COATS IV Rural Deployment Assistance

Analysis and Recommendations for Optimization and Deployment of WeatherShare and Related Web-Based Projects

by

Daniell Richter
Research Associate

And Douglas Galarus
Senior Research Associate, Program Manager -
System Engineering, Integration and Development

Western Transportation Institute
College of Engineering
Montana State University

A report prepared for the

State of California, Department of Transportation
Division of Research and Innovation

April 13, 2011

DISCLAIMER

The opinions, findings and conclusions expressed in this publication are those of the authors and not necessarily those of the California Department of Transportation or Montana State University. Alternative accessible formats of this document will be provided upon request. Persons with disabilities who need an alternative accessible format of this information, or who require some other reasonable accommodation to participate, should contact Kate Heidkamp, Assistant Director for Communications and Information Systems, Western Transportation Institute, Montana State University, PO Box 174250, Bozeman, MT 59717-4250, telephone number 406- 994-7018, e-mail: KateL@coe.montana.edu.

ACKNOWLEDGEMENTS

The authors wish to thank the California Department of Transportation (Caltrans) and the University Transportation Centers Program of the Office of Research, Development and Technology, Research & Innovative Technology Administration at the U.S. Department of Transportation for funding this research. The authors also thank Sean Campbell and Ian Turnbull of Caltrans and David Veneziano of WTI for their input to this work. Shaowei Wang, formerly of WTI, contributed to early development of this report and was the prior lead for programming on the systems covered in this work.

TABLE OF CONTENTS

1. Introduction 1
2. Methodology 2
3. System Hosting Requirements 5
   3.1. Networking and Storage 5
      3.1.1. WeatherShare 5
      3.1.2. ICM/OSS 13
      3.1.3. AWOS/RWIS 18
   3.2. System Resources 22
   3.3. Software 22
   3.4. Summary of System Hosting Requirements 23
4. Software Revision 24
   4.1. Software changes 24
      4.1.1. Client Code 24
      4.1.2. Server Code 24
   4.2. Database Improvements 25
   4.3. Conclusions 25
5. Documentation 26
6. Support 27
7. Hosting Alternatives and Recommendations 28
   7.1. Host Recommendation 29
   7.2. General Recommendations 29

LIST OF TABLES

Table 1: WeatherShare Current Conditions Download Sizes 6
Table 2: Current Conditions Download Sources 6
Table 4: NDFD Forecast Weather Elements Not Currently in WeatherShare 8
Table 5: NDFD Retrieval Sources 9
Table 6L NDFD Weather Variables 10
Table 7: Approximate WeatherShare Layer Sizes 11
Table 8: WeatherShare Usage for the Year 2010 11
Table 9: WeatherShare Networking Summary 12
Table 10: OSS/ICM Data Download 14
Table 11 OSS/ICM Data Sources 15
Table 12 Approximate ICM Page Sizes 15
Table 13: Approximate OSS Page Sizes 16
Table 14: ICM Usage for 2010 16
Table 15: OSS Usage for 2010 17
Table 16: ICM and OSS Networking Summary 17
Table 17: AWOS/RWIS Data Download 18
Table 18: AWOS/RWIS Data Sources 19
Table 19: Approximate AWOS/RWIS Page Sizes 20
Table 20: AWOS/RWIS Usage for 2010 21
Table 21 AWOS/RWIS Networking Summary 21
Table 22: Total Networking Bandwidth 23
Table 23: Total Data Storage 23

1. INTRODUCTION

Through California Oregon Advanced Transportation Systems (COATS) area research efforts, several web-based research products have been developed, including the WeatherShare system, the Integrated Corridor Management Clearinghouse (ICM) system, the One Stop Shop for Traveler Information system (OSS) and the Integration of Aviation Automated Weather Observation Systems (AWOS) with Roadside Weather Information System (AWOS/RWIS). These systems have been recognized as valued products in improving transportation services. Currently these products are being used while running in a laboratory environment and many of the factors related to transitioning them to a production/deployment environment have not previously been fully addressed.

One of the major decisions that need to be made when moving a web based product from the laboratory environment to the production environment is that of determining where to host the system. Host selection, and pricing, is dependent on the system resources and capabilities needed including bandwidth in and out of the host, storage space, and processing power required. To make this decision, accurate information about the current system configuration and usage must be compiled. Since usage of web based products can be variable and dependant on numerous factors including increased advertising of the site, increased capabilities added to the site, and in the case of a site such as WeatherShare, weather conditions that can cause spike in usage, estimates of typical current usage as well as future usage should be made.

Software that was developed in a research environment should be reviewed and hardened to run efficiently in a production environment. Best practices should be applied to ensure maximum efficiency of data manipulation and web page display.

In the first part of this document, we enumerate the considerations necessary to take a web based system from a laboratory environment to a production environment. In the second part we will apply this knowledge to the WeatherShare, ICM, OSS, and AWOS/RWIS systems to determine the available options to transition these systems to a production environment.

2. METHODOLOGY

The transition of a web based product from the laboratory environment to the production environment requires that decisions be made and steps taken to ensure a reliable and maintainable final system. These steps are outlined below.

- Host requirements. One of the requirements of a web based product is that the software be hosted on an Internet-connected server. In the development or laboratory environment the system is typically hosted on a server affiliated with the product developer. For the production environment there are a number of options available, including leaving the system in the development environment, using a traditional hosting option, or using a cloud hosting option similar to Amazon's Elastic Compute Cloud. Selection of the best option is dependent on the needs, present and future, of the web based product in question, as well as the cost.
- Networking. To properly evaluate the hosting requirements and cost, estimates of the system's bandwidth requirements should be made. The two components that make up the overall bandwidth are the data going into the server and the data going from the server to clients.
    - The data coming into the server is easier to quantify in our situation since it is retrieved by the system from predefined sources at predefined intervals. It can still vary due to format and coverage changes at the source. At this point, user requests (typically HTTP GET) are negligible in comparison to content data retrieved by the system from external sources.
    - The data going from the server to the clients can be a little harder to characterize due to its variability both in terms of number of users and the requests that each user may make. The number of concurrent users and the number of pages accessed by each user can have a great impact on the amount of outgoing data. To characterize the data going from the server to the clients we will estimate the amount of data transferred to the client for each web page served up as well as the amount for each page update and the update frequency. This data can then be used to estimate user bandwidth usage based on one or more typical uses. Note that usage can vary greatly depending of the system. WeatherShare, for instance, tends to show greater usage during times of bad weather. Consideration of how typical use, worst case use, and potential growth of the user base should all be taken into account when estimating host bandwidth requirements.
- Storage. Similar to bandwidth, storage space needed on the host system also should to be determined prior to selecting a hosting option. There are several components that make up the storage requirements. The software storage size is the cumulative size of all the executables, configuration files and scripts that are needed to gather data and display the web pages. The data storage can be classified as static and dynamic. The static are reference data that is unchanged barring growth of the scope of the system. The dynamic data is regularly updated as part of the normal operation of the system and for the case where archived data is saved and will keep growing. The amount of data that is saved and how much data is archived and for how long are all parameters that are needed when determining the host storage requirements.

