R&D Additional Information Form — Draft

Project: Adaptive query-plan selection under per-tenant skew for a multi-tenant analytics layer · **Status:** confirmed
Generated: SAMPLE — fictional company, not a real claim · **Software:** MergedClaimed

1. Claimant company

Company name	Helix Analytics Ltd
Companies House number	14872103
Incorporated	2022-06-14
SIC codes	62012, 63110
Registered office	Unit 3, Hoxton Works, 128 City Road, London EC1V 2NX
UTR	0123456789
VAT number	GB412889006
PAYE reference	120/HA12345
Accounting period	2024-07-01 → 2025-06-30
R&D scheme	rdec_merged

2. Project

Project name	Adaptive query-plan selection under per-tenant skew for a multi-tenant analytics layer
Field of science / technology	Computer science — distributed query optimisation
Project start	2024-08-05
Project end	2025-05-21
Resolved status	resolved

ADVANCE SOUGHT

> The project sought an advance in the field of database query optimisation by developing a per-tenant adaptive query-plan selector that operates within an OLTP/OLAP hybrid workload at sub-50ms decision latency. Existing public knowledge (Postgres' planner, Apache Calcite's volcano model, Snowflake's published 2023 papers on plan caching) does not address the specific challenge of plan selection where per-tenant data skew distributions change every 5–15 minutes.

BASELINE (STATE OF PUBLIC KNOWLEDGE BEFORE THIS PROJECT)

> Prior to the project, the publicly available state of the art on tenant-aware query planning relied on: (a) static cost models (Postgres, MySQL), (b) machine-learned selectivity estimators trained on global statistics (BAO, Neo from MIT 2019), and (c) Snowflake's published 2023 paper on plan caching with

> global-only statistics. None of these handle the per-tenant skew shift problem documented in our test bed.

TECHNOLOGICAL UNCERTAINTY

> The competent professional and team identified three areas of scientific or technological uncertainty: (1) whether learned selectivity estimators could be fine-tuned per-tenant within a 50ms inference budget on commodity hardware; (2) whether the per-tenant skew shift detector could distinguish genuine workload changes from sampling noise without producing > 2% false-positive plan flushes; and (3) whether the resulting system could maintain throughput parity with the static-plan baseline at the 99th percentile, given the additional decision overhead.

APPROACH TAKEN

> The team's approach combined three experimental tracks. Track 1 (Aug–Oct 2024) prototyped a per-tenant LightGBM selectivity estimator with online retraining every 2,000 queries. Track 2 (Sep–Dec 2024) built a CUSUM-based skew shift detector with a tunable sensitivity threshold. Track 3 (Jan–May 2025) integrated tracks 1 and 2 into the production query path behind a feature flag and ran a two-month A/B test at 5% of customer traffic. The final adopted system reduced p99 query latency by 31% relative to the static-plan baseline while keeping throughput within 2% of baseline. Source: PRs #418, #422, #441, #456, #473 and issue #389 in the cited evidence trail.

3. Competent professional

Name	Dr Anjali Khatri
Role	Head of Data Platform, Helix Analytics Ltd
Years of experience	11+
Qualifications	PhD Computer Science, Imperial College London (2014) — Distributed Systems
Email	—

4. Qualifying expenditure

Staff costs (salary + ER NI + pension × % time)	£186,500.00 (gross £298,400.00)
Externally Provided Workers (65% UK PAYE)	£27,300.00 (gross £42,000.00)
Subcontractor costs (65% unconnected UK)	£15,600.00 (gross £24,000.00)
Cloud computing & data licences	£21,080.00 (gross £38,400.00)
Software licences (R&D-only tools)	£6,000.00 (gross £10,000.00)
Total qualifying expenditure	**£256,480.00**

ESTIMATED CREDIT (CALCULATOR ONLY — VERIFY WITH YOUR ACCOUNTANT)

R&D scheme	rdec_merged
Headline rate	20%
Estimated gross credit	£51,296.00
Estimated net credit (after 25% corporation tax for RDEC)	£38,472.00

5. Evidence trail

Contemporaneous records linked to this project, captured at the moment the work happened. Source: customer's GitHub installation. HMRC's GfC3 guidance (2023) explicitly prefers this kind of contemporaneous evidence over reconstructed narratives.

PULL REQUESTS (5)

- **#418** Per-tenant LightGBM selectivity estimator (prototype) — akhatri, opened 2024-09-12, merged 2024-09-18 (+2840/-412)
- **#422** Online retraining loop with 2k-query window + ablation harness — rkonkar, opened 2024-10-03, merged 2024-10-09 (+1620/-188)
- **#441** CUSUM-based skew shift detector — sdebnath, opened 2024-11-14, merged 2024-11-22 (+980/-76)
- **#456** Integrate selector behind feature flag — A/B at 5% traffic — akhatri, opened 2025-01-20, merged 2025-02-04 (+1410/-322)
- **#473** p99 throughput regression fix — adaptive batch sizing — rkonkar, opened 2025-03-11, merged 2025-03-17 (+460/-192)

ISSUES (2)

- **#389** Spike: feasibility of LightGBM inference under 50ms p99 — akhatri, opened 2024-08-06, closed 2024-09-11 [research, spike]
- **#401** Baseline plan-cache hit rate per tenant — measurement plan — sdebnath, opened 2024-09-24, closed 2024-10-02 [measurement]