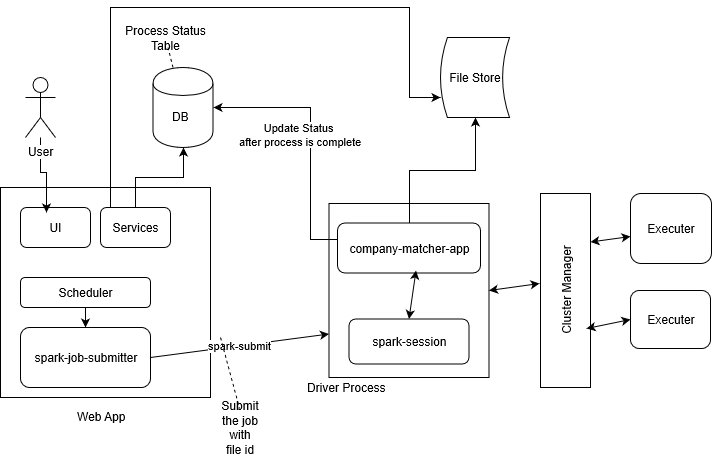
**Proposed Solution**:

1. User uploads the file. File id is generated and store in file storage (could be network drive, shared file system or s3 bucket). Request status is PENDING for newly submitted user requests (DB: REQUEST\_STATUS\_TABLE)
2. Job Scheduler which triggers in every 30 seconds (configurable time), picks up the PENDING requests from DB and submits to the spark (spark-job) programmatically. It also changes the request status from PENDING to PROCESSING
3. Once job is submitted to spark, company matcher app starts processing file. File id, location are provided as an argument to matcher-application. Once the process is complete matcher-application updates the DB to mark the request status as either PROCESSED or ERROR (REQUEST\_STATUS\_TABLE)
4. Request status UI shows the data from REQUEST\_STATUS\_TABLE.

Sample REQUEST\_STATUS\_TABLE Table data:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Request ID** | **Request Status** | **Request File Content** | **User** | **Created On** | **Generated Data** |
| 1 | Complete | [blob] | Riyaz S | 09-Feb-20 | [blob] |
| 2 | Processing | [blob] | Riyaz s | 10-Feb-20 |  |



**Response to follow-up Questions:**

1. Since the propose solution will run on big-data processing platform (spark), the