DATA PROCESS
and
Data Warehouse Projects
in TURKSTAT

31.03.2014
(Muscat, Oman)
Information Technologies Department

- Harzemli Software Development Group
- Management Software Group
- Network and System Management Group
- IT Support Group
- Data Management Group
TurkStat Data Architecture
TurkStat Databases

Production Database:
Current data is stored and always transaction in it.

Institutional Database (Enterprise Database):
The main repository of TurkStat’s harmonized historical data.

Dissemination Database:
Aggregated time series data created from TurkStat institutional databases.

Up to now there are
62 institutional databases
59 dissemination databases

Oracle DBMS is used for all database
Production Databases

- Mainly data entry systems
- Database for web applications
- OLTP (Online Transactional Processing)
- Surveys data are stored in this database
- Huge amount of data entry and frequent data manipulation (inserts, updates, deletes)
- Various but not costly SQLs
- Supports thousands of concurrent users
- We developed a new generic system (Harzemli) that produces surveys generic using DDL. Regional Offices use this program on the netbooks.
Institutional (Enterprise) Databases

- Institutional database has analysed, harmonized and controlled consistency eventual time series data.

- Only for internal users.

- User roles are defined for the statistical units on the newly created institutional databases.

- The statistical units can access only their own data using SAS codes, and just have select right.

- They can never delete, update, insert data by themselves.

- It is necessity for all departments to work in coordination with IT Department during the creation process of institutional databases.
Dissemination Databases

• Dissemination databases are the summary aggregated data repository which are created from institutional databases according to the information request.

• The dissemination database is served on the TurkStat web page.

• They are READ-ONLY system.

• For external users, only confidential data is available.

• It is free of charge, shown on web page.

• Since it is open to whole world, supports many concurrent users.
• There are aggregated time-series data about 62 subjects.

• End users can easily get desired data with the help of interface programs and dynamic reports.

• The current web-based programmes are developed by our programmers with ZK framework and Oracle reports.

• We are developing a new Central Dissemination Project that will show all dissemination data by one software programme.
The main benefits of distinction to these 3 database is:

To maintain different business processes on different servers.

An advantage is performance. While we are making a bulk insertings or queries on instutional database, dissemination database doesn’t effect these procedures. So the performance would not be decreased.

Another advantage is only aggregated data are served to internet users and access to micro data is prevented
Data Harmonization

The main subject of the Institutional databases is harmonization of different data.

After careful investigations and analyses, it has been seen that our main problems come from using different classifications from year to year and projects to projects.

Such as:
- Branch of economic activity
- Occupational group
- Educational status
- Marital Status
- Administrative Structure
- Dictionary tables
Data Warehouse in Turkstat

Being holded in our relational institutional database

Designed as star schema model

ROLAP (Relational OLAP) which means that all OLAP cubes are on relational database
Data Warehouse in Turkstat (Cont’d)

Being held in our institution at the scope of European Union projects, Creating data warehouse process was started in 2006, with **SAS**

And a prototype data warehouse consists of 1985, 1990 and 2000 population censuses were formed

**Oracle BI suite** is used.
Data Warehouse in Turkstat (Cont’d)

8 data warehouses / OLAP projects with 48 data marts:
• 2000 Population Census
• 1990 Population Census
• 1985 Population Census
• Labour Force Statistics
• Municipal Waste Statistics
• Municipal Water Statistics
• Foreign Trade Statistics
• Demographical Statistical

Data warehouses in Institutional database. There are different schemas for each Data warehouse.

SQL Procedures are used as ETL, to producing DW data.
Oracle BI Suite

OBIEE 11g includes:

OBIEE Administration Tool

OBIEE Interface

BI Publisher
OBIEE Interface
OBIEE Interface
OBIEE Interface (pivot table)
## OBIEE Interface (table and chart)

### Table

<table>
<thead>
<tr>
<th>S91.Asansörlerden memnunum</th>
<th>S91.Asansörlerden memnunum kod</th>
<th>Sayı</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bölgede bulunmayan kriter</td>
<td>999</td>
<td>43344</td>
</tr>
<tr>
<td>Cevapraz</td>
<td>99</td>
<td>9883</td>
</tr>
<tr>
<td>Pırrım Yok</td>
<td>6</td>
<td>11792</td>
</tr>
<tr>
<td>Katlıyorum</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Katlıyorum</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Kesinlikle Katlıyorum</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Kesinlikle Katlıyorum</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>Orta</td>
<td>3</td>
<td>32696</td>
</tr>
</tbody>
</table>

### Graph

- **Sayı**
  - **Bölgede bulunmayan kriter**: 999
  - **Cevapraz**: 99
  - **Pırrım Yok**: 6
  - **Katlıyorum**: 2
  - **Katlıyorum**: 4
  - **Kesinlikle Katlıyorum**: 1
  - **Kesinlikle Katlıyorum**: 5
  - **Orta**: 3

---

[Image of the OBIEE Interface with table and chart]
OBIEE Interactive Dashboards

Simply containers for reports and other content (like chart, maps..)

Allows multiple reports to be displayed in a tabbed interface

Used to provide highly summarized and graphical content
An Example Interactive Dashboard
Summary

- Databases of the TurkStat are Production, Institutional and Dissemination
- Separating databases to 3 database is useful both performance, management, confidentiality
- We have some Data Warehouse Projects developed by Oracle BI Suite
- Oracle BI suite has Administration Console, Presentation Interface, Reports Building tool
- Dashboards are containers that can include various and dynamic reports, maps, charts etc., and give very easy and valuable data to the users