Pylon

University of Cambridge

Desmond Zee (Eng) - Team Lead - dz386@cam.ac.uk - coffeemochii78
James Carver (Phys) - FDE - jjsc2@cam.ac.uk - jamesagainn
Dominic Henderson (Eng) - AI Engineer - dgkdh2@cam.ac.uk - cdk1_16336
Nikolaus Nieweth (Phys) - GTM - cnn28@cam.ac.uk - cnn1_1

Problem Focus

This submission addresses **Problem Statement 2 – Compute Energy Convergence**. The system models multi-region data-centre compute assets and their operational characteristics, including server-level and rack-level energy use, thermal constraints, and available compute capacity. **Grid-side forecasts** for carbon intensity, net demand, renewable generation, and wholesale prices are aligned with data-centre telemetry to create a unified, time-indexed representation of both compute supply and grid conditions. Workloads are scheduled into **"workload windows"** (region × time slots) such that:

- runtime, capacity, thermal, region, and deadline constraints are always satisfied,
- inference and training workloads are shifted away from high-carbon or **high-stress grid periods**,
- HVAC and power draw associated with compute is redistributed into lower-carbon, lower-price windows,
- all allocatable compute slots and shifts are surfaced as **Beckn catalog items**, with full immutable auditability.

Solution Overview

The solution is a multi-agent compute scheduler built on a Python backend with Supabase as the state store, and HTML frontend. The mid-layer ontology functions as a digital twin of compute capacity, demand, grid conditions, and future forecasts, updated continuously via a structured ETL pipeline. External data sources (carbon intensity, price, demand, renewable forecasts) and internal telemetry (capacity, utilisation, thermal limits) populate relational tables across regions and time in a single aggregated location.
Three core server-side agents operate over this state with full Beckn protocol:

- **BAP Agents**: handles users requests for compute tasks,
- **BPP Agents**: Assess state of data centres and current energy capacities,
- **BG Agent**: Core mid-layer ontology managing communications, data-flow, privacy civil liberties, and database monitoring.

Each data centre is represented as a **Beckn BPP**, with a user-facing **BAP** interacting through a **Beckn Gateway**. The **BAP** aggregates catalogs from multiple **BPPs** and selects optimal schedules via a neural-net ranking function.

Technical Architecture

We adopt a tech stack that breathes and lives with the data, providing valuable data-driven insight whilst respecting privacy and civil liberties through granular data access rules. Leveraging Beckn protocol, Pylon is able to generalise to any number of BAP and BPP agents for large scaling efficiencies. Multi-Agent communication increases async productivity.

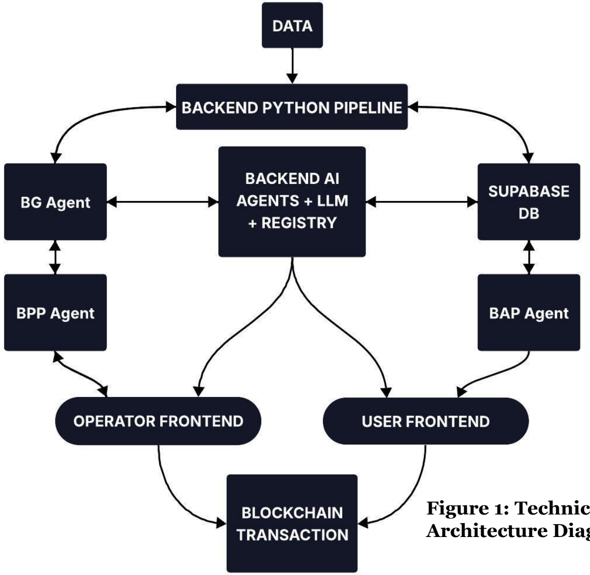

Figure 1: Technical Architecture Diagram

Agent Workflow

A. **User Frontend (BAP integrated**) receives compute workload → persists to Supabase → database trigger fires notification.
B. **Beckn Gateway (BG)** consumes notification → fetches live grid signals (carbon intensity, demand, price, renewable mix) + datacenter capabilities → LLM orchestrator generates n+1 suitability decision contexts → broadcasts Beckn-compliant catalog.
C. **Operator Frontend (BPP integrated)** ingests LLM output → generates weight assignments per datacenter (contextual trade-off matrices capturing cost/carbon/latency) → serves weights via endpoint.
D. **BAP** polls weights → stores assignments in decision graph → blockchain verifies workload placement immutably → agent logs execution trace to audit trail.

