AN EVALUATION PLAN FOR THE CONCEPTUAL DESIGN OF THE FLORIDA TRANSPORTATION DATA WAREHOUSE

PHASE-1-

FINAL REPORT

CENTER FOR **A**DVANCED **T**RANSPORTATION **S**YSTEMS **S**IMULATION
(CATSS)

University of Central Florida Account No.: 16-50-706
Contract No.:**BC355 RPWO# 4**

By

Haitham Al-Deek, Ph.D., P.E.
Associate Professor of Civil & Environmental Engineering, Associate Director of CATSS for ITS Programs, and Director of the Transportation System Institute

and

Amr Abd-Elrahman, Ph.D.
Research Engineer
Transportation Systems Institute
Department of Civil and Environmental Engineering
University of Central Florida
P.O. Box 162450
Orlando, Florida 32816-2450
Phone (407) 823-2988 Fax (407) 823-3315
E-mail: haldeek@mail.ucf.edu

Prepared for
The Florida Department of Transportation
March 2002

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle **AN EVALUATION PLAN FOR THE CONCEPTUAL DESIGN OF THE FLORIDA TRANSPORTATION DATA WAREHOUSE**		5. Report Date March 2002 6. Performing Organization Code
7. Author (s) Haitham Al-Deek and Amr Abd-Elrahman		8. Performing Organization Report No.
9. Performing Organization Name and Address Center for Advanced Transportation Systems Simulation (CATSS) Dept. of Civil & Environmental Engineering University of Central Florida 4000 Central Florida Boulevard Orlando, FL 32816-2450		10. Work Unit No. (TRAIS) 11. Contract or Grant No. BC355 RPWO# 4
12. Sponsoring Agency Name and Address Florida Department of Transportation Research Center 605 Suwannee Street Tallahassee, FL 32399-0450		13. Type of Report and Period Covered Final Report 14. Sponsoring Agency Code
15. Supplementary Notes Prepared in cooperation with Florida Department of Transportation.		

16. Abstract

The goal of this research project is to establish an implementation strategy for the Florida transportation data warehouse, also known as the Central Data Warehouse (CDW). Information used in this report has been compiled through different sources, including meetings with stakeholders, nationwide visits, electronic surveys, phone inquires, and websites. A three-year implementation plan is suggested. Significant amount of traffic data are currently being collected statewide. However, this study demonstrates that these data lack common definitions and collection/archival strategy. The most urgently needed application in the near term is Advanced Traveler Information Systems (ATIS). Additional applications such as predictive travel time information, incident detection, and data mining are also suggested. Multi-tier application development architecture is recommended. In this architecture, the application presentation layer is separated from other layers such as data and analytical layers. A dimensional database model is recommended for the database implementation. In this model, data are saved in fact and dimension tables. Common dimensions such as TIME and LOCATION are used to link different data types and facilitate across-subject queries. Geographic Information Systems (GIS) will play a major role in the implementation of the Florida CDW. GIS will support the database integration functionality and will provide a graphical interface and spatial query capabilities in the developed applications. Hardware, software and labor requirements for the data warehouse have been investigated and presented in the report along with an implementation budget. Finally, this study proposes the UCF-CATSS as the host of the Florida CDW. This suggestion is based on CATSS' extensive experience in developing and managing the I-4 data warehouse over the last ten years and also due to CATSS' overwhelming success in providing web based historical, real time, and predictive travel time information to commuters and tourists in Central Florida and beyond.

17. Key Words	18. Distribution Statement
Florida Transportation Data Warehouse, Data Archival, Transportation Data Modeling, Data Warehouse Applications	No restriction - this report is available to the public through the National Technical Information Service, Springfield, VA 22161

19. Security classif. (of this report)	20. Security Class. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	101	

Form DOT F 1700.7 (8_72) Reproduction of completed page authorized

DISCLAIMER

The opinions, findings and conclusions in this publication are those of the authors and not necessarily those of the Florida Department of Transportation or the US Department of Transportation. This report does not constitute a standard specification, or regulation. This report is prepared in cooperation with the State of Florida Department of Transportation.

ACKNOWLEDGEMENT

This report has been prepared as part of Contact No.: BC355 RPWO# 4, "An Evaluation Plan for the Conceptual Design of the Florida Transportation Data Warehouse". The report has been prepared by the Center for Advanced Transportation Systems Simulation (CATSS) at the University of Central Florida (UCF).

The authors would like to acknowledge the financial support of the Florida Department of Transportation (FDOT) Central Office in Tallahassee, Florida. In particular, the authors would like to thank Mr. Liang Hsia, FDOT ITS Engineering Administrator, and Mr. Chester Chandler, State ITS Program Manager, for their role in supporting and managing this research. Special thanks are also due to Mr. George Gilhooley, Director of Operations, FDOT District-5-, who served as the technical manager of this project. Mr. Gilhooley has been very supportive and very instrumental from the early start of this project. We would like to thank Dr. Harold Worrall, Executive Director of the Orlando-Orange County Expressway Authority (OOCEA) for his advice and cooperation in this research project. The authors appreciate the help of Mr. Fred Ferrell, FDOT District-5- Traffic Operations Engineer, Ms. Anne Brewer, FDOT District-5- Assistant Traffic Operations Engineer, and Ms. Jennifer Heller, FDOT District-5- Traffic Operations. Finally, special thanks go to Mr. Gregory Floyd at FDOT District-5 Regional Traffic Management Center (RTMC) for his sincere and diligent support throughout the entire period of this project.

EXECUTIVE SUMMARY

The goal of this research project is to develop a plan for the conceptual design of the Florida transportation data warehouse, also known as the Central Data Warehouse (CDW). The conceptual plan sets the stage for immediate implementation of the CDW and identifies the most important and urgently needed applications of the CDW. This study reflects recent developments in database design and communication technology as applicable to the State of Florida.

The UCF-CATSS research team had extensive meetings with stakeholders including the Florida Department of Transportation (FDOT) District-5-, the Orlando-Orange County Expressway Authority (OOCEA), and the Florida Turnpike District-8-. Also, the CATSS research team made visits to national universities and agencies with previous experience in data warehousing. An electronic survey was conducted with participants from all of the eight FDOT districts. Meetings and/or telephone calls were conducted with transportation engineering and planning consultants who are interested in this study including PB Farradyne, PBS&J, and Kimley-Horn & Associates. Information was also collected through websites, reports, and recent publications provided by software and hardware vendors.

A three-year implementation strategy for the data warehouse is suggested in this report. There are three phases for implementation of the CDW, this does not include the present Phase-1-. Phase-2-, which is proposed to start immediately, will cover the I-4 corridor, the I-10 corridor, the I-95 corridor under surveillance in Miami and Jacksonville, and the OOCEA and the Turnpike toll roads. Phase–3- covers large parts of the state highway and arterial system, especially in highly populated counties and cities in Central Florida and beyond. The final Phase-4- covers statewide data collected from transit systems, emergency management centers, and seaports.

The estimated total budget of the project can be divided into the start up cost and the yearly cost. The start-up cost is the cost of software, hardware, and fiber network communication requirements, in addition to personnel needed to run the data warehouse. This is estimated at $575,000 as direct cost. The project yearly budget is mainly to cover labor, on-campus hardware maintenance, and dedicated T-1 connections monthly expenses. This is expected to be approximately $320,000/year starting in the second year (or Phase-3- of the project). This yearly cost will continue for as long as the CDW is operational. These estimates exclude the cost of extending FDOT fiber optic network to the foot doors of UCF, i.e., cost of the off-campus fiber network to connect the FDOT Regional Traffic Management Center (RTMC) to UCF-CATSS facilities is not included in the above estimates. It is assumed that the FDOT D-5- will come forward with their present plans to complete the off-campus fiber connections. Furthermore, these estimates are for the primary fiber connection only. Plans for a redundant fiber connection are being investigated.

This study demonstrated that most of the traffic management software deployed at the FDOT districts' Traffic Management Centers (TMCs) use special and proprietary definitions for the collected data, which are not compatible across the districts. These data definitions need to be mapped to the ITS standard definitions, such as the Traffic Management Data Dictionary (TMDD) and the ITS Data Registry (ITS-DR). Although, this will add more pre-processing and reformatting steps to the data warehouse, it will facilitate data expansion in the near future as more agencies start to implement ITS standards and produce extensive data for the CDW.

<u>It is found that ATIS is the most important and urgently needed application of the CDW.</u> In addition, the UCF-CATSS team suggests that traffic monitoring data, dynamic message signs data, incident data, and lane closure data be included in the implementation of Phase-2-. These data types represent the backbone of a statewide ATIS application. Moreover, many of the stakeholders expressed their pressing needs for developing applications that collect and disseminate these types of data. Applications for short-term and long-term travel time prediction, incident detection, data mining, and historical data dissemination are also recommended.

This study suggests two different database replication techniques to copy the data (collected and logged by remote sites) into the Florida CDW database. Both commercial and in-house solutions were reviewed. It is suggested to use in-house development of database replication solutions due to the high cost associated with current commercial options. The study stresses the need for data filtering and fusion. Initial filtering should only filter up-normal data. In the mean time, more robust data filtering techniques will be developed and applied as needed.

A multi-tier approach for application development is suggested for the CDW. This approach identifies at least three separate application development layers or tiers namely: the display or presentation tier, the logical tier, and the data tier. Using multi-tier architecture will provide more portability and reusability to the developed applications. Dynamic query applications such as On Line Analytical Processing (OLAP) are also recommended in this study. The successful deployment of these applications requires the selection of appropriate database model. Multi-dimensional database model is proposed for the data warehouse implementation. In this model, data are organized in fact and dimension tables. A sample of multi-dimensional models is presented in this research for loop detectors, Dynamic Message Signs (DMS), and incident data. Common dimensions, such as TIME and LOCATION are proposed to provide the integration between different data types employed by the data warehouse. This will allow querying all data types in the data warehouse based on common time or location information. Moreover, different aggregation levels are suggested for the data. For example, loop detectors data can be averaged for certain road links, transportation corridor(s), or the entire regional transportation network. In the mean time, the same loop detectors data can also be aggregated for different time intervals.

The Florida CDW will fully integrate GIS to provide the LOCATION dimension in the database model. GIS will also provide significant part of the data warehouse functionality in addition to its traditional role in providing the graphical interface and spatial queries for the developed applications.

The major recommendation of this study is to host the Florida CDW and its activities at the UCF's Intelligent Transportation Systems (ITS) lab within CATSS facilities. In this capacity, the UCF-CATSS will act as the *custodian* of this data warehouse and will be the *distributor* of this data to public and private sectors. The UCF-CATSS has a proven record and extensive solid experience in accumulating, filtering, and processing traffic (loop detector data) and incident data on I-4. These significant efforts started back in 1992 and continued through the present date. The UCF-CATSS has been disseminating real time traffic information on a 39-mile stretch of I-4 for the past three years, see the CATSS website www.catss.ucf.edu.

The feedback of the above website demonstrates that thousands of travelers benefited from this travel information service. Furthermore, the FDOT website has a link to the CATSS I-4 website. Currently, there are some efforts to use the UCF-CATSS website by news media. Moreover, the UCF-CATSS has developed and successfully implemented a robust short-term traffic prediction model on I-4. This model has been thoroughly tested on the I-4 corridor, and proved to be very accurate at the 95% confidence level. The model is capable of projecting travel times and delays using the most recent information collected from loop detector stations and transmitted to the UCF-CATSS lab. Furthermore, the UCF-CATSS has started the process of extracting travel time data from E-PASS/SUNPASS transaction data, visit www.catss.ucf.edu/oocea.

In conclusion, this study presents the guidelines for achieving a fully functional Florida data warehouse. The Florida Central Data Warehouse (CDW) will be distinguished nationwide by its state-of-the-art technologies in software development and database design, and its complete and diversified transportation data sets. Finally, this implies immediate implementation of the recommendations and guidelines provided in this study.

ACRONYMS

AASHTO	American Association of State Highway and Transportation Officials
API	Application Programming Interface
ATIS	Advanced Traveler Information Systems
ATMS	Advanced Traffic Management System
AVI	Automatic Vehicle Identification
C2C	Center-to-Center
CATSS	Center for Advanced Transportation Systems Simulation
CCTV	Closed Circuit Television
CDW	Central Data Warehouse
CORBA	Common Object Request Broker Architecture
DATEX-ASN1	Data Exchange for Abstract Syntax Notation 1
DBMS	Database Management System
DDB	Distributed Database
DMS	Dynamic Message Signs
DOT	Department of Transportation
ER	Entity Relationship
FDOT	Florida Department of Transportation
FHP	Florida Highway Patrol
FHWA	Federal Highway Administration
GIS	Geographic Information Systems

GUI	Graphical User Interface
IEEE	Institute of Electrical and Electronics Engineers
IMS	Internet Map Server
ISO	International Organization for Standards
ITE	The Institute of Transportation Engineers
ITS	Intelligent Transportation Systems
ITS-DR	ITS Data Registry
IVR	Interactive Voice Response
JDBC	Java Database Connectivity
N/A	Not Applicable
NTCIP	National Transportation Communication for ITS Protocols
ODBC	Open Database Connectivity
OLAP	On-Line Analytical Processing
OOCEA	Orlando-Orange County Expressway Authority
ORB	Object Request Broker
RAID	Redundant Array Independent Disks
RTMC	Regional Traffic Management Center
RTMS	Remote Traffic Microwave Sensors
SDE	Spatial Database Engine
SQL	Structures Query Language
TBD	To Be Continued
TMC	Traffic Management Center
TMDD	Traffic Management Data Dictionary

UCF	University of Central Florida
VID	Video Image Detectors
WAP	Wireless Application Protocols

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1.
INTRODUCTION

The Center for Advanced Transportation Systems Simulation (CATSS) was contracted by the Florida Department of Transportation (FDOT) to perform an evaluation plan for the conceptual design of the Florida Transportation Data Warehouse. The rapidly growing deployment of ITS technology statewide, combined with the need to integrate and employ this data in useful applications, was the key motivation behind initiating this study. Moreover, the deployment of the ITS data warehouse market packages is an essential element of the national ITS architecture. The data warehouse market package involves collecting data from multiple agencies and sources, filtering and processing the data, providing the required metadata, and developing end-user applications. These applications should be able to support planning and decision making requirements in addition to providing the general public with needed information.

Data warehousing projects normally utilize well-established, and highly functional databases. In contrary, the current concept of a transportation data warehouse, which involves collecting, archiving, and analyzing data from multiple sources, and generating powerful and useful applications, is fairly new. The following are examples of challenges that face the implementation of the data warehouse:

- Available transportation data are collected by many agencies with no common standards.
- Data organization ranges from data logged in plain files, such as some road maintenance data, to those managed by commercial databases such as transit and toll transactions data.
- Many data types collected by the same system are not integrated. Moreover, no database model is utilized to describe the relationships between data types. For

example, detectors data and incident data collected by the same agency might have totally separate database schemas.

- Data currently available to support needed short-term applications are insufficient. Therefore, a strategy should be developed to make use of existing data in developing multi-phase applications that have the ability to expand when more data become available.
- Administrative and cross jurisdiction problems exist when trying to collect data maintained by different agencies. Agencies hesitate in sharing the data due to cost, liability, and lack of sufficient manpower.
- A database model that links and integrates the different transportation data types needs to be developed.

This study will investigate these problems and suggest a Florida data warehouse implementation strategy in the light of the current developments in ITS standards, software, hardware, and communication technology.

1.1 STUDY OBJECTIVES

This study commenced on August 31st, 2001 with the following three main objectives:

1. Establish a multi-year and multi-phase implementation plan for the data warehouse project.

2. Provide a budget for the data warehouse implementation that includes equipment, hardware and software cost, and cost of human resources.

3. Suggest database design elements and application requirements for the data warehouse.

Achieving these objectives will enable the immediate launching of the Florida data warehouse project. However, the study also extended its recommendations and objectives to provide:

- Recommendations for future traffic management software development.

- Recommendations for deployment of ITS equipment.
- Introducing state-of-the-art developments in data warehousing technology.
- Reviewing different ITS data types available and planned by FDOT districts.
- Initiating discussions for many of the current and future FDOT projects such as the statewide travel information system and regional geographic information system.

1.2 STUDY APPROACH

One of the challenges in conducting this study is the amount of data that needs to be collected and reviewed in a very short time frame. The following is a list of the most relevant sources of information for this study:

- Information gathered in Florida's stakeholders meetings and visits to national transportation centers, universities, local and state agencies.
- Response to electronic surveys designed by CATSS team, communication over the Internet, and phone interviews with the eight Florida districts.
- Transportation publications and case studies nationwide.
- Communications with major traffic management software developers currently used by Florida traffic management centers.
- Phone inquiries, reports, and quotes provided by commercial software and hardware vendors.
- Data warehouse publications in the transportation and business fields.

A phased implementation of the Florida data warehouse is followed in this study. The results of the study are presented in this report in the following order:

1. Summary of investigated data requirements including sources, archiving methodology, and data types.

2. Summary of investigated data processing and applications requirements. Candidate data warehouse applications and their development architecture. This includes applications for initial data collection, filtering and reformatting as well as end-user applications.

3. Database modeling and schema.

4. Software, hardware, communication, and labor needs for the Florida CDW project.

5. Suggested overall implementation time frame and budget.

CHAPTER 2.
INVESTIGATING DATA REQUIREMENTS

The objective of this section is to investigate data requirements in the data warehouse implementation strategy. Statewide and national information were collected through stakeholders' meetings, electronic surveys, nationwide visits, and phone and email inquiries. Candidate data types for the first data warehouse implementation phase are then identified and presented.

2.1 ITS DATA INQUIRIES

Meetings have been conducted with different stakeholders participating in the data warehouse ITS market package. An electronic survey designed to investigate different aspects of current and planned ITS data in addition to expected data warehouse applications was used to collect information from all eight Florida districts. Follow up phone interviews were conducted with each of the districts' ITS engineers based on their response to the electronic-survey. In these interviews, more information was requested. Phone or email enquiries were conducted with major traffic management software developers. Visits were also conducted to nationwide institutes and organizations to explore their experience with transportation data warehousing.

The following summarizes the outcomes of the state stakeholder meetings, electronic survey results, and nationwide visits.

2.1.1 Stakeholder Meetings

The FDOT District-5- Regional Traffic Management Center (RTMC):

In a three-day workshop conducted in Orlando on June 23, 26 and 28, 2001, the statewide regional ITS architecture specified the FDOT D-5 RTMC as the Central Florida regional data center, see http://www.jeng.com/florida/Default.htm. All Central Florida local data will be accessed through the RTMC center.