- Processing. The processor utilization should be measured and documented. Although web based applications don't tend to be thought of as processor intensive, some of the operations that are used to manipulate the data prior to presentation can present a load on the processor. This load should be quantified to be sure that adequate processor power is allocated by the host. Where possible, processor utilization data should be separated by functional use.

- Software. All software required by the host to run the system needs to be documented. This includes operating system type and version as well as any system or third party modules that are required. License requirements or restrictions need to be evaluated. Potential hosts must support these software requirements.

- Software revision. During the research and development phase of a web based project the emphasis is on meeting the needs of the users and the research requirements of the project. There are typically many changes in the software as the data to be presented is refined and the presentation methods are modified to more precisely meet the needs of the users. During this cycle of changes it is easy for the underlying code to become fragmented and inefficient. Prior to production deployment the software should be reviewed, and modified/improved (refactored) where necessary. The goals of this review should be to simplify the code for ease of understanding and support, optimize the code for improved performance, use all available techniques to maximize the efficiency of the data transferred into and out of the server, and to review any third party products used with respect to licensing requirements. Improvements could include:
  - Compression of information passed between client and server and only passing information that is requested by the client to minimize the necessary bandwidth and increase page display speeds.
  - Removing redundant processing on the server to minimize the load and process new data as efficiently as possible.
  - Reviewing storage requirements to remove redundant storage, re-evaluate storage and archiving requirements and speed up data access.
  - Consolidate and minimize the size of scripts passed between the server and client to reduce the necessary bandwidth an increase the speed at which pages are rendered.
  - Review data update intervals for data retrieved by the server to be sure it synchronizes efficiently with the refresh intervals of the source data.

- Documentation. The documentation developed for research projects is generally sufficient to demonstrate why a particular outcome resulted. However, details are often lacking regarding the "nuts and bolts" that would be essential in a production deployment (i.e. an active, publicly advertised website) of the research product, including information such as source code and user documentation.

- Product support. For many of the research products being considered for implementation/deployment, product support is a vital aspect of product viability. One shortcoming of the traditional product licensing model is that product support can vanish if a licensee goes out of business. How can product support be sustained while maintaining a viable business model? Similarly, how can a web application be maintained long term in the absence of a mechanism to continually handle support needs?

3. SYSTEM HOSTING REQUIREMENTS

In this section we will examine WeatherShare, ICM, OSS, and AWOS/RWIS, using the guidelines presented in the previous section to come up with some quantifiable recommendations for the deployment of these systems. The networking bandwidth and storage requirements of each system will be examined separately, but since these systems run on a common platform with shared resources, the processing and software requirements will be handled together.

3.1. Networking and Storage

The WeatherShare system retrieves weather data, both current and forecast, from various sources and consolidates it to be displayed on a Google maps based web display. The system currently covers the state of California. Forecast data is displayed for a twenty-four hour period.

3.1.1. WeatherShare

The networking requirements of the WeatherShare system consist of the combination of the data being downloaded into the server from external data providers and the data being served out for the display of web pages to the users. The frequency and to a lesser degree amount of data downloaded into the system is controlled by the system and is quantifiable. The data being served out of the system is dependent on the amount and actions of the users and can be estimated based on past usage.

3.1.1.1. Data into WeatherShare

The WeatherShare system brings in current weather conditions data from four sources: CalTrans, MADIS, MESOWEST and the National Weather Service. This data is retrieved at regular intervals, processed, and saved on the WeatherShare server for display. While the CalTrans data covers only the state of California, the MADIS and MESOWEST data cover available stations in the continental US. The MADIS and MESOWEST data is processed in the WeatherShare system and only the California data is saved for display. The following two tables summarize the current weather conditions data that is imported into the WeatherShare system with their approximate download file sizes.

Source	Frequency	File Size	Coverage
CalTrans RWIS Atmospheric Data	Once every 15 minutes	7.6 MB text file	State of California
CalTrans Surface Data	Once every 15 minutes	10.7 MB text file	State of California
CalTrans District 3 XML RWIS Data	Once every 15 minutes	0.03 MB total in 10 text(XML) files	District 3 Sites: Bryte_Bend, CHP, Dollar, Expo, Floriston, Pioneer, Rampart, Richards, Rt_99Mcconnell, YoloEast
MADIS	Once every 15 minutes	0.24 MB total in 3 zip files	Continental US
MESOWEST	Once every 15 minutes	0.43 MB zip file	Continental US, about 12,000 stations
National Weather Service 24-Hour Precipitation Data	Twice a day at 17:45 and 21:45 UTC	0.06 MB zip/tar file (NetCDF format)	Continental US

Table 1: WeatherShare Current Conditions Download Sizes

Data Set	Source
CalTrans RWIS data	http://www.dot.ca.gov/travel/dist_03/weather/AtmosphericData.txt
CalTrans Surface Data	http://www.dot.ca.gov/travel/dist_03/weather/SurfaceData.txt
CalTrans D3 XML feed	http://www.dot.ca.gov/travel/dist_03/weather/data/*Sitename*_obs.xml
MADIS	ftp://pftp.madis-data.noaa.gov/LDAD/mesonet/netCDF/ ftp://pftp.madis-data.noaa.gov/point/maritime/netcdf/ ftp://pftp.madis-data.noaa.gov/point/metar/netcdf/
MESOWEST	http://mesowest.utah.edu/data/mesowest.dat.gz
National Weather Service 24-Hour Precipitation Data	http://water.weather.gov/precip/p_download_new/*year*/*month*/*day*/nws_precip_*date*_nc.tar.gz

Table 2: Current Conditions Download Sources

The current conditions data downloads equate to approximately 19 MB every 15 minutes, 76 MB per hour, and 1824 MB per day for WeatherShare.

Current data is displayed for California only; however, data for the Continental US is downloaded as part of the MADIS, MESOWEST, and National Weather Service data sets, so the display of a wider geographic area could be made with no increase in data downloads for these datasets. RWIS data from Oregon is brought in for ICM and adds approximately 0.36 MB per hour of data. Additional state RWIS data could be downloaded with the amount of additional data per hour varying depending on the number of RWIS stations for the particular state.

WeatherShare downloads various forecast data sets from the National Weather Service National Digital Forecast Database (NDFD). Currently 11 datasets are downloaded and each dataset contains 2-7 days (depending on weather element) of forecast data. WeatherShare currently only processes and displays one day of forecast data. These datasets are updated once per hour by the National Weather Service at approximately the top of every hour. WeatherShare downloads this data once per hour. This data covers a region that includes the state of California and most of the state of Nevada that is defined by a box with the north east corner at latitude=43.380247, longitude =-114.374609 and south west corner at latitude=31.000916, longitude= -124.029987. The following table summarizes the forecast data imported into WeatherShare with approximate download file sizes and forecast time periods.

