

PORTAL

A Case Study

Dr. Kristin Tufte
(tufte@cecs.pdx.edu)

Mark Wong
(markwkm@postgresql.org)

Linux Plumbers Conference 2009

September 23, 2009

Overview

- ▶ What is PORTAL?
- ▶ How PORTAL works
- ▶ Improving PORTAL

What is PORTAL?

Portland Oregon Regional Transportation Archive Listing (PORTAL) is an implementation of the U.S. National ITS (Intelligent Transportation Systems) Architecture's Archived Data User Service for the Portland metropolitan region.¹

¹<http://portal.its.pdx.edu/>

What is in the PORTAL Database?

Since in 2004, nearly 1 terabyte of data and over 7 million records of...

- ▶ Loop Detector Data (bulk of data size is here)
- ▶ Incident Data
- ▶ Bus Data
- ▶ Weather Data
- ▶ VMS Data

What can we do with this data?

From the loop detector data:

- ▶ Timeseries plots for occupancy (length of time a vehicle is positioned over a detector)
- ▶ Traffic volume
- ▶ Traffic speed
- ▶ Vehicle Miles Traveled (VMT)
- ▶ Vehicle Hours Traveled (VHT)
- ▶ Travel time
- ▶ delay over a highway or station

PORTAL Web Site

- ▶ <http://portal.its.pdx.edu/>
- ▶ Graphical display of archived data
- ▶ Performance reports, traffic counts, freight data, ...
- ▶ Raw data

Daily Dashboard

The screenshot shows a web browser window with the following elements:

- Browser Address Bar:** `http://portal.its.pdx.edu/archive/dashboard.php`
- Page Header:** **PORTAL** logo and **Portland Oregon Regional Transportation Archive Listing**
- Left Navigation Menu:**
 - Info**
 - Welcome
 - News
 - User Info
 - People
 - Project Summary
 - Our Servers
 - Products
 - Comments
 - Portal Facts
 - Logout
 - Archive**
 - Timeseries
 - Grouped Data
 - Data Fidelity
 - Raw Data
 - Monthly Data
 - Weather
 - Oblique Plots
 - Travel Time
 - WIM Data
- Main Content Area:**
 - Date:** September 01 2009
 - Buttons: [view report](#), [view fine-grained](#), [refresh](#)

Daily Dashboard

- People
- Project Summary
- Our Servers
- Products
- Comments
- Portal Facts
- Logout

Archive

- Timeseries
- Grouped Data
- Data Fidelity
- Raw Data
- Monthly Data
- Weather
- Oblique Plots
- Travel Time
- WIM Data
- Performance
- Dashboard**
- Congestion
- Google Maps
- SVG Maps
- Bivariate Plots
- Google Traffic
- Vehicle Classification
- Trimet Analysis
- Incident Reports
- Quality Control

Daily Dashboard For Sep-21-2009

TOTAL VMT

7,003,117

TOTAL VHT

141,334

CONGESTION %

18.75 %

TTI

1.49

AVG TT

157 min

AVG TS

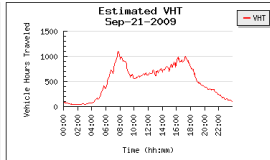
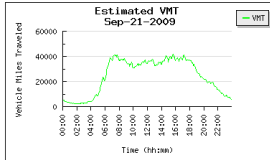
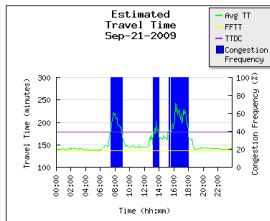
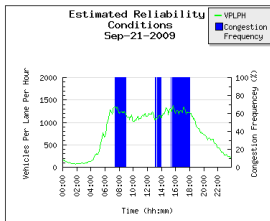
52 MPH

95th TT

219 min

95th TS

38 MPH



Daily Dashboard

File Edit View History Bookmarks Tools Help

http://portal.its.pdx.edu/archive/dashboard.php?month=09&day=21&year=2009&action=view+report