A meeting was held at UCF with the FDOT D-5 and with the Turnpike District on October 24th, 2001. The meeting was attended by the two districts' directors and their assistants. Several meetings were also held with the FDOT RTMC personnel. These meetings were part of continuous consultation and cooperation efforts between UCF-CATSS and FDOT D-5. During these meetings, current data formats, data accessibility, and storage strategies were discussed. This includes loop detectors data, Dynamic Message Signs (DMS) data, incident history data, and lane closures data. Currently, FDOT D-5 is employing the MIST traffic management software under windows operating system. A Sybase database is used to archive 15-minute summaries of detectors data. These data are archived nightly and saved in text file format. The data are backed up and periodically moved to offsite storage. In the mean time, incident history data are archived weekly.

No current strategy to fully disseminate historical data is implemented by the RTMC. The UCF-CATSS has historical loop detector data for the I-4 detectors since 1993 with some intermediate periods of data discontinuities. The UCF system reads loop detectors data parsed into a text file every 30 seconds. The data are read, filtered, and then inserted into SQL SERVER database. Real-time, predictive, and historical traffic data dissemination applications were developed. These applications query the database and output the results through end-user web-based interface.

The FDOT D-5 expressed an essential need for a data warehouse to maintain the different types of available data, integrate them with other data sources, and provide front-end applications that facilitate data access. Another application was discussed with FDOT D-5 for preparing loop detectors' data summary in a format appropriate for the SPS statistical software, currently used by FDOT. The data will significantly reduce the cost and safety hazards associated with performing manual traffic counts by contractors or FDOT personnel in many FDOT projects.

The Orlando-Orange County Expressway Authority (OOCEA):

A meeting was conducted with the OOCEA Executive Director, Dr. Harold Worrall, on October 3rd, 2001. In this meeting, current status and future plans of the OOCEA travel information data extraction and dissemination were discussed. The OOCEA's vision of a real-time traveler information system that includes I-4 loop detector data, the OOCEA and Turnpike toll roads, and arterials was discussed.

A pilot project conducted by PBS&J to derive travel time information from main-line toll plazas transactions was discussed. More options are currently under investigation by the authority to use transaction data from toll collection ramp plazas and to install more Automatic Vehicle Identification (AVI) readers throughout the toll road system for true travel time measurements [1]. A sample of the pilot project speed and travel time data were acquired through OOCEA's password-protected (and under development) confidential website developed by PBS&J. The data were investigated together with transaction data previously processed by the UCF-CATSS team. See link www.catss.ucf.edu/oocea to check the work completed by UCF on OOCEA's data. This was completed with a small sample of one week worth of transactions data. It is presently working off line on CATSS website. The next step is to investigate resources needed to expand and integrate this website with the I-4 website to become one functional travel time information system for Central Florida and Orlando area travelers.

The OOCEA has made preliminary commitments to partner with the FDOT to co-locate its operations at the FDOT D-5 RTMC [2]. This will allow OOCEA data to be accessed through a connection to the RTMC as part of regional access to local transportation data that feeds into the RTMC.

The Turnpike District:

A meeting was conducted with the Turnpike District ITS consulting team led by Transcore, Inc. on Veterans Day, November 12th, 2001. In this meeting, the need to extract real time travel speed data was discussed. Initially, toll plazas' transactions data will be used until more probe sensors are installed. The fiber optic network is essential for real time collection of the transactions data and the probe data. DMS messages need to be extracted from realistic travel time data. The data fusion problem was also discussed. The need to integrate the Turnpike travel time data with other data sources for a statewide traveler information system was discussed. A one-week sample of transaction data was requested from the Turnpike District by the UCF team, however this has not been received from the Turnpike District until the time of writing this report. The Turkey Lake office plans to merge their two offices: office of tolls and office of operations. This will make data processing easier for the planned CDW. Presently, their main operations office is in Pompano Beach.

2.1.2 Visits to National Universities and Other States

Visits were conducted to universities where an existing partnership between the university and the state DOT resulted in a successful implementation of a CDW similar to the one that is being established in Florida. In these meetings, data sources, processing techniques, and hardware/software configurations were discussed. Details of the visits to the University of California at Irvine, the University of California at Berkeley, and the Washington State Transportation Center (TRAC) are summarized in Appendix A of this report. In summary, Berkeley is well ahead of all these systems. The Berkeley system is almost entirely data driven with very little simulation. The UC Irvine data warehouse is not open to public even though they have been working on it for many years now. On the other hand, the UC Berkeley system is open to anyone in academia and states. Private sector can also use the data with special agreements. The State of Washington Transportation Research Center (TRAC) shares ITS offices with the University of Washington and this center is another successful example. Real time traffic information is provided through Washington State DOT website. The current source of data is 2500 single loop detectors on the freeway system. Note that FDOT D-5 has double loops, which enables measuring vehicle speeds rather than estimating them as in the case of TRAC. TRAC plans to include arterials and signal cycles, but this is a pie in the sky at the present time. Probes have to be really large in number to be significant in providing travel time measurements. Loop detectors data is available "off-line" with 20 seconds resolution for planning purposes. In short, none of the CDWs visited nationwide has sufficient probe data to work with on toll roads or arterials. It is rather interesting to know that these data warehouses were the most advanced in the nation. We have a unique situation in Florida. In Orlando alone, there are more than 300,000 SUNPASS transponders and a large number of SUNPASS transponders exists statewide. These transponders can be utilized for tracking probe vehicles and measuring travel time on arterials and toll roads if additional tag readers are installed in appropriate places such as traffic signals.

2.1.3 Florida Fiber Optic Network

Efforts to establish the Florida fiber optic network are underway. In the mean time, regional fiber optic network installations are gaining ground support by counties and cities that are in need for integrated traffic control systems. As mentioned before, the regional ITS architecture identified the RTMC as the hub for Central Florida data. Data collected by Central Florida counties and cities will be accessed through the RTMC center. Therefore, a fiber optic connection between the CDW to be located at UCF-CATSS facilities and the RTMC must be established.

Two meetings were conducted on December 5th 2001 and January 24th 2002 at the Koger center in Orlando to discuss regional fiber optic connections between UCF-CATSS and RTMC. The meetings were hosted by F. R. Aleman & Associates, and attended by the UCF-CATSS research team, representatives of Orange, Seminole, and Broward counties, and the city of Orlando. The OOCEA was represented by the authority's consultant PBS&J in these meetings. During the two meetings, current plans to connect UCF with FDOT D-5 RTMC through Orange and Seminole counties fiber optic networks were discussed. Due to the importance of the meetings' outcomes and their impact on the overall project budget, the results are presented in details in the 'Development of Implementation Schedule and Proposed Budget' section of the report.

2.1.4 Electronic Survey Results

A survey was designed to explore current ITS data collected and archived by each of the eight Florida DOT districts. The survey was delivered as an email attachment to each district's traffic operations engineer and each district's ITS engineer. A unique response was received from each district. The survey also included questions about future ITS implementations, archiving techniques, and needed applications. More emphasis was devoted to freeway detectors data, AVI data, and DMS data. Survey responses were followed by detailed phone interviews with the district representatives. The districts were given a second chance to comment on their earlier responses when they saw the first draft of the survey results. All of the districts' comments were incorporated in the tables of this section.

Tables 1, 2, and 3 summarize the electronic survey results. The tables show lists of current and planned traffic monitoring detectors data, DMS and CCTV data collected by each Florida district. Information about transactions data collected and archived on the Turnpike and OOCEA toll roads are also shown in Table 4. Appendix B presents a more detailed summary of ITS resources available at each district.

The information presented in these tables shows the significant amount of statewide ITS data currently collected on the major highways. Within the next few years, these data sources are expected to grow an order of magnitude of the present day's database. The majority of the data are collected by District-5's RTMC, District-2's Jacksonville TMC, and District-6's Miami TMC. The tables also show that different technologies are utilized to collect the same type of information. For example, the speed, volume, and occupancy data are collected using the inductive loops technology on the I-4 corridor and using the Video Image Detectors (VID) technology on I-95 and I-10. The same type of data is planned to be collected using Remote Traffic Microwave Sensors (RTMS) in many parts of the I-95 corridor within the next few years. Moreover, in many cases, different equipment manufacturers use different data definitions and formats while collecting the same data parameters.

Table 1. Summary of Freeway Detectors Data

District	Detector Type	Number of Units	Location of Existing Units	Location of Units Under Construction	Location of Planned Units	Archived (Yes/No)	Archiving Location	Database	Accumulation
1	Inductive Loops	TBD			I-4 I-75	Yes	TBD	TBD	TBD
2	VID	41	I-10 I-95 I-295			Yes	Jacksonville TMC	Oracle	Initially 15 min
	RTMS	TBD			I-95	Yes	Jacksonville TMC	Oracle	TBD
3	TBD	TBD			I-10	TBD	TBD	TBD	TBD
4	RTMS	TBD			I-95	TBD	Broward County TMC	TBD	TBD
	RTMS	156			I-95	TBD	TBD	TBD	TBD
5	Inductive Loops	70	I-4			Yes	UCF-CATSS	SQL Server	30 sec
							Orlando RTMC	Sybase	15 min
	Inductive Loops	30	I-4			Yes	Orlando RTMC	Sybase	15 min
	Inductive Loops	10	I-4 I-95			Yes	Daytona TMC	Sybase	15 min
	Inductive Loops	13		I-95		Yes	Orlando RTMC	Sybase	15 min
	Inductive Loops	16			I-4	Yes	Orlando RTMC	Sybase	15 min
6	Inductive Loops	3	I-95			Yes	Miami	Plain Text	1 minute
	Inductive Loops	12			I-95	Yes	Miami	Plain Text	1 minute
	VID	35	I-95			Yes	Miami	Plain Text	1 minute
	RTMS	92			I-95 SR826	Yes	Miami	Plain Text	1 minute
7	VID	~60			I-275 I-4	Yes	Tampa RTMC	TBD	TBD
8	TBD	TBD			Mainline Turnpike	Yes (Future)	Pompano & Turkey Lake TMC	Sybase	TBD

Note: Hashed cells indicate "not applicable."

Table 2. Summary of Dynamic Message Signs Data

District	Number of Units	Location of Existing Units	Location of Units Under Construction	Location of Planned Units	Archived (Yes/No)	Archiving TMC	Database
1	TBD			I-4 I-75	TBD	TBD	TBD
2	8	I-10 I-95 I-295			Yes	Jacksonville	Oracle
	TBD			I-95 I-295	Yes	Jacksonville	Oracle
3	4	I-10			No	N/A	N/A
	TBD			I-10	TBD	TBD	TBD
4	16			I-95	TBD	TBD	TBD
	12		I-95		**Yes**	Broward County	TBD
	22		I-595 SR-84		**Yes**	Broward County	TBD
5	37	I-4			Yes	Orlando RTMC	Sybase
	4	I-95			Yes	Orlando RTMC	Sybase
	4	I-4 I-95			Yes	Daytona	Sybase
	6		I-95		Yes	Orlando RTMC	Sybase
	5			I-4	Yes	Orlando RTMC	Sybase
6	4	I-95			No	N/A	N/A
	10	I-95 SR826 Golden Glades Interchange			Yes	Miami	Oracle
7	2	I-275			No	N/A	N/A
	1	I-175			No	N/A	N/A
	~15			I-275 I-4	Yes	Tampa	TBD
8	20		Mainline Turnpike		Yes	Pompano & Turkey Lake	Sybase

Note: Hashed cells indicate “not applicable.”

Table 3. Summary of CCTV Resources

District	Number of Units	Location of Existing Units	Location of Units Under Construction	Location of Planned Units	Traffic Management Center
1	TBD			I-4 I-75	TBD
2	22	I-10 I-95 I-295			Jacksonville
	TBD			I-95 I-295	Jacksonville
3	4	I-10			FHP
	TBD			I-10	TBD
4	28			I-95	TBD
	~50			I-95	Broward
5	10	I-95 I-4			Daytona TMC
	19	I-4			RTMC
	8			I-4	RTMC
	12			I-95	RTMC
6	40	I-95 SR-826 (Golden Gate Interchange)			Miami
	4	I-95 I-595			SmartRoute TMC
7	3	I-275 I-175			City of St Petersburg
	13		I-275		FHP
	~30			I-4 I-275	RTMC
8	TBD	N/A	N/A	Mainline Turnpike	Pompano & Turkey Lake

Note: Hashed cells indicate “not applicable.”

Table 4. Summary of Automatic Vehicle Identification Data

District	Existing AVI Locations	Planned AVI Locations	Archived	Database	Accumulation
Turnpike	Toll Plazas	Mainline and ramps	Transactions	Oracle	TBD
OOCEA	Toll Plazas	Mainline and ramps	Transactions + Travel time	Oracle	5 minutes for travel time data

Note: Hashed cells indicate "not applicable."

The survey and its subsequent phone interviews showed that most of the traffic management software used by the state TMCs are recently deployed. This software utilizes commercial DBMS to store system parameters and to archive collected traffic data. In other words, the primary use of these DBMS is to keep records of collected data rather than to integrate different data sources and provide data warehousing capabilities. Different archiving strategies and data definitions are applied by each software package. In addition, data logs are produced with different time aggregations. This adds significantly to the complexity of any data warehousing effort. Accordingly, common traffic management software library is recommended for the state of Florida. This was also one of the main recommendations of the TMC software study conducted recently by FDOT [3]. The statewide software should consider data collection/archival needs in addition to traditional traffic operations requirements.

2.2 CANDIDATE DATA TYPES

Although, some ITS technologies are mature enough so that their data element definitions have been standardized and widely implemented by the equipment manufacturers and the software developers, these definitions are not implemented in some of the equipment already operational in the state. Furthermore, many promising ITS technologies are still in the emerging stage and no standards have been applied to them yet. Data archiving can be greatly facilitated if common data definition standards are applied statewide. Integrating data that do not share common standards adds extra burden to the data archiving process. It adds many pre-processing and intermediate steps to prepare the data. The following are efforts to establish standard definitions for ITS data to be included in the data warehouse implementation.

2.2.1 ITS Data Definition Standards

A significant part of the data warehouse implementation is concerned with the transportation data types and formats. A consistent data format is necessary to establish a successful data warehouse, especially when dealing on a state level, and ultimately on regional and national levels. Therefore, current efforts to develop standards for transportation data definitions and formats are described in the next two subsections.

2.2.1.1 Traffic Management Data Dictionary (TMDD)

The Institute of Transportation Engineers (ITE), working cooperatively with the Federal Highway Administration (FHWA) and the American Association of State Highway and Transportation Officials (AASHTO), is leading the efforts on the national level to establish standard transportation data definitions. One of the objectives of this data definitions or data dictionary is to provide a unique identification and description of the data elements used in the communication of messages, information among traffic management centers, and information between traffic management centers and travelers equipped with ATIS technology. A full list of data types with TMDD can be found at http://www.ite.org/tmdd/ [4].

2.2.1.2 ITS Data Registry (ITS-DR)

The Institute of Electrical and Electronics Engineers (IEEE) is currently in the process of developing ITS data registry. The IEEE ITS-DR is a centralized data dictionary or repository for all ITS data elements and other data concepts that have been formally specified and established for use with the US national ITS domain. Its primary objective is to support the clear-cut interchange and reuse of data and data concepts among the various functional areas of intelligent transportation systems. The ITS-DR can be accessed through http://standards.ieee.org/regauth/its/index.html [5] (registration is needed).

Several data elements participating in the Florida data warehouse are currently defined by the TMDD and IEEE ITS-DR. Standards for detectors data, DMS data, and incident management data are listed in these standards among other types of traffic management data.

2.2.2 ITS Data Review

In this section ITS data candidate for the Florida data warehouse implementation are reviewed.

Inductive Loop Detector Data

Inductive loop detector technology is the most prominent technology currently used for traffic data collection on freeways and arterials. Although loop detectors technology is more than three decades old, the technology proved to provide significant information that can be used in many applications. Loop detector data are read many times a second and typically reported back to the TMC at intervals of 20 or 30 seconds. A well-maintained loop detectors system can provide dense occupancy and volume information in addition to speed measurements or estimates along freeways every few seconds.

A configuration of two successive loop detectors forms a speed trap and can provide actual speed data. This information was efficiently used to provide real time or near real time credible travel time information for many users nationwide. In the state of Florida, the technology was successfully applied on the I-4 and part of the I-95 corridors. The main disadvantage of this technology is its relatively high installation and maintenance cost.

Real time travel information systems and incident detection algorithms are two prominent applications that can successfully be developed using loop detectors data. Some applications were also developed for vehicle classification. In these applications, the detector signals are digitized and matched with different vehicle signatures. A digital code representing the matching result is output to indicate vehicle type.

Remote Traffic Microwave Systems (RTMS)

Near term plans exist to deploy RMTS technology in many places along I-95 corridor by FDOT D-2, FDOT D-4, and FDOT D-6. The technology can be effectively used to acquire volume, occupancy, and speed information along freeways. Normally, data are collected and sent with at least 10-second interval.

Video Image Detectors (VID)

VID or Video Image Processor (VIP) technology is now implemented on I-95 in FDOT D-6 and on I-10 in FDOT D-2. The technology utilizes images supplied by video cameras and uses image-processing algorithms to extract traffic information. Traffic volume, occupancy, and speed can be provided by this system. A VID system has the advantage (over loop detectors and RTMS systems) of being able to process more than one lane at once. However, VID's performance can be affected by weather conditions such as rain and by changes in light conditions caused by shadow.

Appendix C lists some data elements for the traffic detector sensors as defined by the TMDD. It should be mentioned here that the purpose of adding this part of the detectors TMDD table to the report is to provide the reader with an example of data definitions. More complete and detailed description of the detectors TMDD can be found at http://www.ite.org/tmdd/tmdd3detectors_1.html. The TMDD definition for detectors data includes a field for detector type in which a code is assigned to each type of detectors discussed in the previous sections. A unique definition of the data, including units and data aggregation information need to be well documented for each utilized type of detectors before the data is being archived or employed in any application.

Automatic Vehicle Identification (AVI)

AVI technology is implemented mainly on toll roads across the state. Both the OOCEA and the Turnpike District collect transactions data using AVI readers. As mentioned previously, OOCEA has finished a pilot project to utilize the transaction data to collect travel time information along mainlines. However, these readers are only installed at the toll plaza locations, which enables computing travel time between two plazas on the mainline or between a mainline plaza and a ramp plaza. OOCEA and Turnpike District have plans to install more AVI readers on ramps or along mainlines to collect more dense travel time information. Travel time or speed data over the link should be aggregated over a defined time period. This information should be provided as metadata, i.e., description and information about the data itself, associated with the archived data. Standards for the probe data elements have been documented in the TMDD. Detailed description of the vehicle probe data definition can be found at http://www.ite.org/tmdd/tmdd3probe_1.html.

