Logstash Setup and Configuration

Note: Logstash is always configured on a single node. It's just a mediator, so doesn't require cluster capabilities but it is always better to assign good computation capabilities.

$ sudo apt-get update

$ sudo apt-get install default-jdk

$ java -version

$ wget https://artifacts.elastic.co/downloads/logstash/logstash-6.2.2.tar.gz

$ sudo tar -xzf logstash-6.2.2.tar.gz

$ cd logstash-6.2.2/

$ sudo vi **sample.conf**

$ ./bin/logstash -f **sample.conf** (if .conf file is located somewhere else, then give full path to the file)

These configuration needs to be written in a ***.conf*** file. For example: **sample.conf**

... (configurations are given ahead)

Use the following to get mysql java connector required for logstash conf file:

$ wget https://dev.mysql.com/get/Downloads/Connector-J/mysql-connector-java-5.1.46.tar.gz

1. **Logstash configuration for fetching data from RDS SQL to elastic cloud**:

```
input {
    jdbc{
        jdbc_connection_string=> "jdbc:mysql://tidyquant-mysql.ctzkiiqc8r0v.ap-south-1.rds.amazonaws.com:3306/San_Francisco_Crime_Classification"
        jdbc_user=> "tidyquant"
        jdbc_password=> "tidyquant"
        jdbc_driver_library=> "/home/ubuntu/mysql-connector-java-5.1.46/mysql-connector-java-5.1.46-bin.jar"
        jdbc_driver_class=> "com.mysql.jdbc.Driver"
        statement => "SELECT * FROM test limit 1000"
    }
}
output {
    elasticsearch {
        hosts => "https://b8bd4f06269136f164b670ae8c497567.us-east-1.aws.found.io:9243"
        user => "elastic"
        password => "eGlSf27hAeYT9k8x8xDDNHVj"
        index => "test_logstash"
        document_type => "test_mapping"
    }
    stdout {
        codec => rubydebug
    }
}
```

2. Logstash configuration for fetching data from S3 to elastic cloud:

```
input{
    s3 {
        bucket => "tidyquant-pipeline-test"
                                access_key_id => "AKIAJUEDFWZNW2NYVJXQ"
                                secret_access_key =>
"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"
                            }
}
filter{
    grok {
        match => {"message" => "%{COMBINEDAPACHELOG}"}
    }
    date {
        match => [ "timestamp", "dd/MMM/yyyy:HH:mm:ss Z"]
    }
}
output {
    elasticsearch {
        hosts => [ "https://b8bd4f06269136f164b670ae8c497567.us-east-1.aws.found.io:9243" ]
        user => "elastic"
        password => "eGlSf27hAeYT9k8x8xDDNHVj"
        index => "s3_logstash"
        document_type => "s3_mapping"
                }
                stdout {
        codec => rubydebug
    }
}
```