

Information Analyst Support System



*XML Underpins a Real-Time
Information Analysis /
Decision Support System*

Steve Rizzi

SAIC

410-266-0993

rizzis@saic.com



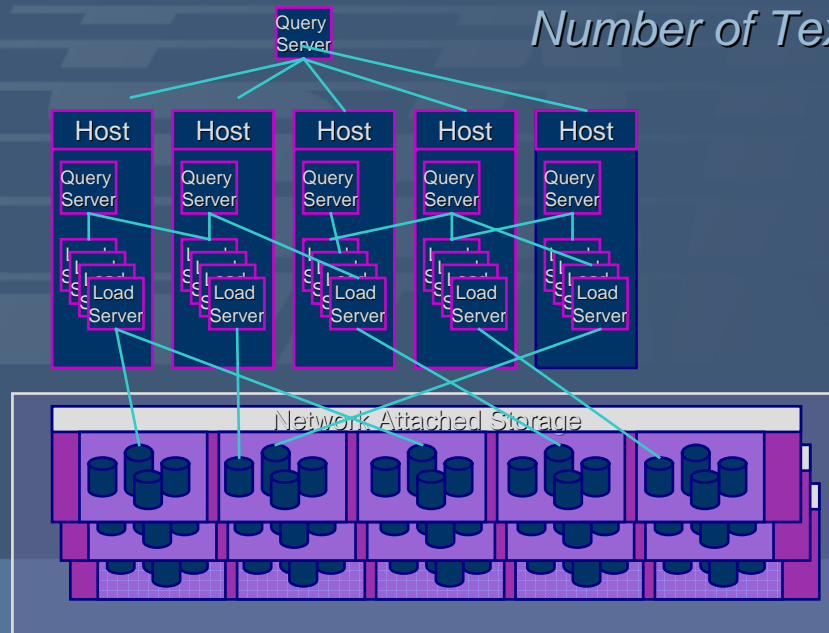
IASS Overview

- Research and analysis system
 - *Time-oriented data, documents, fielded data*
 - *Billions of data records with millions input daily*
 - *24 * 7 availability*
- Multiple database systems
 - *Use 'best' DBMS for each particular data set*
- Messaging architecture
 - *Ties together different technologies*
 - *Uses XML for flexibility and interoperability*



IASS

- A comprehensive set of analytic tools
- Scalable: new capabilities, new data types, mission surge, even supports enterprise requirements
- 4000+ users / 5 billion XML Docs / 2 sec search time
 - *Approximately 50 CPUs!!!*
- tailored data fusion across 4 billion records in 30sec
 - Total number of sources: 186*
 - Total number of data types: 60*
 - Number of relational DB's: 84*
 - Number of Text DB's: 66*



IASS Challenge

- Existing database systems inadequate
 - *Expanding mission requirements*
 - *New sources of XML text information*
 - *High, and increasing, data volumes*
 - *Complex analysis and decision-support*
- Issues
 - *Performance*
 - ✓ Quickly load text documents arriving in bursts
 - ✓ Make data available for querying within seconds of loading
 - *Functionality*
 - ✓ Multiple Languages
 - ✓ Complex searches

Critically important to understand how information will be used by the analyst

Finding The Right Technologies

- XML is lingua franca
- Flexible middleware at the hub
- Evolved from RDBMS with text extensions to full-featured Text DBS
 - *TeraText DBS out-performed other products*
- NAS hardware
- XML accelerator



