IWEEK 3

September 14, 2024

Student Information

Full Name: Deniz Karakoyun
Id Number: 2580678

1 Introduction to Apache Flink

Apache Flink is a powerful framework for stream and batch processing of data. It is highly scalable and fault-tolerant, making it suitable for real-time data analytics. Flink is known for its stateful computations, which are highly useful in complex event processing and other use cases.

2 Installation of Apache Flink

2.1 Prerequisites

Ensure you have Java installed on your machine. You can check the version of Java installed by running the following command:

```
java -version
```

2.2 Download Apache Flink

Visit the official Apache Flink website to download the latest version of Flink. Alternatively, you can use the following `wget` command to download the version directly:

```
# Download Flink 1.17.1
wget https://dlcdn.apache.org/flink/flink-1.17.1/flink
    -1.17.1-bin-scala_2.12.tgz

# Extract the archive
tar -xvzf flink-1.17.1-bin-scala_2.12.tgz

# Move to the Flink directory
cd flink-1.17.1
```

2.3 Running Flink

Start the Flink cluster by executing the following command:

```
./bin/start-cluster.sh
```

The Flink web interface will be available at `http://localhost:8081`.

2.4 Submitting a Flink Job

Once the Flink cluster is up and running, you can submit a sample WordCount job using the following command:

```
./bin/flink run examples/streaming/WordCount.jar
```

2.5 Stopping Flink Cluster

To stop the Flink cluster, execute the following command:

```
./bin/stop-cluster.sh
```

3 Conclusion

In this document, we covered the installation steps for Apache Flink and how to submit a simple job. Flink is a robust platform that can scale with ease for both batch and stream processing needs.

4 Basic Codex

In this document, we will cover some basic commands for managing Apache Flink, including starting and stopping the cluster, submitting jobs, and monitoring jobs. Screenshots of the UI and commands will be included to provide a visual guide.

5 Cluster Management

5.1 Starting the Flink Cluster

To start the Flink cluster (which includes the JobManager and TaskManagers), use the following command:

```
./bin/start-cluster.sh
```

After executing this command, the Flink web interface will be accessible at `http://localhost:8081`. Below is a screenshot of the cluster startup process in the terminal:

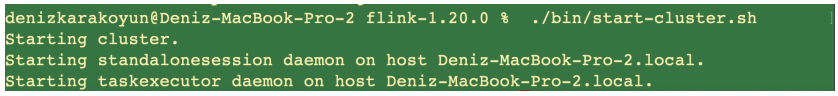

Figure 1: Starting the Flink Cluster in the Terminal

5.2 Stopping the Flink Cluster

To stop the cluster, use the following command:

```
./bin/stop-cluster.sh
```

6 Flink Web UI

Once the Flink cluster is running, you can access the Flink Dashboard via a browser at `http://localhost:8081`. Here, you can monitor the running jobs, task managers, and cluster metrics. Below is a screenshot of the Flink web interface:

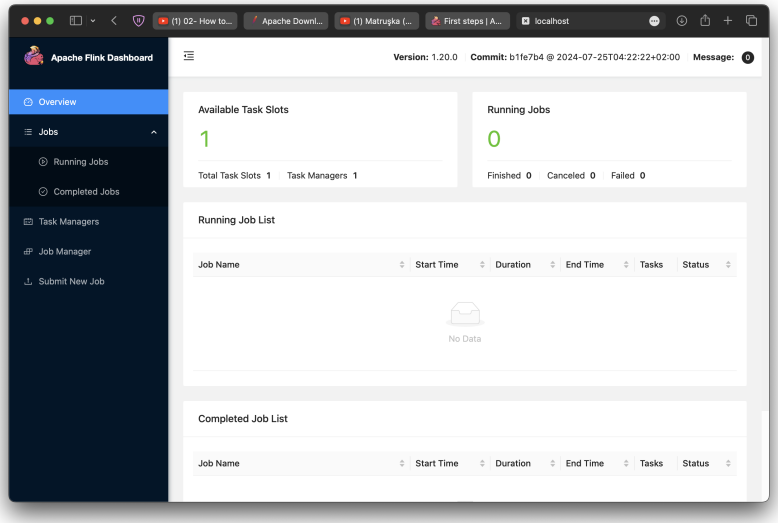

Figure 2: Flink Web UI Dashboard

7 Job Management

7.1 Submitting a Job

To submit a job to the Flink cluster, use the following command:

```
./bin/flink run <path-to-jar>
```

For example, to run the WordCount job included with Flink, use:

```
./bin/flink run examples/streaming/WordCount.jar
```

7.2 Listing Running Jobs

To list all running jobs on the cluster, use the following command:

```
./bin/flink list
```

7.3 Canceling a Job

To cancel a running job, use its Job ID from the list command:

```
./bin/flink cancel <job-id>
```

For example:

```
./bin/flink cancel d8f60a8b1
```