Dynamic Message Signs (DMS)

DMS was one of the ITS technologies that benefited from the development of ITS communication standards. All of the FDOT districts have DMS currently functioning. However, most DMS implemented in Florida use proprietary standards and lack a clear and unified archiving strategy. Furthermore, no algorithm is currently utilized to automatically update the messages based on collected traffic monitoring data. Some traffic management software used in the state of Florida provides the capabilities to log historical DMS messages. In other cases, DMS data are not archived, or access to the logged data is only available through the DMS propriety software.

DMS data includes DMS text, duration, status, device, and location information. The definition of these data elements is documented in the National Transportation Communications for ITS Protocol (NTCIP) standards, discussed later in the report, http://www.ntcip.org/library/documents [6] and in the IEEE-DR standards http://standards.ieee.org/regauth/its/index.html [5]

Highway Advisory Radio (HAR) messages can also be archived in a similar way to DMS data. Both types of messages are mainly characterized by message text, start time, duration, and location. In Florida, the Turnpike district has this service currently functional along the Turnpike mainline. The service is operated from the district's Pompano TMC. Recently, the Turnpike district started to log the broadcasted HAR messages into a plain text format.

Incident data

Typically, one of the objectives of the traffic management software is to provide incident management capabilities. Some of the traffic management software used in Florida provides incident detection algorithms based on collected real time traffic monitoring data. These algorithms use real time detectors data and do not utilize any archived historical data. Moreover, some of this software, such as the MIST system used by FDOT D-5, archive the incident data and allow for manual input of several incident properties. However, TMCs still depend mostly on manual procedures for incident detection. These procedures range from travelers or road rangers information to using CCTV camera feeds to the traffic operation centers. Incident data are defined in the TMDD and can be found at: http://www.ite.org/tmdd/tmdd2nameanddefinition.html

Closed Circuit Television Camera (CCTV)

Currently TMCs are reluctant to store any CCTV data due to liability issues. Although no plans currently exist to store or archive CCTV data, this type of data plays a major role in assisting the traffic operators. Moreover, CCTV is considered a basic element in recent ATIS applications. Additionally, due to the rapid development in the image processing and computer vision fields, CCTV data may be used in the future to extract more valuable traffic information. Recently, most commercial Database Management Systems (DBMS) software provides the tools for archiving and retrieving CCTV still images.

Road Maintenance and Lane Closure Data

Most of the traffic management software either handles or plans to handle this type of data as part of the road incident data. Some districts, such as FDOT D-7, have started a travel information service covering five interstate roads in five counties in the Tampa Bay area. Other districts keep road maintenance and lane closure records in plain files. Although different techniques are used for collecting and archiving this type of data, lane closure and road maintenance data is considered one of the major data widely available in the state.

2.3 DATA INVESTIGATION SUMMARY AND RECOMMENDATIONS

The information presented in the previous sections revealed the significant amount of ITS data currently collected throughout the state. The discussions also showed the large amount of ITS facilities that are currently in the process of implementation (or planned to be implemented in the near future). ITS data are expanding vertically by continuously implementing new ITS technology and horizontally by increasing deployment coverage of existing technology. This is very clear in the planned implementation of AVI sensors by the OOCEA and the Turnpike district, the deployment of RTMS equipment on I-95, and the expansion of I-4 detectors, CCTV, and DMS deployment.

In summary, many types of ITS data are currently implemented statewide. However, no mechanism is currently deployed to monitor statewide ITS field installations, hardware, software, and data inventory. Furthermore, no common standards are being followed throughout the districts for data archiving procedures. The following, conclusions and recommendations are based on the investigated data requirements:

- Major state corridors are candidates for immediate data warehousing implementation. The I-4 and the I-95 corridors enjoy significant deployment of ITS resources. Additionally, the I-10 corridor has several ITS installations in the Jacksonville area. The FDOT D-3 has deployment plans for ITS equipment in the I-10 and I-75 corridors.

- Investment may be needed to fill out gaps in areas lacking ITS data in order to achieve more complete and useful applications based on archived ITS data. This is very obvious for the OOCEA and Turnpike roads, which need more AVI units to produce accurate travel time information.

- Data archiving and application development efforts have already started in the I-4 corridor. These UCF-CATSS efforts need to be expanded and combined with other statewide efforts to enable the start of the Florida CDW.

- For the next phase of the Florida data warehouse, ITS data collected by FDOT D-5 RTMC, OOCEA, FDOT D-2 Jacksonville TMC, FDOT D-6 Miami TMC, and the Turnpike District can be immediately utilized.

- Fiber optic connections between the CDW (to be established at the UCF-CATSS facilities) and the regional data centers, such as the FDOT RTMC, are necessary for the data warehouse implementation.

- There is an urgent need for statewide traffic management software that takes operations requirements into consideration and also considers all data archiving necessities including archived data formats, strategy, and database. In addition, a mechanism should be initiated to manage and monitor the rapid deployment of statewide ITS resources.

- Data items collected by the traffic management software should follow the ITS data definition standards presented in the Traffic Management Data Dictionary and ITS Data Registry. The ITS equipment deployed in Florida should also follow these standards. It is very important to standardize the data archival process and to facilitate communication between the TMCs and the CDW.

- A statewide GIS layer linked to ITS data and equipment inventory database will be very helpful in showing locations of current and planned ITS resources. It will also clearly show gaps in ITS equipment coverage. The data warehouse can play a major role in this area as it will keep track of deployed ITS resources.

- Data definition in any data warehouse implementation should follow the TMDD and the ITS-DR standards. Developed applications may not employ all the data defined in the data dictionary. Accordingly, selected data fields may be chosen from the whole set of archived data to satisfy the needs of each specific application. For data types that are not defined in the ITS standards yet, efforts should be devoted to analyze the currently collected and planned data to come out with a unified statewide definition for these data items. The efforts should depend on major manufacturers proprietary standards, application needs, and future expectations.

CHAPTER 3.
INVESTIGATING PROCESSING AND APPLICATIONS REQUIREMENTS

In this part of the report, data pre-processing and end-user applications are discussed. Other items related to data collection and applications development are also presented.

3.1 DATA PRE-PROCESSING

Any data warehouse development should be preceded by a pre-processing step in which data are filtered and reformatted. This step is extremely important for transportation data warehouses, where data are collected from different agencies, using different techniques, and mainly through field equipment subject to operational environment. Nevertheless, data collected from different agencies still needs to be fused together in order to facilitate application development. These two major pre-processing steps are discussed in the following sections. Figure 1 shows an outline of freeway traffic monitoring data pre-processing and fusion steps.

3.1.1 Data Filtering

Faulty data should be recognized and removed from the data set before they are entered into the data warehouse. Although, identifying and removing up-normal data may provide a cleaner data set, this process does not guarantee the perfection of the rest of the data. Data may be cleaned based on simple removal of nonsense data such as negative speed values or speeds more than 150 miles/hour. However, any used threshold should be documented and kept as part of the data warehouse metadata.

The development of a data warehouse should significantly improve data filtering techniques by increasing the knowledge based on which decisions are made for erroneous data. Algorithms need to be developed to filter out bad data within the developed end-user applications. The historical data in the data warehouse should allow the establishment of temporal and spatial models to provide the information infrastructure for the filtering algorithms.

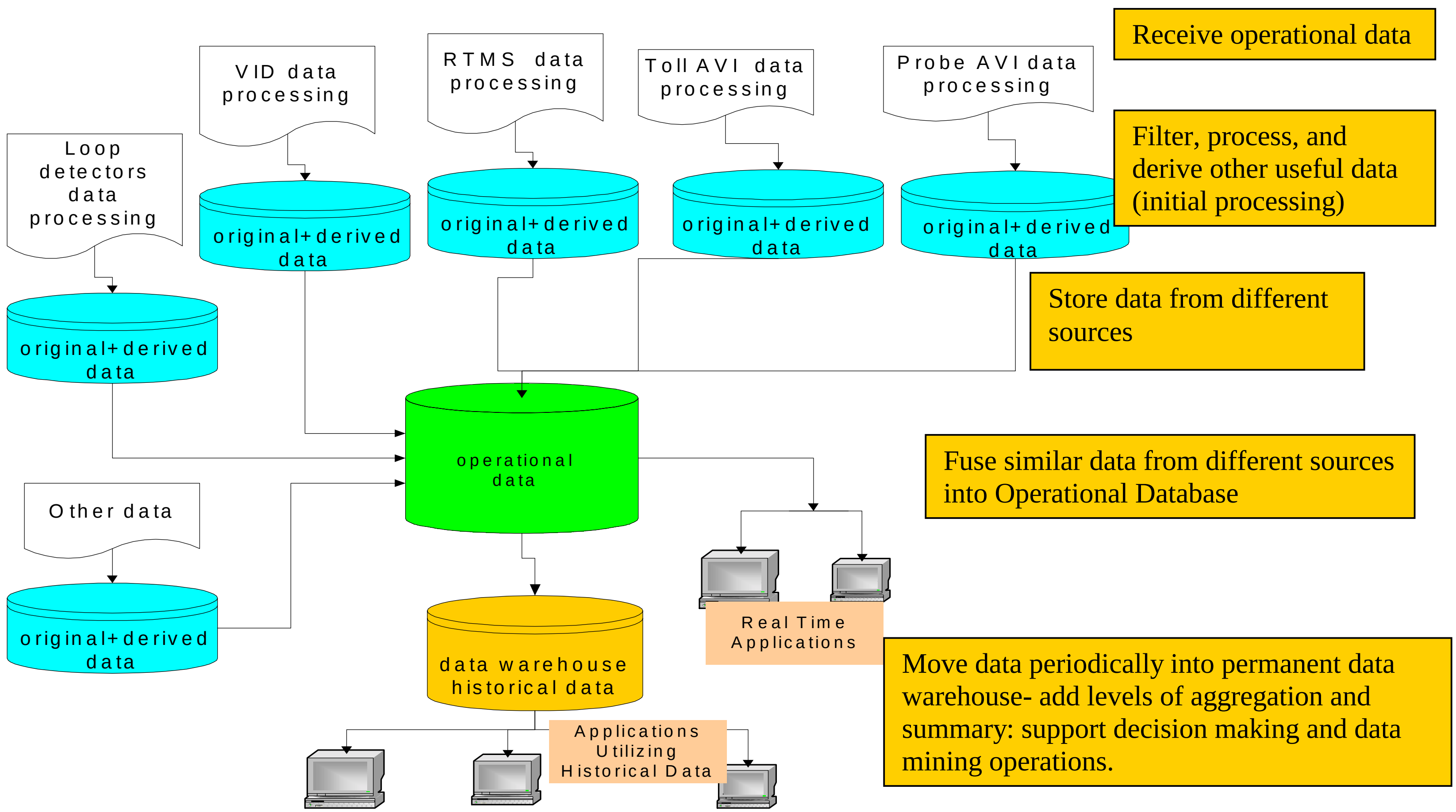

Figure 1. Road-Monitoring Sensors Data Fusion

Each developed application for the data warehouse should have a built-in capability to apply more filtering and cleaning procedures on the data before they can be used in the application.

3.1.2 Data Fusion

Two types of data fusion are needed in the data warehouse implementation. In the first type, similar data acquired from multiple sources should be linked together. For example, detectors data acquired from different TMCs should be combined together as well as with probe data to facilitate developing a regional or statewide traveler information system. Similar data might have different aggregation levels. For example, some detector data may be logged in 15-minute intervals while others are logged in 5-minute intervals. Data from different sources should be brought to the same aggregation level in order for them to be successfully applied in a single application.

In the second type of fusion, different types of data should be fused together to enable queries across all data subjects in the data warehouse. For example, detectors and incident data should be fused together to allow for queries, such as the list of all incidents and road speeds on certain road links during a given time interval. In this case, both detectors and incident data should share the same time and location properties. The database implementation of the data warehouse should facilitate this type of data fusion or data integration. This type of data fusion will be discussed in details in the database design and implementation section of the report.

3.2 INVESTIGATING PROCESSING AND APPLICATION REQUIREMENTS

The 'ITS Information Management' and 'Traffic and Travel Management' user service bundles list many applications that a central transportation data warehouse should provide. Studying these applications before proceeding in the data warehouse implementation is crucial in designing the data warehouse, since any developed database model should take into consideration future applications. While it is difficult to develop all possible applications before the data warehouse is completed, it is possible to come up with a wish list for the future phases. The outcomes of the meetings conducted with different stakeholders, the answers received through the electronic survey, and the review of applications developed nationwide revealed several applications that have been suggested by CATSS team for the implementation phases of the data warehouse. In the following paragraphs, a brief description of these applications is introduced.

3.2.1 Delivery of Advanced Traveler Information Systems (ATIS)

Currently, transportation data utilized by ATIS ranges from simple incident and event reports to travel time/speed data extracted from traffic monitoring sensors. CCTV cameras are used by ATIS through either TV or web broadcasting. The optimal goal here is to produce a statewide real time origin/destination application that depends on real time ITS data collected along corridors and major arterials. This means that many of the available web-based routing application, such as http://www.mapquest.com, should ultimately use dynamic ITS data to provide realistic travel time and routing information.

Other types of data include probe data from transit systems, cellular calls, rangers…etc. Lane closure, road maintenance, and weather data are also common types of data disseminated by ATIS. Travel information data are disseminated via many techniques. Radio broadcasting and web-based travel information services are the most common methods for travel time data dissemination. Other techniques include toll free and 511 phone numbers with Interactive Voice Response (IVR) systems. Some systems use voice recognition techniques to offer the user automatic navigation through several types of services. Wireless and kiosk devices, text messages, email alerts and in-vehicle information are also common ATIS dissemination techniques.

Travel information data are mostly provided by the public sector. This includes travel time data, incident and weather data…etc. A study published by US Department of Transportation (DOT) and PBS&J showed that different business models were tested to generate revenue from ATIS services [7]. The study showed that none of these revenue sources so far has the ability to sustain a profitable ATIS business and there is a need for the public sector to fund these efforts at least in the early implementation stages.

In Florida, two cases of ATIS deployments currently exist. These cases are part of efforts conducted by FDOT to develop a statewide ATIS system. A system that not only provides different types of real time information for the users but also extends the services to provide long-term and short-term predictive information and routing capabilities. In the following section, two sample ATIS efforts conducted in the state of Florida are introduced.

The Southeast Florida Metropolitan Area:

In May 2001, the southeast Florida ATIS was developed as part of the SunGuide program http://www.sunguide.org, managed by FDOT D-6. The SunGuide program is a cooperative program that involves agencies from Broward, Miami-Dade, and Palm Beach counties. One of the objectives of this program is to provide real time travel information services for south Florida travelers. Users can get the real time travel information through the www.smartraveler.com website or through the SmartTraveler assigned phone numbers. As part of the SmartTraveler service, provided by SmartRoute Systems, several CCTV cameras were installed on major highways and intersections within District–6-. Four of these cameras are installed on I-95 and I-595. One of the cameras is installed on a helicopter to broadcast aerial traffic conditions footage. Figure 2 shows a snapshot of the SmartTraveler website for Broward county area.

The I-4 Data Warehouse:

The UCF-CATSS is currently hosting inductive loops and incident data for about 39 miles of the I-4 corridor. Although, the MIST traffic management software used by the RTMC produces 15-minute logs for the loop data, UCF is receiving the loops data using a file parsed by the MIST software every 30 seconds. A dedicated T-1 connection was recently established between UCF-CATSS and FDOT D-5 RTMC to facilitate this operation. Before the T-1 connection was established, a less reliable point-to-point dialup connection was used. UCF hosts the data in SQL 2000 server Database Management System (DBMS) software. This enabled the development of 24/7 real time origin/destination travel time web-based application. In this application, the user can choose his/her origin and destination ramps on I-4 and receives real time, predictive, or historical travel time between the two selected ramps or points on the mainline [8]. The prediction part of the model allows the user to get the travel time information up to 30 minutes in the future. Moreover, real time and historical speed, occupancy, and volume information are also disseminated for each station, with user defined time aggregation level. In addition, the developed applications can provide the user with the loop detectors' health status or the percentage of time a given sensor is producing data.

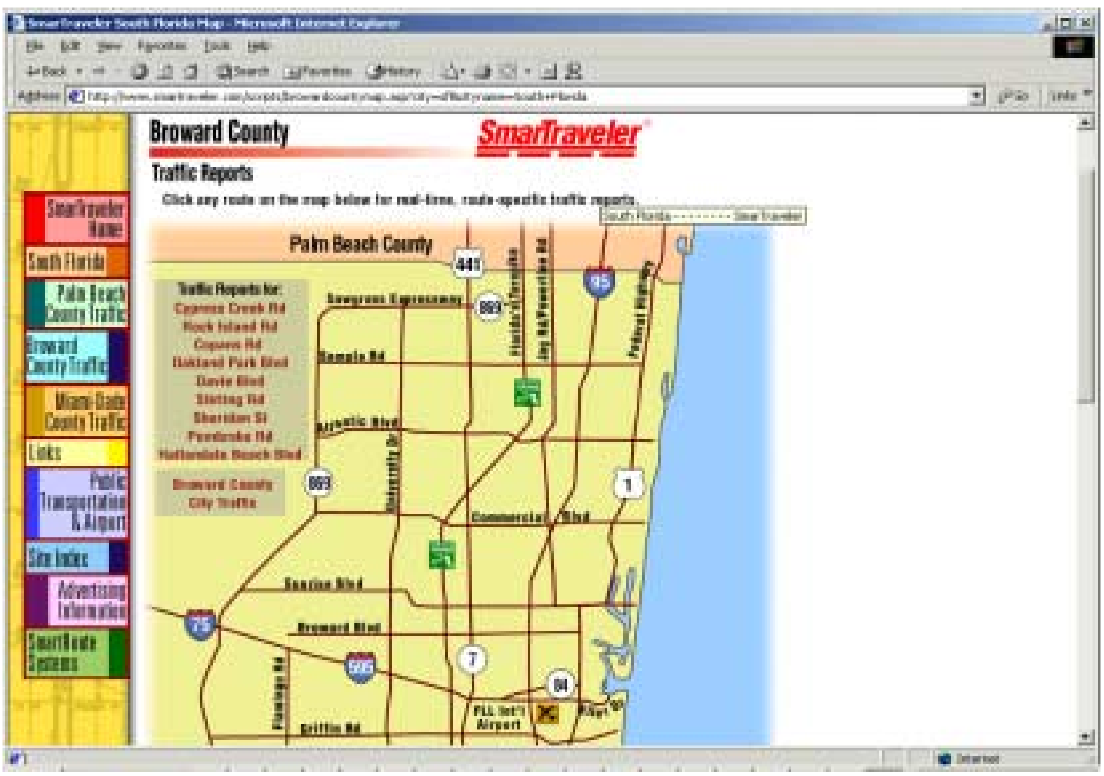

Figure 2. Snapshot of the SunGuide ATIS Website

3.2.2 Predictive and Historical Traveler Information Systems

Although predictive travel time information can be considered a part of ATIS, it can still represent a stand-alone application. As more travelers plan their trips ahead of time, the need for short-term and long-term travel time prediction increases. Data stored within the data warehouse can provide the historical heritage on which a predictive model can be built.

