Book My Cab: Hop In, Let's Go!

Umesh Shukla

A Platform to book Cabs
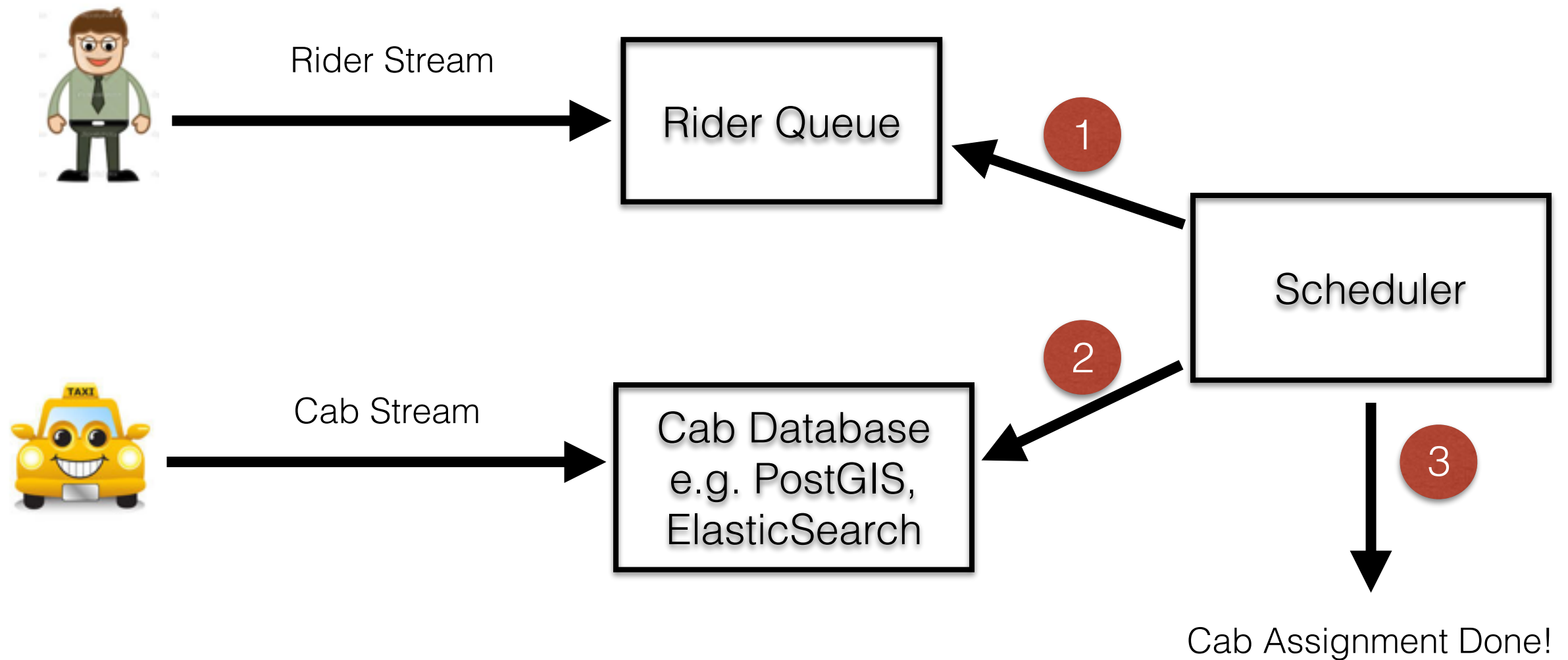
Send me a cab asap →

