Data-Intensive Transportation Decision Making

Kristin Tufte, Basem Elazzabi, Nathan Hall, Morgan Harvey, Kath Knobe, David Maier, Veronika Megler

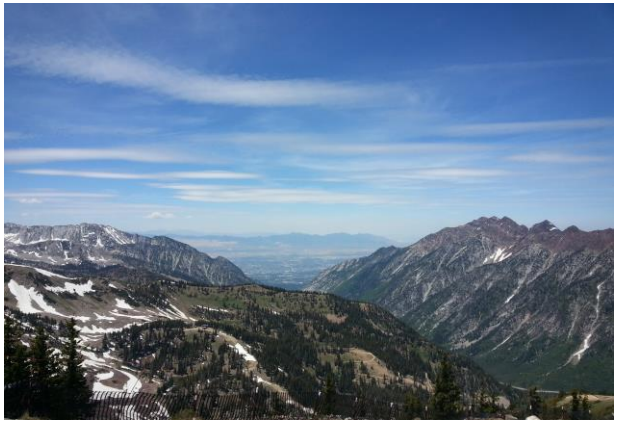
Portland State University
Maseeh College of Engineering and Computer Science

RICE Unconventional Wisdom

CnC-2014, Sept 18, 2014, Portland, OR

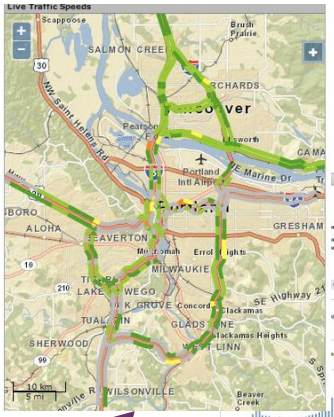
A bit about me...

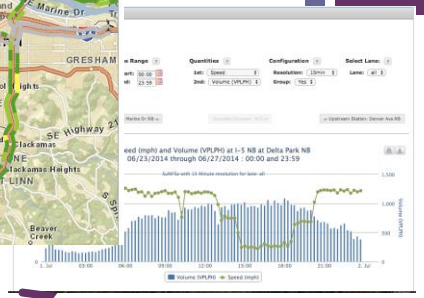
- PhD Database Management – Univ of Wisconsin Madison
- Data Management System Design and Implementation
  - Paradise (1990-1996) - sold to NCR/Teradata
  - NiagaraST Data Stream System (1996-2014)
  - S-Store – Streams + OLTP (2013-…)
- Transportation Data Management
  - Portal Transportation Data Archive (2004-…)
  - Portland Observatory (2013-…)
  - BikePed Portal (2014-…)



Portal Data Archive

- Portland-Vancouver Transportation Data Archive
  - Established in 2004 – Happy 10th Portal!
  - Policy of Open Data
  - Publicly-funded (Thanks to NSF, FHWA, Metro, RTC, OTREC)
  - Focus on open-source software (PostgreSQL, PostGIS, OpenLayers, HighCharts)
  - ~3 TB PostgreSQL Database





Portal Data Archive

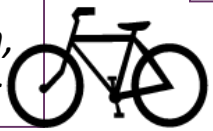
Speed, Count, Travel Time, Weigh-in-Motion, *Variable Speed*

Travel Time, Traffic Signal, Bicycle Count, Pedestrian Push-Button

Ons, Offs, On-Time Performance

Freeway
ODOT, WSDOT, Lane County

Arterial
City of Portland, Clark County, Clackamas County, *Washington County, Gresham, Tigard, Beaverton, Vancouver*