The FDOT D-5 in cooperation with the UCF-CATSS was able to provide a short-term prediction model that provides estimates for travel time information up to 30 minutes in the future. The model was extensively tested and gave excellent results. Figure 3 shows a snapshot of the web page used to get predictive travel time along the 39-mile section of I-4.

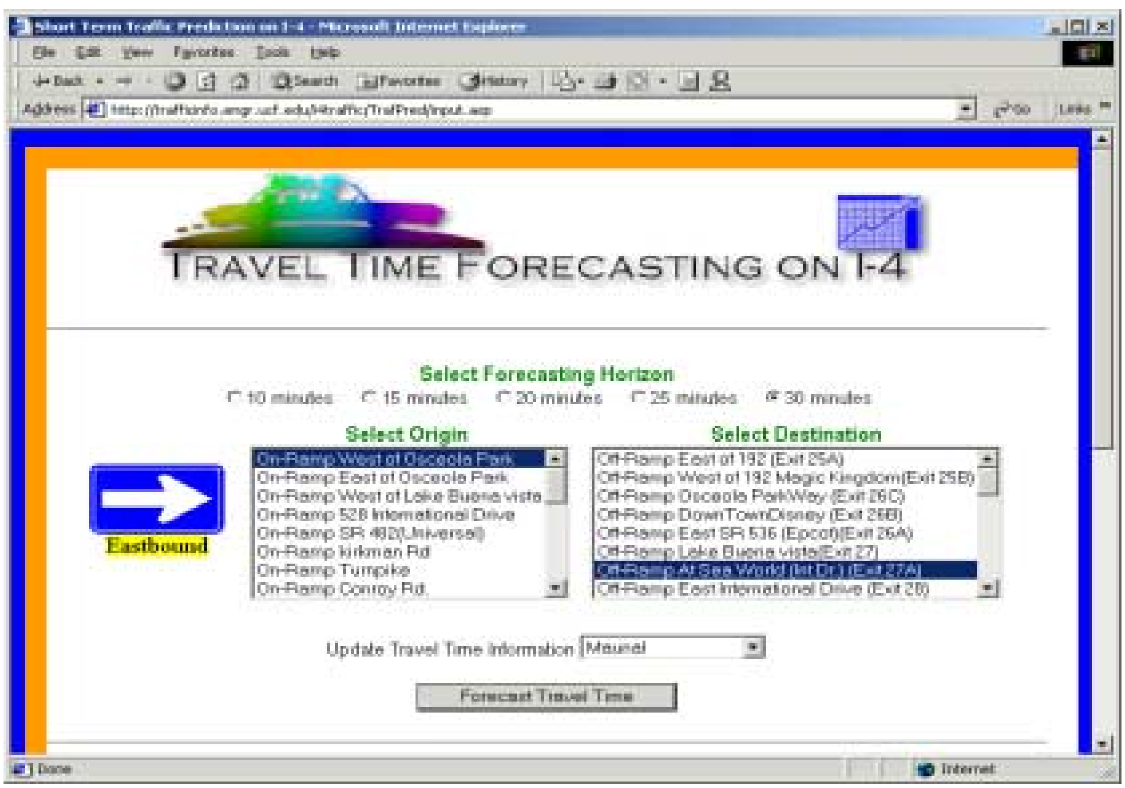

Figure 3. Snapshot of the I-4 Predictive Travel Time Application.

More efforts are currently underway at CATSS to develop a long-term predictive model that utilizes the historical I-4 data archive. The long-term model may be used for both travel time prediction and incident detection applications. Figure 4 shows a plot of speed data at one of the loop detector stations on I-4 using a two-hour time duration for all Tuesdays in August 1998.

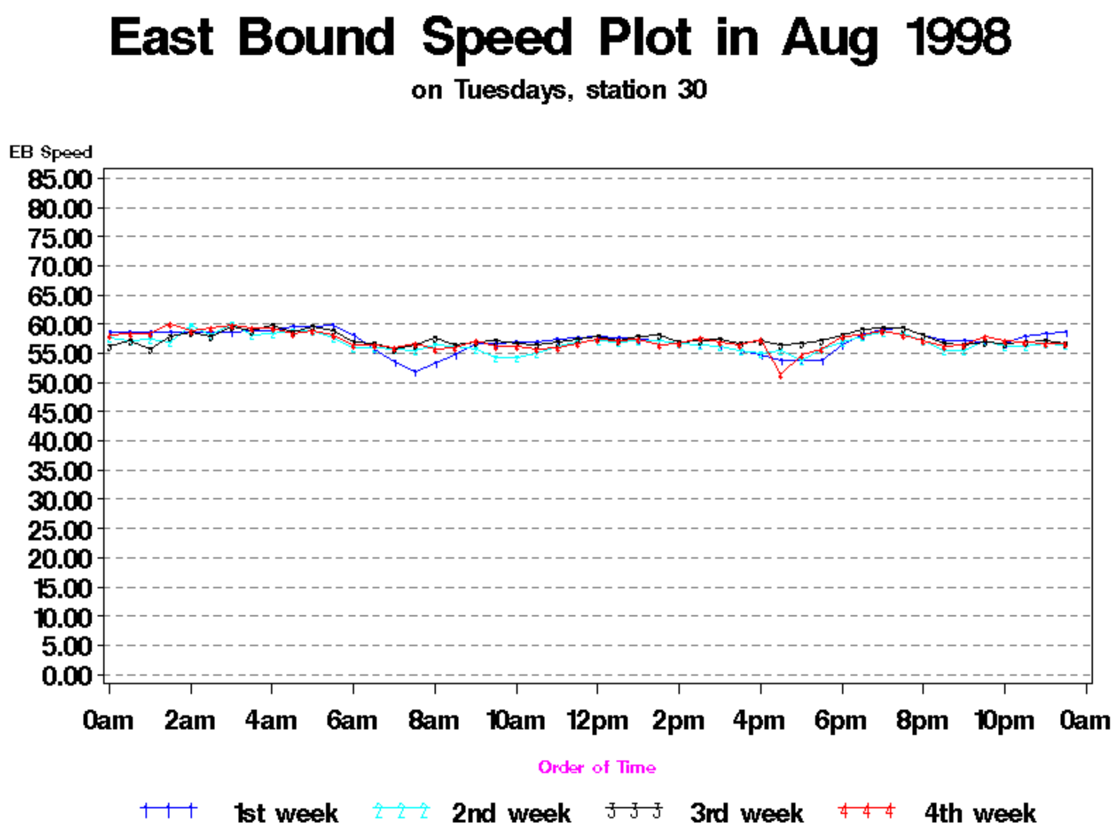

Figure 4. Speed Pattern for an I-4 Loop Detectors Station

3.2.3 Incident Detection Applications

Many of the traffic management software currently deployed at the state's major TMCs have the capability to flag out potential incidents based on real time detectors data. One of the Problems with currently used incident models is the need to fine-tune the algorithm parameters in order to avoid false alarm or late incident detection. However, these models use only real time traffic monitoring data in their incident detection algorithms.

Therefore, new models need to be developed to use the data warehouse historical data in addition to the real time traffic monitoring data. In these models, typical traffic patterns could be identified based on the available historical data. Severe deviations from these patterns may raise the alarm for potential incidents. Both spatial and temporal data distribution should be utilized in these models. Historical incidents data are necessary in developing these models as they provide the model verification means. They can also be included in the model with certain probability to simulate what actually happens in reality.

3.2.4 Using Historical Data for Transportation Planning Applications

Historical data archived within the data warehouse provide a treasure of information for transportation planners. One of the goals of any data warehousing projects is to support planners and decision makers' applications. A data warehouse should not only provide the user with an image of the archived data but also facilitate the tools to query, retrieve, display, and distribute these data. With the recent evolution in the data-warehousing field, many tools have been developed to facilitate dynamic querying of the data.

The data warehouse application development tools can be efficiently used in the transportation field to assist applications ranging from ITS equipment performance evaluation to long-term transportation planning and decision-making processes. For example, the decision maker can select specific equipment type for future deployment based on its historical performance information provided by the data warehouse. Figure 5 shows a snapshot of an application developed by the UCF-CATSS to show the historical health status report of the I-4 loop detectors.

3.2.5 Data Mining and Knowledge Discovery Applications

The purpose of data mining applications is to extract implicit and potentially useful information from the data. Generally, data analysis and data mining applications have direct impact on the decision making and planning applications. Different data mining techniques should be used to explore the historical data in the data warehouse. For example, long-term traffic patterns can be identified using data mining applications. Factors concerning traffic safety, highway performance, and effect of ITS technology deployment can be extracted from stored historical data using data mining algorithms. Normally, data mining applications employ machine learning, statistical, and visualization techniques.

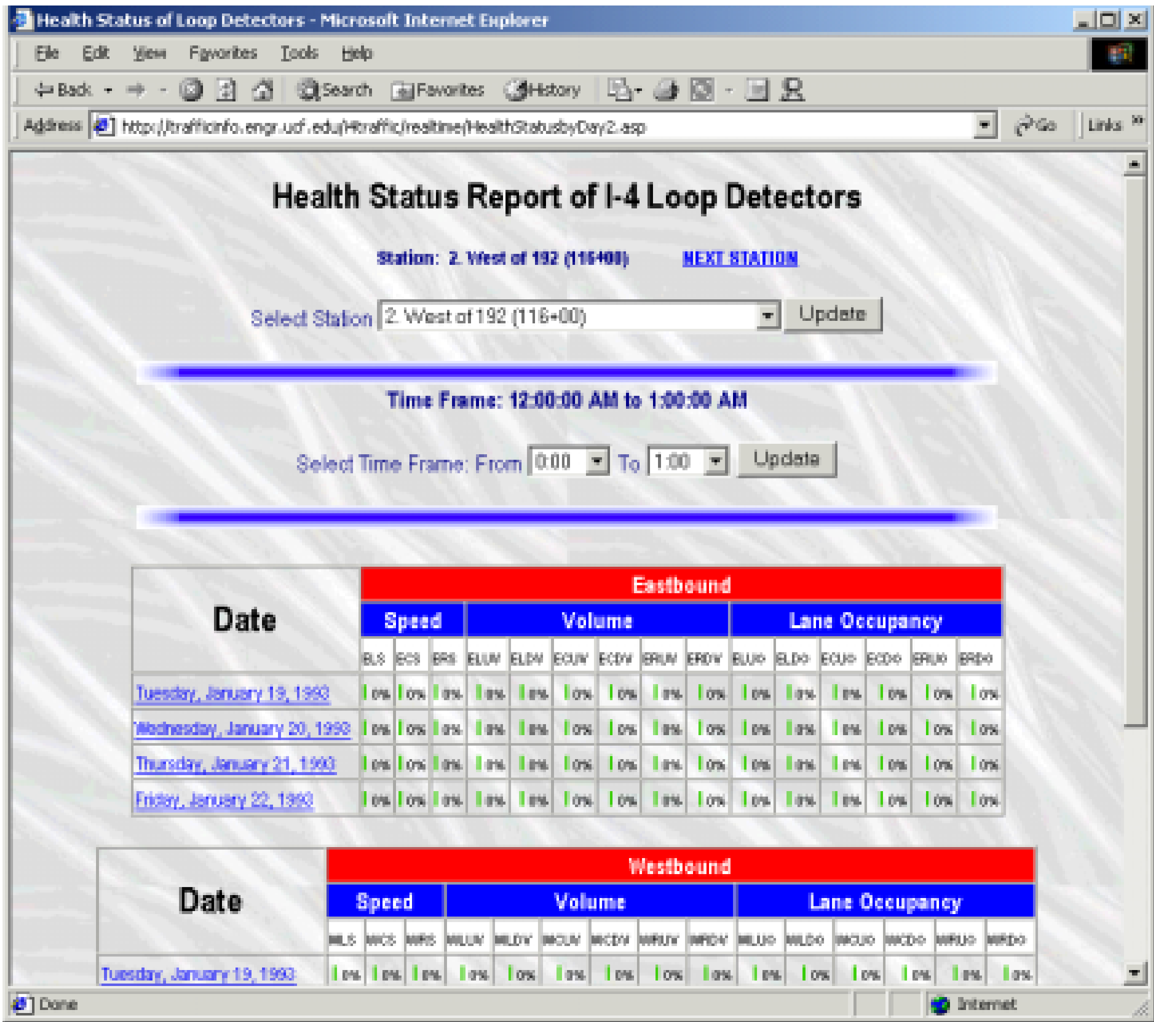

Figure 5. Snapshot of the I-4 Loop Detectors Health Status Application

3.2.6 Bulk Data Distribution Applications

In the meetings conducted by the UCF research team with the state stakeholders, explicit requests to design applications for data distribution were made. The development of such applications is important for planners and researchers. Additionally, these applications are needed in special events and emergency management situations. For example, many stakeholders complained about the lack of fast and reliable historical information required by federal government and/or the media in case of hurricane evacuation. This type of information should be provided by the developed applications and be handy for decision-makers.

Applications should be designed to allow the users to identify and select their data needs from the database. In the mean time, mechanisms should also be created to allow efficient delivery of the data. One of the objectives of these mechanisms is to prevent clogging the network with large data volumes. Off-line data distribution through CD ROMs or tapes still represents the best solution for bulk data distribution. Fortunately, most commercial databases now support bulk data distribution by creating snapshots of the database at specific moments of time, which in turn can be distributed through data distribution media. It is essential to mention here that a well-maintained metadata for the database elements is the basis for any successful implementation of bulk data distribution application.

3.3 APPLICATIONS DEVELOPMENT STRATEGIES

Several types of data warehouse pre-processing and end-user applications were discussed in the previous sections. Pre-processing applications normally work directly on the data and do not need fancy Graphical User Interface (GUI). The major requirements needed for these types of applications are to efficiently perform the needed tasks and to comply with the general coding standards. Programming languages used in this case should be one of the high-level programming languages such as C++ or Java. It is recommended here to use Java due to its cross-platform implementation capabilities.

On the other hand, end-user application requirements extend from efficiently performing the needed task to provide an easy to use and familiar GUI that is able to work on different platforms with minimum installation needs. In this context, web-based application development is currently gaining momentum in the data warehousing environment. With these applications, the user does not need any specific application installed on his/her machine other than his/her familiar web browser. Functional applications are provided to the user within the web-browser itself. This will significantly reduce the needed training cost and raise the application portability.

Applications of this kind depend on multi-tier software development architecture. Normally, the first tier, namely the display or interface tier is an application that works within the user web browser [9]. The last tier is the data stored in the data warehouse database. Generally, applications that need complex SQL queries and demand analyzing of large query results need intermediate tier(s) called analytical or logical tier(s) to facilitate data querying, processing, and formatting between the first and last tiers. In this intermediate tier, many high-level languages such as C++, or Java can be used to perform intermediate complex analytical processing tasks on the data apart from the presentation interface and the data itself. Multi-tier software development architecture will provide better performance, flexibility, and scalability to the data warehouse project. Figure 6 represents a diagrammatic representation of three-tier software architecture.

Another feature that must be provided by the data warehouse applications is the On-Line Application Processing Tools (OLAP). Using these tools, the user can dynamically query the data warehouse database. This will significantly assist planners and decision makers by providing on-the-fly reporting capabilities. There is no need to write static applications that only perform certain queries and output certain reports and charts. A data warehouse that applies this technology has no limits for information extraction other than what the user thinks of.

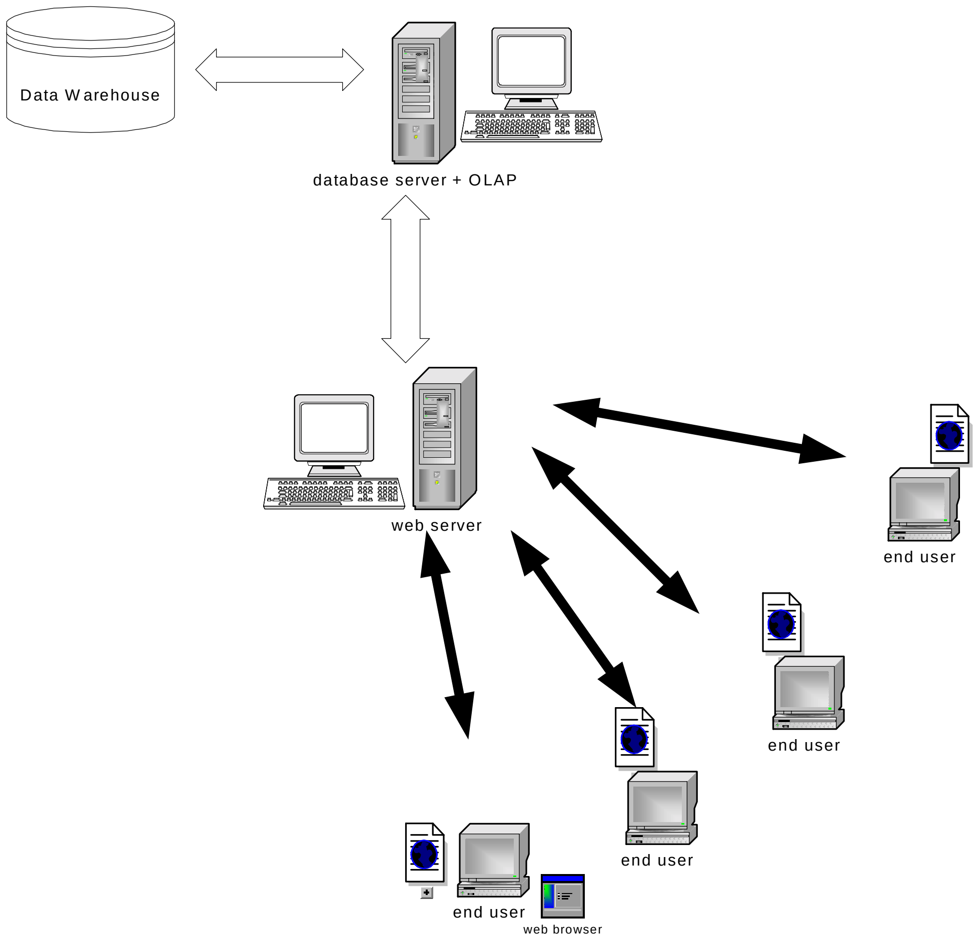

Figure 6. Three-Tiers Application Development.