IASS and XML

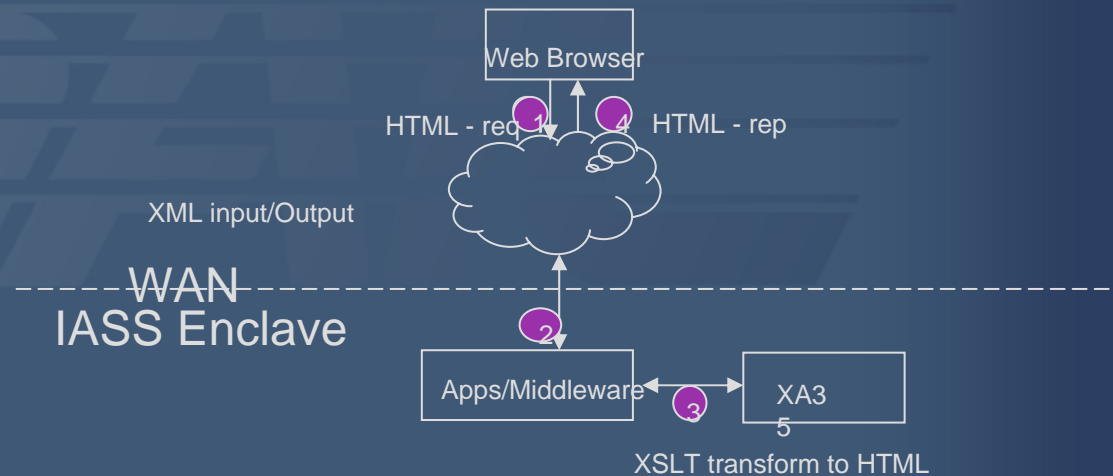
- Widely varying complexity in DTDs
 - *Customer did not select encoding format for all data*
 - *XML mark-up performed on some data sets to facilitate internal processing and data fusion*
- Customer required full text searching of XML document content
- Selected Text DBS is tightly coupled to XML
- Hardware 'assist' provided by XML accelerator



DataPower

IASS draws on the hi-speed performance of Data Power hardware in the sorting of every result set and the presentation of responses to every user request

- IASS is highly distributed using commodity hardware.
- Distributed hardware/processing, using 1U, 2U, and 4U processors, delivered acceptable performance
- The DataPower XA35 was added to reduce the bottlenecks related to transformations and sorting



Benefits

- The DataPower XA-35 provides 10-50x increased performance in XSLT transformations
- Integrates well with industry standard load-balance software & hardware delivering the scale required for most enterprise systems
- Supports all W3C standards related to XML processing
- Simple installation – designed as a 1U rack mount device with a simple web-based or CLI management interface
- No spinning media allows for a rugged reliable system
- Operates in 3 modes Co-Processor, proxy, and in-line; Homebase uses the DataPower XA-35 in both proxy and Co-Processor modes.

Languages

El más grave atentado desde el derrocamiento de Sadam
Gravísimo atentado, el más grave desde el derrocamiento de Sadam. El presidente rotatorio del Consejo de Gobierno de Iraq, Ezedin Salim, ha resultado muerto en un atentado perpetrado contra su residencia, situada junto al cuartel general de la Coalición en Bagdad, según informaba a primera hora la cadena de televisión Al Yazira.

Lunes, **17 mayo 2004**

AMIGOT NEWS / INFORDEUS

El responsable del Consejo de Gobierno designado por Estados Unidos murió junto a al menos ocho personas, tras la explosión de un coche bomba en un control en Bagdad, confirmó después el viceministro de Exteriores, Hamed al Bayati, a Reuters.

Abdul Zahra Othman Mohamed, también conocido como Izedin Salim, estaba esperando para acceder al principal edificio del complejo, cuando se produjo la explosión. Tropas estadounidenses en el lugar confirmaron que al menos ocho personas habían muerto. Había varios vehículos en llamas y una columna de denso humo negro se elevaba hacia el cielo. Las tropas estadounidenses bloquearon la zona. La explosión se escuchó en todo el centro de Bagdad. Varios testigos dijeron que la explosión destruyó varios coches que hacían cola en el control para entrar a la Zona Verde, un área que pertenecía a uno de los complejos palaciegos de Sadam Husein y es ahora principal sede de la coalición.

El **6 de mayo**, un suicida mató a cinco iraquíes...