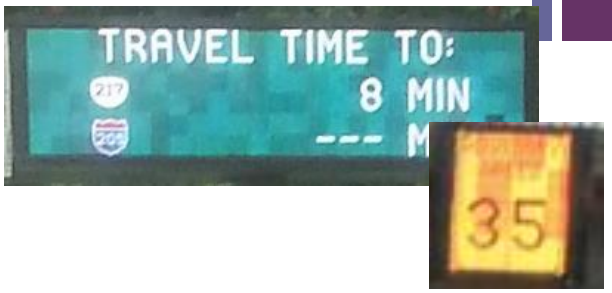


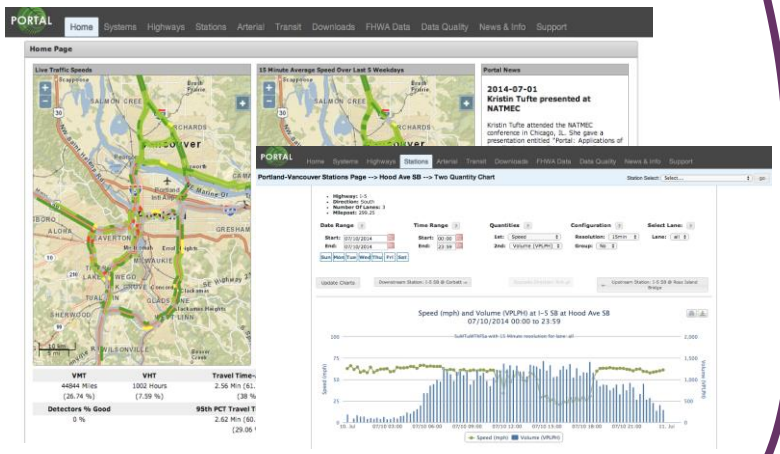
Transit
TriMet
C-TRAN

Other
Weather, Weigh-in-Motion, *Variable Speeds*

Uses of Portal

- Advanced Traffic and Incident Management Project Development
  - Research -> Planning -> Implementation -> Evaluation
- Education
  - Cloud Data Management
  - Civil & Environmental Engineering Curriculum
- Agency Performance Reporting
- Travel Model Usage
- ODOT Bottleneck Analysis
- Powell Blvd Signal Analysis
- Portland Arterial Concept of Operations





Portal Archive

Why Transportation Decision Making?

- **Increased Volume and Variety of Transportation Data**
  - Improved sensor and mobile device technology
  - Realization that data is too valuable to delete
    - ADUS - ITS Program Plan addendum 1998
  - Decreased storage costs

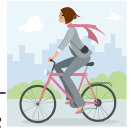
- **Increased Demand for Data-Driven Decision Making (MAP-21)**

Freeways
- Inductive Loops
- High-Definition Radar
- Third-Party

Transit
- AVL/APC
- GTFS

Safety
- Incident Reports
- Crash Reports

Arterial
- Measured Travel Time
- Traffic Signal Systems



Bicycle/Pedestrian
- Automated
- Manual

Ideas credit: David DeWitt, Microsoft/UW-Madison, SQL Server PASS Talk 2011

Why Transportation Decision Making?

- **Increased Volume and Variety of Transportation Data**
  - Improved sensor and mobile device technology
  - Realization that data is too val
    - ADUS
  - Decreased
- **Increased I**

- **Variety**
  - Occurs in many domains
    - Education, healthcare, …
  - Not lots of good existing solutions

Freeways
- Inductive Loops
- High-Definition Radar
- Third-Party

- AVL/APC
- GTFS

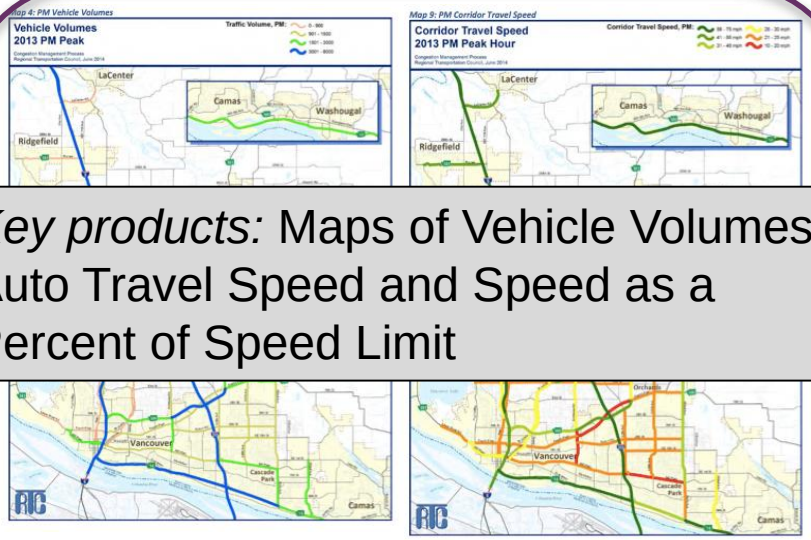
Safety
- Incident Reports
- Crash Reports

- Measured Travel Time
- Traffic Signal Systems

Bicycle/Pedestrian
- Automated
- Manual

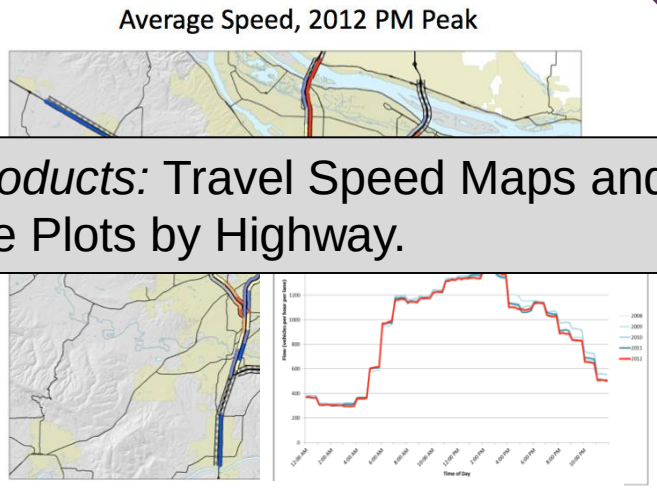
Ideas credit: David DeWitt, Microsoft/UW-Madison, SQL Server PASS Talk 2011

A Story of Three Reports…



Key products: Maps of Vehicle Volumes, Auto Travel Speed and Speed as a Percent of Speed Limit

RTC Congestion Management Process Monitoring Report (Vancouver, WA)



Key products: Travel Speed Maps and Volume Plots by Highway.