3.4 CENTER-TO-CENTER (C2C) COMMUNICATIONS

In the data warehouse implementation, most of the communications between the data warehouse and the data sources are two-way communications. Any center that supplies the data warehouse with data expects, in return, to acquire information from the data warehouse. Common communications and data definitions are needed to standardize this process. The American Association of State Highway and Transportation Officials (AASHTO), the Institute of Transportation Engineers (ITE), and the National Electrical Manufacturers Association (NEMA) are now jointly cooperating to develop the NTCIP standards. The main objective of these standards is to establish the communication rules (protocols) and the vocabulary (objects) that allow field equipment and centers to communicate with each other as a system. Two C2C communication standards are being defined by NTCIP. The first standard (NTCIP 2305) is called Common Object Request Broker Architecture (CORBA) and the second (NTCIP 2304) is called Data Exchange for Abstract Syntax Notation 1 (DATEX-ASN1).

In the Florida CDW, the majority of the data are collected from the TMCs and logged in commercial databases. A full center-to-center implementation will enable effective transfer of data and messages between the data warehouse and the centers. This is very essential to optimize the performance of the data warehouse and the developed applications. For example, an incident detection application in the data warehouse should be able to communicate directly with the TMC to alarm for potential incident. Also, the data warehouse should be able to provide actual real time delay and incident messages to the DMS through the TMC.

Unfortunately, only a few of the current traffic management software deployed in the state support DATEX or CORBA implementation. Among the investigated traffic management software in Florida TMCs, it was found that only the MIST system deployed at FDOT D-5 RTMC (the largest TMC in Florida) supports DATEX standard. The SunNAV system deployed at the Turnpike District is the only traffic management software that is CORBA based. In the following, a brief overview of both CORBA and DAETEX standards is introduced.

3.4.1 DATEX-ASN

DATEX-ASN is a network protocol of standards for exchanging encoded ASN data. The DATEX-ASN standards are intended to provide interoperability for communications between management subsystems within ITS community. DATEX is currently developed by the NTCIP C2C working group in cooperation with the International Organization for Standards (ISO). The DATEX standards are procedural and do not directly support object oriented programming. The standards assume small bandwidth, and hence it is suitable for communications among field devices and between field devices and traffic management centers. Real time data transfer is currently supported by DATEX-ASN standards. The steps needed to implement the DATEX-ASN standards can be found at: http://www.ntcip.org/library/documents/pdf/ap-datex_980102_w2.pdf.

3.4.2 CORBA

CORBA is an object-oriented standard designed especially for distributed systems. One of the major advantages of using CORBA as object-oriented standard is its reusability and complexity of abstraction properties. These properties are supported by the object-oriented inheritance and encapsulation characteristics. A CORBA implementation can be used to integrate systems whether or not there is a database in the systems.

A CORBA implementation is enabled through several software components. The Object Request Broker (ORB) is the component that enables objects to make requests and receive responses independently of the used platform or operating system. Many ORBs are currently available commercially or as freeware. CORBA services supply the support functions for basic operations of distributed network. The CORBA naming service, notification service, group service, and synchronization service are the basic services supplied by most available ORBs. The naming service is a key service as it provides a distributed name, which references created objects. Many other important services, such as the query service, which supports database querying, and the persistent state service, that allows objects to maintain their object state, are currently available by many commercial ORBs.

The CORBA standards are still immature. Additionally, no estimate for CORBA maturity currently exists. Most of C2C communications currently developed using the CORBA technology are still in the experimental stages. The NTCIP C2C working group is currently developing the base transportation object models for the CORBA standards. First priority was given to the Advanced Traffic Management System (ATMS) object models. These models will include primitive classes such as location, identification, and date/time; base classes such as center, event, device, vehicle, and person; and traffic management classes such as dynamic message signs, traffic signals, traffic congestion, and events/incidents. More information about the ITS C2C CORBA standards and developed base object models can be found at http://www.ntcip.org/restrict/ [10] by clicking at the center-to-center profile hyperlink.

3.5 PROCESSING AND APPLICATION REQUIREMENTS SUMMARY AND CONCLUSIONS

Initial data processing and suggested end-user applications for the data warehouse project were discussed in the previous sections. The following recommendations can be made:

- Original data received from different sources should be stored independently after initial data filtering. These data should keep their original spatial and temporal resolution. Any database implementation should keep track of the original data source and data quality as part of the data warehouse database.

- Tests should be performed independently and periodically, as part of maintenance plans, on the field equipment to guarantee the quality of the data. The results of these tests should be included in the data warehouse metadata. In addition, the data warehouse implementation should keep track of the data quality fields provided by many of the field equipments as a measure for the data and equipments quality.

- A decision on common data aggregation levels should be taken. All data of the same type should be brought to this aggregation level to facilitate data fusion. Courser granularity levels may then be developed. Fused data with different granularity levels are part of the data warehouse database and should be separated from original individual data.

- Suggested applications for the data warehouse implementation includes:
  1. Establishment of regional and statewide ATIS.
  2. Development of incident detection algorithms based on real time and historical traffic monitoring data.
  3. Development of data mining and knowledge extraction applications.
  4. Facilitating bulk data distribution applications

- With the status of the data currently collected (or planned to be collected) in the state of Florida, any developed travel information service should at minimum, provide the following information through the data warehouse implementation:
  1. Real time speed and/or travel time information for the road links included in the ATIS corridor/network.
  2. Real time incident data and information. These data maybe be collected through incident detection algorithms, CCTV monitors, road rangers, cell phone calls…etc.
  3. Lane closure information. Closure data for road maintenance, special events, and incidents that might influence current or future traffic performance should be provided.
  4. Detailed GIS map for the implementation area. The GIS map should include major road networks and ITS installations. The GIS library should also include hyperlinks for incidents, major events, road maintenance, DMS, and CCTV data.
  5. A website should be developed in the early implementation stages of the data warehouse to examine the collected data and the ability to provide the

appropriate information to the user. Once the information is collected in the data warehouse and tested through a test website, ATIS service provider(s) may be contracted to add more information dissemination tools, such as phone services, radio advisory, email alerts, wireless technology…etc.

- Application development should consider multi-tier application architectures, in addition to typical client/server architecture. Multi-tier applications with web-based interface offer familiarity, portability, and reduction in training costs.

- OLAP applications should be developed in the Florida data warehouse to support planning and decision-making. The developed applications should run in an intranet type of environment. This will be facilitated by the Florida fiber optic network.

- Security measures should be considered by assigning the appropriate permission privileges to the users. This can be applied at the network, application and/or database levels.

CHAPTER 4.
DEVELOPMENT OF DATABASE MODEL AND PROPOSED CONFIGURATION

The purpose of this section is to develop a database model and implementation strategy for the data warehouse. The database model and implementation strategy will consider the data definitions and application requirements presented in the previous sections. In addition, software and hardware configurations are suggested in this section.

Although several states started to develop their transportation data warehouses few years ago, only a few of these developments took advantage of the recent advances in database and data warehousing technology. Due to the uniqueness of the transportation data warehouse, many database related issues need to be discussed thoroughly before any guidelines for the database development are established. In the following sections, the basic issues and guidelines for the data warehouse database development are discussed.

4.1 TRANSACTIONS PROCESSING VERSUS DATA WAREHOUSING

One of the basic steps in any database design project is to establish the conceptual schema of the database, using a high-level data model. Two types of data processing are currently used in database applications namely transaction processing and data warehousing. Transaction processing databases are generally highly normalized. This means that the data has to be split in many tables with very few or no redundancy. This suits the objectives of the transactions processing environments, where data are changing dynamically, and simple database queries are expected. In the data warehousing projects, databases are mostly populated with historical data. The sole purpose here is to query the data and extract meaningful information. A highly normalized database needs significant processing overhead in order to accomplish the complex queries expected in the data warehousing applications.

Transaction processing and data warehousing environments implement their own database models that effectively support their specific expected functionality. These models are called the Entity-Relationship (ER) model and the Multi-Dimensional model, respectively. Due to the role each of the ER and Multi-Dimensional models is playing in the database design, separate sections are assigned to explain them individually. Table 4 lists the main differences between transactions processing and data warehousing environments.

Table 5. Summary of Differences between Transaction Processing and Data Warehousing Environments.

Transaction Processing	Data Warehousing
ER schema is mostly used	Multi-dimensional schema
Volatile data. Data are updated dynamically. This includes data insertion and deletion.	Non-volatile data. Data are historical, virtually no data change once the data is in the data warehouse
Client/server applications with pre-defined processes	Supports On-Line Analytical Processing (OLAP) applications-Dynamic Querying
Database is highly normalized	Less normalized database. Data are aggregated with different granularity levels.

4.1.1 Entity-Relationship Model

Over the years, the ER high-level conceptual database model was used for most relational database designs. In this model, detailed descriptions of the database entity types, attributes, relationships, and constraints are expressed. The main objective of the highly-normalized ER model is to eliminate data redundancy by creating more tables. The normalization process involves analyzing the database schema and modifying it by further grouping the attributes until they satisfy certain functional properties.

In the data warehouse database design, the ER model will play limited role. However, the ER model plays a major role in the original database sources. Data stored by the traffic management software, transit systems, and emergency management systems are usually stored in databases that utilized ER models.

4.1.2 Multi-Dimensional Model:

Recently, with the development of the data warehousing technology, another type of schemas evolved to support the functionality of the data warehouse. In a data warehouse, where data update is rare and many query variations are expected, multi-dimensional models are predominantly used [11]. In this model, data are arranged in fact and dimension tables. The phenomena being described, e.g. detectors speed, occupancy, and volume data, are stored in the fact tables, while other properties such as time, location, and source are organized in dimension tables. The primary key of the fact table is a combination of the primary keys of the dimension tables. Database normalization is less restricted in the multi-dimensional model to give the advantage of fast data access. Variations of the multi-dimensional schemas are used to describe more complex data structures and to support different data granularities. Figures 7,8, and 9 show sample multi-dimensional models for detectors, DMS, and incident data, respectively.

Figure 7 shows suggested 4 dimensions for the detectors data. Different types of detector data such as loop detectors, VID, and RTMS are fused together in this model. The volume, speed, and occupancy data are stored in a fact table. Typically, the fact table is expected to continue to grow as new data are received. Each record in the fact table is linked to each of the DATE, TIME, LOCATION, and SOURCE dimensions. In this context, a record in the fact table, i.e. volume, speed, occupancy, and detector health status data, is defined by certain, date, time, location and source. The same concept is also shown in Figure 8 for the DMS data, and in Figure 9 for incident data. The DATE, TIME, and LOCATION dimensions should be shared by all types of transportation data. This will give the power of linking all data subjects together and allow cross-subject queries and analysis, which is considered one of the major characteristics of a data warehouse.

More detailed discussion on common dimensions in the multi-dimensional database model will be given in the following section. More dimensions may be needed to add more descriptions to the data. In addition, modifications to the shown fact tables and dimensions may be needed in the implementation stage.

