Apache Spark

Sensors → Kafka → [Stream / Map / Batch] → MongoDB → OSRM → Website / Mobile App

OSM → OSRM

Technologies

Apache Kafka is a distributed streaming platform. mXpress utilizes Kafka in order to stream legacy sensor data into Apache Spark.

Apache Spark is a distributed cluster computing framework that facilitates our batch and streaming pipeline. Spark allows for extremely large data sets to be taken and processed in memory and across multiple different servers. This not only allows for processing time to be significantly reduced, this allows mXpress to be scalable across hundreds if not thousands of nodes in a larger application.

HDFS or Hadoop Distributed File System is the underlying file system that Spark interfaces with. It allows for a single set of data to be stored and accessed across multiple nodes.
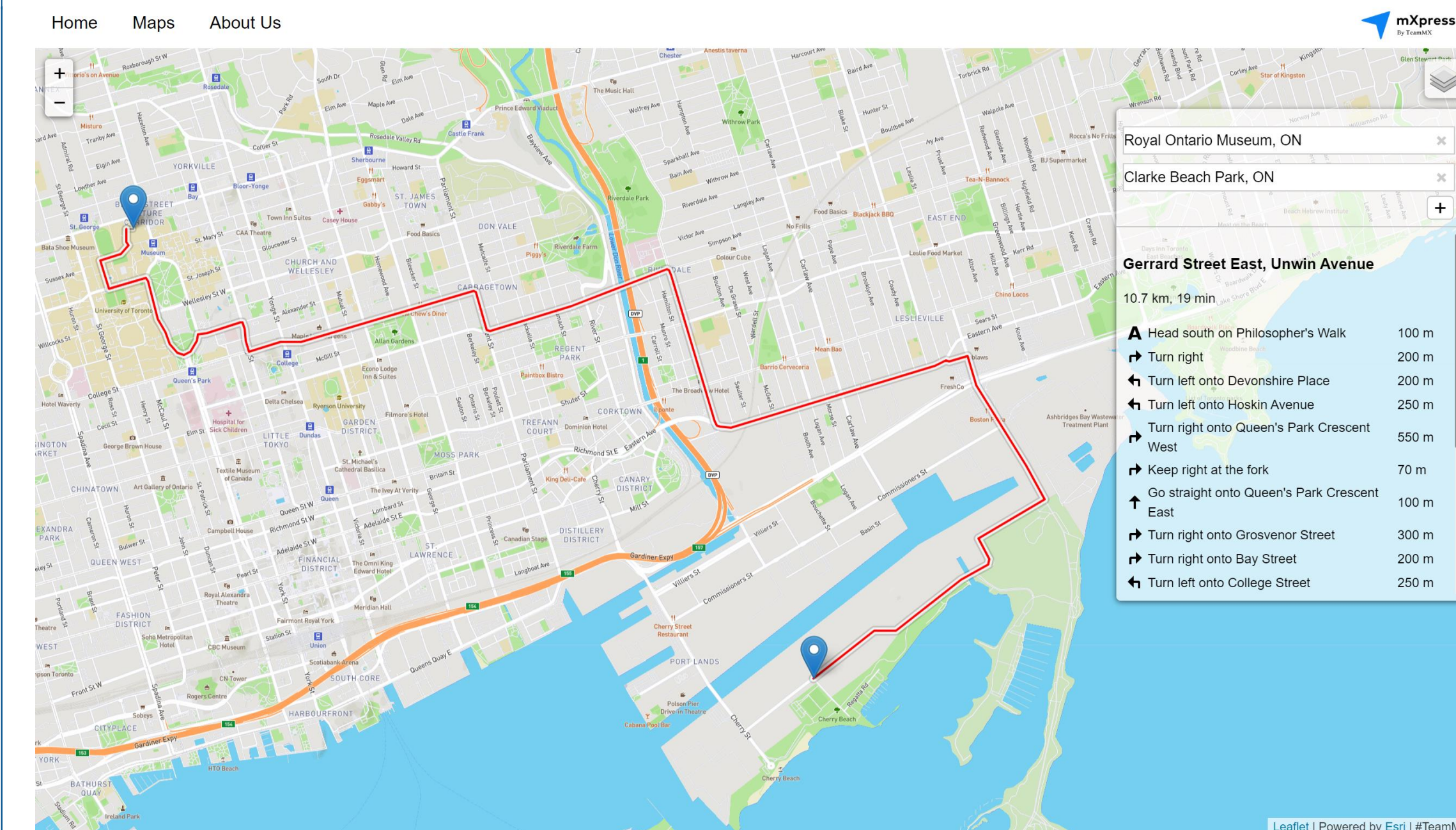
MongoDB is a intermediary database, it exists as a buffer, constantly accepting new data from Spark and delivering it as required to OSRM.

