Geographic Resource Dispatch

a tool to facilitate optimization of distributed computing resources based on real-time market signals

INSIGHT

The Problem

Businesses increasingly rely on leveraging large data sets and distributed computing resources for processing.

Geographic distribution of these resources leads to variations in pricing and capacity of the nodes that can be leveraged to optimize output.

Inherent latency and reliability of network communications presents a challenge to reliable determination of the optimum dispatch scenario.

The Solution

A prioritized job queue, with metadata on the expected compute-hours and urgency for completion.

Real-time data streams (weather, energy cost) and stream processing (cooling energy) to determine resource capacity and pricing per compute-hour.

Optimization algorithm to generate dispatch queue for submitting jobs across the geographically distributed resource nodes.

Monitoring and visualization to facilitate manual interaction and analysis.

The Approach

Streaming architecture (Kafka) and stream processing (Kafka Streams) to accommodate real-time ingestion and processing use case.

Simulated data streams (Python) from static store (S3) of representative data to facilitate development for maximum flexibility.

- Streaming interval
- Geographic distribution

Log key data for metrics and visualization (Dash) of market optimizations.
(database to be determined)

The Technology
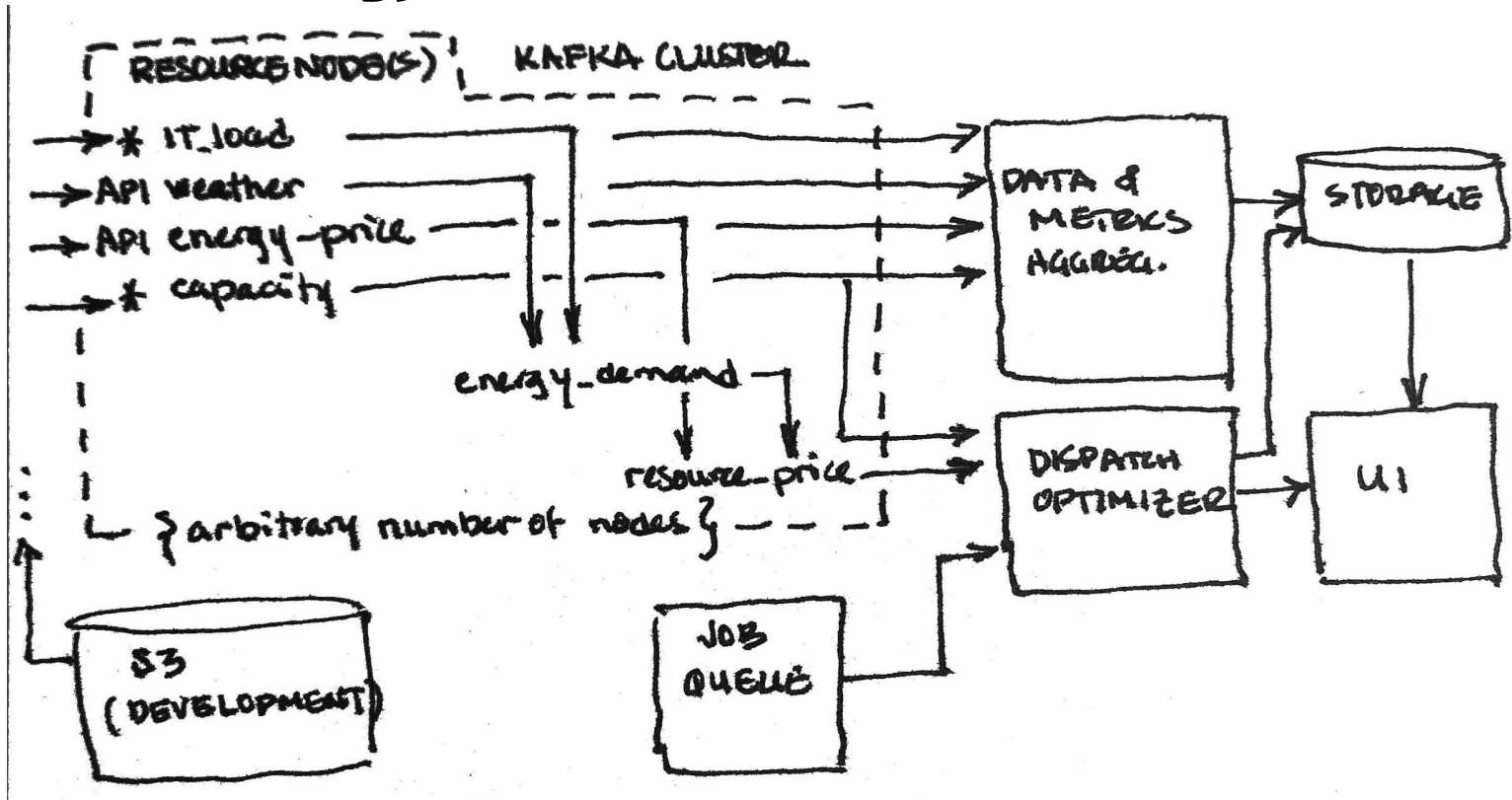
The Future

- Stabilize data pipeline
- Scale in time (shorter time intervals)
- Scale in space (more resource nodes)
- Time coordination of distributed streams (latency, node failure)
- Metrics and logging
- UI and visualization

- Ingestion of "real" data processing jobs

The Fellow

Charles Simchick

charles.simchick@gmail.com
linkedin.com/in/csimchick
github.com/simchuck/

Mechanical Engineer
Energy & Sustainability
Data Advocate

INSIGHT