Data Set	**Weather Element**	**Approx. File Size**	**Temporal Coverage**
temp	Forecast temperatures	1.20 MB bin file	7 Days
pop12	12 Hour probability of precipitation	0.20 MB bin file	7 Days
wspd	Forecast wind speed	1.00 MB bin file	7 Days
wgust	Forecast wind gust speed	0.61 MB bin file	3 Days
wdir	Forecast wind direction	1.30 MB bin file	7 Days
td	Forecast dew point temperature	1.20 MB bin file	7 Days
sky	Forecast sky cover percentage	0.61 MB bin file	7 Days
qpf	Forecast six hour amount of precipitation	0.15 MB bin file	3 Days
snow	Forecast snow amounts	0.42 MB bin file	2 Days
wx	Forecast weather conditions	0.39 MB bin file	7 Days
rhm	Forecast relative humidity percentage	0.94 MB bin file	7 Days

Table 3: NDFD Forecast Data Download Sizes

The NDFD data downloads equate to approximately 8 MB per hour and 192 MB per day for WeatherShare coverage of the state of California.

As stated above, this data includes forecasts for the next 2-7 days (depending on weather element) so the displayed forecast data could be extended out beyond the current 24 hours with no impact on the incoming bandwidth. The geographic region for forecast data can be expanded by using the National Weather Service predefined sectors for the NDFD. Using the Pacific Northwest and Pacific Southwest sectors (which cover all of Washington, Oregon, Idaho, California, Nevada, Arizona, Utah) would increase the download data size to an estimated 12.19 MB per hour and 293 MB per day while downloading the data for the entire continental US would increase the download data size to an estimated 32.26 MB per hour and 774MB per day. In addition there are weather elements not currently being brought into the WeatherShare system that we may wish to add. The following table shows some elements not currently in the

WeatherShare system along with estimates of the approximate data download sizes. Note that the sizes of these files are dependent on the weather conditions that are currently being forecast. The numbers below reflect a snapshot from 24 January 2011.

Weather Element	**Temporal Coverage**	**Current Area**	**Pacific Northwest / Southwest Grids**	**CONUS**
Convective Hazard Outlook	7.5 Days	12 KB	31 KB	95 KB
Probability of Tornadoes	1.5 Days	0.25 KB	15 KB	48 KB
Probability of Hail	1.5 Days	0.25 KB	2 KB	0.33 KB
Probability of Damaging Thunderstorm Winds	1.5 Days	0.25 KB	15 KB	49 KB
Probability of Extreme Tornadoes	1.5 Days	0.25 KB	2 KB	0.33 KB
Probability of Extreme Hail	1.5 Days	0.25 KB	2 KB	0.33 KB
Probability of Extreme Thunderstorm Winds	1.5 Days	0.25 KB	2 KB	0.33 KB
Total Probability of Severe Thunderstorms	3.5 Days	10.6 KB	16 KB	47 KB
Total Probability of Extreme Severe Thunderstorms	3.5 Days	0.50 KB	2 KB	0.66 KB

Table 4: NDFD Forecast Weather Elements Not Currently in WeatherShare

Including all these additional elements with the current geographic boundaries would add 24.6 KB to each download interval and an extra 590.4 KB per day. Extending the area to cover the Pacific North and Southwest would add 87 KB to each download interval and 2088 KB per day. Covering the continental US would add 240.98 KB to each download interval and 5784 KB respectively for each download interval.

Area Covered	Source
Current Area	http://ndfd.weather.gov/cgi-bin/ndfd/gribcut?lat1=31.000916&lon1=-124.029987&lat2=43.380247&lon2=-114.374609&var=*var*(see variable table below)*
Pacific North West	http://weather.noaa.gov/pub/SL.us008001/ST.opnl/DF.gr2/DC.ndfd/AR.pacnwest/VP.001-003/ds.*var**.bin for days 1-3 http://weather.noaa.gov/pub/SL.us008001/ST.opnl/DF.gr2/DC.ndfd/AR.pacnwest/VP.004-007/ds.*var**.bin for days 4-7 *var** see variable table below
Pacific South West	http://weather.noaa.gov/pub/SL.us008001/ST.opnl/DF.gr2/DC.ndfd/AR.pacswest/VP.001-003/ds.*var**.bin for days 1-3 http://weather.noaa.gov/pub/SL.us008001/ST.opnl/DF.gr2/DC.ndfd/AR.pacswest/VP.004-007/ds.*var**.bin for days 4-7 *var** see variable table below
CONUS	http://weather.noaa.gov/pub/SL.us008001/ST.opnl/DF.gr2/DC.ndfd/AR.conus/VP.001-003/ds.*var**.bin for days 1-3 http://weather.noaa.gov/pub/SL.us008001/ST.opnl/DF.gr2/DC.ndfd/AR.conus/VP.004-007/ds.*var**.bin for days 4-7 *var** see variable table below

Table 5: NDFD Retrieval Sources

Weather Element	Variable Name	Temporal Coverage
Maximum Temperature	maxt	168 Hours
Minimum Temperature	mint	154 Hours
12-hour Probability of Precipitation	pop12	168 Hours
Quantitative Precipitation Forecast	qpf	72 Hours
Sky Cover	sky	168 Hours
Snow Amount	snow	48 Hours
Temperature	temp	168 Hours
Dewpoint	td	168 Hours
Wind Direction	wdir	168 Hours
Wind Speed	wspd	168 Hours
Weather	wx	168 Hours
Significant Wave Height	waveh	120 Hours
Apparent Temperature	apt	168 Hours
Relative Humidity	rhm	168 Hours
Wind Gust Speed	wgust	72 Hours
Probability of Tornadoes	ptornado	36 Hours
Probability of Hail	phail	36 Hours
Probability of Damaging Thunderstorm Winds	ptstmwinds	36 Hours
Probability of Extreme Tornadoes	pxtornado	36 Hours
Probability of Extreme Hail	pxhail	36 Hours
Probability of Extreme Thunderstorm Winds	pxtstmwinds	36 Hours
Total Probability of Severe Thunderstorms	ptotsvrtstm	84 Hours
Total Probability of Extreme Severe Thunderstorms	ptotxsvrtstm	84 Hours
Convective Hazard Outlook	conhazo	180 Hours

Table 6L NDFD Weather Variables

WeatherShare also downloads National Weather Service Alerts for California. The National Weather Service updates these alerts every few minutes. WeatherShare downloads this data 4 times per hour at 15 minute intervals. The size of these downloads are dependent on the number of active alerts, and typical sizes range from 0 KB to about 70 KB. Then, corresponding to a 70 KB file download every 15 minutes, there would be 1680 KB downloaded per day. Currently WeatherShare only downloads alerts for California. Alerts for each state can be downloaded separately with file sizes varying dependent on the number of active alerts, as with California. Alerts for the continental US can be downloaded with download sizes reaching 800KB per interval, or about 19,200 KB (18.75 MB) per day, again depending on the total number of alerts.

3.1.1.2. Data out of WeatherShare

To quantify the data being served out of WeatherShare we will examine the prior year's usage and the size of the various web pages (layer downloads) that comprise the system. From this we will estimate some per user bandwidth amounts.