Metro Portal Annual Report

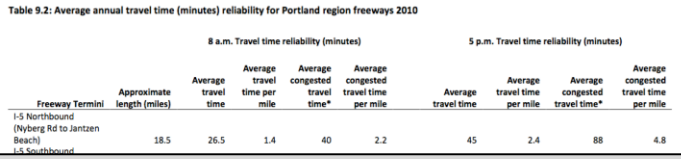


Table 9.2: Average annual travel time (minutes) reliability for Portland region freeways 2010

		8 a.m. Travel time reliability (minutes)				5 p.m. Travel time reliability (minutes)			
Freeway Termini	Approximate length (miles)	Average travel time	Average travel time per mile	Average congested travel time*	Average congested travel time per mile	Average travel time	Average travel time per mile	Average congested travel time*	Average congested travel time per mile
I-5 Northbound (Nyberg Rd to Jantzen Beach)	18.5	26.5	1.4	40	2.2	45	2.4	88	4.8
I-5 Southbound									

Key product: Table of travel time reliability

Metro Performance Measures Report (Portland, OR)

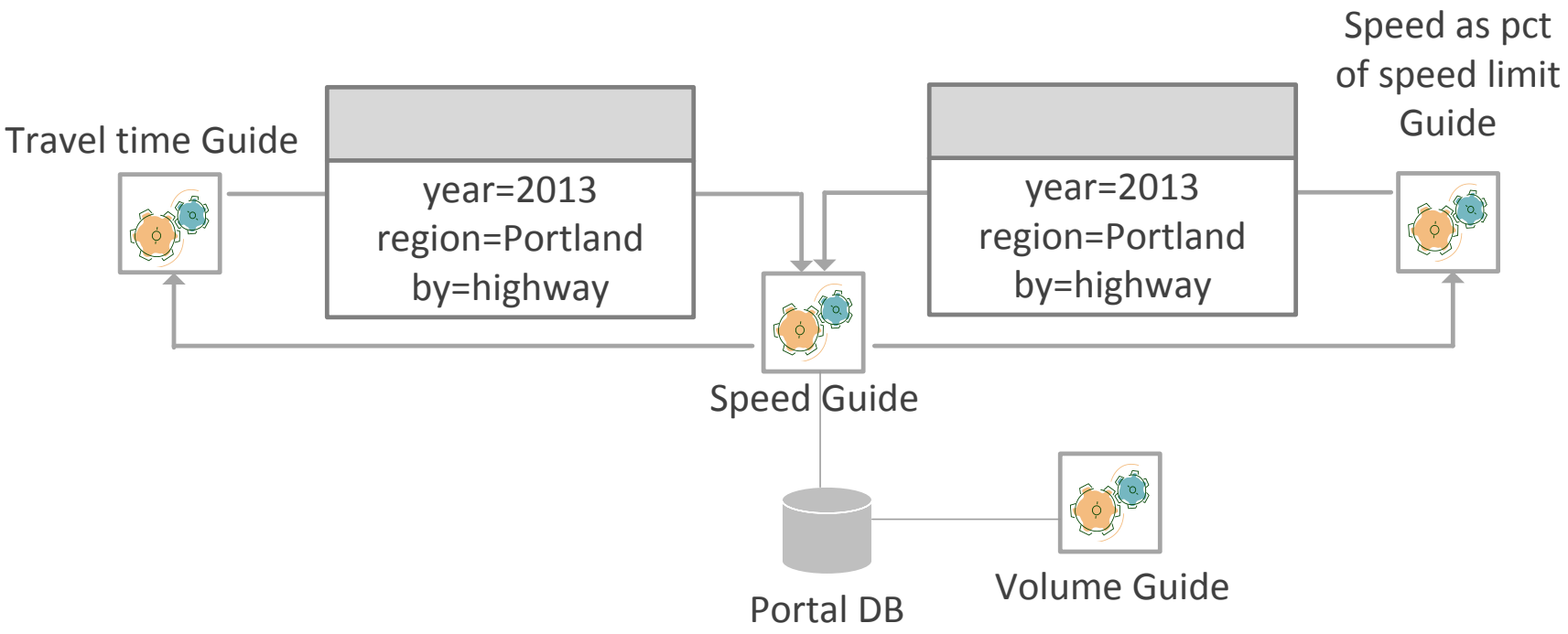
Comparisons & Observations

- Common Data Inputs to Key Products: Speed and Volume
  - Travel time is based on speed
- Similar types of Selections: Peak period; Mid-week days
  - Consistency across reports desirable in certain cases
- Recurrent Decision-Making Tasks
  - Similar decisions / reports for different locations and time periods
    - Annual Reports
    - Portland Reports, Vancouver Reports
  - Consistency across analysis and decisions is desirable in certain cases
- Currently require significant human time & effort to produce