Google

Weather

Precipitation: 0.02 in
High: 84.92 °F
Low: 50 °F
Conditions: Clear

Fidelity

Good: 1,512,214 (68.37 %)
Inhibited: 373,513 (16.89 %)
Bad: 326,113 (14.74 %)

KEY:

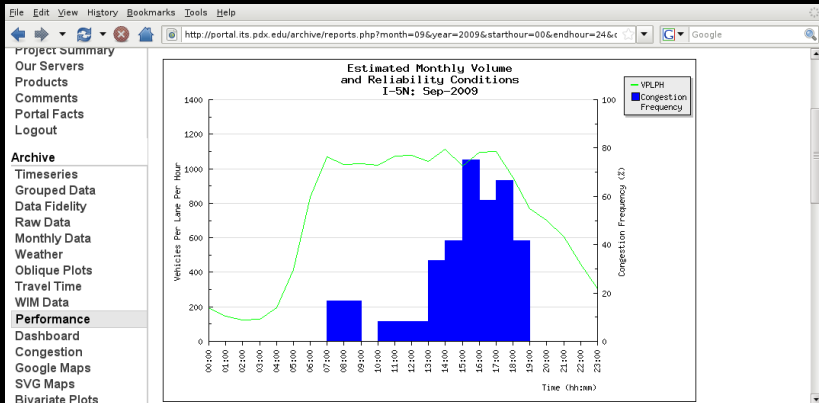
Hwy Len - Length of Highway network being queried.
Avg TT - Average Travel Time across network.
Avg TS - Average Travel Speed across network (Hwy Len/Avg TT).
FFTS - Free Flow Travel Speed (Assumed to be 60 MPH)
FFTT - Free Flow Travel Time (Hwy Len/FFTS).
95th% TT - 95th Percentile Travel Time.
95th% TS - 95th Percentile Travel Speed.
Congestion % - Percentage of time slices that are congested. ***
Congestion Frequency - 0% or 100% - Indicates if time slice is congested or not. ***
TTI - Travel Time Index (Avg congested TT in Window/FFTT).
TTDC - A Travel Time is considered Congested if equal to or above TTDC. (1.3*FFTT).
VMT - Vehicle Miles Traveled on network.
VHT - Vehicle Hours Traveled on network.
*** A reading is considered congested if the travel time is greater than or equal to 1.3*FFTT

Performance Report - Reliability

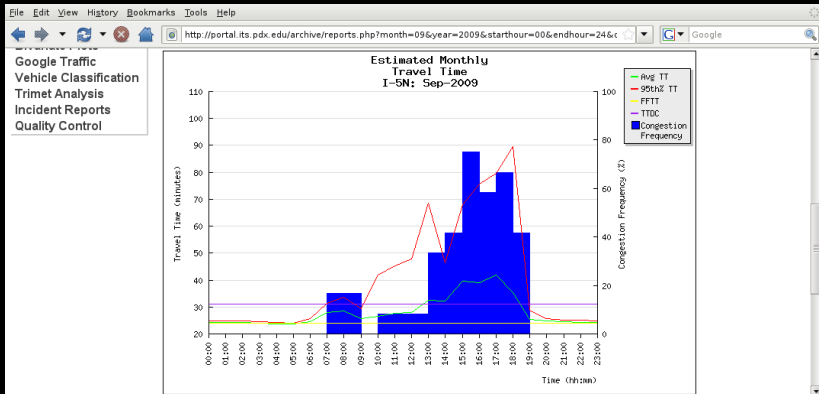
The screenshot shows a web browser window with the following elements:

- Browser Address Bar:** `http://portal.its.pdx.edu/archive/reports.php`
- Page Header:** **PORTAL** logo and **Portland Oregon Regional Transportation Archive Listing**
- Left Navigation Menu:**
 - Info
 - Welcome
 - News
 - User Info
 - People
 - Project Summary
 - Our Servers
- Main Content Area:**
 - Date:** September 21, 2009, 00 to 24
 - Include Hours:** 00 to 24
 - Include Days:** all
 - Data Resolution:** 1 hour
 - Highway:** I-5 NORTH
 - Station:** I-205 NORTH mile 3.55 - Stafford to I-205 NB
 - All Highways:** (radio button selected)
 - Report Type:** monthly
 - Buttons:** view plot, view table, refresh, generate csv