The following table outlines the data served out of the WeatherShare server for various layers. Note that this only includes data from the WeatherShare server and not data served up from other sources such as Google Maps. Note that the Recent Conditions and Forecast layers display a relatively static amount of information since the number of stations for Recent Conditions varies little and the number of mileposts for which forecasts are displayed does not vary at all. The Alert layers are dependent on the number of active alerts, which has greater variability. The icons displayed for Recent Conditions and Forecast layers are about 0.5 KB each so the display and update size is dependent on the number of different icons that need to be displayed. The table reflects the worst case where the icons have not been cached for the initial display. The information listed in the following table is representative of the data sizes.

Display Type	Initial Display Size (nothing cached)	5 Minute Refresh Size	Weather Layer Switch
Recent Conditions (any type)	340 KB	3-7 KB	4-54KB depending on layer and distinct icons displayed
Forecast (any type)	348 KB	5-13 KB	200 KB
Alert (any type)	280 KB	5-70 KB	5-70 KB

Table 7: Approximate WeatherShare Layer Sizes

The table above gives some basic estimates about how much data would be served out of the WeatherShare host to satisfy some basic page requests. To get a better idea about how the system is being used we can look at the logs from the previous year. The following table summarizes the usage statistics for the year 2010.

Month	Unique Visitors	Number of Visits	Monthly Bandwidth	Bandwidth per Visit
January 2010	701	1102	695.52 MB	646.29 KB/visit
February 2010	261	510	299.10 MB	600.55 KB/visit
March 2010	226	481	239.97 MB	510.86 KB/visit
April 2010	329	607	264.62 MB	446.40 KB/visit
May 2010	142	335	129.06 MB	394.49 KB/visit
June 2010	83	272	111.59 MB	420.10 KB/visit
July 2010	91	238	85.22 MB	366.64 KB/visit
August 2010	80	257	121.28 MB	483.21 KB/visit
September 2010	106	306	107.63 MB	360.18 KB/visit
October 2010	154	331	144.68 MB	447.6 KB/visit
November 2010	576	833	328.57 MB	403.9 KB/visit
December 2010	809	1206	549.95 MB	466.95 KB/visit

Table 8: WeatherShare Usage for the Year 2010

The 2010 data shows that usage of the WeatherShare system tends to be greater during the bad weather season as would be expected. These statistics can be a bit misleading as they tend to average out different behaviors. For instance, some of the unique visitors may have ended on a WeatherShare page and then immediately left as it was not what they were looking for. The number of visit totals show that some visitors are returning to the site, as these totals are greater than the totals of unique visitors. The bandwidth per visit statistic, combined with the layer size information in Table 5, implies that visitors are accessing multiple layers and/or staying through multiple refresh cycles on layers. The overall bandwidth can be useful in determining the requirements for a host system, and the bandwidth per visit could be used to project how additional users may affect the system. For instance for a high end, quick estimate you could take the existing usage for WeatherShare, which covers one state, and multiply times four to get a high end estimate for WeatherShare covering four states. Using the busiest month (January) this would mean about 695.52 MB/month times 4 for a total of about 2782 MB/month. This would likely be an over-estimate because of the size of California relative to other states, but could be used as a reasonable upper bound. The biggest problem with this sort of estimate, though, is that it does not account for increased usage in California.

3.1.1.3. Summary of WeatherShare Network Requirements

One requirement for evaluating hosting options for a web based product is knowledge of the bandwidth usage of the system, and in particular the monthly bandwidth into and out of the system. Numbers may not be exact due to the variable nature of the data coming into the system and the variable number of users. Thus, in this section we attempt reasonable estimates of the amount of data going out of the system.

For the current configuration of WeatherShare, which covers the state of California only, bandwidth estimates are shown in the following table. For the data out of WeatherShare we use the maximum monthly bandwidth from the last year’s (2010) data.

Data in	Data out	Total
62,779 MB (61.30 GB) /Month	696 MB (0.68 GB) / Month	63,475 MB (61.99 GB) /Month

Table 9: WeatherShare Networking Summary

3.1.1.4. WeatherShare Storage

The WeatherShare system's storage requirements cover the combination of the WeatherShare specific program space (excluding system code), static data space, and dynamic data space. The WeatherShare system currently uses approximately 37 MB of storage space for the programs that run the system. The static data consists of the support data files such as icons and other images, as well as downloaded files that get replaced during each download interval. These static files use 18,857 MB of storage space. The dynamic data space consists of the data that the WeatherShare system downloads and archives. WeatherShare archives recent conditions data that is downloaded to the system in a MySQL database. This database containing the archived data as of February 10, 2011 used 29,162 MB and grows by an average of about 39 MB per day.

Increasing the geographic region covered by WeatherShare will increase the data storage requirements. For recent conditions data there will be at least an additional 100 bytes for each station added and 50 bytes of data for each station sensor reading. If the additional sensors are archived, as is done with current sensors, there will be an additional 35 bytes per sensor reading per update interval. The static data that is made up largely of the NDFD forecast data would increase from approximately 18,857 MB to about 33,795 MB. Note that this is just an estimate and could be reduced by decisions on how the data is mapped and stored.

3.1.2. ICM/OSS

The Integrated Corridor Management System (ICM) and One Stop Shop (OSS) both utilize data from WeatherShare, and that data is not included in this section. They also share and display most of the same data so they are handled together in this document. The formal coverage area for both ICM and OSS is Northern California and Southern Oregon. However, some of the datasets extend to cover either all of Oregon or all of California.

The networking requirements of the ICM and OSS systems consist of the combination of the data being downloaded into the server and the data being served out for the display of web pages to the users. The frequency and amount of data downloaded into the system is controlled by the system and is quantifiable. The data being served out of the system is dependent on the amount and actions of the users and can be estimated based on past usage. These systems don't archive data, except for a log file, so the storage requirement consists primarily of static data and code space.

3.1.2.1. Data Downloaded into ICM and OSS

The ICM and OSS system download Closed Circuit TV (CCTV) images, changeable message sign (CMS) messages, California Highway Patrol (CHP) incidents, chain requirements, and road construction information from CalTrans and road weather information system (RWIS), CCTV images, changeable message sign (CMS) messages, chain requirements, and road construction information from Oregon's Tripcheck system. The following table summarizes the data that is downloaded into the OSS/ICM system with approximate download file sizes.

Source	Retrieval Interval	File Size	Coverage
CalTrans CMS	Every 5 minutes	3 KB text file	CalTrans District 2
CalTrans Chain Control	Every 15 minutes	14 KB text file	CalTrans District 2
California Construction Closures	Every 15 minutes	15-30 KB text file	CalTrans District 2
CHP Incidents	Every 5 minutes	80-110 KB XML file	All of California
California CCTV	Every 15 minutes	2660 KB in 103 jpg files (ranging from 3KB- 58KB)	CalTrans Districts 1, 2 and 6
Oregon CMS	Every 15 minutes	2 XML files at 28 KB each	All of Oregon
Oregon Chain Control	Every 60 minutes	132 KB XML file	All of Oregon
Oregon Construction Closures	Every 15 minutes	406 KB XML file	All of Oregon
Oregon RWIS	Every 15 minutes	93 KB XML file	All of Oregon
Oregon CCTV	Every 15 minutes	3800 KB in 173 jpg files (ranging from 16 KB-30 KB per jpg file)	All of Oregon

Table 10: OSS/ICM Data Download

For OSS/ICM data, there is approximately 28.82 MB downloaded per hour, or approximately 692 MB downloaded per day.