Cab Booking System

Send me a new rider asap →
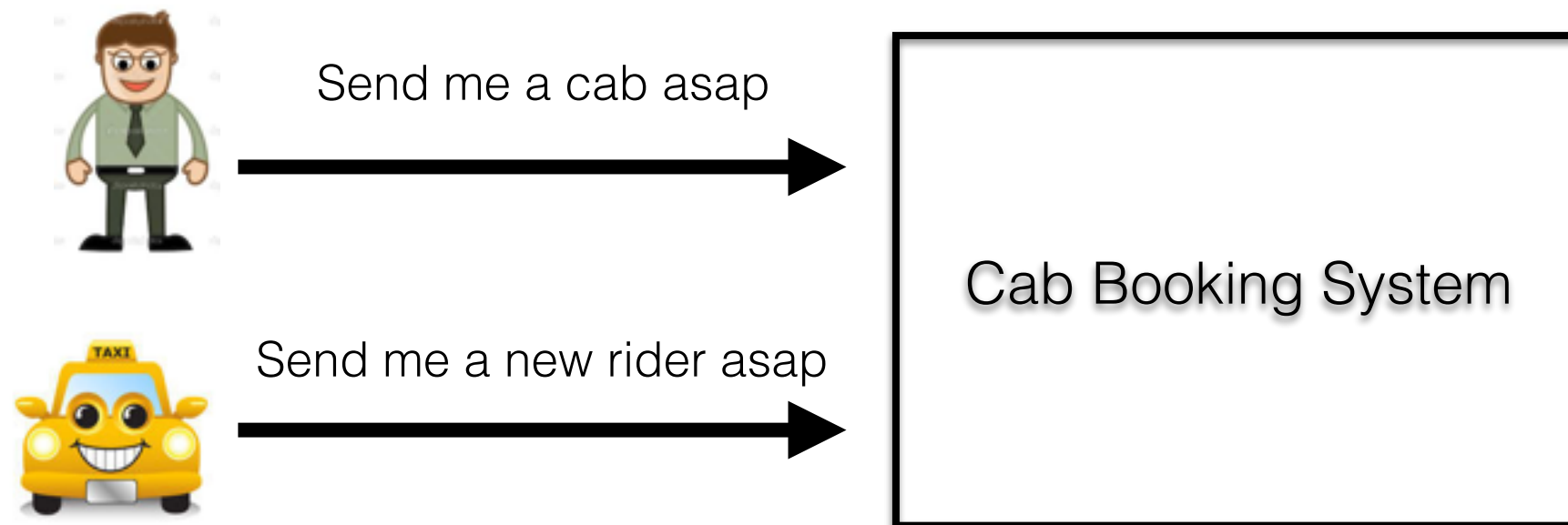
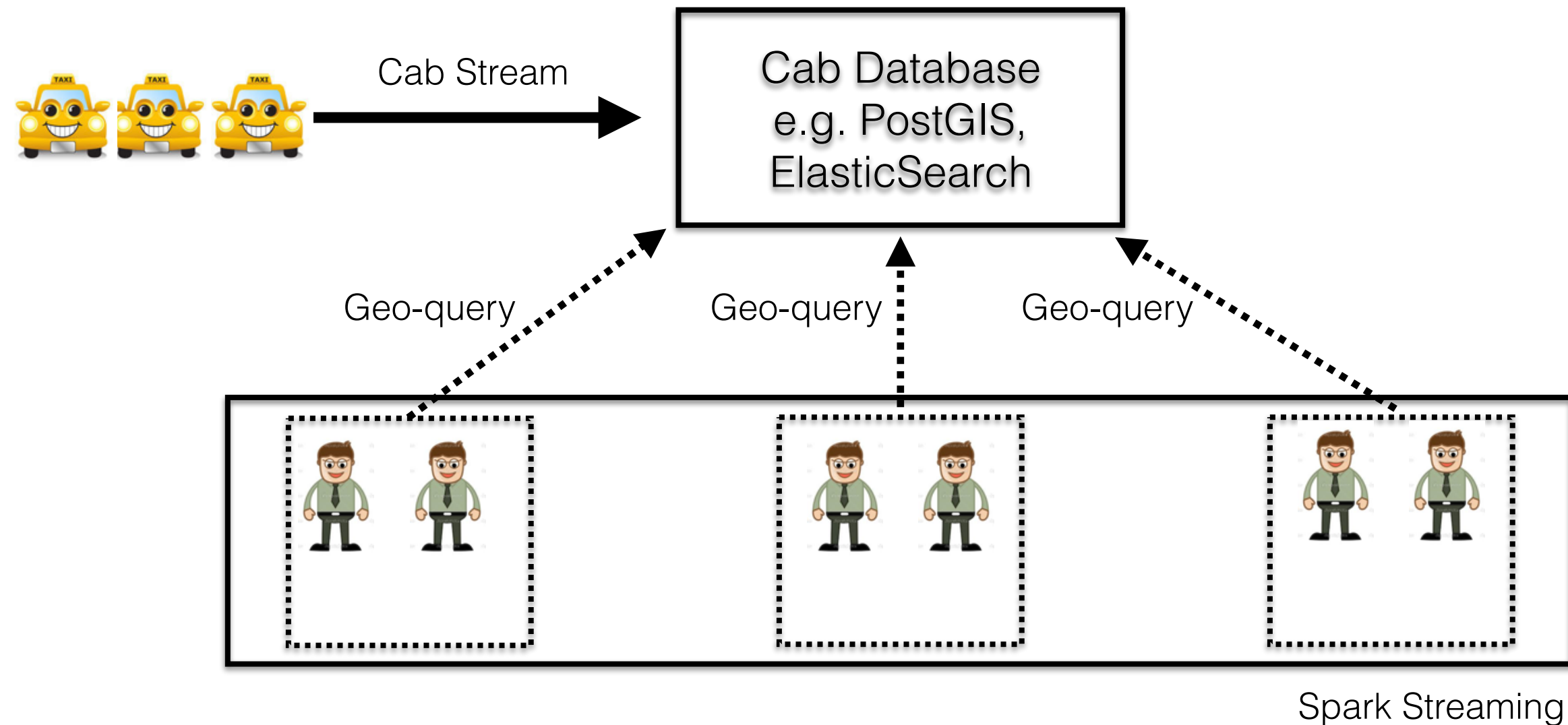How