

Apache Druid allows clustering of the data and pre-analytics by putting the data into respective buckets by a particular parameter (time, location and such).

"Original storage" is in there to keep pure data as we might need it to do analytics and possibly rerun the pipelines for data optimization in any case of failure or requirement of additional data, creation of the new index and etc.

"original storage" can be anything suitable enough for our needs. It can be a data lake (might be S3 based), can be Elastic (or similar in GCP/Azure), can be NoSQL with RPC or Graph connection on top. This will allow us to quickly analyze the incoming data filtering on a particular flight.