IASS Text Database Requirements

- Handle a wide variety of languages and hierarchical document structures
- Provide users with access to documents within seconds of loading
 - *Satisfy broad search requirements*
- Manage large volumes of structured and unstructured text documents
 - *Scale easily to support growing number of users and data feeds*
 - *Use storage resources efficiently*
- Robust query capabilities
- Full record-level security with role-based access control

TeraText DBS Functionality and Performance

- Immediate availability
- Query response time
- Scalability
- Storage efficiency



TeraText DBS Functionality and Performance

- Immediate availability
- Query response time
- Scalability
- Storage efficiency
- Rich Boolean query language
- Customizable to meet special language, document structure, and functional requirements
- Full XML support
 - *Built in parser*
 - *XPath, XSLT*
- Interoperability and standards compliance
 - *Z39.50, ODBC, XML, SGML, Unicode, CCL*



Text DBS Search Capabilities

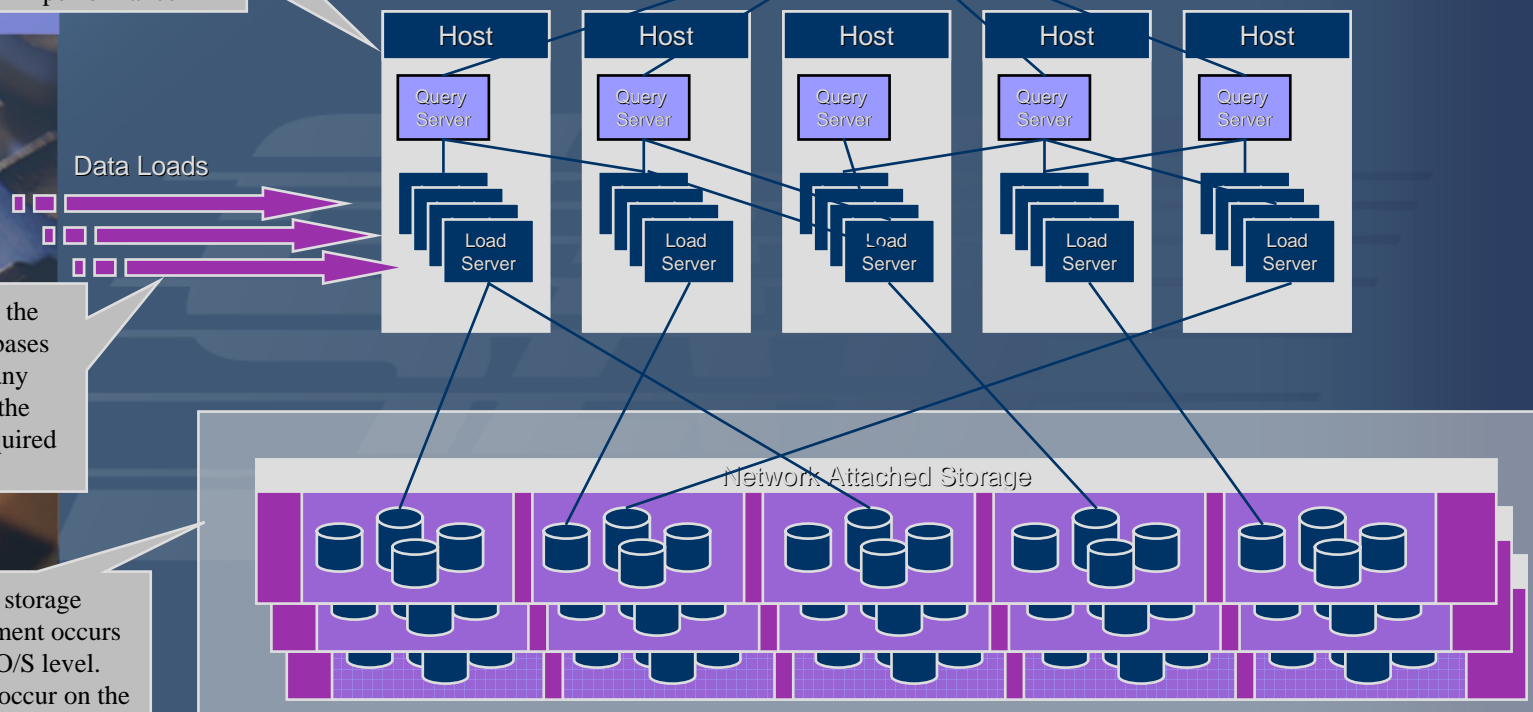
- Full text and fielded
- Proximity operators (near, within, same, order)
- Range operators (string, numeric)
- Fuzzy match, stemming, weighted
- Limit operations
- Custom case folding, punctuation stripping, transformations, expansions, modular lexing and relevancy algorithms, etc.
- Boolean operators (and, or, not)
- Wildcards (#, #n, ?, ?n)
- Relevance ranked search
- Index scan operations
- Hit highlighting
- Saved searches and result set caching
- High speed sort
- Index sets
- **Manages permissions, authentications, security**

