

A Spatiotemporal Data Cube for Large-Scale Traffic Sensor Data Analysis



Afsin Akdogan, Fendy Santoso, Ugur Demiryurek, Cyrus Shahabi
Integrated Media Systems Center
University of Southern California

iCampus/MediaQ

iWatch

IT ✓

Introduction

Ever-growing data

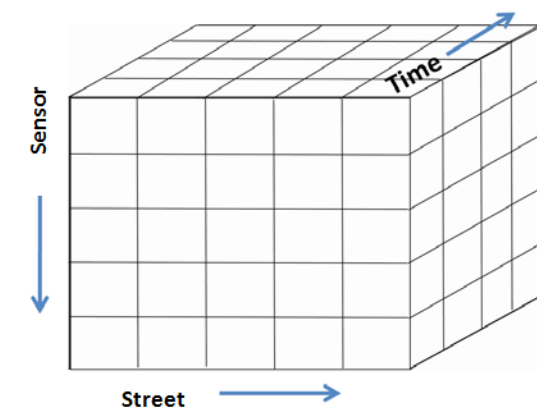
- The amount of traffic sensor data is rapidly growing. 4500+ sensors in Los Angeles freeways generate data in every 30 seconds.



Motivation

Spatiotemporal cube

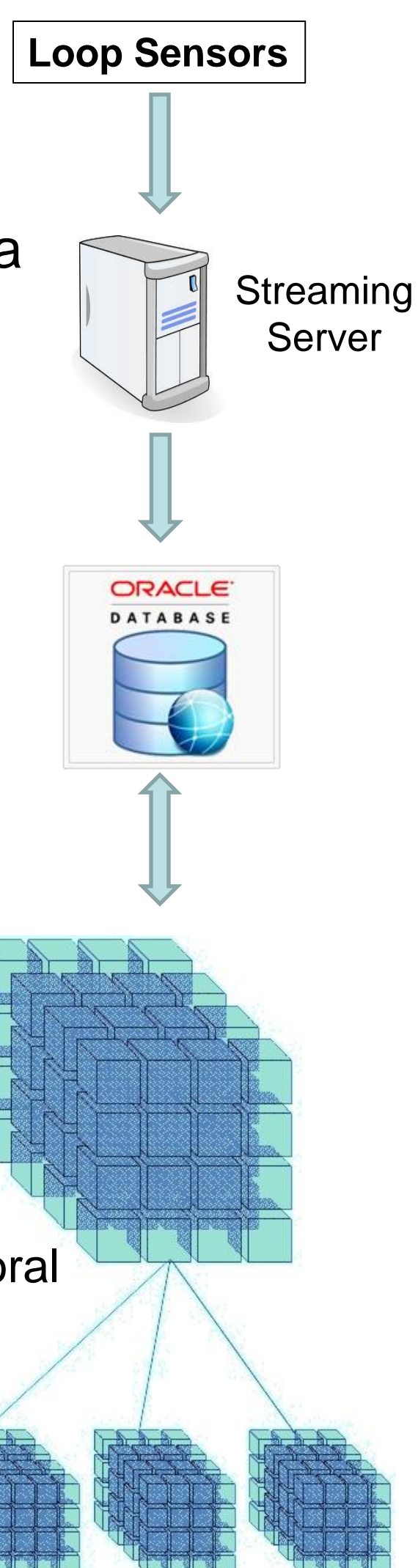
- Provide **scalable** traffic data analysis using both spatial and temporal dimensions. E.g., sensor, freeway, direction, latitude, longitude, year, month, day, etc.
- Efficient processing of various types of queries. E.g., traffic patterns for a given day, day of week, etc.
- Fast incremental update