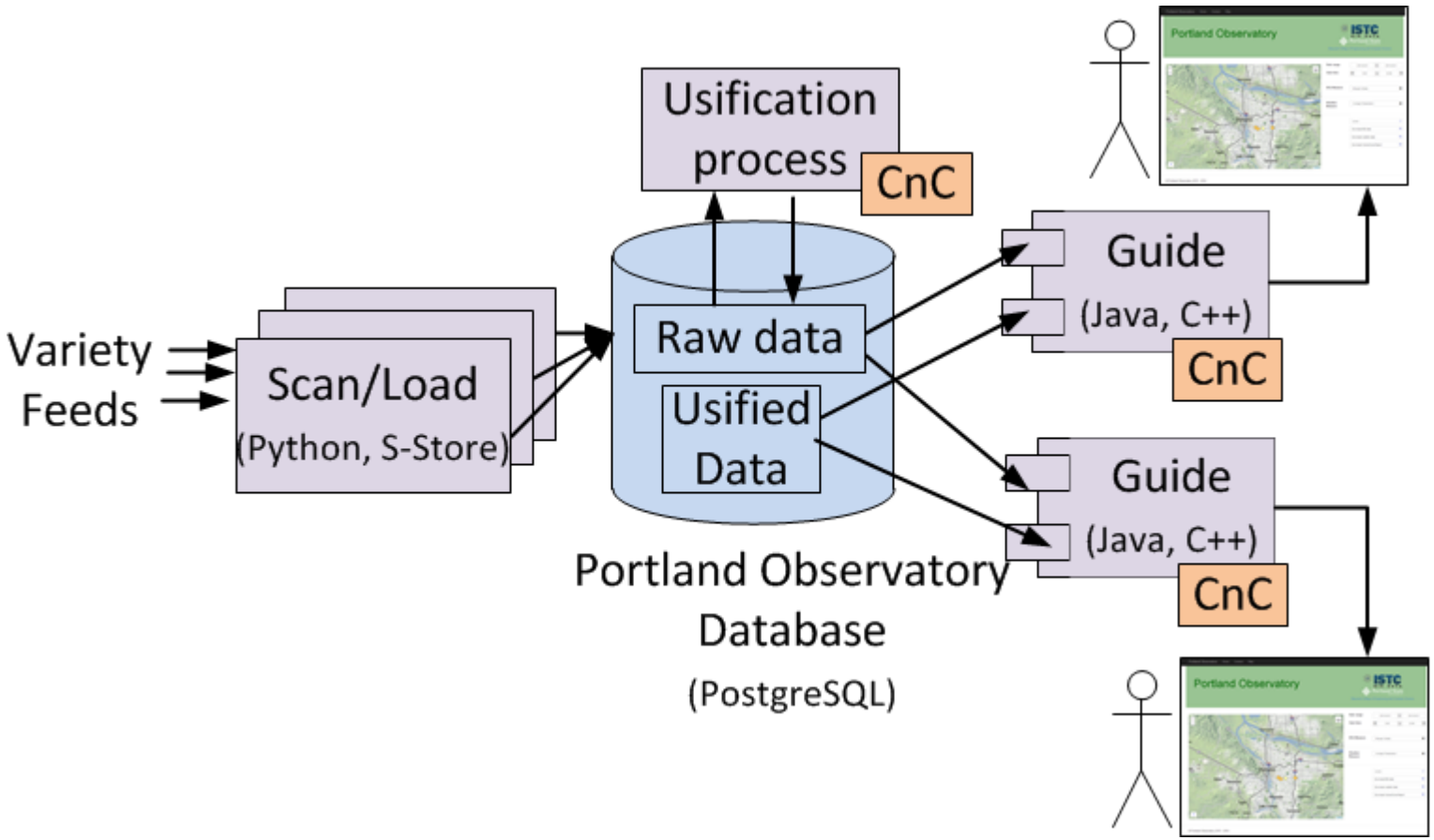
From Data to Decisions…

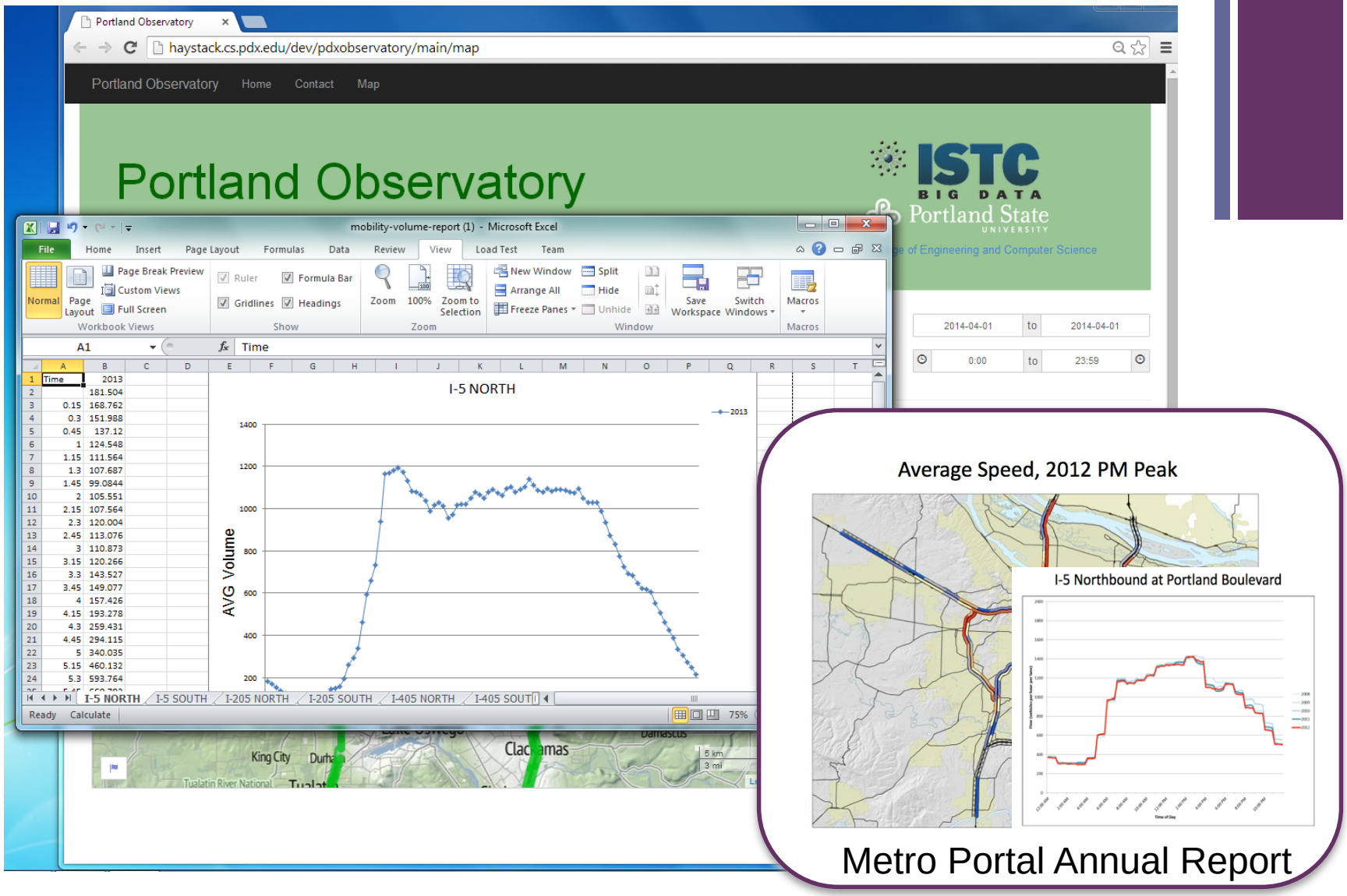
- Data stored across disparate locations and systems
  - Getting the data is sometimes the hardest part
- Decisions and analysis are difficult to document
  - Capturing human judgment and manual manipulation of data
- Can we apply existing technologies?
  - Data Integration – Traditional options are heavy-grade; not effective for reports generated annually
  - Database + SQL queries – Useful; but don't record judgments or reasoning behind judgments
  - Traditional data management techniques may be overkill (Data Integration) or may not capture the required information (SQL Queries)