Every grid signal, datacenter attribute, and workload requirement is typed, linked, and queryable. Our AI Agents reasons over this **semantic richness**, generating contextual optimization weights that reflect **real-world constraints, not statistical artifacts**. Blockchain seals those decisions **immutably**. The result: AI that produces explainable, auditable, operationally meaningful recommendations—not just plausible-sounding guesses. We've decoupled LLM reasoning from data governance, enabling **intelligent orchestration without sacrificing transparency or trust**.

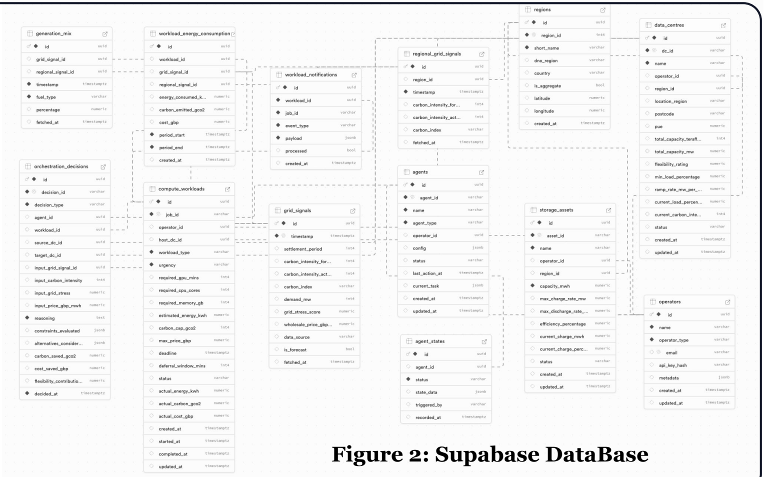

Figure 2: Supabase DataBase

Business Model & Impact

The Arbitrage: We monetize grid inefficiencies by treating compute workloads as tradable flexibility assets. Data-centre operators pay tiered SaaS (base MW subscription) + performance revenue share: energy-cost savings from carbon-aware scheduling + ESO market access (DFS, P415, wholesale). We operate an internal queue market—Priority (premium tier) vs. Flexible (discounted)—where our agents arbitrage tenant demand against external energy/flex markets in real-time.
Revenue Scales with Deployment: Each integration captures transaction value on arbitrage spreads and market-access events. No marginal cost per workload—profit compounds with scale.
Fast Deployment and Upskill: Leveraging FDEs, we build operational advantage into the platform itself, creating switching costs while competitors still build POCs.
Operational Impact: Utilities unlock granular, predictable flexibility. Big data is transformed from numbers to real interpretable semantics. Data centres unlock new revenue and ESG differentiation. Tenants cut per-inference costs and carbon. Grid peaks and curtailment collapse as AI workloads migrate into low-carbon windows. We're infrastructure-as-arbitrage: working to solve everyone's problem simultaneously.

References

1. National Grid Carbon Intensity Forecast (48h): https://api.carbonintensity.org.uk/intensity/fw48h
2. National Grid ESO Demand Forecast: https://api.nationalgrideso.com/api/3/action/datastore_search?resource_id=59f7e1fa-4dbd-43b0-8ea4-d884ef68553c
3. Renewable Generation Forecasts (Wind/Solar): https://data.nationalgrideso.com/renewables/embedded-wind-and-solar-forecasts
4. Elexon BMRS Energy Price Forecast (IPF): https://www.bmreports.com/bmrs/?q=api
5. Beckn Protocol Core Specifications (BAP/BG/BPP roles; search/on_search, select/on_select, confirm/on_confirm, status flows)

Declarations

All external data sources and libraries are appropriately licensed and referenced.
No confidential or proprietary third-party data has been included.

Signatory: Desmond Zee
Date: 23 November 2025