

Comtrade

Technology Overview

August 2005

Comtrade Components

- UN Comtrade:
 - Web Application for Data Dissemination
 - ComtradeBatch for batch data request
- Comtrade Management:
 - User/Usage Management
 - UN Comtrade Site Management
- Comtrade Web Services:
 - Data Exchange Project (with World Bank, ECLAC, FAO, UNIDO, ITC, WTO, OECD)
 - Disseminate Data in SDMX format
- CoprA:
 - Common Comtrade Data Processing (Joint Development Project with OECD)

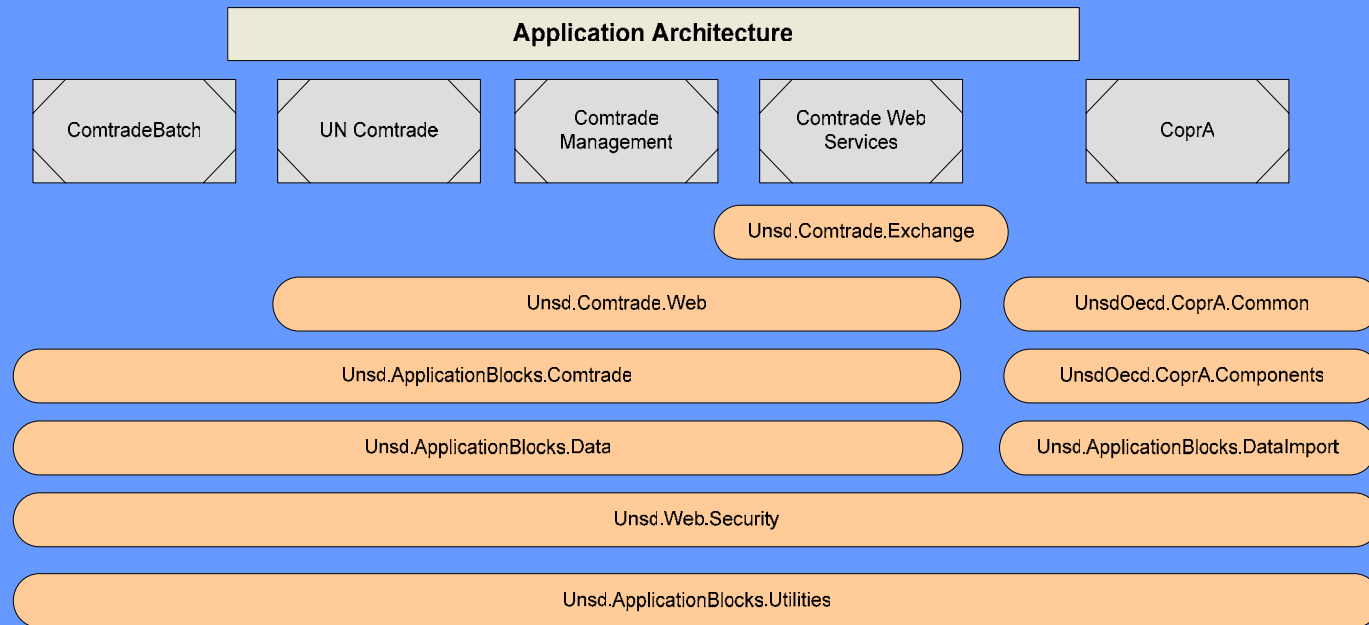
Software & Hardware

- Windows 2000 Advanced Server SP4 (with .NET Framework 1.1)
- IIS 5.0
- Microsoft SQL Server 2000 Enterprise Edition
 - Enterprise Edition for partition view feature
 - Heavy use of SQL Agent
- HP Proliant ML530/G2, Server Dual PIII Xeon 3 GHz, 6GB RAM, 1.2TB disks, RAID 5

Development Approach

- Web based application
 - Easier to deploy and maintain
 - Can reach many audiences
- Business rules in database layer
 - More reliable, do not rely on other layers to run the business
 - Faster, no additional traffic or overhead
- Reusable .NET Classes
 - Improve the development speed
 - Reduce code redundancies