Hierarchical Report Structure

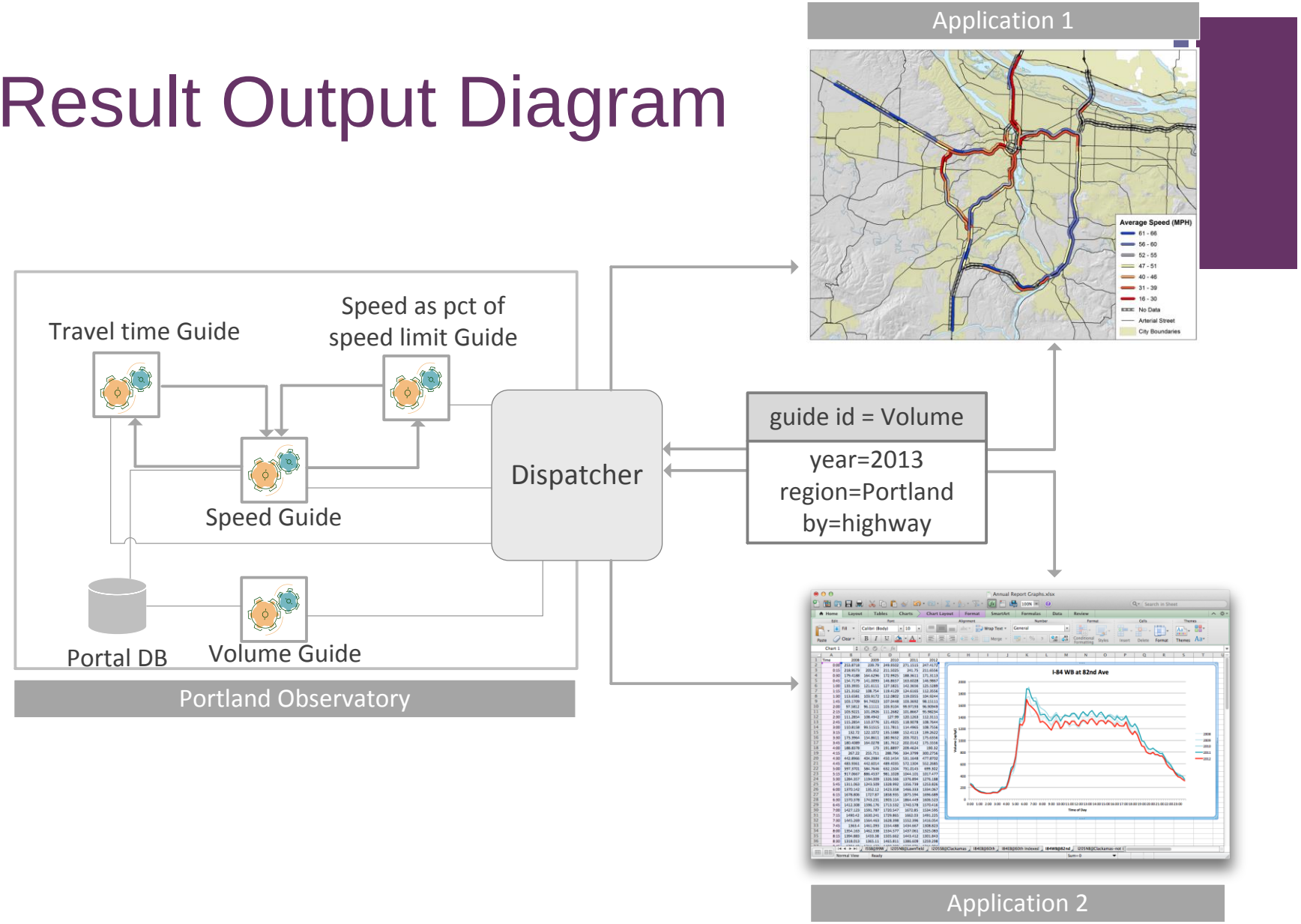


System Architecture




Portland Observatory
haystack.cs.pdx.edu/dev/pdxobservatory/main/map
Portland Observatory
Home
Contact
Map
Portland Observatory
ISTC
BIG DATA
Portland State
UNIVERSITY
mobility-volume-report (1) - Microsoft Excel
I-5 NORTH
AVG Volume
2013
2014-04-01
to
2014-04-01
0:00
to
23:59
Average Speed, 2012 PM Peak
I-5 Northbound at Portland Boulevard
Metro Portal Annual Report

Result Output Diagram



Summary

- Data-Driven Decision Making is the Future
- Opportunities found through examination of three performance reports
  - Shared Data Access
  - Capturing Recurrent Decision-Making and Report-Generation Processes
- Implemented a Guide Prototype
  - Front end Guide Capture using a web interface and
  - Back end Guide Instantiation using CnC and the PostgreSQL relational database.