IASS TeraText DBS Implementation

Single point access using either C++ or Java API can touch any or all databases

Scalability and load balancing occurs at the host level. Sun E420R 4 CPU hosts. Low cost/high end performance

Application Adapters (C++/JAVA API)

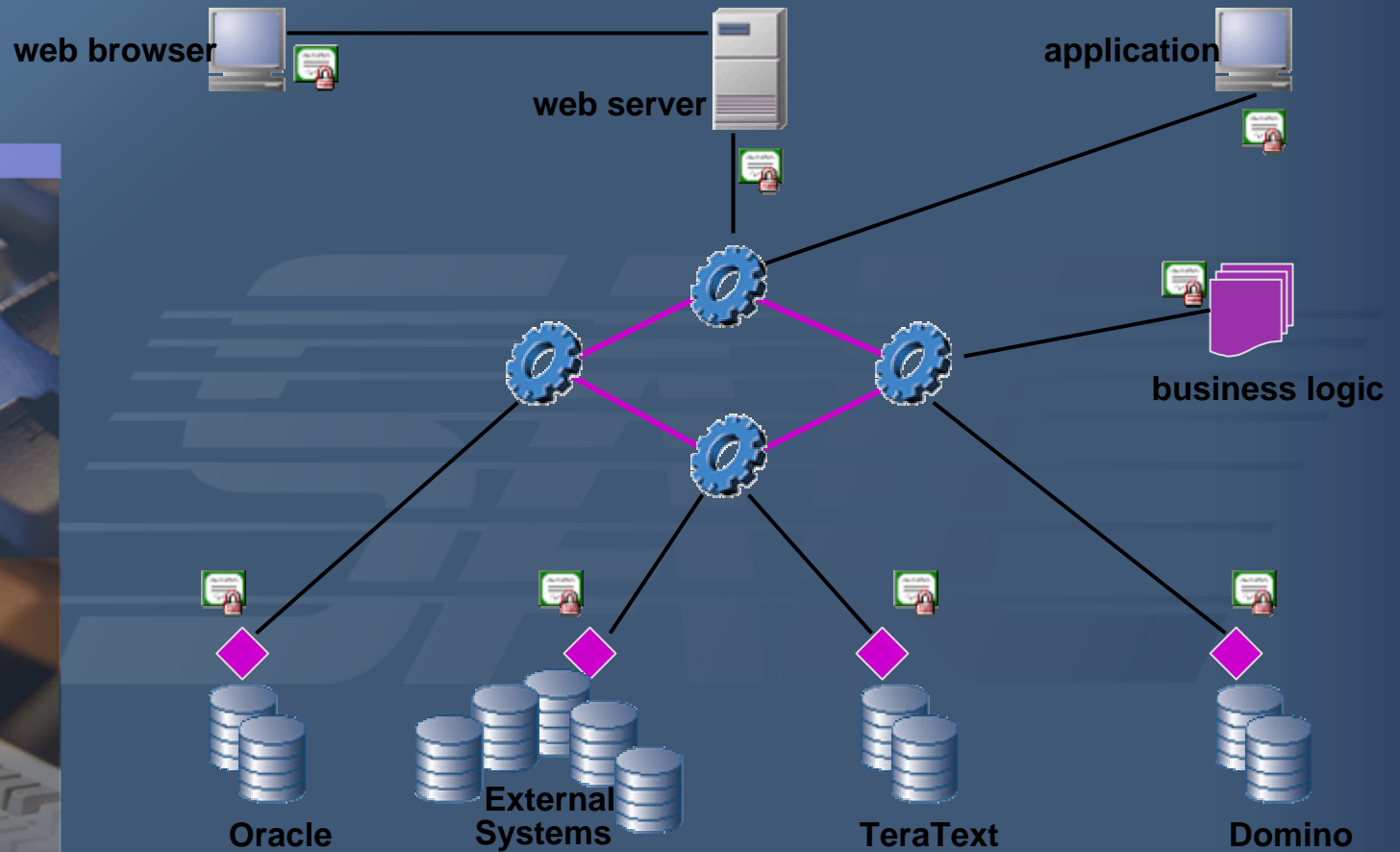


Loading into the physical databases
Can target any database in the network as required

Data storage management occurs at the O/S level.
Backups occur on the storage devices



Message-based Architecture (multi-Broker)



Analyst Driven Data Fusion

- Federated query plus business logic:
 - *Understand the data types*
 - *Understand the relationships*
 - *Compose more complicated services from more atomic ones*
- Institutionalize the knowledge / methods of expert users



Analyst Driven Data Fusion

Query: *assassinated Iraqi leader, May 2004*



Analyst Driven Data Fusion

Query: *assassinated Iraqi leader, May 2004*

Results:

[Suicide Bomb Kills Top Iraqi Official](#)

Suicide Bomb Kills Top Iraqi Official Abdel-Zahraa Othman Was The Current Head Of The Iraq Governing Council May 17, 2004 7:12 am

[Head of Governing Council killed in car bombing](#)

... A US soldier secures the site where a car bomb exploded in Baghdad and killed Abdel-Zahraa Othman. By Ramzi Haidar, AFP. May 17, 2004

[Suicide bomb kills Iraqi council chief](#)