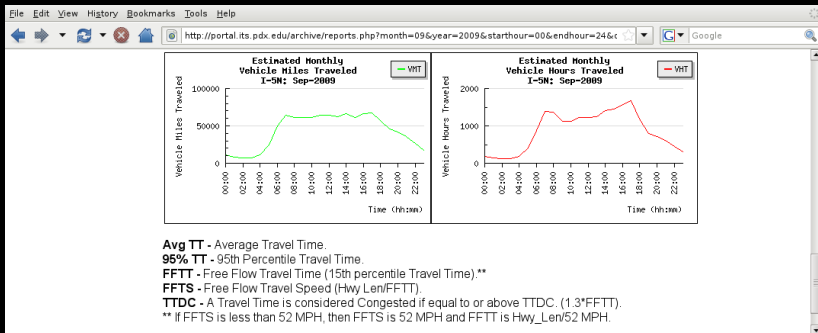
Performance Report - Reliability



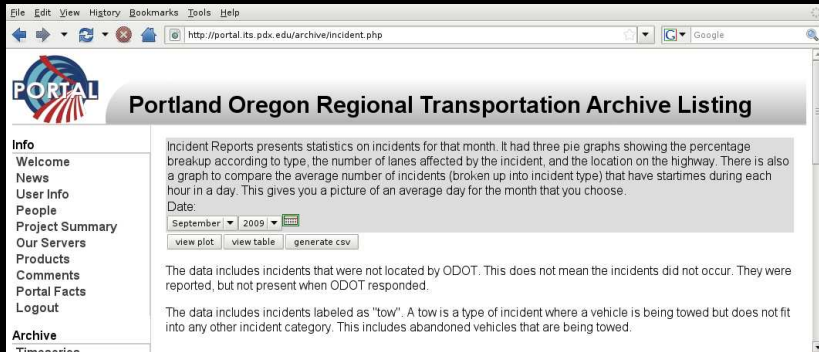
Performance Report - Reliability



Performance Report - Reliability



Speed Plot with Incident Reports



File Edit View History Bookmarks Tools Help

http://portal.its.pdx.edu/archive/incident.php

Google



Portland Oregon Regional Transportation Archive Listing

Info

- Welcome
- News
- User Info
- People
- Project Summary
- Our Servers
- Products
- Comments
- Portal Facts
- Logout

Archive

Timeseries

Incident Reports presents statistics on incidents for that month. It had three pie graphs showing the percentage breakup according to type, the number of lanes affected by the incident, and the location on the highway. There is also a graph to compare the average number of incidents (broken up into incident type) that have starttimes during each hour in a day. This gives you a picture of an average day for the month that you choose.

Date:

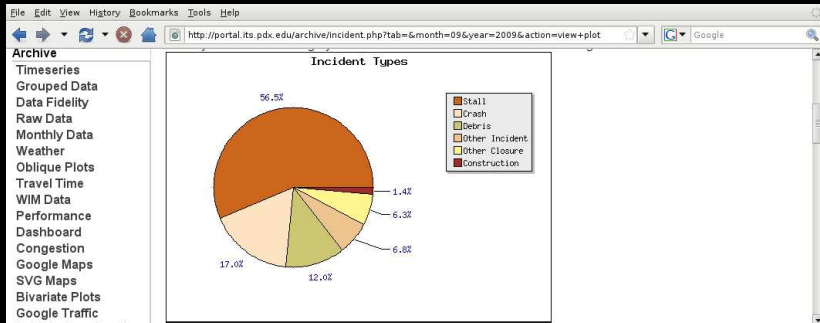
September 2009

view plot view table generate csv

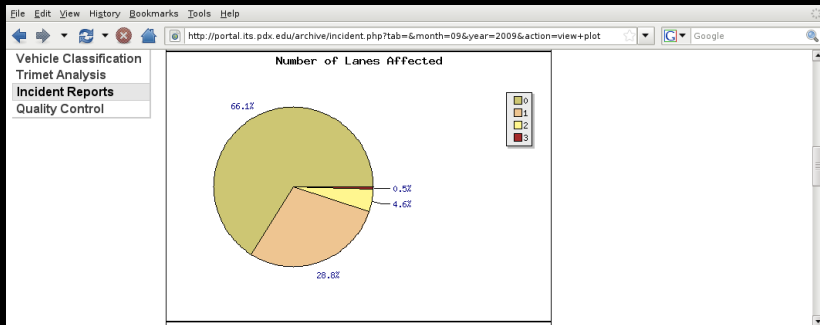
The data includes incidents that were not located by ODOT. This does not mean the incidents did not occur. They were reported, but not present when ODOT responded.

The data includes incidents labeled as "tow". A tow is a type of incident where a vehicle is being towed but does not fit into any other incident category. This includes abandoned vehicles that are being towed.

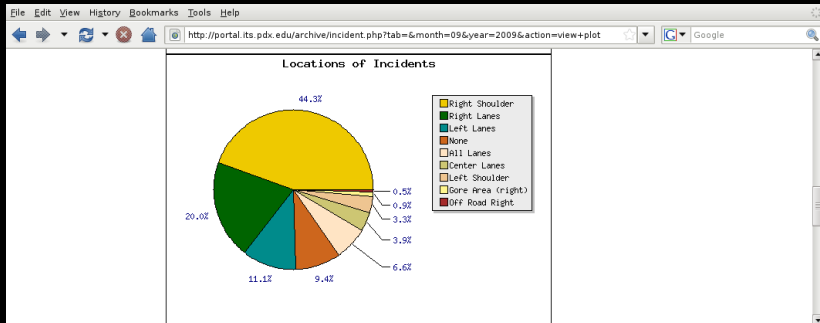
Speed Plot with Incident Reports



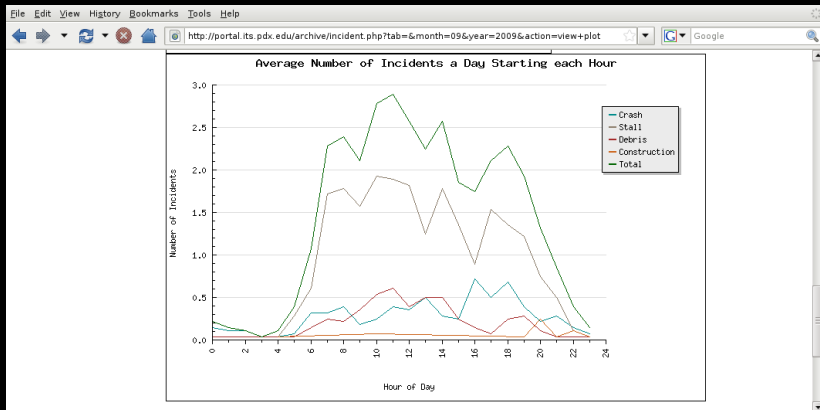
Speed Plot with Incident Reports



Speed Plot with Incident Reports




Speed Plot with Incident Reports



Time Series

File Edit View History Bookmarks Tools Help

http://portal.its.pdx.edu/archive/timeseries.php



Portland Oregon Regional Transportation Archive Listing

Timeseries plots display one of: volume, speed, occupancy (length of time a vehicle is positioned over a detector), VMT (Vehicle Miles Traveled), VHT (Vehicle Hours Traveled), travel time, and delay.

Plots can be made for a highway or a detector station. Highway plots are contour plots. In these plots, the y-axis represents the freeway and mileposts along the freeway and the x-axis represents time. The plot shows the condition of the highway throughout the day through color with green indicating near free-flow conditions and red indicating congestion. Plots with a data resolution of 20-seconds are supported for volume, occupancy and speed only. Incidents may be displayed on the highway plots, this display shows the time and location of incidents for that day. More information about the incidents is given in a table below the plot.