Application Architecture



Legend



.NET Class



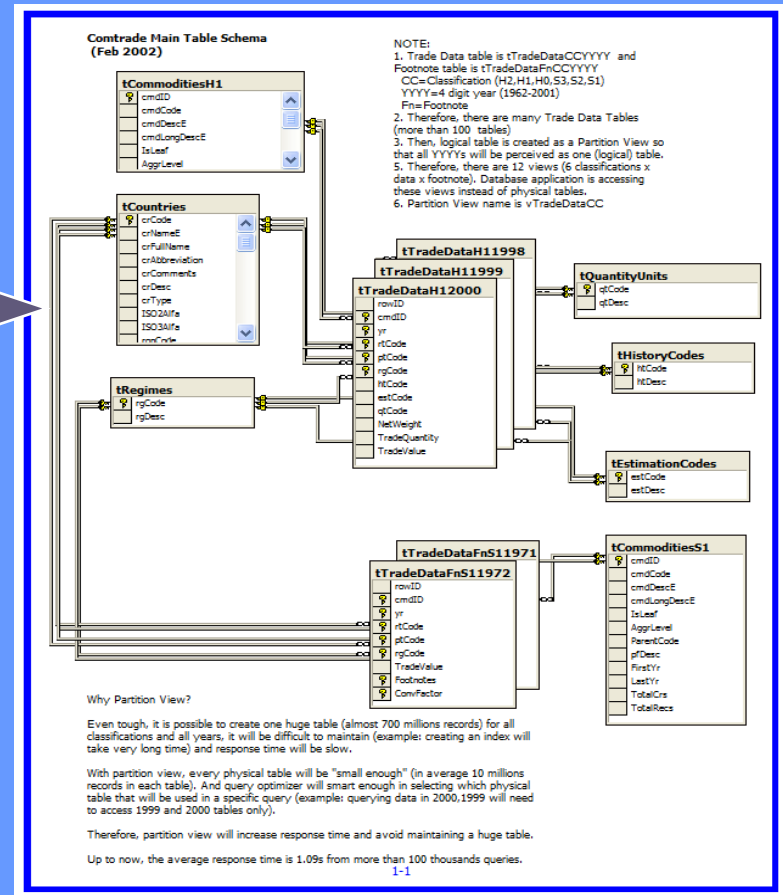
Comtrade Dissemination
Database Server



Comtrade Processing
Database Server

Database Design

- Star Schema (OLAP Style)
- Partition View for the main table (see next page for detail)
- In addition of partition view, indexes had been created for the main tables
- Current statistics:
 - 990 million records
 - DB Size: 300 GB
 - Growth: 100 million records/year



Why Partition View?

- Even tough, it is possible to create one huge table (almost 990 million records) for all classifications and all years, it will be difficult to maintain (example: creating an index will take very long time) and response time will be slow.
- With partition view, every physical table will be "small enough" (in average 10 millions records in each table). And query optimizer will smart enough in selecting which physical table that will be used in a specific query (example: querying data in 2000,1999 will need to access 1999 and 2000 tables only).
- Therefore, partition view will increase response time and avoid maintaining a huge table.
- Up to now, the average response time is 0.49s from more than 3 million queries.

UN Comtrade & Comtrade Management

<http://unstats.un.org/unsd/comtrade>

The Challenges

- To create web application that has a reasonable response time from a large database.
- To provide user with many ways of data query and extraction
- To provide user with personalization (such as customize download format)
- To be able to run batch jobs

Highlighted Features

- ComtradeBatch
 - Run in the background, No Timeout.
- Notification and AutoBatch
 - User will be notified and query will be run automatically if there is a new / updated data
- My Comtrade
 - Provide personalization. User can create custom country or commodity groups. The aggregation is done by the application.

Highlighted Features

- Full Text Search
 - Application tries to map text to codes. Example: in Express Selection, user can type USA instead of the code 842.
- Comtrade RSS
 - User can be informed for any new / updated datasets using standard format
- Comtrade SDMX
 - To promote SDMX, Comtrade offers user to download free 1000 records/day in SDMX format.