Data Set	Retrieval location
CalTrans CMS	http://www.dot.ca.gov/research/its/data/cms/cmsStatusD02.txt
CalTrans Chain Control	http://www.dot.ca.gov/dist2/chainup/docs/exportSummary.txt
California Construction Closures	http://www.dot.ca.gov/travel/dist_02/lcs/lane_closures_d2.txt
CHP Incidents	http://media.chp.ca.gov/sa_XML/sa.XML
California CCTV	http://www.dot.ca.gov/dist1/d1tmc/hwypix/*siteid.jpg* http://www.dot.ca.gov/dist2/cctv/*site/site.jpg* http://dot.ca.gov/dist6/cctv/images/*siteid.jpg*
Oregon CMS	http://www.tripCheck.com/TTIPv2/TTIPData/DataRequest.aspx?uid=10*xx*&fn=dmsstatus http://www.tripCheck.com/TTIPv2/TTIPData/DataRequest.aspx?uid=10*xx*&fn=Dmsstatus-sw
Oregon Chain Control	http://www.tripCheck.com/TTIPv2/TTIPData/DataRequest.aspx?uid=10*xx*&fn=rw"
Oregon Construction Closures	http://www.tripCheck.com/TTIPv2/TTIPData/DataRequest.aspx?uid=10*xx*&fn=incd
Oregon RWIS	http://www.tripCheck.com/TTIPv2/TTIPData/DataRequest.aspx?uid=10*xx*&fn=rwis
Oregon CCTV	http://tripcheck.com/roadcams/cams/*sitename.jpg*

Table 11 OSS/ICM Data Sources

3.1.2.2. Data out of ICM/OSS

To quantify the data served out of ICM and OSS we will examine the size of the various web pages that comprise the system as well as looking at the past years usage. From this we will estimate some per user bandwidth amounts.

The following tables outline the data served out of the ICM and OSS servers respectively for various screens. Note that this only includes data from the ICM/OSS server and not data served up from other sources such as Google Maps. Note also some of the data sizes change depending on conditions such as number of incidents, construction closures, or chain control notices. The information listed in the following table is representative of the data sizes.

Display Type	Initial display size (nothing cached)	5 Minute Refresh Size
ICM Screen	900 KB	150-200KB
CCTV image click	14-40 KB	NA

Table 12 Approximate ICM Page Sizes

Display Type	Initial display size (nothing cached)	5 Minute Refresh Size	Weather Layer Switch
DOT Field Elements	1330 KB	600-735 KB	NA
Current Weather	26 KB	4-5 KB	7-40 KB depending on layer selected
Forecast	180 KB	4-5 KB	40-190 KB depending on layer selected
Mountain Passes	5.1 KB	NA	NA
Vista Points	7.1 KB	NA	NA
Rest Areas	5.2 KB	NA	NA
Truck Scales	4 KB	NA	NA
Trip Planner	10 KB	8 KB	7-190 KB depending on layer selected
Route Details	370 KB	1-12 KB	7-190 KB depending on layer selected

Table 13: Approximate OSS Page Sizes

To better understand how the system is being used we can look at the logs from the previous year. The following table summarizes the usage statistics for 2010. Note that the totals in the beginning of the year include WTI usage and testing, which inflates the values. In particular for the months of January and February the top 3 most-frequent visits were from WTI computers. Further, this site was only advertised to a handful of prospective users, and was not advertised to the general public.

Month	Unique Visitors	Number of Visits	Monthly Bandwidth	Bandwidth per Visit
January 2010	13	54	78.94 MB	1496.89 KB/visit
February 2010	10	59	96.75 MB	1679.12 KB/visit
March 2010	15	52	11.23 MB	221.09 KB/visit
April 2010	10	57	45.36 MB	814.87 KB/visit
May 2010	18	91	145.39 MB	1636.04 KB/visit
June 2010	17	52	22.94 MB	451.78 KB/visit
July 2010	11	43	10.56 MB	251.59 KB/visit
August 2010	17	48	37.74 MB	805.14 KB/visit
September 2010	16	48	47.87 MB	1021.14 KB/visit
October 2010	11	50	201.00 MB	4116.54 KB/visit
November 2010	34	70	75.62 MB	1106.19 KB/visit
December 2010	25	68	86.21 MB	1298.17 KB/visit

Table 14: ICM Usage for 2010

Month	Unique Visitors	Number of Visits	Monthly Bandwidth	Bandwidth per Visit
January 2010	7	19	17.02 MB	917.54 KB/visit
February 2010	22	138	1351.68 MB	10003.14 KB/visit
March 2010	44	174	341.32 MB	2008.69 KB/visit
April 2010	28	110	246.05 MB	2290.46 KB/visit
May 2010	23	102	379.35 MB	3808.39 KB/visit
June 2010	26	58	314.86 MB	5558.93 KB/visit
July 2010	11	27	94.23 MB	3573.86 KB/visit
August 2010	18	55	213.42 MB	3973.45 KB/visit
September 2010	25	75	209.71 MB	2863.24 KB/visit
October 2010	15	32	81.28 MB	2601.05 KB/visit
November 2010	22	60	262.36 MB	4477.68 KB/visit
December 2010	57	109	207.99 MB	1953.94 KB/visit

Table 15: OSS Usage for 2010

3.1.2.3. Summary of ICM and OSS Network Requirements

One requirement for evaluating hosting options for a web based product is knowledge of the bandwidth usage of the system, specifically the monthly bandwidth into and out of the system. Exact numbers may not always be determined because of the variable nature of the data coming into the system and the variable number of users. In this section we have attempted to determine reasonable estimates.

Note that OSS in particular was under development during the beginning of 2010, so bandwidth reflects use by WTI during development and testing. For the data out of ICM, we use the maximum monthly bandwidth from the last year's (2010) data. For data out of OSS we will throw out the highest number, which is most likely a result of WTI activity and use the next highest.

System	Data in	Data out	Total
ICM	20,751 MB (20.27 GB) per Month	200 MB (.20 GB) per Month	20,951 MB (20.46 GB) per Month
OSS	NA	379 MB (.37 GB) per Month	63,492 MB (62.00 GB) per Month

Table 16: ICM and OSS Networking Summary

3.1.2.4. Storage

The ICM and OSS system's storage requirements consist of the combination of program space (excluding system code), static data space, and dynamic data space. The ICM and OSS systems currently use approximately 340 KB and 750 KB respectively for the user programs that run each system. The static data files take up about 13 MB and 105 KB of storage space respectively. The dynamic data space consists of a log file for ICM only that keeps track of the systems data activity this currently takes up about 19.69 MB and grows at a rate of about .08 MB per day.

