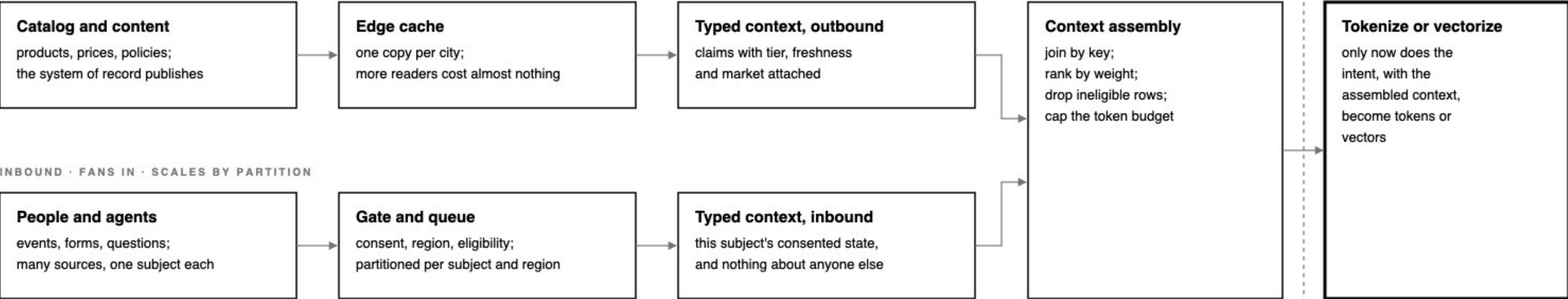


CONTEXT AND DATA STREAMS — SCALE BY DIRECTION, TOKENIZE INTENT LAST

outbound fans out and scales by cache; inbound fans in and scales by partition; both are typed, gated and ranked before anything becomes tokens or vectors

Dashed line = the tokenization line. Heavy box = where intent becomes tokens.

OUTBOUND · FANS OUT · SCALES BY CACHE



BEFORE THE LINE: TYPED, FILTERABLE, CHEAP · AFTER IT: TEXT AND VECTORS, PAID PER TOKEN

the tokenization line

