

MoSCoW Requirements

Oliver Little

2011802

October 3, 2022

Functional / Non-Functional	Must (M) / Should (S) / Could (C)	Requirement Description
Data Processing		
F	M	The system must allow users to filter datasets according to custom criteria.
F	M	The system must allow users to join two datasets together according to custom criteria.
F	M	The system must allow users to group portions of datasets together by unique keys.
F	M	The system must allow users to apply custom functions to each row of a dataset.
F	M	The system must allow users to apply a reducer function to a dataset.
NF	S	The complexities of the system should be hidden from the user; the operation of the system should be identical whether the user is running the code locally or over a cluster.
Cluster		
F	M	The user should be able to index specific columns of the data to allow these rows to be accessed more quickly by the cluster.
F	S	The cluster should be able to handle node failures by restarting them and requesting new work.

NF	S	The cluster should feature some form of load balancing to ensure that nodes are not idling if there is still work to be performed.
Main Nodes		
F	M	The main node must delegate work to the cluster nodes to perform the calculation efficiently.
F	M	The main node must collect the results from the cluster nodes to produce the final output for the user.
F	M	The main node must handle cluster node failures to ensure that the calculation completes successfully anyway.
Automated ETL Workflows		
F	S	The user should be able to define workflows, defined as the below requirements in this section (<i>Automated ETL</i>), using some kind of configuration file.
F	S	The system should receive files sent from the user, ingest them into the system automatically, and start a calculation when the files are sent.
F	S	The user should start calculations manually, schedule them for a specific time, or schedule them at a regular interval.
F	S	The system should output the results of a calculation to a user-defined location.
F	S	The system should raise alerts when an unrecoverable failure occurs during a calculation.
F	C	The system could send alerts via email to the user when an unrecoverable failure occurs during a calculation.
F	C	The system could monitor a specific folder location for changes, then initiate a calculation when files in that location are changed.
Dashboard		

F	S	The system should have a dashboard to display basic information about user-created workflows (<i>from Automated ETL</i>), including source and destination files, and previous executions.
F	C	The system could allow the user to start or schedule workflows from the dashboard.