3.1.3. AWOS/RWIS

The networking requirements of the AWOS/RWIS system consist of the combination of the data being downloaded into the server and the data being served out for the display of web pages to the users. The frequency and amount of data downloaded is controlled by the system and is quantifiable. The data served out of the system is dependent on the amount and actions of the users and can be estimated based on past usage. However, since this system is in the prototype stage, this data is a best guess estimate. Similar to ICM and OSS, this system doesn't archive data.

3.1.3.1. Data into AWOS/RWIS

The AWOS/RWIS system utilizes data from WeatherShare, and that data is not included in this section. The data downloaded into the ASOS/RWIS system consists of AWOS/ASOS weather data, Pilot Reports, Terminal Aerodrome Forecasts, links to Radar images, links to Satellite images, and Winds and Temperature aloft forecast data. The following table summarizes the data that is imported into the AWOS/RWIS system with their approximate download file sizes and frequencies.

Source	Retrieval Interval	Approximate File Size	Coverage
AWOS/ASOS Weather data	Every 60 minutes	0.5 KB per file 40 KB total	107 Stations in California
Pilot Reports	Every 5 minutes	10 KB, depending on number of reports	California/Nevada region defined by SWLat= 32/SWLon=-125 and NELat=42.5 and NELon = -113.75
Terminal Aerodrome Forecasts	Every 60 minutes	8.8 - 50 KB, depending on number of reports	107 Stations in California
Radar: NWS CONUS Merged Reflectivity Composite	Every 5 minutes	4.5 KB	CONUS
Satellite: IR Temperature	Every 5 minutes	1.3 KB	CONUS
Satellite: Water Vapor	Every 5 minutes	1.3 KB	CONUS
Satellite Visible	Every 5 minutes	6.4 KB	CONUS
Wind Aloft and Temperature Aloft	Every 60 minutes	27,000 KB	CONUS

Table 17: AWOS/RWIS Data Download

Data Set	**Retrieval location**
AWOS/ASOS Weather data	ftp://tgftp.nws.noaa.gov/data/observations/metar/decoded/*stationname.txt*
Pilot Reports	http://adds.aviationweather.noaa.gov/pireps/index.php?stationList=&distance=200&SWLAT=32&SWLON=-125&XNELAT=42.5&XNELON=-113.75&NHOURS=2&HAZARD=ALL&NULLREP=NEG&MAXREP=SEV&text=Get+text"
Terminal Aerodrome Forecasts	ftp://tgftp.nws.noaa.gov/data/forecasts/taf/stations/
Radar: NWS CONUS Merged Reflectivity Composite	http://wdssii.nssl.noaa.gov/tiles/MergedReflectivityQCComposite_00.00/index.kmz
Satellite: IR Temperature	http://wdssii.nssl.noaa.gov/tiles/IR_band4_00.00/index.kmz
Satellite: Water Vapor	http://wdssii.nssl.noaa.gov/tiles/WaterVapor_00.00/index.kmz
Satellite Visible	http://wdssii.nssl.noaa.gov/tiles/Visible_00.00/index.kmz
Wind Aloft and Temperature Aloft	ftp://ftp.ncep.noaa.gov/pub/data/nccf/com/nam/prod/nam.*date./.filename*

Table 18: AWOS/RWIS Data Sources

For the AWOS/RWIS web application, approximately 26.72 MB of data is downloaded per hour, corresponding to 641.28 MB per day.

3.1.3.2. Data out of AWOS/RWIS

Display Type	Initial Display Size	5 Minute Refresh Size	Layer Switch
Recent Conditions	200-260 KB	5-18 KB	10-40 KB depending on layer selected
NDFD Forecast	175-185 KB	4-15 KB	140-190 KB depending on layer selected
NWS Alert	25-45 KB	20-30 KB	NA
Caltrans CCTV	25 KB	25 KB	14-40 KB for each CCTV image clicked
AWOS/ASOS (METAR)	3-5 KB	3-5 KB	NA
Pilot Reports (PIREPS	4 KB	1-5 KB	NA
Terminal Aerodrome Forecasts (TAF)	2-4 KB	1-4 KB	0.1 KB for each forecast marker clicked
Radar: NWS CONUS Merged Reflectivity Composite	10 KB	10 KB	NA
Satellite	14 KB	14 KB	14 KB
Wind Aloft	150-160 KB	28-30 KB	61-180 KB
Temperature Aloft	83-100 KB	65-75 KB	60-100 KB

Table 19: Approximate AWOS/RWIS Page Sizes

To better understand how the system is being used, we can look at the logs from the year 2010. The following table summarizes the usage statistics for 2010. Note that some of the numbers early in the year represent WTI usage and testing since system development was still in progress. In particular, for the months of January and February the most frequent visits were by WTI and in April and May the top two out of three frequencies were from WTI prior to a project wrap-up meeting.

Month	Unique Visitors	Number of Visits	Monthly Bandwidth	Bandwidth per Visit
January 2010	40	87	187.56 MB	2207.54 KB/visit
February 2010	29	113	59.28 MB	537.21 KB/visit
March 2010	40	103	66.78 MB	663.86 KB/visit
April 2010	122	194	104.58 MB	552.01 KB/visit
May 2010	41	99	47.79 MB	494.28 KB/visit
June 2010	28	55	18.18 MB	338.43 KB/visit
July 2010	34	68	14.65 MB	220.58 KB/visit
August 2010	20	37	9.30 MB	257.45 KB/visit
September 2010	60	76	19.21 MB	258.82 KB/visit
October 2010	47	76	26.19 MB	352.91 KB/visit
November 2010	34	51	12.95 MB	259.96 KB/visit
December 2010	40	72	35.04 MB	498.34 KB/visit

Table 20: AWOS/RWIS Usage for 2010

3.1.3.3. Summary of AWOS/RWIS Network Requirements

One requirement for evaluating hosting options for a web based product is knowledge of the bandwidth usage of the system, and in particular the monthly bandwidth into and out of the system. While exact numbers are not always possible due to the variable nature of the data coming into the system and the number of users and thus the amount of data going out of the system, this section attempts to come up with some reasonable estimates.

Note that the AWOS/RWIS system was under development during the beginning of 2010 so bandwidth reflects use by WTI during development and testing. This system is currently in the prototype stage so usage numbers are hard to quantify. For the data out of AWOS/RWIS we use the maximum monthly bandwidth from the last years (2010) data. For data out of AWOS/RWIS we will throw out the highest number which is most likely a result of WTI activity and use the next highest.

Data In	Data Out	Total
20,751 MB (20.27 GB) per Month	200 MB (.20 GB) per Month	20,951 MB (20.46 GB) per Month

Table 21 AWOS/RWIS Networking Summary

3.1.3.4. AWOS/RWIS Storage

The AWOS/RWIS system’s storage requirements consist of program space (excluding system code) and static data space. This system does no archiving. The AWOS/RWIS system currently takes up about 10 MB of storage space for the user programs that run the system. The static data files take up about 567 MB of storage space.

3.2. System Resources

In this section, we will look at the processing requirements for current integrated system, which will give us a better understanding the server load for running all four systems.

Currently all the applications are hosted on a Dell PowerEdge 2950 2U rack-mount server, with two Quad-Core Intel® Xeon™ 3.0 GHz X5450 CPUs, two 300GB hard drives in a RAID 1 array and 16 GB memory.