Highway: Station: Lane: HOV:

From: To:

Date: Include Hours: to

Data Resolution: Quantity:

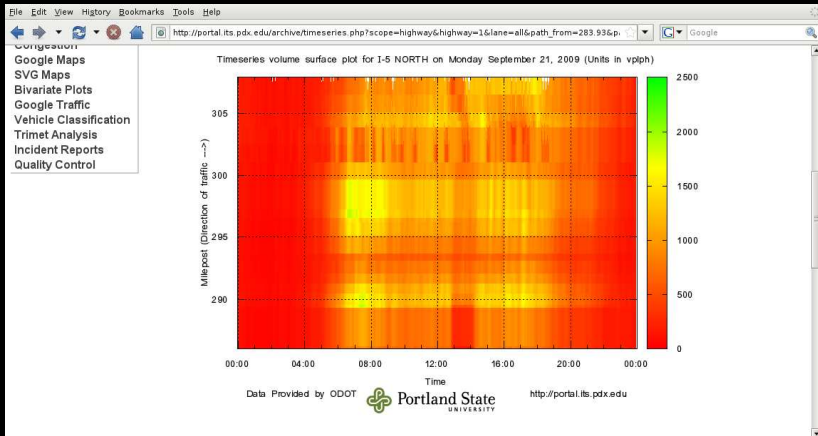
Travel Time Messages: VMS Messages: Zmin Zmax

The display of incident data includes incidents that were not located by ODOT. "Not located" does not mean the incidents did not occur, it means that the incidents were reported, but no one was present when ODOT responded.

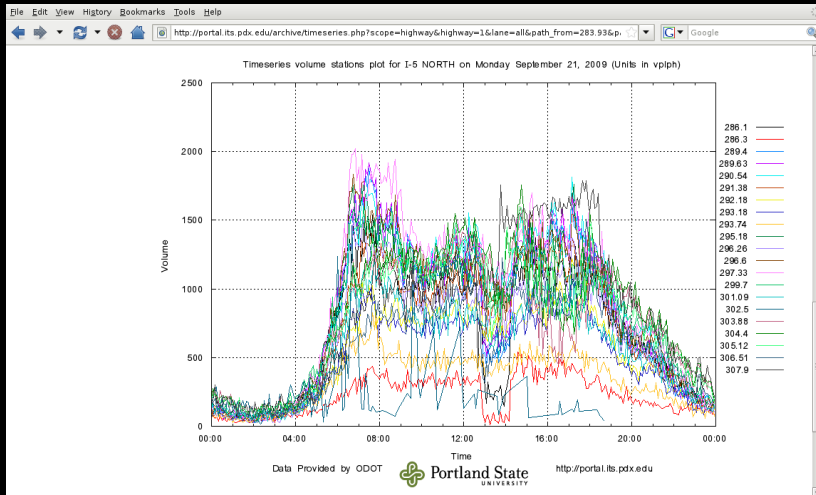
Note that effective Thursday November 15th 2007 Timeseries graphs are plotted with direction of travel going up.

- Info
- Welcome
- News
- User Info
- People
- Project Summary
- Our Servers
- Products
- Comments
- Portal Facts
- Logout
- Archive
- Timeseries
- Grouped Data
- Data Fidelity
- Raw Data
- Monthly Data
- Weather
- Oblique Plots
- Travel Time
- WIM Data

