Query Cost Isolation Under Shared Multi-Region SQL Joins

Prepared for Hannah at PostHog (posthog.com) - 2 analytical spend bottlenecks and 3 tenant-level guardrails

OBSERVED SIGNAL & CITATION

Expanded cross-source SQL warehouse joins across shared US and EU analytical tiers.

Source: https://posthog.com/docs/data-warehouse

PRACTITIONER CONTEXT

Self-serve SQL joins across external billing tables and raw event partitions raise scan variance sharply when heavy tenants share ClickHouse compute pools. The patterns below focus on isolating wide-scan workloads before they erode event margins.

1. TWO CONCRETE EXPOSURE & BOTTLENECK PATHS

[PATH_1] Unbounded cross-source joins exhausting shared memory pools

Mechanic: Ad-hoc joins between external synced tables and unpartitioned event ranges trigger broad granule reads that spike tail latency for neighboring accounts.

Why here: Warehouse SQL supports arbitrary joins across synced CRM and Stripe tables inside the same analytical tier. [Evidence: F-01, F-02]

[PATH_2] Cross-region backfill retries inflating storage and compute burn

Mechanic: Incremental sync replays across US and EU clusters duplicate part merges when schema drift causes silent retry loops overnight.

Why here: Dual-region operation across US and EU clouds doubles merge overhead whenever external connectors change schemas. [Evidence: F-01, F-03]

2. THREE PRACTICAL CONTROLS (FULL PATH COVERAGE)

Per-tenant read-byte budgets with automatic pool spillover (Covers: path_1)

Route queries projected above 50 GB scanned into an isolated low-priority pool with strict memory limits before execution.

Pre-execution partition pruning validator on external joins (Covers: path_1)

Require time-bound predicates on event tables joined against external tables and surface estimated scan cost prior to run.

Idempotent sync checkpointing with schema-diff circuit breakers (Covers: path_2)

Pause connector retries after two identical schema mismatch failures and emit structured margin alerts per cluster.

3. DIAGNOSTIC QUESTIONS FOR YOUR TEAM

1. What share of monthly ClickHouse compute spend currently comes from the top 5 percent of warehouse SQL tenants?
2. Do ad-hoc joins against synced external tables execute on the same worker pool as core funnel dashboards?
3. How quickly can engineering detect a silent connector retry loop before it triggers redundant part merges?