OSRM or Open Source Routing Machine is the routing engine. OSRM takes the road speed data provided by Spark and uses it to calculate the most efficient route possible when queried by leaflet.

Leaflet is a JavaScript library for interactive maps. Leaflet allows for online maps to be quickly and easily built using JavaScript. The leaflet-based interface allows the user to interact with the map and a plugin called leaflet routing machine facilitates routing requests. When a request is made leaflet routing machine queries OSRM for a route and then displays it on the map.

Map interface

Home   Maps   About Us

Royal Ontario Museum, ON
Clarke Beach Park, ON

Gerrard Street East, Unwin Avenue
10.7 km, 19 min

A	Head south on Philosopher's Walk	100 m
	Turn right	200 m
	Turn left onto Devonshire Place	200 m
	Turn left onto Hoskin Avenue	250 m
	Turn right onto Queen's Park Crescent West	550 m
	Keep right at the fork	70 m
	Go straight onto Queen's Park Crescent East	100 m
	Turn right onto Grosvenor Street	300 m
	Turn right onto Bay Street	200 m
	Turn left onto College Street	250 m

Leaflet | Powered by Esri | #TeamMX

Project Overview

mXpress is a big data framework to efficiently route vehicles in congested cities. It utilizes Apache Spark and Apache Hadoop to create a hybrid model of traffic patterns using historical data combined with real time streaming data. A front-end using leaflet and OSRM provides a responsive interface for user route generation and mapping. mXpress uses infrastructure sensors in the city of Toronto to gather data and create the most optimal route based on real world conditions. The technology behind the project is designed to be adaptable to a variety of use cases that require vehicle routing, including emergency services, ride sharing, food delivery, and transit networks.

Infrastructure
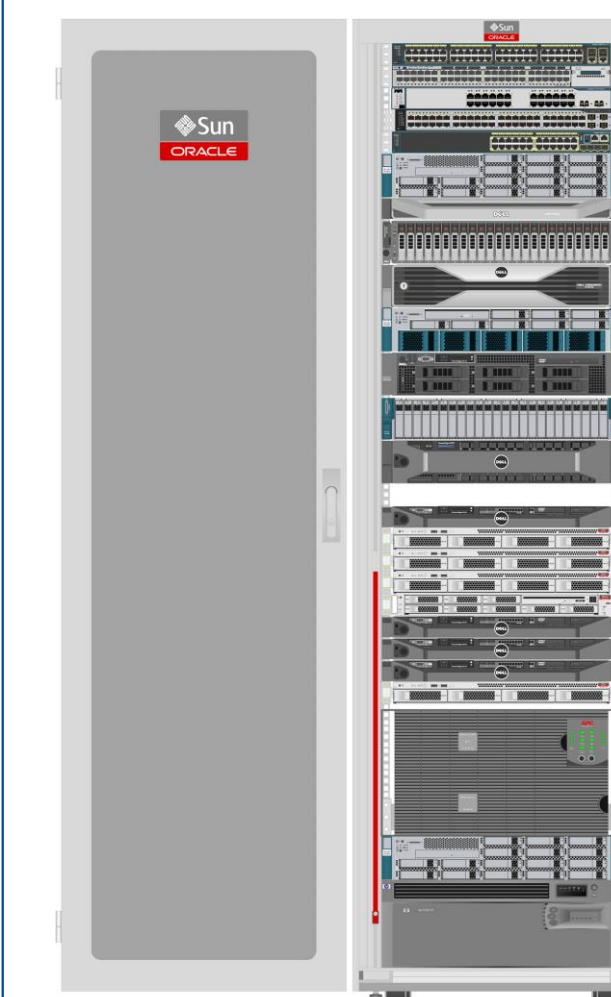
The infrastructure that comprises the backend of mXpress has an approximate cost of $2000. Average power consumption of the servers and networking equipment is around 3A on 240V service, averaging a 24/7 load of 518.4 kWh/month. In Regina, this costs around $71/month of residential electricity. In contrast, a similarly performant cloud deployment of mXpress would cost between $2000 and $15000 monthly depending on the service level required.