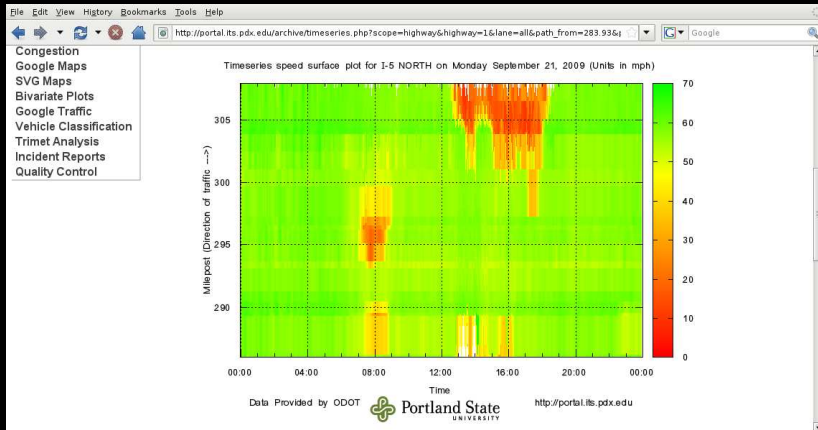
Surface Plot - Volume



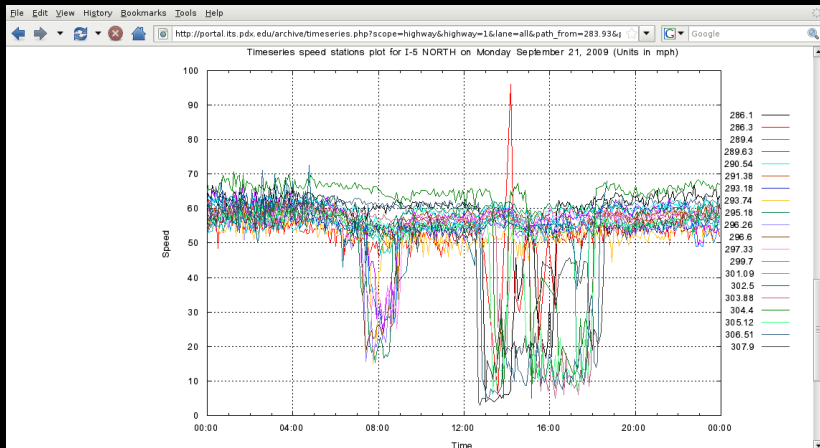
Time Series - Volume



Surface Plot - Speed




Time Series - Speed



Grouped Data - Travel Time

File Edit View History Bookmarks Tools Help

http://portal.its.pdx.edu/archive/grouped.php?scope=highway&highway=1&lane=all&path_from=283.93&path_to= Google



Portland Oregon Regional Transportation Archive Listing

Info

- Welcome
- News
- User Info
- People
- Project Summary
- Our Servers
- Products
- Comments
- Portal Facts
- Logout

Archive

- Timeseries
- Grouped Data**
- Data Fidelity
- Raw Data

Highway: I-5 NORTH I-205 NORTH mile 3.55 - Stafford to I-205 NB All Highways: Lane: all HOV:

From: Wilsonville to I-5 NB(283.93) Jantzen Beach NB(307.9) To:

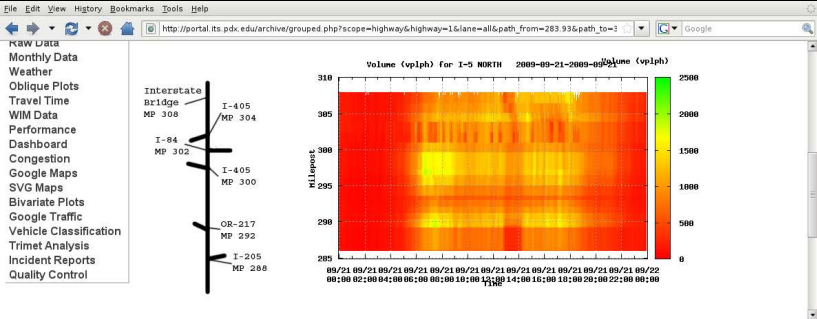
From Date: September 21 2009 To Date: September 21 2009

Quantity: volume Group Results by: do not group For Groups Show: mean

Include Days: all Include Hours: 00 :00 to 24 :00 Data Resolution: 5 minute Zmin Zmax

There is an option to include HOV data. There is only one High Occupancy Vehicle(HOV) Lane in Portland. It is located on I-5 north from Going St (MP 303) to Columbia Blvd (MP 306). It is limited to vehicles with at least two passengers. It is in operation only Monday through Friday, 3pm to 6pm at other times of day it is not limited.

