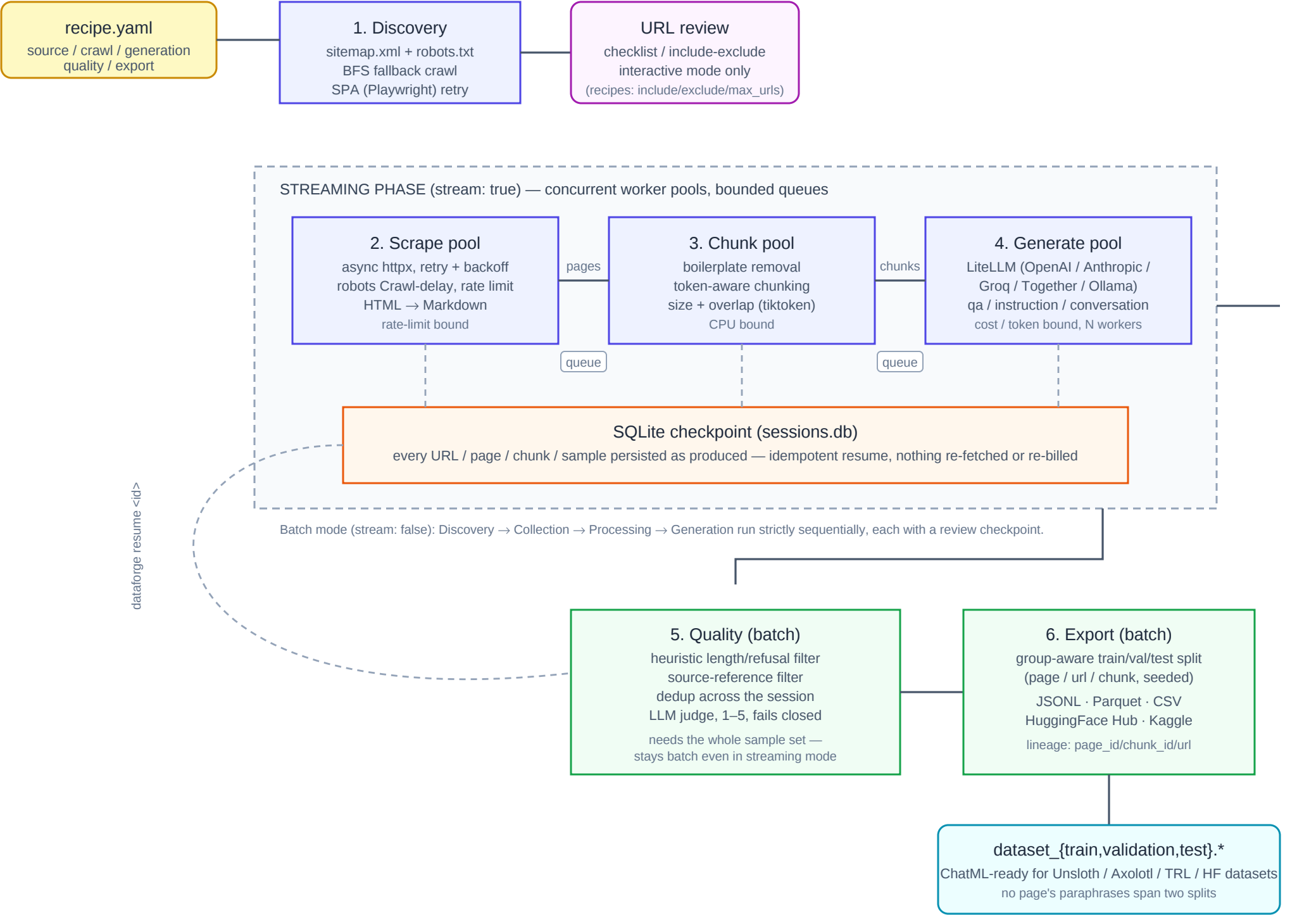


DataForge — Pipeline Architecture

Recipe (YAML) in → checkpointed, scored, split dataset out. Streaming mode overlaps collection / processing / generation.



LEGEND

- streaming / stage agent
- batch-only stage (needs full sample set)
- persisted state (SQLite)
- data flow
- - - - checkpoint / resume