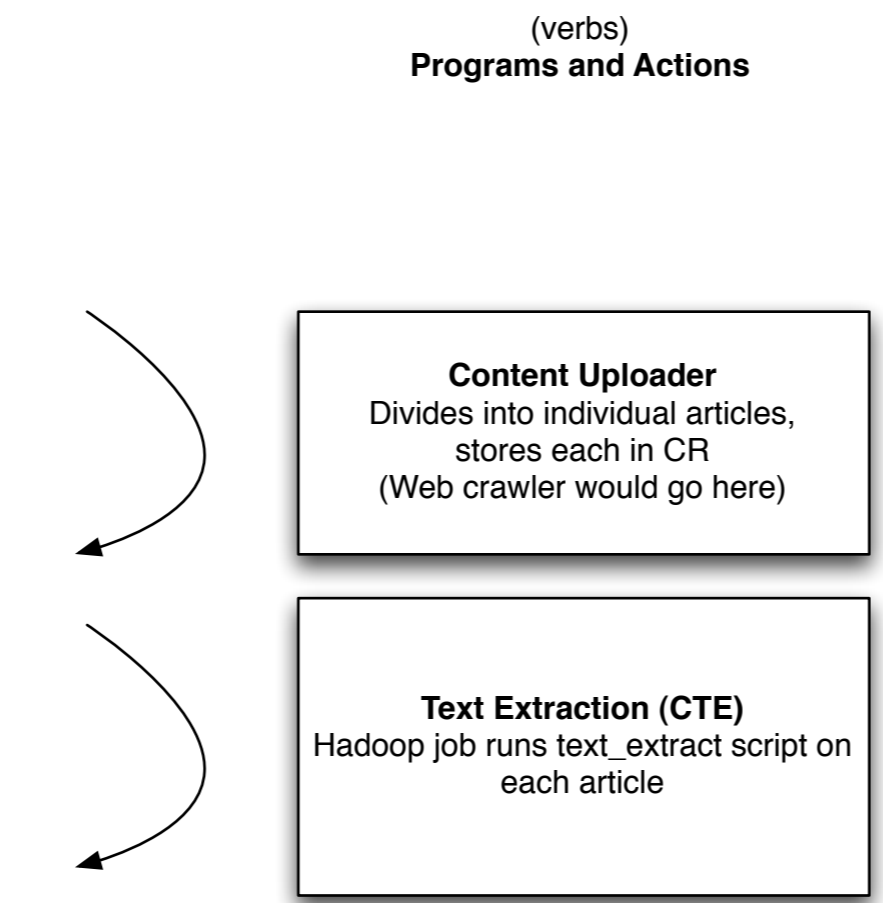
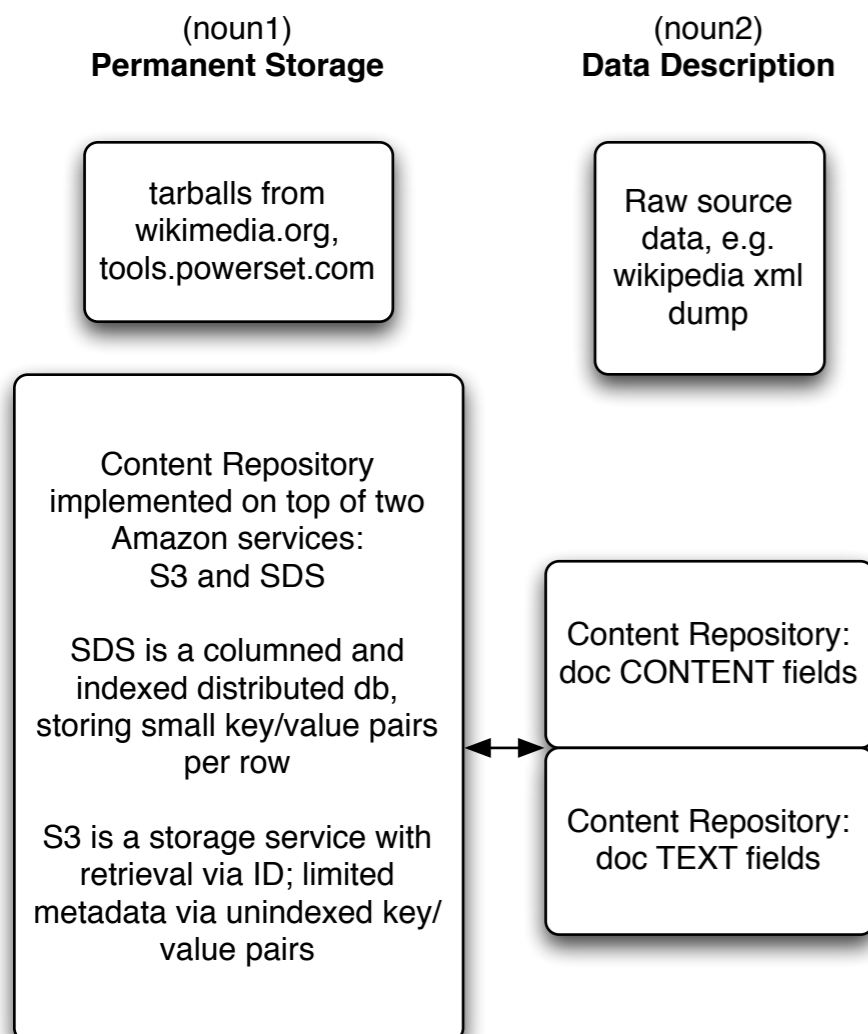


Update of the "January 2007" diagram on <https://wiki.powerset.com/twiki/bin/view/Engineering/IndexingDiagram>
 Brendan O'Connor, 7/29/2007



----- START CoralScript Pipeline processing phase -----

Below this line is the CoralInterpreter CoralScript Pipeline: per-sliver, per-article processing on a single local machine. Data has hierarchical organization: Project => (Slices) => Workunits => Slivers => Articles => Sentences. One Workunit = One Hadoop job. The upper levels of hierarchy are necessitated by limitations of Hadoop. The Hadoop MapTask consists of sequentially executing all pipeline steps on a single Sliver. The sequence of pipeline steps and their parameters are defined by a CoralScript. Bulk of processing occurs here: mostly Parsing, and secondarily Semrepping.

(indirect objects) **Resources**

Rule-based system

Knowledge resource