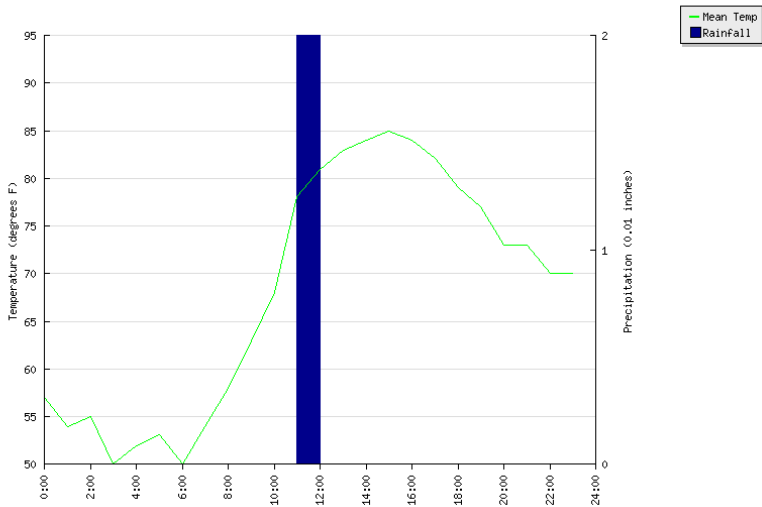
Grouped Data - Travel Time



Weather Popup

<http://portal.its.pdx.edu/archive/weather/popup.php?station=1113&scope=highway&frommonth=09&fromday=21&fromyear=2009>

Hourly Weather for Sep-21-2009



4:00 - Clear | 8:00 - Clear | 12:00 - Clear | 16:00 - Clear | 20:00 - Clear

Mapping - Speed by Month

File Edit View History Bookmarks Tools Help

http://portal.its.pdx.edu/archive/svg_maps.php?map=sidebyside&first_frommonth=07&first_fromday=01&first_fr...

Google

News
User Info
People
Project Summary
Our Servers
Products
Comments
Portal Facts
Logout

Archive
Timeseries
Grouped Data
Data Fidelity
Raw Data
Monthly Data
Weather
Oblique Plots
Travel Time
WIM Data
Performance
Dashboard
Congestion
Google Maps
SVG Maps
Bivariate Plots
Google Traffic
Vehicle Classification
Trimet Analysis
Incident Reports
Quality Control

METRO

[Portland State

This material is based upon work supported by the National Science Foundation and is not necessarily reflect the views of the funding agency.

○ Current Speed Map ○ Multi-day Map ● Side by Side Comparison ○ Subtraction Comparison

First_from Date: July 01 2009
Second_from Date: August 01 2009

First_to Date: July 31 2009
Second_to Date: August 31 2009

Include Days: all

Include Hours: 00:00 to 24:00

Data Resolution: 1 hour

View Maps

Mapping - Speed Subtraction

File Edit View History Bookmarks Tools Help

http://portal.its.pdx.edu/archive/svg_maps.php?map=subtract&first_frommonth=07&first_fromd. Google

Project Summary
Our Servers
Products
Comments
Portal Facts
Logout

Archive

- Timeseries
- Grouped Data
- Data Fidelity
- Raw Data
- Monthly Data
- Weather
- Oblique Plots
- Travel Time
- WIM Data
- Performance
- Dashboard
- Congestion
- Google Maps
- SVG Maps**
- Bivariate Plots
- Google Traffic
- Vehicle Classification
- Trimet Analysis
- Incident Reports
- Quality Control

First_from Date: July 01 2009
First_to Date: July 31 2009
Include Days: all
Include Hours: 00:00 to 24:00
Data Resolution: 1 hour

Second_from Date: August 01 2009
Second_to Date: August 31 2009
Include Days: all
Include Hours: 00:00 to 24:00
Data Resolution: 1 hour

