HDFS

Data in form
of CSV files

Feature
Extractor

Feature extraction in
Flink environment

1	1	26.5	603	.
1	2	34.8	1204	.
1	3	27.3	1044	.
1	4	41.2	896	.
.	.	.	.	.

Sample N x M
feature matrix