**QUESTION / ISSUE:** The New York State Comptroller's Office (OSC) is expanding its data mining functions. As part of this process, OSC is requesting information that other states may have related to data warehouses and data mining. Specifically, we are interested in the following:

1. Does your state maintain a data warehouse of operational and/or expenditure data?
2. If yes, what type of data is stored in the warehouse and how often is it refreshed? (e.g., financial, program performance, statistics, other?)
3. What database management system is the warehouse built on? (e.g., Microsoft SQL Server, Oracle, SAP HANA, dBASE, FoxPro, IBM DB2)
4. What software tools are used to access data, build reports, and perform data analysis and data mining?
5. Who are the stakeholders for your state data warehouse and what are their roles?
6. Is there a specific state entity or representative which could be contacted for further discussion?

See below for data warehouse examples in New York State:

- Statewide Financial System
- AnalyzeNY, the Statewide Financial Data Warehouse
- State Education Department
- NYSED - Information and Reporting Services
- NYS - Department of Health
- Medicaid Data Warehouse
- Office of Children and Family Services
- Connections

<table>
<thead>
<tr>
<th>State</th>
<th>Comments</th>
</tr>
</thead>
</table>
| Georgia    | The Georgia Department of Audits and Accounts has developed specific data warehousing and data mining practices for the databases it maintains. These policies apply to large data sets involved with our financial and healthcare audits, but can be used for any audit we perform (performance or financial). Other state agencies may maintain additional databases, but we cannot comment on their content or design. We recommend New York State Comptroller's Office contact the Georgia Technology Authority, the State Accounting Office of Georgia and the Board of Regents of the University System of Georgia for information regarding other potential datasets. Their contact information is as follows:  
Georgia Technology Authority  
47 Trinity Avenue S.W.  
Atlanta, Georgia 30334  
Main Office: 404.463.2300  
Website: [www.gta.georgia.gov/](http://www.gta.georgia.gov/)  
State Accounting Office of Georgia  
200 Piedmont Avenue SE, Suite 1604, West Tower  
Atlanta, Georgia 30334-9010  
Main Office: 404-656-2133  
Website: [www.sao.georgia.gov](http://www.sao.georgia.gov)  
Board of Regents of the University System of Georgia  
270 Washington Street, SW  
Atlanta, Georgia 30334  
Main Office: 404-962-3050  
Website: [www.usg.edu](http://www.usg.edu) |
Our answers to New York’s questions for the databases we maintain are as follows:

1. State organizations are required by law to submit payments and obligations data to us which is published on the State’s transparency website. However, this information is not audited. Additionally, we do maintain a database repository of Medicaid claims information for use in healthcare audits as well as support and services provided to Georgia’s Medicaid Fraud Control Unit. This database repository contains raw data that is subsequently extracted, analyzed, and reported via SQL and other tools.

2. State salary and travel data used in the transparency website is refreshed annually. Medicaid claims and related information (e.g., provider data, member data, etc.) is received on a weekly basis and uploaded to the database monthly.

3. The transparency website and Medicaid claims databases are built on Oracle.

4. We typically do not perform analysis on the transparency database. However, we do use a variety of software tools to analyze the Medicaid claims database; these include Oracle Sql Developer, Pentaho and Qlik. For audits of outside datasets, we use ACL, ActiveData, SQLDeveloper, R (base + ggplot2, plyr, etc…), Microsoft Office Excel/PowerPivot, PERL scripts, Jasper iReports, and MS Access.

5. For the transparency website, the stakeholders are the legislators and general public who may have an interest in this information. For Medicaid claims, the Healthcare Audit Division staff and Georgia’s State Medicaid Fraud Control Unit are the stakeholders and they use this data to conduct audits and investigations.

6. Our Department’s IT Division manages the Oracle environment and handles data loading, schema creation and performance tuning. Our various divisions manage data analysis/cleaning, request data from auditees and the dissemination of reports to management/staff. If New York is interested in additional information on the transparency website, then Lynn Bolton, Director of our IT Division can be reached at 404-657-9978. For additional information on the Medicaid claims database, please contact Kirby Davis of our Healthcare Audit Division at 404-657-7024.

Michigan

Our software assistance section typically is split between financial/accounting information, and departmental program/performance data functions, so we have attached a set of responses from each sub-section.

Program Data Warehouse
1. Yes, Michigan has numerous departments that utilize the state's data warehouse, including the Department of Community Health (DCH), the Department of Human Services (DHS), and the Department of Treasury (DOT).

2. The type of data varies by department, I am most familiar with the DCH and DHS data sets, which both contain extensive data regarding clients, providers, eligibility, and benefit payments. The DOT module contains a variety of both business and individual tax information. The warehouse information is typically refreshed nightly on a limited basis and fully on a weekly basis.

