**GreenQuery - Sustainable BigQuery Optimization**  
*Reboot 2025 | Solution Design | Lloyds Technology Centre*  
**Date:** 27 June 2025

### **Problem Statement**

BigQuery is a powerful data warehouse solution, but without proper monitoring, it can become both expensive and energy-inefficient. Inefficient queries lead to increased compute costs and elevated carbon emissions. **GreenQuery** addresses these challenges by offering predictive, sustainable, and cost-efficient query optimization.

### **Vision**

Empower teams with actionable insights into BigQuery usage, enabling them to forecast costs and CO2 emissions before execution. GreenQuery provides intelligent, ML-driven recommendations to optimize queries and reduce resource wastage.

### **High-Level Architecture Solution**

**GreenQuery Architecture Flow:**

BigQuery  
 ↓  
Log Sink  
 ↓  
BQ Dataset  
 ↓  
BQ ML → Looker Studio → Users

### **Step-by-Step High-Level Solution**

1. Ingest BigQuery audit logs into centralized datasets.
2. Apply SQL transformations to identify inefficient or costly queries.
3. Estimate query cost and CO2 emissions using the Dry Run API.
4. Use BigQuery ML for classification, forecasting, and anomaly detection.
5. Serve dashboards and reports via Looker Studio for different personas (Executives, Analysts, Engineers).
6. Offer optimization suggestions (e.g., partitioning, filtering, clustering, caching).
7. Alert users about cost anomalies via Slack or Email.
8. Provide scheduled exports and APIs for external integration.

### **Key Features**

* Real-time cost and carbon estimation using Dry Run
* Query monitoring and optimization recommendations
* Forecasting with BQML (usage trends, classifications)
* Persona-based Looker Studio dashboards
* Alerts for inefficiencies and cost anomalies

### **Predict Query Cost Before Execution**

By using BigQuery’s **dry\_run** mode, users can estimate the amount of data processed, potential cost, and carbon footprint **before** execution. This helps teams catch expensive patterns early and optimize queries before deployment.

### **How Other Applications Can Utilize GreenQuery**

* **CI/CD Validation:** Integrate into data pipelines to classify and catch inefficiencies pre-deployment.
* **Query Advisory APIs:** Enable external systems to fetch optimization suggestions via APIs.
* **UI Embedding:** Embed the estimator into tools like AppSheet or Google Forms.
* **Alerts and Dashboards:** Teams can consume alerts and visual insights for monitoring and governance.

### **KPIs Enabled**

* Estimated cost and CO2 saved
* Average cost per query and per department
* Forecasted emissions vs. departmental budget
* Efficiency score and leaderboard rankings

### **Enhanced High-Level Solution Design**

#### **Creativity, Originality & Innovation**

GreenQuery uniquely blends cost optimization with sustainability by translating query behavior into measurable carbon footprint metrics. It innovates by combining **dry run estimations** with **ML-based insights** to foster environmental responsibility in data operations.

Gamification (efficiency leaderboards), modular dashboards, and real-time cost feedback make it engaging and impactful.

#### **Design Quality - Seamless Tech Integration**

* Fully serverless and GCP-native (BigQuery, Cloud Functions, Looker Studio)
* Modular SQL logic
* ML models for segmentation, anomaly detection, and efficiency scoring
* Clean integration interfaces across components

#### **User Experience - Ease of Use & Engagement**

* Persona-specific dashboards
* Dry Run estimator for cost/CO2 preview
* Optimization tips linked to actual queries
* Drill-down filters for interactive, self-service analytics

#### **Scalability & Maintainability**

* BigQuery offers native scalability
* Cloud Functions automate dry runs, alerts, and advisory
* Fully managed, no server maintenance
* Easy to extend with new models or features

#### **Functionality & Extensibility**

* Real-time and historical query insights
* Dry run estimators (live and logged)
* Forecasting with ARIMA (BQML)
* Efficiency scores and gamification

**Future Extensibility:**

* Gemini-based query rewriting
* CI/CD guardrails for query cost budgets
* Integration with ServiceNow, Microsoft Teams, or Slack
* Geo-specific emissions factors
* Real-time cost blocking for sensitive workflows

#### **Multi-User & Customer Usability**

* Supports GCP IAM integration for auditability
* Dashboards and alerts scoped by user and department
* Turnkey deployment across projects and organizations
* No coding needed for end-users

### **Conclusion**

**GreenQuery** delivers an actionable, scalable, and sustainable solution for BigQuery optimization. Built entirely on Google Cloud, it supports team-level accountability, enables efficient query usage, and aligns with Lloyds’ sustainability and cost-efficiency goals.