Examination of system usage shows that the average CPU usage for the year 2010 was about 8% of one core or 8% of a total of 800% available. The maximum for this period was 116% of the 800% available. This maximum seems to be an aberration, possibly due to development work. For the month of January 2011, the average CPU usage was 7.66% with the maximum being 66.29%. Looking at the system load averages for the same period (load being defined as the number of processes waiting to run) the average for the year 2010 was 0.39 processes with the max load being 15.01. These figures indicate that there is relatively little time in which a process that is ready to run is waiting for something else to finish.

Generally, maximum system usage appears to correspond with two system events: system backups and the processing of the WeatherShare NDFD forecast data. The spikes related to system backup occur at midnight and noon each day when a backup process is run. These spikes could probably be minimized by looking at different more efficient backup procedures. The spike related to the processing of WeatherShare NDFD forecast data occurs hourly. This processing involves processing a large amount of data and generating .png raster files for each weather element. This load could potentially be reduced by refactoring the code and possibly migrating it from a scripting language (PHP) to a compiled language such as "C".

Memory usage averages approximately 2 GB when looking at the average from February 2010 to February 2011. This would imply that the current amount, 16GB, is adequate and possibly could be reduced when considering external hosting options.

3.3. Software

The WeatherShare, ICM, OSS, and AWOS/RWIS systems are designed to run on a Linux, Apache, MySQL, Perl/PHP platform (known by the acronym LAMP). The system is currently running under the following versions of the various components:

- Debian Linux 5.0
    - Linux Kernel 2.6.26-1
- MySQL 5.0.51a
- PHP 5.2.6-1
- PERL 5.8.8
- Apache 2.2.9

While the system should run under different versions than those stated above, further testing would be required to ensure compatibility.

3.4. Summary of System Hosting Requirements

The following tables summarize the network and storage needs of the combined WeatherShare, ICM, OSS and AWOS/RWIS systems. This data can be used to evaluate the offerings from various hosting options.

	Data in	**Data out**	**Total**
WeatherShare with Current Coverage Area	62,779 MB (61.30 GB) per Month	696 MB (0.68 GB) Per Month	63,475 MB (61.99 GB) per Month
ICM	20,751 MB (20.27 GB) per Month	200 MB (0.20 GB) Per Month	20,951 MB (20.46 GB) per Month
OSS	See ICM.	379 MB (0.37 GB) per Month	379 MB (0.37 GB) per Month
AWOS/RWIS	20,751 MB (20.27 GB) per Month	200 MB (0.20 GB) Per Month	20,951 MB (20.46 GB) per Month
Total	104,281 MB (101.84 GB) per Month	1,475 MB (1.44 GB) per Month	105,756MB (103.28 GB) per Month

Table 22: Total Networking Bandwidth

	Program Space	**Static Data**	**Dynamic Data**	**Total**
WeatherShare	37 MB	18,857 MB	29,162 MB + 1170 MB per month	48,056 MB + 1170 MB per month
ICM	198 MB	13 MB	19.69 MB + 2.4 MB /month	230.69 MB + 2.4 MB per month
OSS	99 MB	NA	NA	99 MB
AWOS/RWIS	148 MB	567 MB	NA	715 MB
Total	482 MB	19,437 MB	29,181.69 MB + 1172.4 MB per month	49,100.69 MB (49 GB) + 1172.4 MB per month

Table 23: Total Data Storage

The current hardware configuration has been running and shown to have adequate processing and memory capacity and could be used as a base line for host system requirements. We note that while this system has adequate spare processing and memory resources, expansion of the web applications in terms of geographic coverage or functionality may necessitate an increase in storage capacity.

4. SOFTWARE REVISION

Prior to deploying any software system in a production environment the software should be reviewed and reworked where necessary to ensure reliability and efficiency. In this section we will examine WeatherShare, ICM, OSS, and AWOS/RWIS, and make general recommendations for revisions required to make each system production ready regardless of host platform. This should not be considered a complete examination of the systems studied but rather examples of some of the aspects that should be examined further.

4.1. Software changes

The software modules that make up each system should first be reviewed and re-worked with an eye toward ensuring cleaner and more efficient code. Web based systems such as those examined in the scope of this report are made up of client code that gets downloaded to the client machine for the display of the web pages and server code that runs on the server to obtain and process the data that will be eventually be displayed.

4.1.1. Client Code

The client code for these systems consists of HTML and javascript code used to render the web pages. There are a number of best practices that should be applied when examining and reworking this code for a production environment. Some of these practices that should be applied to the systems examining in this document are listed below:

- To cut down on the bandwidth used and increase the speed in rendering pages system Apache web server settings should be changed to enable compression on text files sent to the client.
  - For WeatherShare there are 6 JavaScript files and a CSS text file that would be compressed.
  - For OSS there are 7 JavaScript files and a CSS text file that would be compressed.
  - For ICM there are 4 JavaScript files and a CSS text file that would be compressed.
  - For AWOS/RWIS there are 7 JavaScript files and a CSS text file that would be compressed.
  - In addition, there are multiple XML files for which the file size could be reduced via compression.
- Each of the systems use a main JavaScript file called mapcodes.js to control the display of data on the Google map. For each system, this script should be examined and re-worked as necessary. Particular attention should be paid to removing unused code and ensuring that the timing of refreshes is correct and that the refreshes are handled in a logical and efficient manner.
- Combine multiple javascripts into one file. This will cut down on the overhead of multiple HTTP requests.

4.1.2. Server Code

Each system has multiple server side scripts that are used to access and process data that will be presented. Review should include the following:

- General clean up of the code by including appropriate inline documentation (comments) and eliminating unnecessary code.
- Cleaning up code directories to eliminating old and unused code files.
- Examination of the run interval timing of the scripts to ensure that they are appropriate with the data update intervals of the sources.
- The WeatherShare system generates png files using php scripts; the use of a compiled language such as “C” should be considered to improve efficiency of these CPU intensive operations. We note here that this should be done if significant speedup will result – this is uncertain without further testing. Library calls should involve roughly the same amount of processing, and if most of the time spent in these routines is to library calls, then a significant speed-up would not be expected.

4.2. Database Improvements

The WeatherShare system uses a MySQL database for storing a variety of data, some dynamic and some static. An example of the dynamic data stored is the historical information kept for all the stations current weather data. Examples of static data are the boiler plate information about each station and the table of milepost information for the Caltrans post-mile data. The operation of the database affects system resource usage, in particular the amount of disk space used and the CPU load. The following should be reviewed to ensure the database is production ready.

- Query efficiency
  - Examine and optimize queries and the use of indexes. Make sure all indexes that are created are needed and used for queries. Unused indexes affect the size of the database and have an adverse affect on inserts to the database. Note a balance needs to be struck between query efficiency and insert efficiency.
- Archiving
  - WeatherShare archives all station observation data for historical displays. The amount of data that is necessary for historical purposes should be reviewed and options of other sources for the historical data researched.
- Backup methods
  - The current system for backing up the database causes a spike in the load on the system, and its ease of use for restoring data is unproven. Backup procedures should be created and used that ensure reliable backups and ease of restoring in the event of a data loss. At present, all backups are full backups. Incremental or differential backups should be investigated.