to do design a scalable real-time cab booking system?
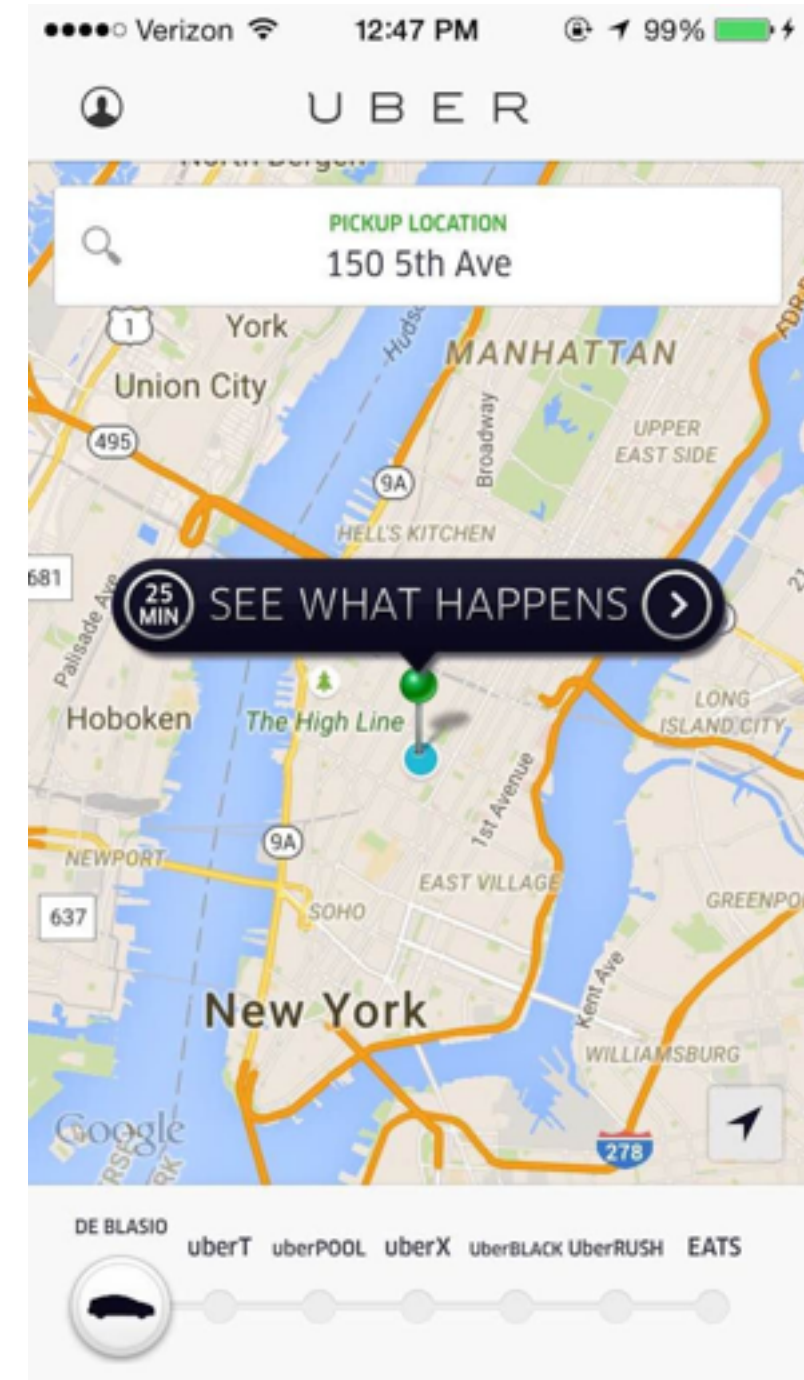
Demo