3. The data warehouse utilizes the Teradata management system.

4. Various software tools are used to access the data. BI/Query is used by DCH and DOT, while DHS typically uses the Teradata Assistant (we use both).

5. I can't speak to who the stakeholders are, and what their roles may be, I'm sure this varies greatly by department. For our purposes, the data warehouse has allowed us to quickly capture pertinent data for many of our audits, both performance and financial.

6. The Department of Technology, Management and Budget oversees the data warehousing functions.
<table>
<thead>
<tr>
<th>State</th>
<th>Comments</th>
</tr>
</thead>
</table>
|       | Financial Data Warehouse  
1. Yes  
2. Financial, personnel, purchasing, and user security data are stored in the statewide financial data warehouse called Management Information Database (MIIDB). MIDB financial data is refreshed weekly and daily towards year end processing. Purchasing and personnel data daily is refreshed bi-weekly. User security is refreshed weekly/daily from various systems.  
3. MIDB (statewide financial data warehouse) uses ORACLE.  
4. We use Business Objects web version to access data, build reports, and perform data analysis on data for MIDB.  
5. All state departments are stakeholders.  
6. Neil Miros, DTMB MIDB Manager, mirosn@michigan.gov |
| Minnesota | The Minnesota Office of the State Auditor oversees local government finances. However, after doing a quick Google search, it appears that the State of Minnesota has a data warehouse: http://www.mmb.state.mn.us/system/ia  
"Minnesota Management & Budget Information Access (IA) Data Warehouse provides authorized agency administrative and agency business users with a collection of data for compiling information and generating reports. The IA Data Warehouse integrates data from several operating systems to provide information that is more complete and easier to access. It is a data warehouse providing current and historical information from MAPS, which provides accounting and procurement data; and SEMA4, which provides human resource, labor distribution, leave, and payroll data."  
If you click on the above link, you will see more information on the warehouse, the software used, and more.  
Another good link is: http://www.swift.state.mn.us/swift-data-warehouse. See content below.  
SWIFT Data Warehouse Overview  
The SWIFT Data Warehouse provides State of Minnesota agencies with a business intelligence tool (Oracle Business Intelligence Enterprise Edition) to query and present financial and human resources data in a variety of formats including reports, charts, and dashboards. The SWIFT Data Warehouse is being implemented for state agency use as the various data subject areas are developed, tested and become ready for production. The following categories of data are planned to be included in the SWIFT Data Warehouse:  
HCM - (Human Capital Management) - includes the State of Minnesota's Human Resources and Payroll data from SEMA4  
FSCM - (Financial and Supply Chain Management) - includes the State of Minnesota's financial and procurement records from SWIFT (from Fiscal Year 2012 forward)  
In terms of a contact for our state, I recommend contacting Carolyn Parnell, Commissioner and Chief Information Officer, (651) 556-8007, or Ed Valencia, Deputy Commissioner and Chief Technology Officer, (651) 556-8029. |
| Missouri | 1. Yes, for the statewide accounting system  
2. Various levels of financial data, pay data from the HR system, summary data, and some system data such as chart of account elements are stored in the warehouse. The data are refreshed nightly during the business week.  
3. DB2 |
<table>
<thead>
<tr>
<th>State</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4. Pre-developed reports reside in the warehouse, and users connected to the warehouse can either develop specific reports to perform data analysis or download the data to their computers. Also, Access via open database connectivity reporting and Focus reporting are used.</td>
</tr>
<tr>
<td></td>
<td>5. All state agencies are allowed to access the warehouse after obtaining the proper authorization. Stakeholder roles are varied and include basically all state government operations.</td>
</tr>
<tr>
<td></td>
<td>6. The Office of Administration – Information Technology Services Division maintains the warehouse; the website address is <a href="http://content.oa.mo.gov/information-technology-itsd/">http://content.oa.mo.gov/information-technology-itsd/</a>.</td>
</tr>
<tr>
<td>New Jersey</td>
<td>1. Yes, for purchasing, expenditures, payroll and Medicaid</td>
</tr>
<tr>
<td></td>
<td>2. Financial – daily; Payroll – weekly; Purchasing – weekly; Medicaid – daily</td>
</tr>
<tr>
<td></td>
<td>3. Oracle</td>
</tr>
<tr>
<td></td>
<td>4. SAP’s Business Objects Web Intelligence XI</td>
</tr>
<tr>
<td></td>
<td>5. OIT (NJ Office of Information Technology), and Department of Treasury Medicaid, Department of Health</td>
</tr>
<tr>
<td></td>
<td>6. OIT (NJ Office of Information Technology) Medicaid, Department of Health</td>
</tr>
<tr>
<td></td>
<td><strong>NJ OIT Response:</strong></td>
</tr>
<tr>
<td></td>