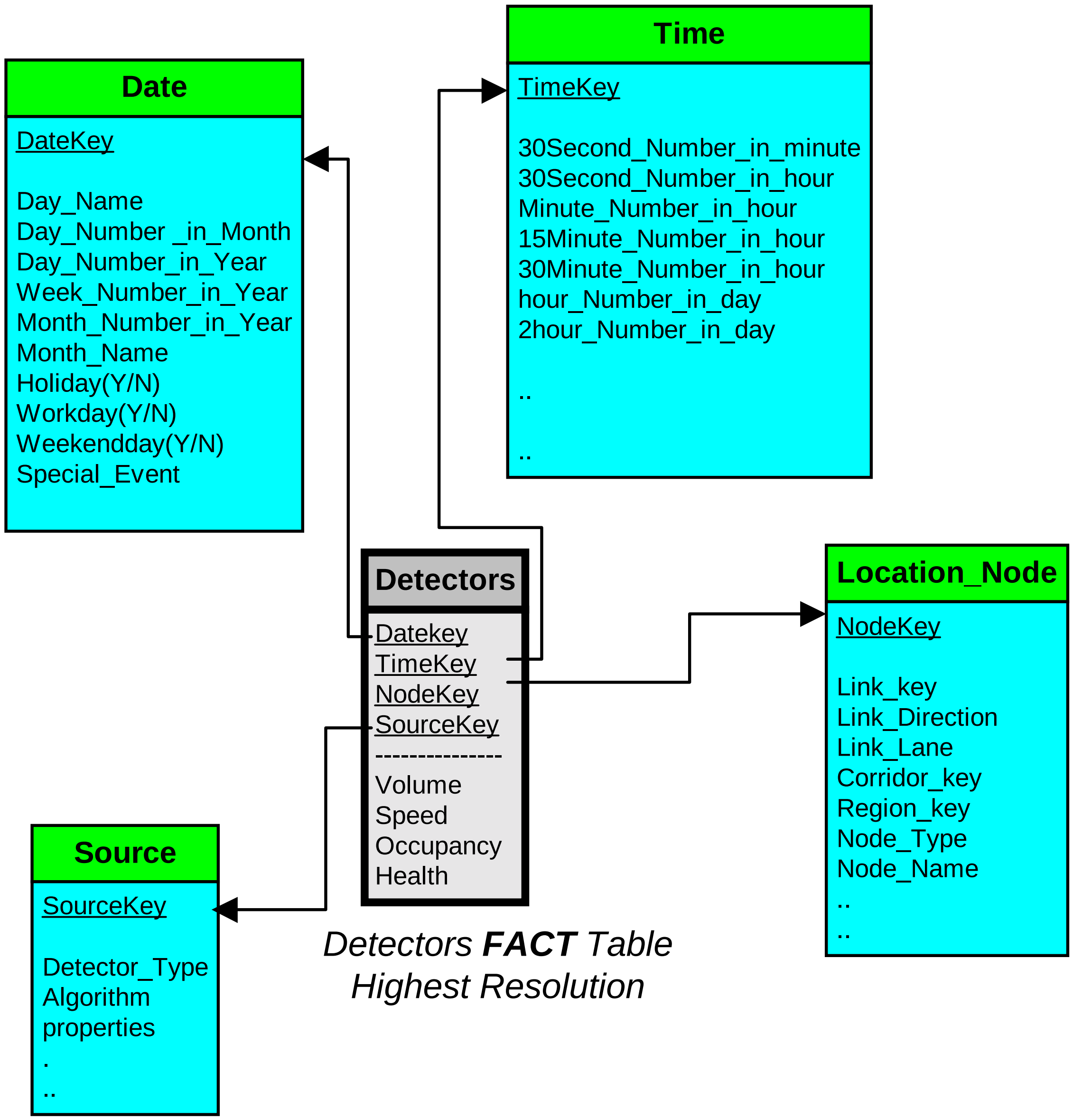

Figure 7. Sample of Multi-Dimensional Implementation for Detectors Data

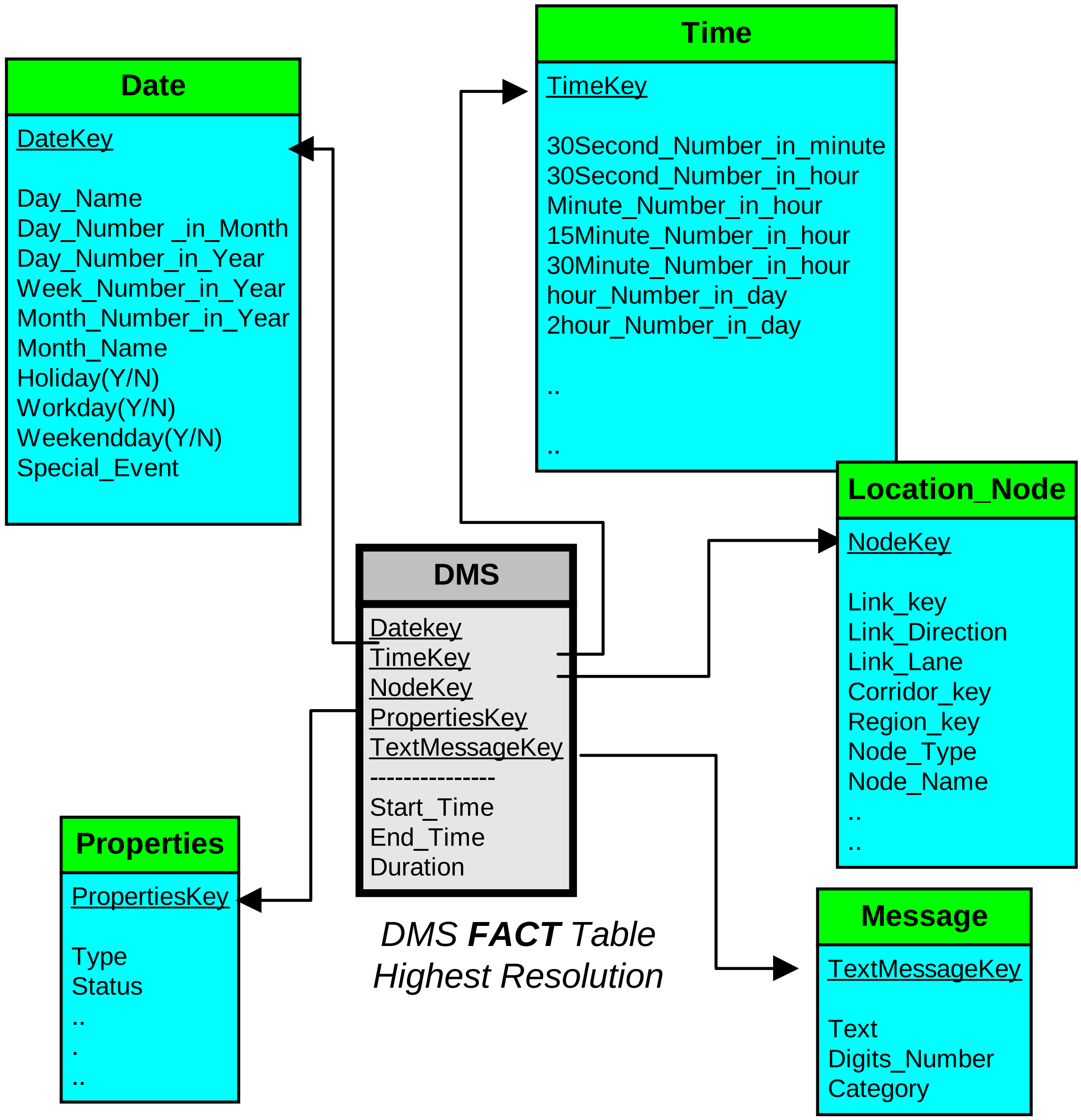

Figure 8. Sample of Multi-Dimensional Implementation for DMS Data

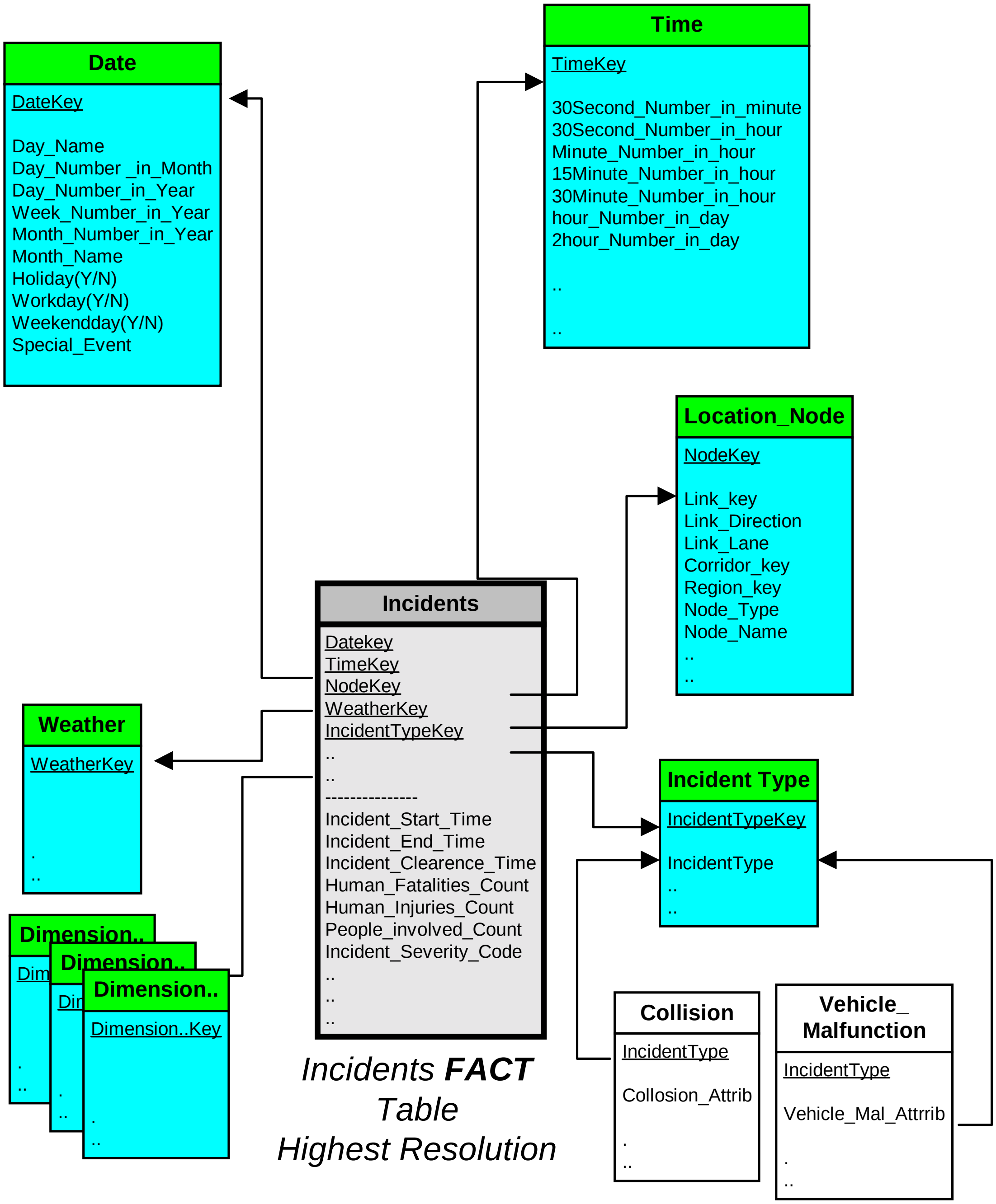

Figure 9. Sample of Multi-Dimensional Implementation for Incident Data

4.1.2.1 Common Dimensions in the Multi-Dimensional Model:

In the Multi-Dimensional model, different data types or subjects are connected to each other using common dimensions. Common dimensions such as DATE, TIME, and LOCATION facilitate across subject queries. For example, a query that tabulates accidents casualties at certain road section within a specified time period against corresponding average section speed needs common DATE, TIME, and LOCATION dimensions for both speed and incident data types. Both speed data and incident data are represented in fact tables, while the date, time, and location information corresponding to each row in the fact table are linked to common date, time, and location dimension. A phased implementation of the data warehouse can be realized using common dimensions that link old and new data subjects and facilitate queries across all these subjects. More examples of common dimensions in a transportation data warehouse would be PASSENGER and VEHICLE. In the following the TIME, DATE and LOCATION dimensions are discussed.

<u>The TIME dimension:</u>

The DATE and TIME dimensions are the basic dimensions in any data warehouse implementation. The data warehouse, as a tool to archive data with time, can only be realized if theses dimensions are shared by all of the data warehouse subjects (fact tables). Each row stored in the fact tables should have a time stamp. This time stamp links data from different sources together. Basically, the TIME and DATE dimensions could be joined in one dimension. However, in the case of the Florida CDW, where details go down to the one-minute resolution or less, it is preferably to split them into two separate dimensions.

<u>The LOCATION Dimension:</u>

Another very important dimension in the transportation data warehouse implementation is the LOCATION dimension. Since most transportation data have location property, it becomes very important to share this property between all transportation data subjects. Spatial information such as road network sections and nodes should be provided as common dimension(s) to all fact tables involved in this phase of the data warehouse. This will enable querying all the data in the data warehouse for location-specific information. The LOCATION dimension in the CDW should be provided through Geographic Information System (GIS). The significant role of GIS in the Florida data warehouse will be discussed later in a separate section.

4.1.2.2 Granularity in the Multi-Dimensional Model:

One of the basic concepts of any data warehouse is its ability to aggregate the data into groups. This aggregation allows the user to drill down into details and roll up into the aggregated groups. Although this technique can add redundancy to the database, the expected benefits in facilitating complex queries and reducing the computations time for certain type of applications far exceeds the redundancy drawbacks.

In the Multi-Dimensional model, a fact table should be produced for each aggregation level. For example, a fact table containing average speed, occupancy, and volume data is produced for each time aggregation (e g, 15 minutes, 30 minutes, one hour…etc). The same concept should also be applied for the location dimension. In this case, a fact table should be produced for links, corridors, and regions aggregations. Aggregated data should be represented in the finest details in the dimension table. This will add links between the high-resolution and the aggregated data.

Table 5 shows the different granularity levels suggested for the TIME and LOCATION dimensions. Using these aggregation levels, the user can easily get a list or chart showing the average speed for the period from 5:00 pm to 6:00 pm for Mondays along the year for certain corridor. No complicated SQL statements are needed in this case. Only simple SQL statement applied through OLAP application that supports pivot table analysis can be used. Using this type of modeling and applications, the user can drill down into the data to show averages for 15-minute intervals and roll up to see averages computed over two-hour intervals.

It should be mentioned here that each aggregation level could be desirable only in certain types of applications. For example, the travel information system might utilize a 5-minute aggregated traffic data to provide near real time travel information. On the other hand, a data mining application that studies average speeds over a link will consider speed perturbations within 5-minute intervals as noise and will be more interested in hourly or bi-hourly averages. One of the data warehouse desirable attributes is to provide appropriate aggregation levels for the developed applications.

Table 6. Suggested Granularity Levels for the TIME and LOCATION Dimensions

TIME	LOCATION
5 minutes	Link
15 minutes	Corridor
30 minutes	Region
1 hour	
2 hours	

4.2 GEOGRAPHIC INFORMATION SYSTEMS AND SPATIAL QUERIES

The role of the GIS in the data warehouse should pass the typical use as a presentation and display tool to actually provide significant part of the data warehouse infrastructure and functionality. The LOCATION dimension, common to almost all data subjects in the data warehouse, should be totally integrated with GIS. GIS by definition have the capabilities to describe spatial relationships such as adjacency and connectivity. These spatial relationships are considered difficult to model and describe using only traditional commercial relational DBMS. Integrating the GIS capabilities within the CDW enables complex spatial queries to be applied to the data warehouse. Furthermore, having a GIS that includes not only transportation features but also other geographically-related data will have a great impact on emergency management situations such us fire and flood evacuations.

GIS is also extremely important to ensure the integrity of the spatial features in the database. For examples, GIS have highly optimized algorithms to recognize an incident location as existing node already stored in the database. Applying this functionality will ensure that this location is automatically linked to the existing nodes tables in the database. Figure 10 shows the effect of such functionality on database integrity and functionality.

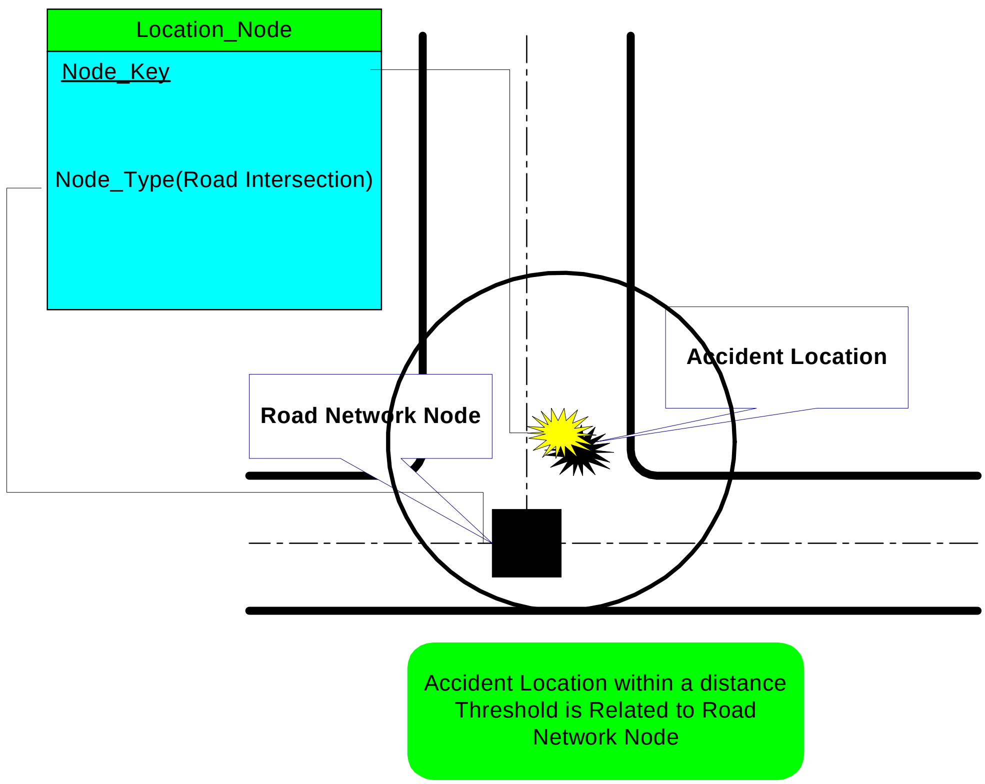

Figure 10. Automatic Linking of Accident Location to Roads' Intersection Nodes Using GIS Capabilities

4.3 CENTRAL VERSUS DISTRIBUTED DATABASE IMPLEMENTATION

While early database implementations moved towards centralized databases and ended up with huge local databases, the recent advances in network and data communications are encouraging for practitioners in the distributed database management field. A Distributed Database (DDB) can be defined as a collection of multiple logically interrelated databases distributed over a computer network [12]. The software system that manages a distributed database while making the distribution transparent to the user is called a distributed database management system.

Database distribution leads to increased complexity in the system design and implementation. On the other hand, distribution can improve system reliability and performance, and facilitate expansion. Generally, for a data warehouse implementation, a centralized database management system remains the most common development architecture. In reality, the distributed data warehousing technology is still in its early stages. In addition, the amount and diversity of received data and the excessive steps needed to filter and reformat this data support the use of centralized data warehouse.

Currently, in the state Florida, no common traffic management software is used. Moreover, the used software does not follow a well-defined archiving strategy. All this combined with the additional labor, training, and resources cost expected when applying distributed data warehouse encourages the implementation of centralized data warehouse at least in the first implementation phases of the Florida CDW.

In the mean time, the efforts to develop a statewide operations concept that adopt common traffic management software should continue. One of the objectives of these efforts should be to develop common strategies for data archival and maintenance. Data should be cleaned, formatted, archived, and managed by a database management system. Furthermore, metadata should be developed and archived within the system. Once this system is working statewide, a distributed data warehouse that makes use of all the distributed data and resources could be developed.

4.4 HETEROGENEOUS DATABASE REPLICATION

The development of centralized data warehouse for the state of Florida implies the need to consolidate the data from many sources into one destination site. Since, many of the transportation data agencies participating in the data warehouse are already logged into commercial databases, e.g., Oracle and Sybase, a database to database replication solution should be developed to replicate these data in real time into the CDW database.

Many commercial solutions were developed to replicate data, seamlessly to the user, in heterogeneous database environment. Such solutions were motivated by the commercial needs of distributed database and data warehousing projects. Among these solutions are, OmniReplicator™ from Lakeview Technology Inc., Itware-replicate from Incession Technologies, and SYMBIATOR from Vision Solutions. Database replication strategies can be divided into two major groups:

1. **Using proprietary application to capture changes applied to the database**. In this case, an application is built to capture and replicate changes in the source database into the target one. A similar solution was used by the UCF-CATSS transportation research team for the I-4 data warehouse project. PB Faradayne Inc., the software developer of the MIST traffic management software used at the FDOT-D5 RTMC, modified the software so that a file is parsed by the software having the detectors data every 30 seconds. An application was written by the UCF team to read this file periodically and replicate the data into the I-4-data warehouse. The major disadvantage of this solution was its need to make some modifications in the MIST traffic management software so that it parses the data file every 30 seconds to a specific computer folder accessible to the UCF-CATSS application. Figure 11 shows a diagrammatic sketch of how this solution was implemented.

2. **Using database logs and triggers**. In this type of solution, database log files or triggers are used to automatically detect changes in the source database and extract this information in a queue. The queue is then replicated to the target database. Figure 11 shows a diagrammatic sketch of database replication techniques based on database logs or triggers.

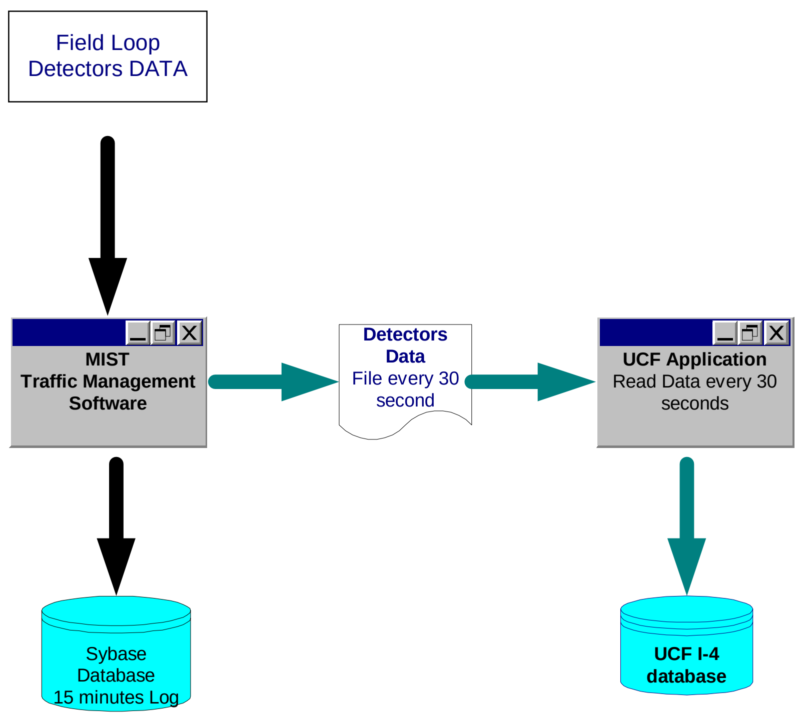

Figure 11. Using Proprietary Applications to Capture Traffic Monitoring

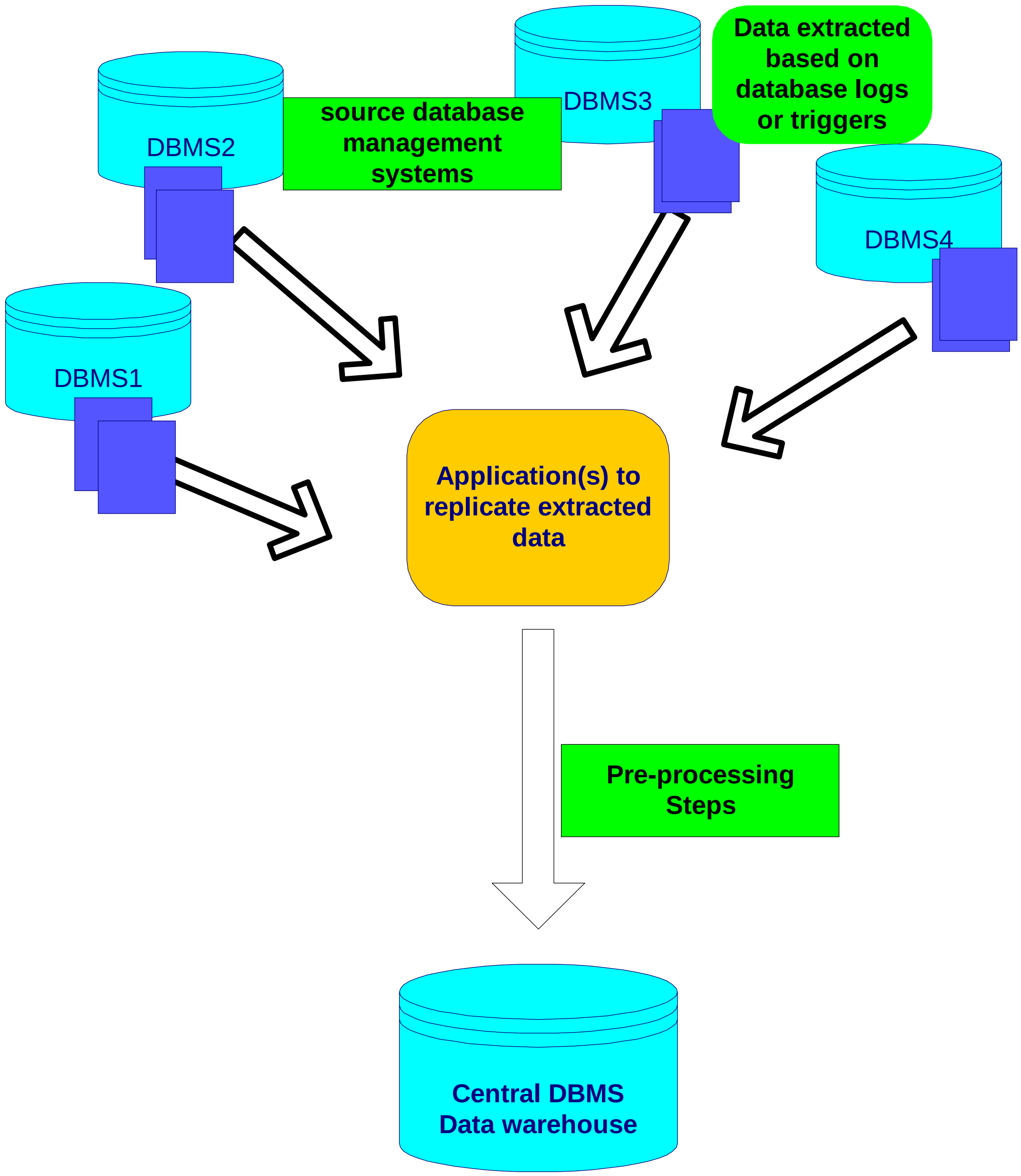

Figure 12. Data Replication Using Database Logs or Triggers

Since commercial database replication solutions are dedicated to the business enterprise area, they are mostly expensive. These solutions are licensed on a per-year and per-server basis, which adds to the project overhead cost and significantly limits the project expandability. On the other hand, any of the previously suggested two solutions can be efficiently developed to capture data received by the source database and consequently copy the data into the CDW database.