CnC + Databases

- Database Strengths
  - Schema
  - Transactions
  - Recovery
  - Very efficient on filters, joins, grouping
- Database Weaknesses
  - Limited expressiveness of SQL
  - Scaling can be limited by transactions and recovery
- CnC Strengths
  - Expressiveness and declarative nature

CnC + Databases: Just dump it all to files…

2013 Mid-weekday Volume by Station for selected stations

```
SELECT  s.name,  extract(year from l.starttime) as year,
        extract(hour from l.starttime) as hour,
        (extract(minute from l.starttime)%4)*15 as min15,
        round(avg(volume)*180) as volume,
        count(*) as numreadings
FROM loopdata_2013 l, detectors d, stations s
WHERE l.detectorid = d.detectorid  AND
        d.stationid = s.stationid AND
        extract(dow from l.starttime) in (2,3,4) AND
        d.stationid in (1059, 1031, 1021, 1092, 1090, 1064, 1079, 1010,
                        1105, 1044, 1100, 1127) AND
       volume > 0
GROUP BY stationid, year, hour, min15
```

Or maybe not…

CnC + Databases - Opportunities

- Language-level - CnC + SQL
  - Give visibility into parts of the decision making process that were previously obscured by “lines of code”
  - Transparency (justification), documentation of data used (privacy issues)
- Systems-level
  - CnC: Declarativeness at a higher level
  - Database: Efficient and pre-programmed filter, join, aggregation
- Current thought: SQL Query inside CnC step

Portland Observatory: Goals

- Find a domain that challenges the third “V” of Big Data: variety
- Align with PDX’s interest in Sustainability
  - Air quality data, Intelligent Transportation Systems
- Help people we work with
  - Planners, researchers, managers, public
- Reach out to other Intel centers and labs
  - ExaCloud, S-Store, Intel-UXR/ISTC-Social

Portland State UNIVERSITY

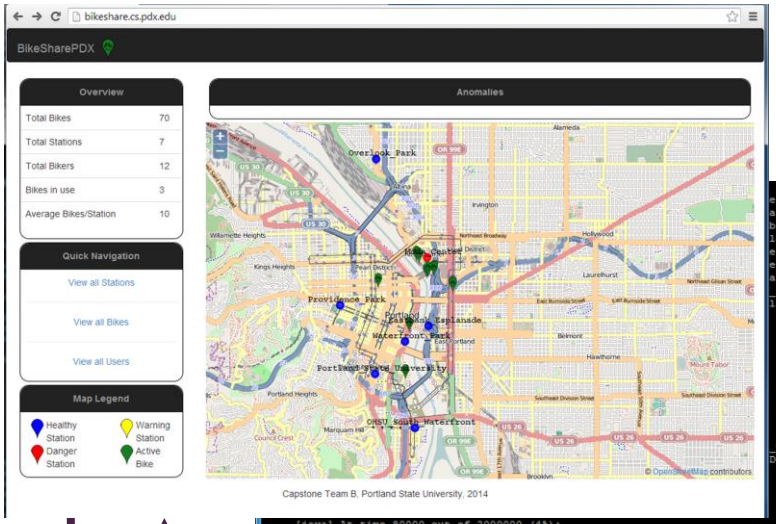
ISTC BIG DATA

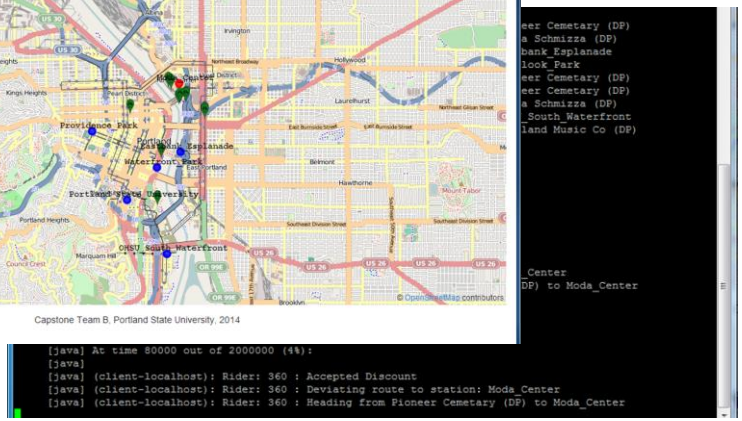
S-Store: Bike Share Demo

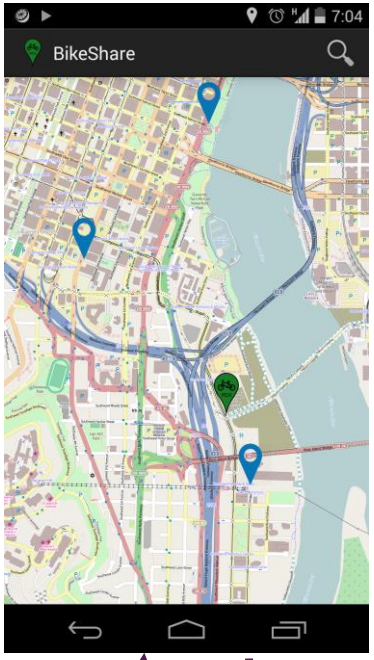
- S-Store: State of the art system for big-velocity applications
  - Combines stream processing and transactional guarantees
  - Built on the H-Store Database System
- Demo implementation of BikeShare system on S-Store
  - Web site and mobile app based on a bike sharing scenario
  - Customers check-out, ride & check-in bikes **(OLTP)**
  - Bike locations arrive as a stream
  - Discounts used to encourage riders to return bikes to stations needing bikes
    - Discounts calculated dynamically based on streaming bike positions **(Stream Processing)**

