Hand-coded scanners

- Widely used in practice, relies on developers to recognize opportunities for optimizations

- Might include extra enhancements such as standardizing input format, stripping comments, etc

- Might choose to simplify processing by lumping keywords together with identifiers, then add an extra processing stage for tokens recognized as identifiers: perform something like a dictionary lookup to check if it is actually a keyword

Input buffering

- Rollback often involves expensive i/o operations: putting characters back into the input stream then re-reading them later

- Common optimization is to read a chunk of input into a buffer then process from the buffer content, read new chunk when buffer is exhausted

- What if rollback takes back to position before start of buffer? Can maintain sequence of "small" buffers, most work happens in current, but have ability to roll back further "cheaply" if necessary