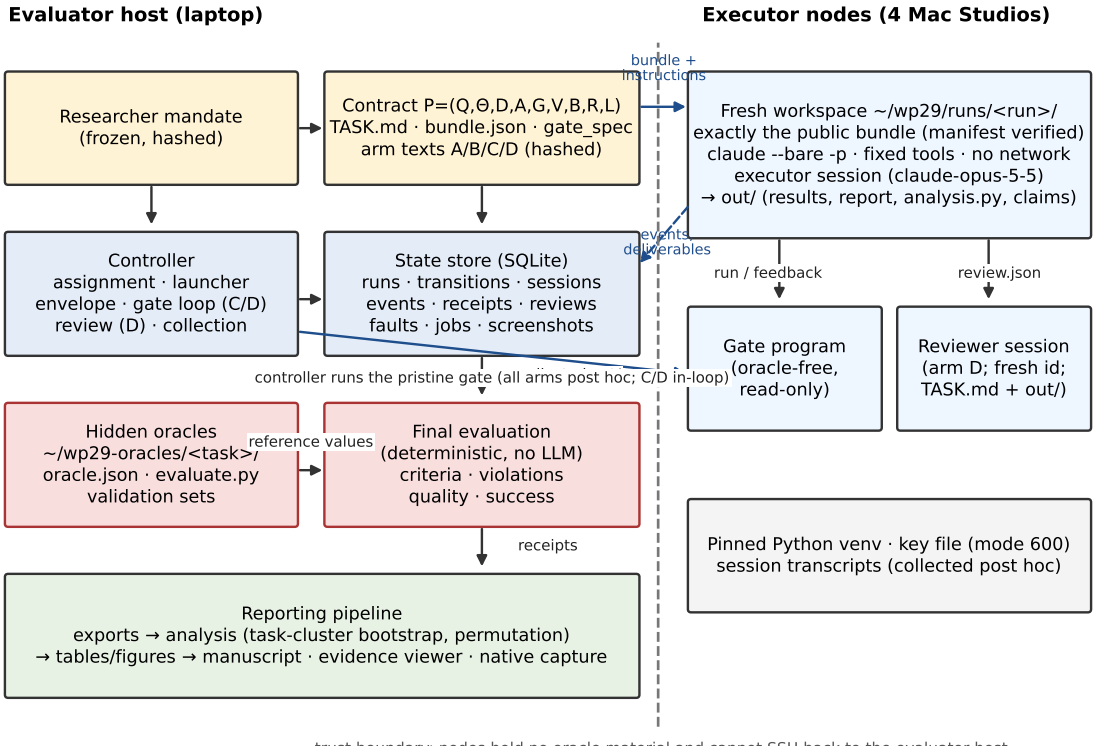

Evaluator host (laptop)
Executor nodes (4 Mac Studios)
Researcher mandate
(frozen, hashed)
Contract P=(Q,Θ,D,A,G,V,B,R,L)
TASK.md · bundle.json · gate_spec
arm texts A/B/C/D (hashed)
bundle +
instructions
Fresh workspace ~/wp29/runs/<run>/
exactly the public bundle (manifest verified)
claude --bare -p · fixed tools · no network
executor session (claude-opus-5-5)
→ out/ (results, report, analysis.py, claims)
Controller
assignment · launcher
envelope · gate loop (C/D)
review (D) · collection
State store (SQLite)
runs · transitions · sessions
events · receipts · reviews
faults · jobs · screenshots
events
deliverables
run / feedback
review.json
Gate program
(oracle-free,
read-only)
Reviewer session
(arm D; fresh id;
TASK.md + out/)
controller runs the pristine gate (all arms post hoc; C/D in-loop)
Hidden oracles
~/wp29-oracles/<task>/
oracle.json · evaluate.py
validation sets
reference values
Final evaluation
(deterministic, no LLM)
criteria · violations
quality · success
receipts
Pinned Python venv · key file (mode 600)
session transcripts (collected post hoc)
Reporting pipeline
exports → analysis (task-cluster bootstrap, permutation)
→ tables/figures → manuscript · evidence viewer · native capture
trust boundary: nodes hold no oracle material and cannot SSH back to the evaluator host