User and Usage Management

- Done in Comtrade Management
- Some of reports are provided by SQL Server Reporting Services
- Comtrade Management provides real-time monitoring of the usage, how many queries in last 20 minutes, from where, via Web Services or Batch, etc
- Used also for answering the feedbacks

Current Activities
16 active users with 48 recent queries
Comtrade Guest (Canada: 67.68.26.160): 10 queries.
Comtrade Guest (Saudi Arabia: 212.138.47.15): 2 queries.
Comtrade Guest (UNKNOWN: 70.184.171.3): 3 queries.
Comtrade Guest (USA: 63.65.196.2): 3 queries.
United Nations (USA: 157.150.194.10): 3 queries.
Comtrade Guest (Saudi Arabia: 212.138.113.16): 1 queries.

UN Comtrade Usage Statistics

71,358 unique users, from 153 countries, have been accessing UN Comtrade, generating 3,155,979 queries and downloading 11,448,106,173 records.

The average response time is 0.49 sec for on-screen data request (3,019,474 requests) and 1.73 sec for data download (95,458 requests). So far, 20,089 batch data downloads have been requested and 17,818 web service* requests have been served.

Usage statistics as of 11/7/2005 (taken from June 2003 until July 2005)

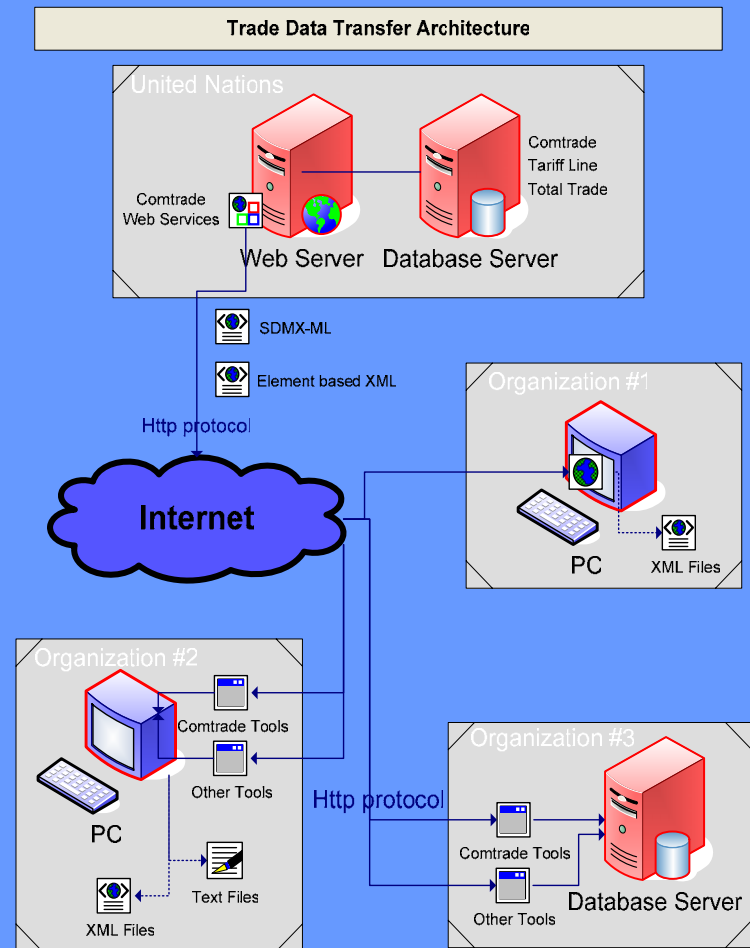
* Public web services were provided at the beginning of 2005

Comtrade Web Services (CWS)

<http://unstats.un.org/unsd/comtrade>

CWS: Overview

- Initially developed for CoprA project
- Open to public as of January 2005
- Suitable for large and regular data transfer
- In SDMX-ML Cross Sectional and element based XML format
- Uses on-the-fly data compression



CWS: Narrative (see previous chart)

- UN Site contains a web server and database server. Comtrade, Tariff Line, and Total Trade Data are available through the web services.
- Web Services can stream compressed trade data in SDMX-ML format or Element based XML. The web services also have some metadata such as the data availabilities and references.
- Organization #1 connects to the internet and obtains trade data using the internet browser. He/she can save-as the data to XML-file.
- Organization #2 uses ComtradeTools (developed by UNSD) to obtain the trade data and convert it into text files. Other tools can also be used.
- Organization #3 set up the automatic one-way synchronization so that the local trade database will be always up to date with UN site.
- All three organizations are authenticated by IP Addresses and Web Service Access Control List.