http://bit.ly/2dquv0B
(Might be down)

Basic Algorithm

How to Design this System in Distributed & Parallel Fashion?

Cab Stream

Cab Database
e.g. PostGIS,
ElasticSearch

Geo-query   Geo-query   Geo-query

Spark Streaming
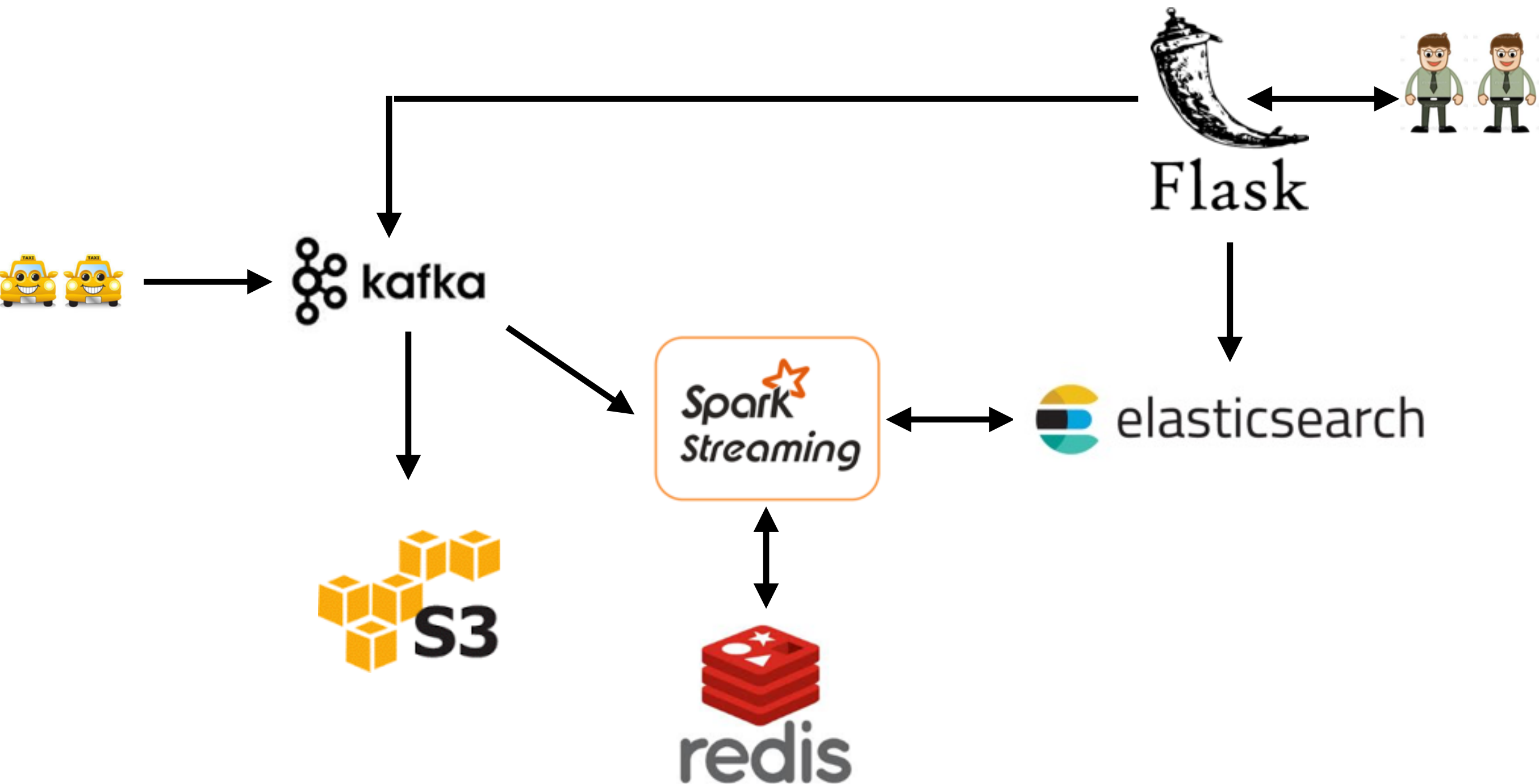
Over Booking Scenario