<td>1. We have an enterprise data warehousing environment supporting multiple lines of business. It has a few dozen subject areas, which include revenue and expenditures.</td>
</tr>
<tr>
<td></td>
<td>2. In the financial area, we start with general ledger details loaded nightly. Throughout the entire warehouse environment just about any reasonable periodicity, frequency, and level of detail/summarization is represented.</td>
</tr>
<tr>
<td></td>
<td>3. The majority of the warehouse environment is built upon Oracle. Microsoft SQL Server is used for some data marts and to support some auxiliary functions. We will be moving towards the use of HDFS (Hadoop) for data reception and archiving, and eventually for general unstructured content.</td>
</tr>
<tr>
<td></td>
<td>4. The enterprise BI reporting platform is SAP BusinessObjects, but other tools are used as well (such as WebFocus, Cognos, SAS, SAP Crystal, and some of the boutique BI products like Tableau). If by data mining you mean as defined by the data management profession (analyzing data through various algorithms to identify patterns and correlations that would then inform the design process for data analytic solutions), we are not at the moment doing this. If by data mining you mean as it is often used in common vernacular (ad hoc reporting and analysis of large volumes of integrated data to produce reports or gain insight), the tool most often used is SAP BusinessObjects, along with some use of SAS. There are about 3,000 users of the BusinessObjects environment. The enterprise ETL (Extract, Transform and Load) platform is the IBM InfoSphere DataStage suite. In addition, materialized views, SQL scripts, UNIX Chron, and other tools support the integration of several thousand files. Much data is extracted and transferred off of large legacy mainframe applications through batch processes that are precursors to the enterprise ETL jobs.</td>
</tr>
<tr>
<td></td>
<td>5. The stakeholders vary by subject area. In several cases, integrated subject areas are the responsibility of steering committees representing the line-of-business agencies that use the domain (examples: Traffic Records, Recidivism and Early Childhood Development). In some cases, a single business unit is the primary stakeholder (example: Office of Management and Budget for the</td>
</tr>
<tr>
<td>State</td>
<td>Comments</td>
</tr>
<tr>
<td>---------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
</tbody>
</table>
| Utah    | 1. Yes  
2. The Utah State Division of Finance maintains a data warehouse of all general ledger (financial) information. This data is refreshed daily (corresponding to the update of the data in the general ledger).  
3. Sybase IQ  
4. Within the Office of the Utah State Auditor, the primary tools are MS Access, MS Excel, and Caseware IDEA. The auditors typically access the data using an ODBC connection within MS Access.  
5. Utah State Division of Finance – Responsible for the maintenance of the data warehouse. All state agencies – Users of the data warehouse.  
6. Wayne Frewin, IT Analyst, Utah Department of Technology Services (801) 538-3691 |
| Washington | 1. The Department of Enterprise Services maintains Enterprise Reporting which allows users the ability to query Agency Financial Accounting System (AFRS) Financial data, Enterprise Contact management System (ECMS) and the Statewide Vendor population. Access to this data is limited to customers with a user ID and password. The data is updated nightly.  
   The Office of Financial Management (OFM) maintains a data warehouse of Washington State Fiscal Information; including budget, spending and revenues, staff, K-12, and some other financial data. See www.fiscal.wa.gov. The website will display data that are routinely updated: for budgets as budget proposals are made public; for revenues, expenditures, and staffing, on a monthly basis as monthly actual data are finalized; for historical data, as those data are updated to reflect changes in budget formats; and for workloads, caseloads, performance measures, and performance audits, as websites that display those data are updated.  
   This agency also maintains multiple operational data marts via their website. The information maintained here includes but is not limited to: economy, educational, expenditures by county, health care, higher education and population information. It appears this data has a variety of update dates. See http://ofm.wa.gov/forecasting/default.asp.  
2. The database warehouses are built on both Microsoft SQL Server, and SAP systems.  
3. Microsoft SQL Reporting Services, Microsoft SharePoint, SAP Web Intelligence, and SAP Dashboard.  
4. There are multiple stakeholders for the data described above however, the most important one is the public as noted during the 2008 regular session, the Washington State Legislature passed Senate Bill 6818 and the Governor signed it into law (RCW 44.48.150). The bill indicated the legislative intent that state revenue and expenditure data be made as open, transparent, and publicly accessible as is feasible with the goal of making government more accountable.  
5. My recommendation would be to contact OFM. See contact information at http://ofm.wa.gov. |