CWS: The Strategies

- **Using a standard data format. SDMX (<http://www.sdmx.org>) version 1.0 has been released and it is being used as one of the format for trade data dissemination;**
- **Leveraging Internet as a low cost transfer medium and for its standards, such as http protocol, web services, etc;**
- **Client platform independent means that client can use any operating systems, any database, and program languages to obtain trade data;**
- **Be able to transfer a very large data. The REST Web Service and internet stream compression play important role here;**
- **The possibility for unattended one-way synchronization.**

CWS: The Implementations

- Architecture: **“Pull” architecture due to the flexibility and tight security (such as UN Server can’t initiate http request to any web sites). The client has to initiate a request.**
- Synchronization Approach: **By comparing the data availability between two sites for creating the list of new/updated datasets.**
- Data Format: **SDMX-ML version 1.0 is currently being used. Element based XML is provided as an alternative.**
- Medium and Protocol: **Internet (with http protocol) via Web Services.**

CWS: The Implementations

- **Compression: The SDMX-ML is compressed during the data transfer. It reduces 70-80% of the bandwidth usage and download time.**
- **Encryption: None at the moment. If it is necessary, the https will be implemented instead of http.**
- **Security: Based on IP Address and Access Control List (to use Web Service).**
- **Tools: A small command line program (beta) has been created that downloads Comtrade SDMX-ML via Web Services and (can) import them into the SQL Server. If it is necessary, a full client program can be developed. *Please bear in mind that this tool is not the only one that is able to obtain trade data. The client can obtain the trade data using an Internet Browser such as Netscape or Internet Explorer.***

CWS: Case Studies

- ECLAC TRADE
 - Client was developed in Flash
 - User accesses UN Comtrade via ECLAC Trade client
 - Commodity and Country aggregations done at UN Comtrade
 - Productivity was greatly improved
- World Bank WITS
 - Not like ECLAC Trade. World Bank downloads all data in all classifications
 - Cross Sectional SDMX-ML is used during data transfer
 - Then, data is stored in World Bank database.
 - User accesses World Bank database not UN Comtrade

CoprA (Comtrade Processing)

<http://unstats.un.org/unsd/comtrade>

CoprA: Overview

- Unified Data Collection (OECD Member States)
- Unified Standards for Data Processing
 - ✓ Countries and Areas
 - ✓ Commodity Nomenclature
 - ✓ Quantity Estimation
 - ✓ Data Checking
 - ✓ Treatment of Confidential Information
- Unified Data Storage, Dissemination, Meta-data Structure

CoprA: Joint Development Project

- OECD (mostly) works on database layer
- UNSD (mostly) work on presentation layer
- OECD uses the same database design with UNSD for their dissemination database
- Data Exchange will be done via Internet using SDMX format (see CWS)

CoprA: Shared Data Processing

- OECD will process trade data of OECD member states.
- UNSD will process the other countries.
- The end-result will be exchanged via Web Services.

CoprA: Processing Steps

- Registration
- Data Preparation
 - Convert National Format to Access or SDMX
 - Import to SQL Database
- Code Normalization
 - Partner Codes, Quantity Units, Trade Flows
 - Convert into USD
- Completeness Check
- Commodity Normalization
- Basic Level Aggregation
- Quantity & Value Checking
- Higher Level Aggregation
- Conversion to other classifications
- Data Upload

Summary

- Comtrade has many sub-applications but shares the same .NET codes and store procedures.
- Due to the size, database design is very important to improve response time.
- With Comtrade Web Services, data dissemination is taken to the next level: automated data exchange/dissemination and the availability of virtual Comtrade database (see ECLAC case study).
- The infrastructures: S/W and H/W are very reliable. There are no major technical problems or performance issues.