Technically, most of these database solutions adopt the Open Database Connectivity (ODBC) and the Java Database Connectivity (JDBC) standards offered by major database vendors. Microsoft's ODBC interface is considered a de facto standard for client/server database connections. Most prominent software vendors provide ODBC as an Application Programming Interface (API) for accessing SQL databases. It specifies a set of functions that allow applications to connect to databases, prepare and execute SQL statements at runtime, and retrieve query results. Typically, Sun Microsystems did not take long to develop its own version of ODBC namely JDBC. JDBC was developed by Sun Microsystems to provide java programmers with cross-DBMS connectivity to a wide range of SQL databases. Sun also developed ODBC-JDBC bridge implementation that allows many of the existing Microsoft ODBC data drivers to operate as JDBC drivers. Both ODBC and JDBC connectivity now offer real time and seamless client server connection between any two major databases.

4.5 INVESTIGATING SOFTWARE/HARDWARE REQUIREMENTS.

Suggested software, hardware, and communication configurations are discussed in this section. Software requirements include DBMS and GIS software. Suggested hardware requirements include platforms and operating systems. A brief description of connectivity needed for the data warehouse is also introduced in this section.

4.5.1 Database Management Systems:

Two candidate DBMS software are recommended. Previous experience showed that both Microsoft SQL SERVER and Oracle software were successfully utilized in transportation data archiving applications. All other competitors were eliminated due to either low performance and/or lack of previous experience in the transportation data archiving field. The following factors are being considered when selecting the appropriate database management system:

- Previous experience from transportation data warehouse case studies and stakeholder needs.
- System scalability, reliability, and security.
- Initial costs of license, hardware, and application development.
- Administration and maintenance cost.

Both Microsoft and Oracle relational databases include data warehousing capabilities. ORACLE though puts a premium for its data warehouse and OLAP extensions. UCF has a pleasant experience with Microsoft SQL SERVER used in the I-4 data warehouse project. Licensing, maintenance, and application development costs favor SQL SERVER while scalability and reliability issues favor ORACLE software under UNIX operating system. However, the advances in windows operating system in addition to low initial and maintenance cost are considered a plus for using SQL SERVER DBMS under windows operating system.

Redundant Array of Independent Disks (RAID) systems should be utilized to improve the reliability and the performance of the database. RAID systems increase database reliability by providing a way of recovering the data in case of disk failure. Redundant storage disk(s) is used to store information about other disks in the system. Moreover, RAID systems improve system performance by optimizing data input/output and physical disk access. Different levels of RAID systems could be applied. Level 5 RAID systems are advised for the data warehouse project. In Level 5 RAID systems, one extra storage disk will be needed to provide the protection against loosing any of the storage disks in the system.

4.5.2 Geographic Information System Software

Many commercial products integrate DBMS and geographic data in order to enable spatial queries and analysis. ESRI is a leading software developing company in the GIS software area. FDOT has been successfully using ESRI products such as ARC/INFO and ARCVIEW in many different projects. ESRI Internet Map Server (IMS) can provide a web-based map interface for the transportation data. On the other hand, The Spatial Database Engine (SDE), an ESRI product that can be integrated with commercial databases, can be used to provide spatial queries capabilities for the data warehouse applications.

4.5.3 Hardware and Operating System

Both Windows NT and UNIX based servers were used in transportation data warehouse implementations nationwide. UNIX platforms are considered more reliable and secure. However, initial labor and maintenance cost associated with the UNIX platforms is considered a major implementation obstacle. Furthermore, the windows operating system is witnessing some significant performance enhancements. These enhancements are accompanied and supported by fast improvements in the PC hardware technology. Therefore, fast Pentium machines with enough RAM, cash memory, storage, and operated with windows advanced server operating system will be considered adequate for the Florida data warehouse implementation.

4.5.4 Connectivity Requirements

Although the statewide Florida fiber optic network is not yet a reality, many parts of the state is currently enjoying fiber optic connections. For example, installations are underway to connect the RTMC center with OOCEA, Orange County, and Seminole County. Also, a study is now under investigation to connect UCF with RTMC. However, a dedicated point-to-point T1 or a simple dial up connection could be leased in the initial implementation stages of the data warehouse, to allow data transfer and application testing. Such a connection has been already established between UCF-CATSS and RTMC. This connection proved to be sufficient in the short-term for transportation data transfer short of CCTV video feeds. There is a need to establish similar connection with the Miami and Jacksonville TMCs (and possibly the Turnpike TMC).

4.6 DATABASE DESIGN SUMMARY AND CONCLUSIONS

The findings of the database design investigation section as well as suggested software/hardware configurations are presented here.

- Both ER and multi-dimensional database models will be utilized in the Florida data warehouse project. ER models should be used in the initial data preparation and when data are frequently updated such as in processing web surveys. On the other hand, a multi-dimensional schema should be used for historical data modeling. This will provide the efficiency needed in the initial data collection and processing steps and the flexibility and efficiency required to query and process the historical data.

- Common dimensions in the Multi-Dimensional model should provide the integration between different data subjects. Examples of common dimensions are the TIME, DATA, and LOCATION dimensions. More common dimensions may include PASSENGER and/or VEHICLE dimensions.

- Multi-dimensional sample representation for traffic monitoring detectors, incidents, and DMS data was introduced.

- The proposed database model should be able to support OLAP applications described in the application requirements part of the report.

- A centralized database for the CDW database is recommended at least for the early implementation stages. Common archiving strategies should be adopted statewide by the TMCs to diminish initial data processing efforts and to set the stages for future distributed data warehouse.

- TMC data stored in local databases need to be consolidated in the CDW database. Two solutions were suggested to replicate data from heterogeneous databases to a central target database. The first solution utilizes proprietary applications to receive the data from the traffic management software and copy the data into the CDW database. The other solution depends on database log files or triggers installed on the source database to identify and queue data changes. Then, the queued data are transferred to the target database. Either one of these two solutions could be adopted for the Florida case. The steps for consolidating transportation data from different TMC into centralized database can be summarized as follows:

  1. Identify different DBMS software, data formats and archiving strategies implemented by the data-generating centers.
  2. Develop software solutions to consolidate the data from the source centers to the centralized database.
  3. Data should be replicated in an operational data warehouse first. At this stage, other processes such as data filtering and reformatting should be applied.
  4. Fuse and aggregate the data into different predefined aggregation levels.
  5. Add the data into the data warehouse database.

- Either ORACLE or SQL SERVER DBMS software could be used for the Florida data warehouse implementation. A RAID system should be implemented to provide protection against failed drives. Periodic database maintenance that includes offline backup of the database should be scheduled. ESRI SDE and IMS software are suggested to provide the GIS component of the data warehouse and to help develop needed applications.

- Windows based platforms that use up to the data hardware configuration in addition to Windows Advanced Server operating system is considered sufficient for the data warehouse implementation.

- Fiber optic connectivity between the different elements in the data warehouse is necessary for successful data warehouse implementation and development of efficient applications.

- Leased point-to-point dial up or T1 connections will be needed for the initial and experimental stages of the data warehouse development. This will be established between the UCF CATSS, FDOT D-2, FDOT D-6, and the Turnpike TMCs.

CHAPTER 5.
IMPLEMENTATION SCHEDULE AND PROPOSED BUDGET

The objective of this section is to present a data warehouse implementation plan and schedule. This includes implementation time frame and budget. For this purpose, a multi-year multi-phase project has been proposed and explained in the following sections.

5.1 MULTI-PHASE IMPLEMENTATION

For the Florida data warehouse, three overlapped implementation phases (Phases 2 –4) are suggested. Each of these three phases has implementation duration of one year. In Phase –2-, detectors data on I-4, I-95, and I-10, toll transactions data from the OOCEA and the Turnpike District will be utilized to provide a single source of travel time information for Central Florida and statewide. Furthermore, DMS and incident data will also be utilized in this phase. Phase-2- can be implemented in two stages. The first stage involves the I-4, the Turnpike, and the OOCEA data. In the second stage, I-10 and I-95 data can be collected and merged with the first stage data. Moreover, once the first stage of Phase-2- is performed, an ATIS can be developed to serve travelers in Central Florida. The ATIS application can then be extended to serve Jacksonville and Miami areas by the end of the second stage of Phase–2.

According to the proposed time frame, detector data, probe data, and traffic signal data on major arterials are to be implemented in Phase-3-. The last implementation phase is proposed to cover other types of data such as transit, emergency management, and commercial vehicles data. This multi-phase development strategy is subject to changes based on data availability, readiness, and on application priorities. Furthermore, overlapped implementation for all phases is necessary. That is preparation for next phase should start before the previous phase is accomplished. Table 6 summarizes the proposed phases, data types involved, and proposed time frame.

Table 7. Proposed Three Implementation Phases for the Florida Central Data Warehouse (CDW)

Implementation Phase	Data Types	Geographic Coverage	Sources	Time Frame
Phase 1	Conceptual Plan			
Phase 2	Detectors data, probe data, DMS data, incident data, lane closure data	I-4, OOCEA, Turnpike	RTMC, OOCEA, Turnpike	1 year
		Freeways (I-95, I-10)	Miami TMC, and Jacksonville TMC	
Phase 3	Detectors data, probe data, DMS data, incident, lane closure data, signal control data	Major arterials (regional implementation is suggested)	Cities and counties (Example: Orange and Seminole Counties and the City of Orlando)	1 year
Phase 4	Data from transit, commercial vehicles, ports, and emergency management centers.	Statewide	LYNX, VOITRAN, Florida ports, emergency management centers	1 year

5.2 SUGGESTED IMPLEMENTATION BUDGET

The budget included in this section covers software/hardware, labor, and connections infrastructure for the three-year implementation plan suggested in the previous section. Proposed software and hardware installations will be the core for future implementation phases. Both fiber optic and T1 connections are proposed to connect major data centers with the data warehouse site. The budget presented in this study is divided into initial cost and yearly budget to run the CDW at UCF. The initial cost represents required hardware, software and connectivity resources. On the other hand, the yearly cost represents mainly labor and T-1 lines' expenses. Tables 7, 8, and 9 list estimated budget for software/hardware and labor/leased-equipment needed for the Florida data warehouse project.

Due to the significant cost of the fiber optic installations in the overall data warehouse project budget, special care should be taken in selecting the Florida CDW site. This selection should benefit from existing fiber optic installations in the state. The distinctive location of UCF in Central Florida on the border between Seminole and Orange counties and is close proximity to the FDOT D-5 RTMC provides the UCF-CATSS with already established fiber optic infrastructure, thus making CATSS the best candidate for the ideal location of the Florida CDW. This will decrease the fiber optic connections cost and provide the potential for physical network redundancy.

Two short links are needed to connect UCF with the Central Florida FDOT fiber network through Seminole and Orange counties. Adding these two links will allow UCF to be connected with FDOT D-5 RTMC through two independent routes. These two missing links are:

1. Colonial Drive link (this will be called the Primary Connection):
   Connecting the UCF-CATSS with the RTMC through the OOCEA fibers on SR408, with a special permission from the OOCEA. The missing link in this connection has a total length of about 2.5 miles. This link extends from the end of SR408 to Alafaya trail along East Colonial Drive plus a small segment on Alafaya trail.
2. Lockwood Blvd link (this will be called the Redundant Connection):
   Connecting UCF with Seminole county fiber optic network. This link is about 2.5 miles long and will bring the Seminole county fibers to the northern UCF boundaries.

   The UCF Telecommunication and Computer Services Department investigated the cost associated with establishing these two missing links in addition to the on-campus connections needed to bring fiber from UCF perimeter to the CATSS lab. The cost estimate included underground conduits, fiber cables, labor…etc. The presented estimates are given under the assumption that FDOT D-5 will come forward with their fiber optic network in time to the UCF perimeter. Table 10 lists the cost for the primary and the redundant connections.

Table 8. Estimated Software Budget

Item	Quantity	Unit Price $	Total Price $
DBMS software	2	15,000	30,000
Microsoft Windows 2000 Advanced Server	3	800	3200
Microsoft Windows 2000 Operating Systems	2	200	400
ESRI SDE	1	1000	1000
ArcIMS and Arcview GIS Software	1	1500	1500
		Total$	**36,100**

Table 9. Estimated Hardware Budget

Item	Quantity	Unit Price $	Total Price $
Servers	3	15,000	45,000
Application development workstations	2	6,000	12,000
Decoders	2	6,000	12,000
Fiber optic equipment (Switches-GBIC Modules)	1	38,000	38,000
T1 connection installation + installation labor**	3	2500	7,500
Routers and interface cards for T1 connections	4	4,000	16,000
		Total$	**130,500**

Table 10. Yearly Budget: Estimated Labor and Lab Software Licenses/Hardware Maintenance

Item	Total Price $
Labor and human resources*	272,000
Maintenance of fiber connection (UCF-RTMC)	10,000
T-1 dedicated connection**	38,000
Total	**320,000**

* The labor includes two months for a full-time faculty to manage this project, two full time research engineers, two graduate students, and two undergraduate students in civil/computer science or engineering majors.
** There are three proposed T-1 lines between UCF and three TMCs, these are the Miami, Jacksonville, and the Turnpike TMCs.
*** The above estimates <u>do not include</u> the university overhead of 5%.

Table 11. Estimate of UCF to FDOT Fiber Optic Connections

Item	Link	On Campus Cost $	Off-Campus** Cost $	Total $
Primary Connection	Colonial Drive link	89,000	371,000	460,000
Redundant Connection	Lockwood Blvd link	317,000	804,000	1,121,000

*The off-campus connection cost is assumed to be the responsibility of FDOT D-5-.

CHAPTER 6.
CONCLUSIONS AND RECOMMENDATIONS

In this study, the UCF research team investigated different sources of information to provide the basis for a thorough implementation strategy of the Florida CDW. The study investigated currently available data, pre-processing and end-user applications, database design, software and hardware configurations, in addition to implementation schedule and proposed budget. A detailed summary and recommendations are provided at the end of each chapter. The following is a summary of the overall conclusions and recommendations in this study.

In addition to Phase -1-, three other phases are suggested for the Florida data warehouse. A one-year implementation period is suggested for each phase. Immediate implementation of this plan is deemed necessary so that the Florida CDW can be fully functional by 2005. In Phase-2-, freeway detectors, DMS, incident, and lane closure/maintenance data collected at the state major TMCs (including FDOT D5-RTMC, FDOT D-6 TMC, FDOT D-2 TMC, Florida Turnpike TMC, and OOCEA) will be warehoused. Data will be collected along I-4, I-95, I-10, Turnpike and OOCEA toll roads.

Significant amount of transportation data are currently being collected statewide. However, this study demonstrates that these data lack common definitions and collection/archival strategy. Accordingly, this study recommends the use of a statewide traffic management software to solve these problems.

The most urgently needed application of the CDW is ATIS. Therefore, Phase –2- of the data warehouse will significantly support the deployment of ATIS in Florida. As stated previously, available traffic monitoring data, DMS, incident, and road closure data as well as statewide CCTV cameras represent the basic elements needed for implementation of ATIS. The ultimate goal is to provide historical, real time, and predictive travel information on state highways and major arterials. This ATIS will eventually be integrated with current web travel information systems such as http://www.mapquest.com to provide real time ATIS for Florida travelers.

It is important to build on the existing ATIS experience in Florida. For example, over the years, the State of Florida has already invested resources in developing the I-4 data warehouse and the traveler information website developed by UCF-CATSS. Other data warehouse applications are also presented in this study such as historical data analysis and dissemination, incident detection and data mining applications. Multi-tier application development architecture is recommended to provide the separation between different application layers such as the presentation, analytical, and data tiers. The presentation tier in this architecture is normally delivered through a web-based interface.

Problems associated with data collection and replication are presented in this research report. Typically, for a centralized data warehouse, data should first be copied and temporarily stored in an operational database. At this stage, data pre-processing operations such as filtering and re-formatting are conducted. Then, the data are aggregated, summarized, and moved to the permanent data warehouse. Two data replication solutions are proposed in this study. In the first solution, data collected at the source database are copied into the central data warehouse using proprietary application that receives data before it is logged into the sources database and copy the data into the target database. The second solution depends on log files or triggers installed on the source database to identify data changes and copy them into the target database. Both replications are appropriate for the Florida CDW.

Based on the previous transportation data warehouse case studies, and based on the evaluation of software performance and cost, the CATSS research team suggests using either ORACLE or SQL SERVER DBMS software under Windows Advanced Server operating system for the Florida CDW. Both software support the Multi-Dimensional data model recommended for the data warehouse. In this multi-dimensional model, common dimensions, such as TIME and LOCATION are suggested to link several data subjects together. The study presented different aggregation levels for each of these two dimensions.