View Map

Speed (mph)

- < -15
- 15 - -10
- 10 - -5
- 5 - 0
- 0 - 5
- 5 - 10
- 10 - 15
- 15 <
- No Data

rtland State UNIVERSITY

on | National Science

METRO

[Portland State U

Other Uses of PORTAL

- ▶ Resource for local transportation professionals
- ▶ Metro RTP
- ▶ Projects
 - ▶ Travel Time
 - ▶ Bottleneck Identification
 - ▶ Data Quality Evaluation
 - ▶ Gap Filling
 - ▶ TriMet Data Analysis
 - ▶ Oregon Freight Data Mart
 - ▶ Incident Autopsy

How PORTAL Works

For the loop detector data:

- ▶ Data from loop detectors on the road are aggregated into 20 second intervals at the device
- ▶ The aggregated data are immediately transmitted to ODOT (Oregon Department of Transportation), then transmitted to PSU data
- ▶ The XML data is transformed into SQL statements to load into the database
- ▶ Scripts aggregating data into 5 min, 15 min and 1 hour intervals are run at regular intervals

Current PORTAL Configuration

- ▶ PostgreSQL v8.1
- ▶ Red Hat Linux
- ▶ 2 quad-core Core 2 processors
- ▶ 32 GB RAM
- ▶ 1 TB of storage on a SAN

Things to try

- ▶ Take advantage of PostgreSQL's Portland Performance Pad
- ▶ Take advantage of the way data is loaded to horizontally partition tables
- ▶ Reduce database size
 - ▶ Data is duplicated for performance for 5 min, 15 min, and 1 hour aggregates
 - ▶ Determine if additional indexes are necessary
- ▶ Experiment with newer versions of PostgreSQL (and Linux)

PostgreSQL Test System

- ▶ PostgreSQL v8.4
- ▶ Gentoo Linux
- ▶ 2 quad-core Core 2 processors (not exactly the same as the production system)
- ▶ 32 GB RAM
- ▶ 25-disk 72GB SAS array

Database Sizes

- ▶ One day's worth of loop data is approximately 165 MB
- ▶ The primary key index is approximately an additional 160 MB
- ▶ The additional index is approximately another 80 MB
- ▶ A month's worth of data and indexes approximately 12 GB of data
- ▶ A year's worth of data and indexes approximately 145 GB of data

Aggregation Size Overhead

- ▶ Total size for a month
 - ▶ table and indexes = 11,821 MB
 - ▶ table and indexes for 5 min aggregates = 493 MB, 4%
- ▶ Total size for the year
 - ▶ table and indexes = 151,935 MB
 - ▶ tables and indexes for 15 min aggregates = 1,913 MB, 1%
 - ▶ tables and indexes for 1 hour aggregates = 478 MB, 0.3%

Aggregation Performance Overhead

- ▶ 5 min aggregate table is updated every 5 minutes
- ▶ 15 min aggregate table is updated every 15 minutes
- ▶ 1 hour aggregate table is updated every hour
- ▶ 5 min, 15 min, and 1 hour aggregate tables take about 15 minutes of time to run per 1 month of data

Scaling users

Running the timeseries query:

▶ Users = Response Time

▶ 1 = 8.6s

▶ 2 = 9.5s

▶ 3 = 12.6s

▶ 4 = 16.6s

▶ 5 = 20.4s

▶ 6 = 24.6s

▶ 7 = 28.5s

▶ 8 = 32.7s

▶ 9 = 36.1s

▶ 10 = 39.7s

A brief look at i/o data

- ▶ Timeseries query doing approximately 0.5 MB/s reads per second
- ▶ fio testing shows 4 MB/s expected from the drive for random read
- ▶ Suggests increasing spindles per table...

Future Work

- ▶ Figure out why oprofile isn't working on the system (or what to use instead)
- ▶ Study more system characteristics when scaling up concurrent database users
- ▶ Experiment with filesystems other than ext2
- ▶ Experiment with increasing spindles

Thank you!