4.3. Conclusions

While the systems running in a research environment have proven their worth, they are generally used by a limited number of users. When migrating systems into a production environment with the potential for a greater number of users it is important to review all the code making up the system and ensuring that maximum efficiency is achieved. This is important for both server side code for efficient utilization of system resources and client side code to improve the user experience.

5. DOCUMENTATION

Systems being moved into a production environment need to have adequate documentation so that the system can be maintained outside of the research environment. The web based systems examined are designed to be intuitive to use and therefore contain only a minimal set of user documentation. They have minimal system/operation documentation. Thorough system documentation is essential for the maintenance and support of a production system.

The following are examples of the types of things that should be included in the system/operation documentation prior to deploying a system in a production setting.

- Data sources. All data sources should be documented including the access methods and any authentication requirements. Update intervals for the data sources should be documented as well as the update intervals for retrieving the data.
- Database. The database structure should be documented in detail as well as authentication and access procedures. Backup/restore routines and policies should be documented.
- Code documentation. A document outlining all the scripts and source code needed to run the system should be created. All dependencies should be documented including system, application and library dependencies.
- Additional system components. All additional or third party software components that are utilized should be thoroughly documented.

6. SUPPORT

When deploying a web based system to a production environment provisions need to be made for ongoing support of the system. Some of the support issues that need to be considered are:

- General system monitoring. The system needs to be monitored periodically for overall operation health. This should include monitoring logs for any system or program errors and ensuring that successful data backups are occurring as intended.
- Data source changes. Systems that import data from external sources are dependent on the location and structure of that data. If/when changes are made by the owner of the external data, code modifications will need to be made to accommodate these changes.
- Provisions need to be made for applying various operating system and security updates. This should include testing on a development platform prior to deploying to a live system.

7. HOSTING ALTERNATIVES AND RECOMMENDATIONS

There are several options for external hosting of a web based product including shared hosting, virtual servers, dedicated servers, and different types of cloud based dynamic servers.

Shared hosting is generally the least costly; however, these options are generally marketed toward smaller sites and don't necessarily meet the system resources needs for our purposes.

Virtual servers divide up the physical resources of a server into multiple virtual servers. While these options tend to be less costly, ranging from $30-$60/month, they generally have smaller available disk, processor and memory capacity options, so they won't be considered here.

Dedicated servers are leased servers at a hosting facility. Some server administration can be done by the provider as an additional service. This option offers a fair amount of flexibility in configuration; however, you are leasing a fixed hardware configuration so you must attempt to take into account future expansion when initially sizing the server. Using the existing development sever as a base line, pricing for a similar system from 1and1.com that includes 2.3 GHz Quad core processor, 8 GB RAM, 1,000 GB of storage and a monthly bandwidth allotment of 4000 GB would cost about $200 per month. A similar offering from Inmotion hosting with 500 GB of disk storage and 2500 GB of bandwidth allotment would cost about $300 per month. While both of these options offer less CPUs and less memory than the current development system, based on the system usage for the current system (see Section 3.2) there should be adequate resources for the current and near term future needs. Both offer a much higher monthly bandwidth usage then we currently require and this should be able to handle increased usage of the sites as well as an increase of data into the system.

There are a few different flavors of cloud based dynamic servers. Cloud based servers logically connect the resources from any number of servers to create a flexible virtual server. These servers offer the ability to re-provision the system resources as required capacity grows or shrinks. Two types of cloud based server options will be examined below.

1and1.com offers their 1and1 dynamic cloud server which offers similar configuration options to their dedicated server, however it allows you to modify your configuration as the need arises. For a Quad-core system with 8 GB of RAM, 500 GB disk space, and 2000 GB per month of bandwidth the cost would be about $240 per month. As an example, raising the disk space to 700 GB and the RAM to 15GB would increase the cost to about $350 per month. While this system offers a fair amount of flexibility in configuration, it is primarily geared toward infrequent modifications and care must be taken in picking a host so that you don't outgrow all the available choices.

A second type of dynamic cloud server is the elastic compute cloud offering by Amazon. The Amazon Elastic Compute Cloud (EC2) is a virtual computing environment that is designed to allow the expansion or shrinking of resources quickly as requirements or usage change. The application can be designed to automatically scale itself if desired and you only pay for the resources that you use. Pricing varies depending on the options chosen and used. Pricing for a large instance, defined by Amazon as: 7.5 GB of memory, 4 EC2 Compute Units (2 virtual cores with 2 EC2 Compute Units each), 850 GB of local instance storage, 64-bit platform, is $297 per month. This includes 200 GB per month of incoming bandwidth and 5 GB per month of outgoing bandwidth. This does not include any dynamic expansion services, which should be investigated further. There are numerous options for monitoring and modifying the system that could be advantageous given the usage patterns of the systems examined with higher usage during times of bad weather and moderate to low usage other times.

7.1. Host Recommendation

The three viable options for external hosting as outlined above are: dedicated server, dynamic cloud server, and elastic cloud server.

The dedicated server option generally is the least costly; however, it is also the least flexible. Expansion or contraction of system resources can only be done by leasing a different dedicated server and may be subject to server availability. The monthly costs are fixed, however, and can be budgeted for on a long term basis.

The dynamic cloud option is more flexible allowing you to start with a base server configuration and then expand or contract resources as needed. The cost for a dynamic system may be more than an equivalent dedicated server; however, this option could allow you to start with a lesser configuration, with a less expensive cost, and increase capabilities when needed. Depending on the amount of resources used over an extended period of time, this long term costs of a dynamic cloud may be equivalent or less than a dedicated server. Depending on the provider there can be ceilings on the amount of resources that can be acquired and they may be subject to resource availability. Also with the variability of costs, depending on the resources used, budgeting for monthly costs can become more complicated.

The Amazon Elastic Compute Cloud option offers the most flexibility and expansion capability. While it is more costly than an equivalent dedicated server, it may allow for a lesser configuration to be used during times of normal activity with expansion only as needed. Since you pay for the resources used, the costs over a longer period of time may not be that much higher than a dedicated server. Due to the size of Amazon, the available options are extensive and need to be fully understood to be taken advantage of. The upper end for available resources is higher than smaller service providers. Again, the variable nature of the costs, due to the expansion or contracting of system resources, can make budgeting for monthly costs more complicated. Ultimately for systems such as the ones being examined here, this may be the most intriguing option as it will allow us to adjust to the variable nature of usage and easily react to added capabilities of the system.

7.2. General Recommendations

Prior to proceeding with external hosting, it is recommended that the general and specific changes documented in this report be further addressed and implemented to optimize the system. Specific recommendations such as implementing server-side compression on text files served by the web server would immediately improve performance for users. Other items presented such as whether or not to archive data must be further discussed to weigh the pros in terms of system performance and scalability with the associated, reduced functionality. Code-refactoring should be implemented as time and funding allow. And, greater effort and emphasis needs to be placed on system documentation and on-going support and maintenance needs.