This study suggests a comprehensive role for GIS in the Florida CDW. GIS should provide the graphical and presentation interface for the data warehouse applications and the data warehousing functionality such as spatial querying and data integration. Dynamic querying applications based on OLAP tools and pivot tables technology are recommended for the Florida data warehouse. Using OLAP applications, the user can easily and dynamically navigate through different data subjects, create various data views, and compute data summaries. This study stresses the need for a complete and extensive metadata to support OLAP applications.

The UCF-CATSS recommends hosting the Florida CDW at its facilities. The UCF-CATSS gained significant experience in collecting, filtering, mining, and warehousing more than ten years of the I-4 data. The center also developed many applications that benefited thousands of Florida travelers. Among these applications is the real time, historical, and predictive point-to-point travel time information on a 39-mile section of the I-4 corridor. System coverage is expected to increase with the current expansion in ITS data deployment on I-4 and elsewhere. In addition, CATSS developed an experimental application with one-week sample of OOCEA EPASS transactions data to extract travel time information between toll plazas.

The detailed project budget presented in the study demonstrates the need for a start up cost of about $575,000 for the first implementation year. This cost includes the project infrastructure such as hardware/software requirements and also the labor cost for the first year. A yearly cost of $320,000 is estimated for each subsequent year to cover labor, the expenses of leased T-1 lines, and on-campus hardware maintenance. Knowing how fiber optic connections are crucial for the data warehouse implementation, two different scenarios to connect the UCF-CATSS with the FDOT fiber optic network have been investigated. The budget given for the data warehouse implementation excludes the cost for extending FDOT fiber optic network to the foot doors of UCF. Finally, the Florida CDW could become a reality if the recommendations in this report are implemented in the near future.

REFERENCES

1. Orlando-Orange County Expressway Authority, (2000). "Probe Data Collection Phase 1 – Concept Study", final report. Prepared by PBS&J, April 2000.

2. Orlando-Orange County Expressway Authority, (2001). "Systemwide ITS deployment--Phase-1", Functional Requirements. Prepared by PBS&J, June 2001.

3. Florida Department of Transportation (FDOT) and Michigan Department of Transportation (MDOT), (2001). "TMC Software Study", final report. Prepared by South Research Institute (SRI), November 2001.

4. http://www.ite.org/tmdd/. Traffic Management Data Dictionary (TMDD) and Message Sets for External Traffic Management Center Communications (MS/ETMCC), Accessed on January 25th, 2002.

5. http://standards.ieee.org/regauth/its/index.html. IEEE Intelligent Transportation Systems Data Registry. Accessed on December 19th, 2001.

6. http://www.ntcip.org/library/documents. The National Transportation Communications for ITS Protocol, ON-LINE RESOURCES. Accessed on December 20th, 2001.

7. U.S. Department of Transportation, (2001) "ATIS U.S. Business Models Review", final report. Prepared by PBS&J, November 2001.

8. Ishak, S., Al-Deek, M., and Abd-Elrahman, A., (2002). "Dissemination of Forecasted Traffic Information", ITE Spring Conference, Palm Harbor, Florida, March 2002.

9. Richard T., (1997). "The Intranet Data Warehouse". John Wiley & Sons, Inc., New York.

10. http://www.ntcip.org/restrict/. The National Transportation Communications for ITS Protocol, ON-LINE RESOURCES. NTCIP Group Status: Center-to-Center Profiles. Accessed on January 2nd, 2002.

11. Inmon W., (1996). "Building the Data Warehouse". John Wiley & Sons, Inc., New York.

12. Elmasri R., and Shamkant B., (2000). "Fundamentals of Database Systems", Addison-Wesley, New York.

APPENDIX A.

VISITS TO NATIONAL UNIVERSITIES AND OTHER STATE DOT CENTERS

Tuesday May 1 to Friday May 4, 2001

University of California at Irvine (UCI) May 2nd, 2001

Impression: ORACLE database compatible with Caltrans system; data transfer through T-1 line; receive loop detector data only with no video; budget of $0.75 million/year.

The SC ATMS testbed started in October 1992. UC, Irvine hired two full time computer science majors, two postdoctoral graduates, and graduate students. A full-time faculty position was introduced UCI spent intensive effort on backing up data on "expensive $60 tapes." They use a 150 Gig/tape storage system, SUN SPARC server 420R Enterprise with 4 processors, 4 Gigs memory, 218 Gig hard drive, cost of $104,000 for SUN server only. ORACLE expert charges Caltrans through UCI about $300,000/year @ 8hrs/day. ORACLE license costs $18,000 with 8 concurrent connections to database, one fully installed and ten web processors. UCI does not provide data on-line for previous years. Caltrans (Ct) D-12 hired 5-6 software programmers. UCI uses PARAMICS for off-line simulation to provide Ct with missing loop data and other recommendations. Key future contact: Fred Yazdan of Ct D-12.

University of California at Berkeley UCB-Main Campus Meeting with the PEMS Group (Prof. Varaiya)–May 3rd, 2001

Impression: ORACLE database with strong emphasis on Unix for scalability reliability, and security; data transfer through DS3 45 Mega bits per second, very fast fiber line connected to the Caltrans WAN; UC Berkeley receives loop detector data from LA freeways Ct D-7 (4000 loops) and ready to accommodate all 12 districts; running yearly budget of $0.5 million/year. UCB is far advanced in the data warehouse approach and more appealing to private sector and Value Added Resellers (VARS). The UCB group suggested to contact Tom Chow of Ct D-7 (LA)

for a future field visit to get more details on data warehouse issues. The UCB set up is ready for a standardized data warehousing system that can tie neighboring Caltrans districts in all CA.

A two-hour demonstration of PEMS public website (Freeway Performance Measurement System) was conducted. Raw data is available to anyone for free. Users can get historical, real time, and predictive travel time and speed information (can specify arrival or departure time of their trips). There are 15 possible routes (shortest path by distance, user preferred route, ..etc). Users can store their own profile. Data is in the process of being disseminated on wireless communications. Caltrans and UCB established the Center for Commercialization of ITS Technologies (Hamed Benouar, Director). The PEMS user-friendly software on UCB's website provides freeway bottleneck analysis, contour maps, and performance of any link/loop detector. Data quality checks, and freeway performance with and without ramp metering is useful for practitioners running the operation so they can see trade offs. It takes two weeks to bring in an entire district (i.e., add it to the existing system as soon as it is loop-detector ready). The powerful Unix enables Berkeley to use 64 CPUs while the PC based system can only support up to 4 CPUs. The amount of data is 2 Gigs/day for the entire California freeways (Southern CA takes up one Gig/day). The cost of DS3 is $1500/month. UCB copies the database from each district and gets it through their DS3 line. Booze Allen, Inc. provided training sessions on PEMS. The audience is Ct employees. Dr. Karl Petty (an EECS with major emphasis on traffic engineering) is the key person in this entire operation. He is a private consultant, formerly a PhD student of Prof. Varaiya, who has been hired by UC Berkeley to develop the Unix software, the web page user-friendly front end, to manage hardware/software connections from Ct districts to UCB, and to manage the ORACLE DBA (Data Base Administrator) running the database engine. UCB Professor Pravin Varaiya is in charge of the PEMS research. A number of graduate students work on this research.

Washington State Transportation Center (TRAC)

Impression: SQL SERVER with emphasis on MS windows and NT, compatible with WSDOT; 3000 (mixed double and single) loop detectors; house WSDOT inside University of Washington TRAC; it is difficult to provide sufficient probe data (even if you have 50,000 of them in an area like San Antonio), yearly running budget of $0.35 million/year or $0.7/2 years.

The WSDOT ITS office is located inside TRAC with Pete Briglia as the key WSDOT person who ran the SMARTREK project. Dan Daily (was not present in our meeting) is the key researcher from the University of Washington. Mark Halenback is the research manager of the data warehouse at TRAC. The budget sources are: WSDOT (50%), FHWA (25%), and consulting firms (25%). TRAC plans to include arterials and signal cycles (presented some interesting ideas about cycle failures and how these can be strong indicators of travel time values on arterials). Probes have to be really large in number to be significant in providing travel time measurements. WSDOT still uses VAX 6420 and 6430 machines with CDs for data backup. Data is transferred by fiber (no technical specifications were provided in the meeting). Loop data is available "off-line" with 20 seconds resolution for planning purposes. TRAC has five engineers, one graphics person, one DBA, and many students. That does not include Mark Halenback. TRAC provided the UCF-CATSS team with some useful websites to demonstrate their operation.

APPENDIX B.

SUMMARY OF ELECTRONIC SURVEY RESULTS FOR FLORIDA DISTRICTS

FDOT D-1:

No traffic monitoring detectors or DMS installed on District -1- freeway system. However, future plans exist to deploy inductive loop detector sensors and DMS along the I-75 from the Hillsborough/Manatee County line to the Collier/Broward County line, and on I-4 from the Hillsborough/Polk County line to the Polk/Osceola County line.

FDOT D-2

District –2- is currently testing new traffic management software at its Jacksonville TMC. The software is based on the NaviGAtor software, owned by Georgia Department of Transportation (GDOT). About 41 VID and 8 DMS are installed on I-10, I-95, and I-295 in the city of Jacksonville area. The VID and DMS data are logged in ORACLE DBMS tables at the Jacksonville TMC. Short-term plans exist to deploy RTMS sensors and more DMS on I-95. Incident data are handled through the NaviGAtor software based on real time detectors data.

FDOT D-3

District –3- has 4 DMS installed at Escambia Bay Bridge and at the Weigh-in-Motion facility on I-10. No data archival currently exists by FDOT for this data. The ITS Master Plan for I-10 and I-110 and the Regional Architecture for I-10 propose more detectors and DMS along I-10 and I-110.

FDOT D-4:

The district does not currently employ any traffic monitoring detectors. However, about 156 RTMS units are being designed for I-95 in Palm Beach County. A total of 34 DMS are now being deployed on I-595 and I-59 in Broward County. These units are expected to be functional during the year 2002. Another 16 DMS units are under construction in Palm Beach County.

FDOT D-5:

ITS data collected within District –5- is considered the largest among all Florida's districts. District –5- employs about 41 DMS along 50 miles of the I-4 corridor. Four more DMS are functioning on I-95 within the district. Loop detectors data collected on Interstate-4 by the FDOT D-5 RTMC covers 39 miles (70 stations) from Lake Marry Blvd to US-192. The UCF-CATSS is currently collecting, filtering and warehousing this data at the UCF ITS lab facilities. A new set of data is received and added to the database every 30 seconds. Access to the data is achieved through a file parsed by the MIST traffic management software currently deployed at the RTMC. The MIST system also logs 15-minute summaries of loop detector data into the Sybase database. The Daytona TMC receives the data from 10 more loop detectors data installed on I-4 and Interstate-95. The RTMC will eventually access this data once the fiber optic connection between the two centers is established.

The RTMC just started to receive data for 11 additional miles (about 30 stations) to the east of Lake Merry Blvd. Work is underway at CATSS to add data from these newly installed stations. In addition, more loop detectors are currently being installed (under construction) along the I-4 corridor from US-192 to US-27. Also, detectors and DMS units are under construction on I-95. The RTMC will receive these data once the system is installed and fully functional.

FDOT D-6:

FDOT D-6 has 3 loop detectors installed on I-95. As part of the I-95 Intelligent Corridor Systems (ICS) phase A, 35 VID were implemented. One-minute detectors data can be acquired and saved through the SunGuide traffic management software installed at the Miami TMC in a plain text format. Currently, FDOT D-6 is operating 14 DMS. Oracle database is used to archive actions, commands, messages, alarms and flags associated with 10 of the deployed DMSs. No log of messages and commands is provided for the rest of deployed DMS. More detectors are planned on I-95 through the ICS I-95 project phase B. About 42 RTMS detectors and 12 loop detectors are planned in this phase. Moreover, approximately 50 RTMS stations will be deployed on SR-826.

Incidents are handled through one of the SunGuide software modules. This module analyzes real time detector data to determine if a suspected incident exists or if an existing incident should be cleared.

FDOT D-7:

SUNPASS transponders are used for toll collection on I-275, SR-682, and SR-589 within the district. However, the toll transactions data are managed by the Turnpike tolls office. Three DMSs, located on I-275 and I-75, are implemented within the district. The DMS software logs all the DMS actions. However, this data is not archived in any database. About 60 VID detectors and 15 DMSs are planned on I-4 and I-275 within District –7-.

A web site (www.TBInterstates.com) was developed and maintained by Fierro and Associates, under FDOT contract, to provide the public with construction information for Tampa Bay area interstates. The information is updated manually based on notifications from projects personnel or changes in the interstates' work zones. A free subscription is allowed to get email alerts of road construction information. Moreover, the service is accessed through mobile phones and wireless devices that support the Wireless Application Protocols (WAP) by pointing to **http://tbinterstates.com/wap**.

FDOT D-8:

Currently, the FDOT D-8 SUNPASS transactions data are not used to extract travel time information between toll plazas. However, the Turnpike office of tolls is beginning to archive tolls transaction data, which will allow toll plazas congestion, volume, and other types of data to be available in the future. There has been a recent deployment of four Portable Roadside AVI Readers. In addition, there are existing plans to deploy more AVI sensors along the Turnpike mainline.

Twenty DMSs are currently under construction along the Turnpike mainline. The Turnpike district just installed the SunNav Release 1.0 traffic management software at the district's Pompano TMC. A Sybase database is used by the SunNav software to archive DMS, incident, and road maintenance data.

APPENDIX C.

TMDD DEFINITION FOR DETECTORS DATA

In this section, some of the fields presented in the TMDD definitions of detectors data are listed. The sole purpose of this appendix is to introduce samples of the TMDD required for the data warehouse implementation.

Descriptive Name:	Definition:	Data Type:	Representation Class Term:	Value Domain:	Valid Value Rule:	Representation Layout:
DETECTOR_StartTime_utc	Start time for collecting traffic count	IA5 String	utc	ANSI NCITS.310	HH=00 through 23; MM=00 through 59; SS=00 through 59. HH represents the hours, MM the minutes, and SS the seconds.	HHMMSS
DETECTOR_Direction_code	Identifies direction of the traffic movement a particular detector is monitoring (e.g., on a roadway such as a reversible lane).	Enumerated	Code	Numeric code	1=N; 2=NE; 3=E; 4=SE; 5=S; 6=SW; 7=W; 8=NW, 0=Any other.	9
DETECTOR_IntersectionApproachName_text	String name or description of the intersection approach being monitored by the detector (e.g., NB Left Turn).	IA5 String	Text	Character	Any set of ASCII characters up to 128	SIZE (1..128)
DETECTOR_LinkIdentifier_identifier	An unique identifier to identify the link that an individual detector is monitoring	IA5 String	Identifier	Identifier	Any set of alphanumeric characters up to 32	SIZE (1..32)
DETECTOR_Status_code	A code which indicates whether a detector is	Enumerated	Code	Numeric code	Valid Value List: 0 = Failed; 1 =	9

Descriptive Name:	Definition:	Data Type:	Representation Class Term:	Value Domain:	Valid Value Rule:	Representation Layout:
	failed or operating				Operational; 2 = Off.	
DETECTOR_MeasurementDate_date	Traffic data collection date.	IA5 String	Date	ANSI X3.30	Valid date only. MM=01 through 12; DD=01 through 31; YYYY=1900 through 9999. YYYY represents the year, MM the month, and DD the day.	YYYYMMDD
DETECTOR_Other_text	Data element used to indicate the presence of additional information. This data element is to be used only when a precedent DETECTOR data element of type code has a value of "1= Other, additional information required".	IA5String	text	character	Any set of ASCII characters up to 256.	SIZE(1..256)
DETECTOR_VehicleQueueLength_quantity	A count of the average number of vehicles in queue, measured by the detector over a specified time period.	Integer	Quantity	SI 10-1997; vehicles	Valid Value Range: VALUE (1..100000).	9999
DETECTOR_LaneNumber_code	The number which indicates the lane a particular detector is monitoring	Bitstring	Code	Numeric code	Valid Value Range: VALUE (1..255); Select one bit per lane. Lanes are numbered from the median out	999

Descriptive Name:	Definition:	Data Type:	Representation Class Term:	Value Domain:	Valid Value Rule:	Representation Layout:
					beginning with 1.	
DETECTOR_MarginalPerformanceFactor_percent	Marginal detector performance factor in percentage used by raffic responsive (TRSP) and 1.5 GC modes in undercount/overcount tests.	Integer	Percent	SI 10-1997; percent	Valid Value Range: VALUE (1..100)	999
DETECTOR_LoopOperationMode_code	Operation mode of a loop detector.	Enumerated	Code	Numeric code	0=Other, no additional information required; 1=Other, additional information required; 2 = Presence; 3 = Pulse.	9
DETECTOR_Identifier_identifier	Unique identification number of an individual detector within a network.	IA5 String	Identifier	Identifier	Any set of alphanumeric characters up to 32	SIZE (1..32)
DETECTOR_StationIdentifier_identifier	Unique identifier for a detector station.	IA5 String	Identifier	Identifier	Any set of alphanumeric characters up to 32	SIZE (1..32)

Descriptive Name:	Definition:	Data Type:	Representation Class Term:	Value Domain:	Valid Value Rule:	Representation Layout:
DETECTOR_Type_code	Code naming the type of a vehicular detector providing traffic data.	Enumerated	Code	Numeric code	Valid Value List: 0=Other; no additional information required; 1=Other; additional information required; 2=Inductive loop; 3=Magnetic; 4=Magnetometers; 5=Pressure cells; 6=Microwave radar; 7=Ultrasonic; 8=Video Image; 9=Laser; 10=infrared; 11=Road tube.	99
DETECTOR_EndTime_utc	End time for collecting traffic count.	IA5 String	utc	ANSI NCITS.310	HH=00 through 23; MM=00 through 59; SS=00 through 59. HH represents the hours, MM the minutes, and SS the seconds.	HHMMSS
DETECTOR_VehicleCount_quantity	The number of vehicles detected by a detector during a specific time period.	Integer	Quantity	SI 10-1997; vehicles	Valid Value Range: VALUE (1..100000).	999999
DETECTOR_SectionIdentifier_identifier	Unique identifier within the network for the section in which	IA5 String	Identifier	Identifier	Any set of alphanumeric characters up to	SIZE (1..32)

Descriptive Name:	Definition:	Data Type:	Representation Class Term:	Value Domain:	Valid Value Rule:	Representation Layout:
	detector resides				32	
DETECTOR_Class_code	User-defined class of detectors for traffic control.	Bitstring	Code	Numeric code	0=Other, no additional information required; 1=Other, additional information required; 2=Stop bar; 3=System; 4=Pedestrian; 5=Adaptive; 6=Call; 7=Extension; 8=Mainline; 9=Reversible lane; 10=Ramp demand; 11=Ramp merge; 12=Ramp passage; 13=Ramp queue .	99