Getting Data to the Mapper

- **The data passed to the Mapper is specified by an *InputFormat***
  - Specified in the driver code
  - Defines the location of the input data
    - Typically a file or directory
  - Determines how to split the input data into *input splits*
    - Each Mapper deals with a single input split
  - Creates a RecordReader object
    - RecordReader parses the input data into key/value pairs to pass to the Mapper

cloudera®
Ask Bigger Questions