Portland State UNIVERSITY ISTC BIG DATA BROWN MIT CSAIL

BikeShare Architecture







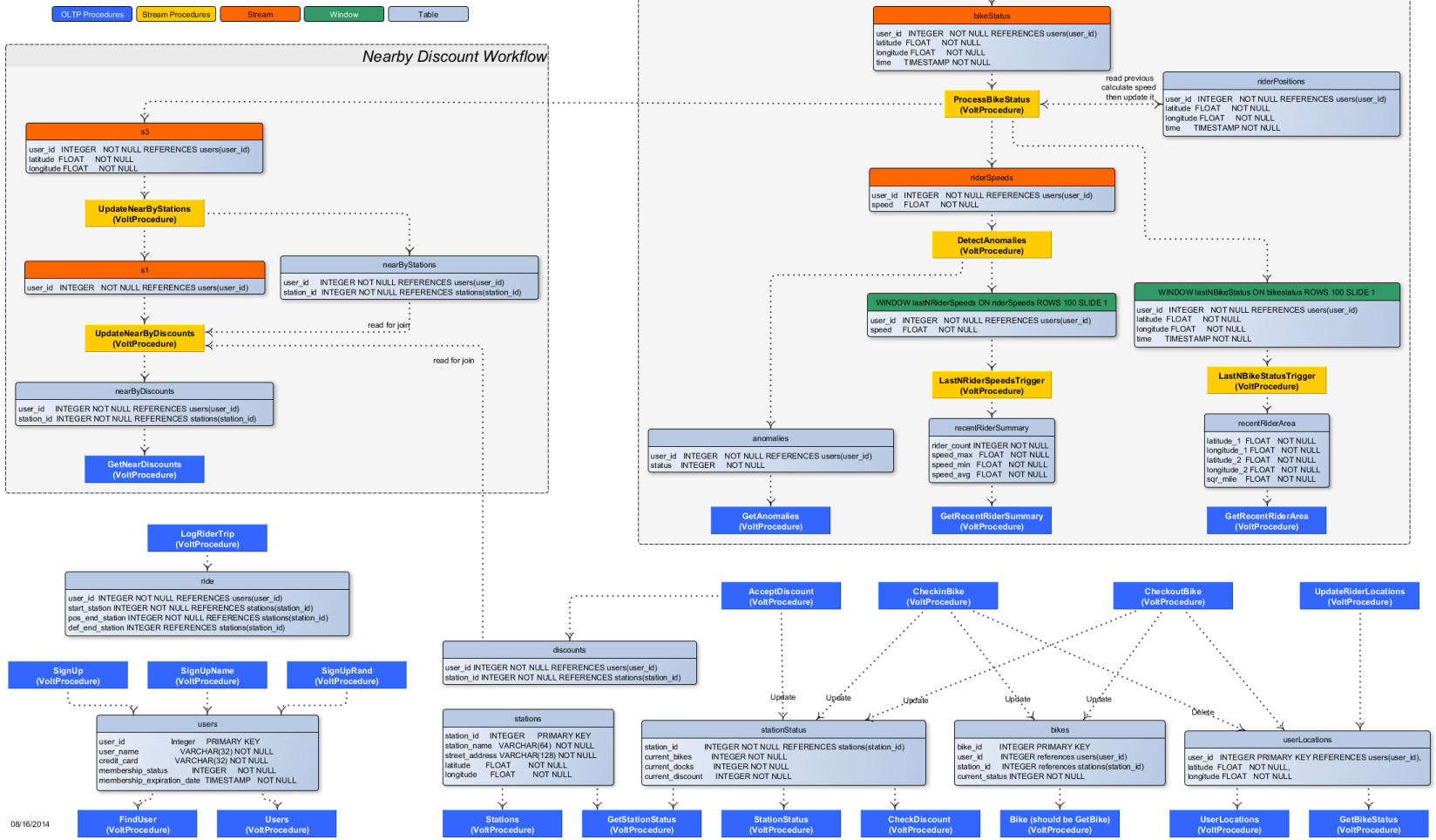
REST API (Flask)

Custom S-Store Client

S-Store

BikeShare Workflow (complete)

S-Store - BikeShare Benchmark
OLTP & Streams Workflows



THANK YOU!!

tufte@pdx.edu