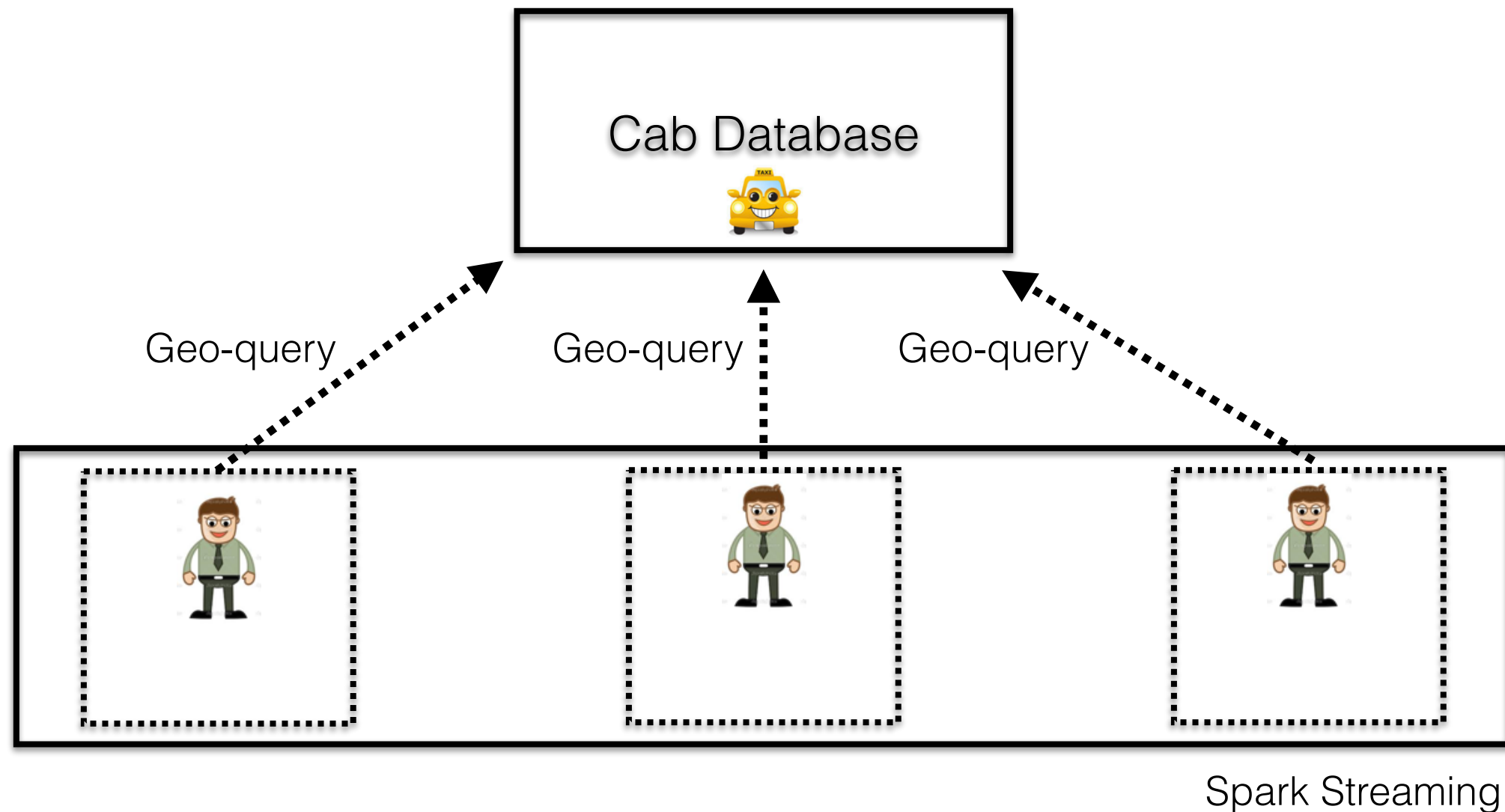
Over Booking Scenario
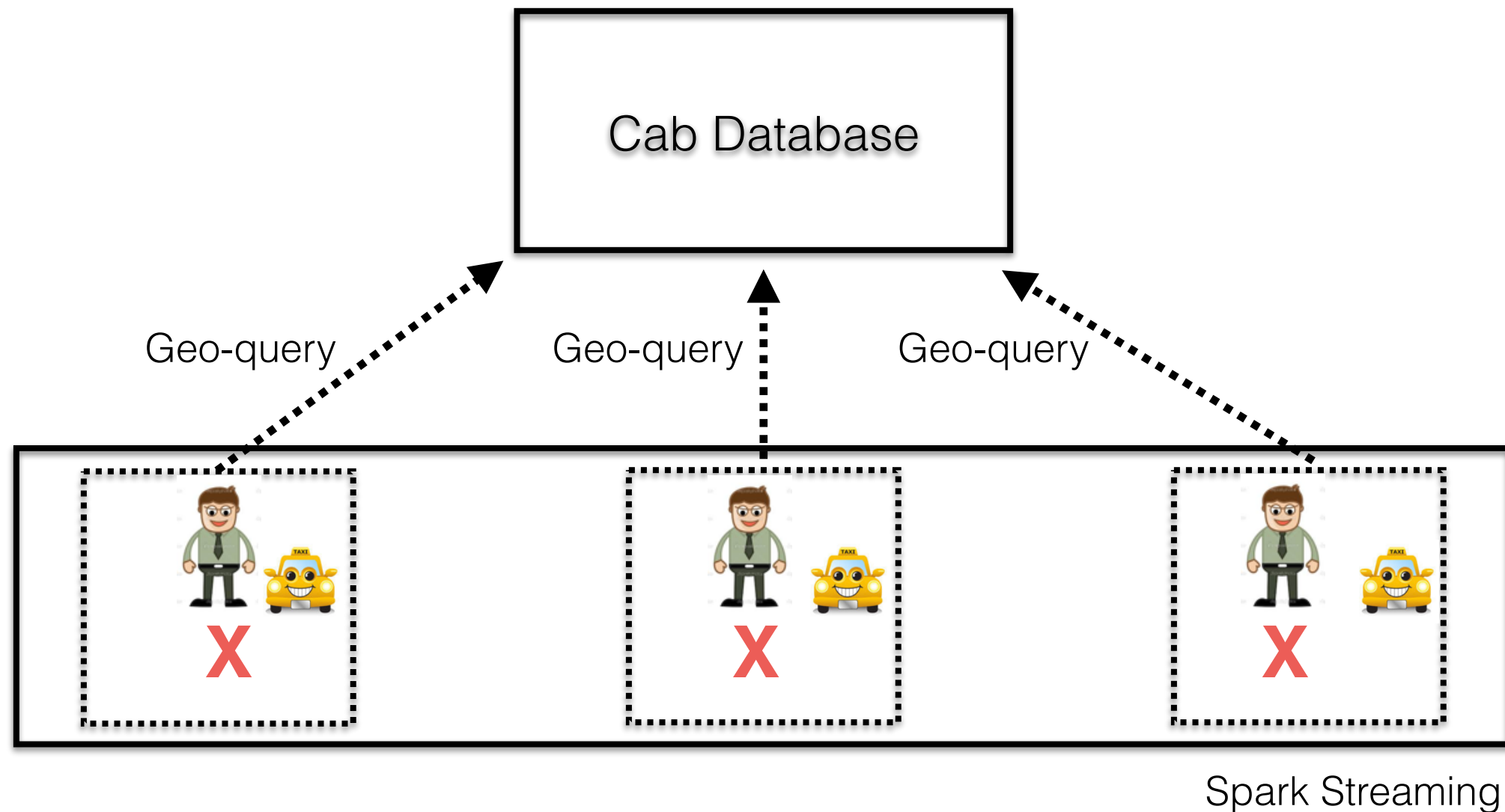
Solution: Use a transactional Cache e.g. Redis

Data Pipeline

Performance Numbers!

- Playback with NYC Yellow Cab data, 4-node AWS cluster

- System completes assignment of ~ 2000 customers and ~2000 drivers in 4-5 s. Spark Streaming Window Size: 5 s

Next Steps

- Take the full advantage of a micro-batch to come up with even more optimal cab assignment algorithm to improve cab utilization e.g. Maximal Weight Matching

- Challenge: More computationally intensive O(N^3) vs O(N^2)