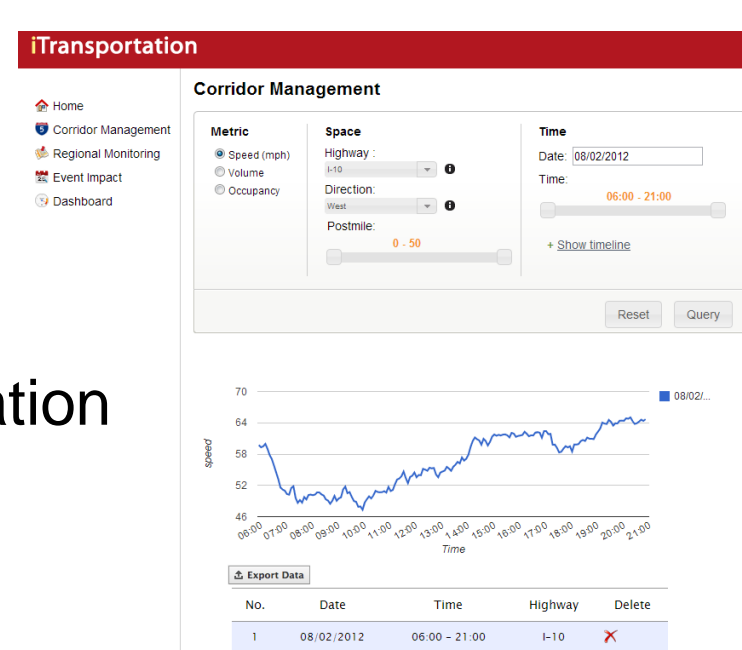
System Architecture

Challenges

- Conceptualizing the **multidimensional** view of data
 - How to break down the spatial features of the highways into directions, streets, latitude and longitude.
 - How many levels of hierarchy is required in the temporal dimension.
- Partitioning technique selection which will **horizontally scale**



iTransportation GUI



Spatiotemporal Cube

Design

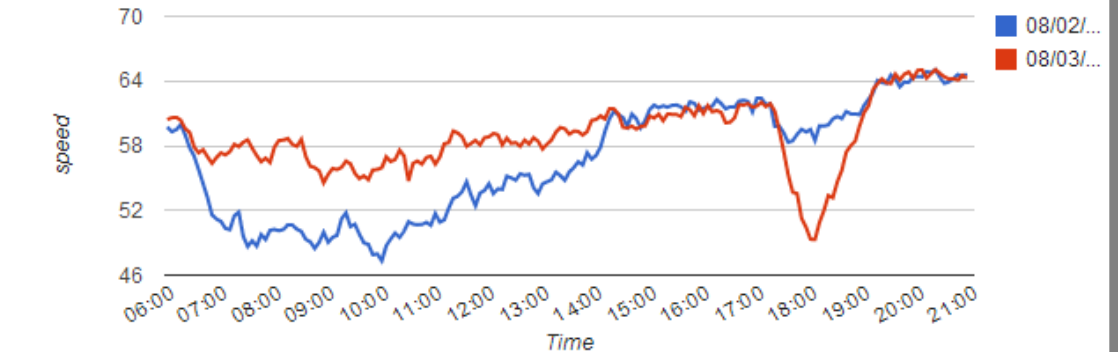
- Day based partitioning enables most of the queries to run on a single partition.
- Space dimension hierarchy
 - direction → highway → sensors
 - latitude, longitude
- Time dimension hierarchy
 - year → month → day → 5 min

Capability

Extensive query support

3 different metrics

- Average speed, volume, occupancy



Spatial dimension

- Spatial window query
- Sensor-level queries: a specific sensor or a list of sensors can be analyzed
- Freeway-level queries: entire highway or a certain segment on it can be analyzed
- Direction integration



Temporal dimension

- Multi-level: year, month, day of week, day, hour, 5-minute
- Time interval selection on any of the levels



Performance Analysis

Dataset

- Los Angeles highway sensor data collected in 2012 (2.5 billion records)

Evaluation

- Historical data store is range partitioned on month and hash partitioned on sensor id
- We report query response time for various types of queries.

Sample Query: Speed pattern on I-405 North from Orange County to I-10, from 6 AM to 10 AM.

- Cube: **1.6 seconds**
- Historical data store: **~85 minutes**

Conclusion and Future Work

- With the right design, OLAP Cube can answer many types of queries used in traffic data analysis.
- R-tree and Cube Integration
 - Building minimum bounding boxes (MBB) on sensors and injecting MBBs into the spatial dimension hierarchy.
 - This extension might provide better results for spatial window queries.