement separator blocks, such as semicolons
2. A statement is checked for having additional SQL statements like drop, create, delete, etc.
3. A statement is checked for having special characters like \*, [], etc

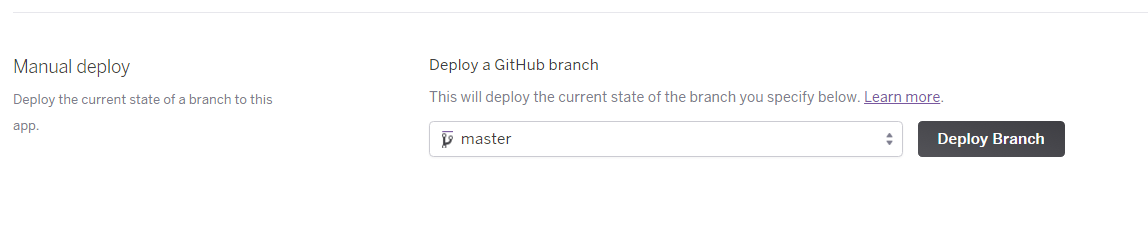
To read more on SQL Injection protection one may refer to <http://larrysteinle.com/2011/02/20/use-regular-expressions-to-detect-sql-code-injection>

## Deployment Instructions

### Deployment to Heroku via GitHub

The deployment instructions assume you have forked the ripository **postgres-jbi**

* Create new Heroku application
* If using external Postgres, make sure the database, attached to your dyno by default is erased and no DATABASE\_URL variable exists in your application environment.
* On the application settings page, navigate to Deploy section
  + Link your GitHub repository
* On the application Settings section, add all three environment variables .
  + 
* Go Back to Deploy section and execute manual deploy of your new application from your github ripository



Once your application is deployed, you may use curl to validate all the REST services. Samples of the curl usage may be found in docs/curl.samples.txt file.

## REST Service Endpoints

|  |  |  |  |
| --- | --- | --- | --- |
| Endpoint | Method | Format | Description |
| srv/search | POST | JSON | A text query passed in JSON body is used to create FTS search request against underlying PostgreSQL database |
| srv/index | POST | JSON | Create index on documents, created based on the selected columns. Columns are added to the ts\_vector with coalesce() function |
| srv/async/index | POST | JSON | Creates index in asynchronous mode. The service should be used when index creation procedure takes enough time to force your synchronous http post request to time out |
| srv/index | GET | JSON | Checks if index with the exact name exists within your database |
| srv/index | DELETE | JSON | Drops the index by name |

## Data Format for JSON messages

### Search Request

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Field | | | **Type** | **Mandatory** | **Description** |
| *table* |  |  | Object | Yes |  |
|  | ***name*** |  | String | Yes | Name of the table against which the FTS search will be executed |
|  | ***columns*** |  | Object[] | Yes | Column name(s) to create a document (unit of searching) |
|  |  | ***name*** | String | Yes | Name of the column within the table |
|  |  | ***selectable*** | Boolean | Yes | Indicates if column will be in SELECT statement predicate |
|  |  | ***tsvectorinclude*** | Boolean | Yes | Indicates if column will be used to create ts\_vector |
| *query* |  |  | String | Yes | User-written text to be used in *plain\_tsquery* function to create search terms |
| *orderby* |  |  | String | No | Specify column name(s) to order the resut set |
| *limit* |  |  | Integer | No | Number of records to return |
| *offset* |  |  | Integer | No | Records offset |
| *configuration* |  |  | String | No | Configuration name to be used to parse and normalize strings. By default it will be set to *english* |

Note: According to the documentation <http://www.postgresql.org/docs/current/static/queries-limit.html>, using LIMIT and OFFSET should be avoided due to the risk of having inconsistent results.

{ “table”: {  
 “name”: “products”,  
 columns:[  
 {“name”:”product\_name”, “selectable”:true, “tsvectorinclude”:true},  
 {“name”:”description”, “selectable”:false, “tsvectorinclude”:true},  
 {“name”:”product\_id”, “selectable”:true, “tsvectorinclude”:false}],

“query”:”sold by pair”,  
 “configuration”:”english”  
 }

### Search Response

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Field |  | Type | Mandatory | Description |
| result |  | Object | Yes |  |
| ***sqlstatement*** | String | Yes | Name of the table against which the FTS search will be executed |
| ***records*** | Object[] | Yes |  |
| ***javatimemls*** | Integer | Yes | Indicates time the java program ran FTS search, this is to collect statement statistics |
| error |  | Boolean | Yes | User-written text to be used in *plain\_tsquery* function to create search terms |
| error\_message |  | String | No | Error message |

### Create Index Request

This is the same format for srv/async/index and srv/index endpoints

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Field | | | **Type** | **Mandatory** | **Description** |
| *name* |  |  | String | Yes | Name of the index to be created |
| *type* |  |  | String | No | Type of the index to be created, GIN will be default type. |
| *table* |  |  | Object | Yes | Name of the table for which index will be created |
|  | ***name*** |  | String | Yes | Name of the table against which the FTS search will be executed |
|  | ***columns*** |  | Object[] | Yes | Column name(s) to create a document (unit of searching) |
|  |  | ***name*** | String | Yes |  |
| *configuration* |  |  | String | No | Configuration name to be used to parse and normalize strings. By default it will be set to *english* |

{"name":"idx\_prddescr",  
 "table":{"name":"products",  
 "columns":[{"name":"description"},{"name":"product\_name"}]}}

### Create Index Response (Synchronous)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Field |  | Type | Mandatory | Description |
| *result* |  | String | Yes | Status of the request |
| *sqlstatement* |  | String | Yes | The SQL statement which was executed |
| *error* |  | Boolean | Yes | User-written text to be used in *plain\_tsquery* function to create search terms |
| *error\_message* |  | String | No | Error message |

{"result":"created",  
 "error":false,  
 "sqlstatement":"CREATE INDEX idx\_prddescr ON products USING GIN (to\_tsvector(\u0027english\u0027,coalesce(description,\u0027\u0027) || \u0027 \u0027 || coalesce(product\_name,\u0027\u0027)))"  
}

### Create Index Response (Asynchronous)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Field |  | Type | Mandatory | Description |
| *result* |  | String | Yes | Status of the request |
| *message* |  | String | No | Message for the request |
| *error* |  | Boolean | Yes | User-written text to be used in *plain\_tsquery* function to create search terms |
| *error\_message* |  | String | No | Error message |

{"result":"scheduled","error":false,"sqlstatement":"check later for index create status"}

### Drop Index Request / Check Index Request

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Field |  |  | **Type** | **Mandatory** | **Description** |
| *name* |  |  | String | Yes | Name of the index to be dropped |

{"name":"idx\_prddescr"}

### Drop Index Response

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Field |  | Type | Mandatory | Description |
| *result* |  | String | Yes | Status of the request |
| *sqlstatement* |  | String | Yes | The SQL statement which was executed |
| *error* |  | Boolean | Yes | User-written text to be used in *plain\_tsquery* function to create search terms |
| *error\_message* |  | String | No | Error message |

{"result":"dropped",  
 "error":false,  
 "sqlstatement":"DROP INDEX idx\_prddescr"}

### Check Index Response

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Field |  | Type | Mandatory | Description |
| *result* |  | String | Yes | Status of the request |
|  | ***location*** | String | Yes | Table location where index belongs |
| *error* |  | Boolean | Yes | User-written text to be used in *plain\_tsquery* function to create search terms |
| *error\_message* |  | String | No | Error message |

{"result":[{"location":"products.idx\_ts3"}],"error":false}