... Abdel-Zahraa Othman, commonly known as Izzadine Saleem, was the second member of the US-appointed council assassinated so far. He ...
May 17, 2004



Analyst Driven Data Fusion

Query: *assassinated Iraqi leader, May 2004*

Results:

[Suicide Bomb Kills Top Iraqi Official](#)

Suicide Bomb Kills Top Iraqi Official Abdel-Zahraa Othman Was The Current Head Of The Iraq Governing Council May 17, 2004 7:12 am

[Head of Governing Council killed in car bombing](#)

... A US soldier secures the site where a car bomb exploded in Baghdad and killed Abdel-Zahraa Othman. By Ramzi Haidar, AFP. May 17, 2004

[Suicide bomb kills Iraqi council chief](#)

... Abdel-Zahraa Othman, commonly known as Izzadine Saleem, was the second member of the US-appointed council assassinated so far. He ... May 17, 2004

Next Steps:

- **Who** is Abdel-Zahraa Othman?
- Who are his known **associates**?
- What other **analysts** are tracking Abdel-Zahraa Othman?
- What **locations** are associated with Abdel-Zahraa Othman?
- What **reports** have been issued recently concerning Abdel-Zahraa Othman?
- etc.



SAIC



SAIC



SAIC



SAIC



SAIC



SAIC



SAIC



SAIC



SAIC



SAIC

Conclusion

- The IASS performance requirements drove us to find alternative solutions
- The IASS functional requirements demanded a rich query language and multilingual support
- XML served as the best choice to structure, store, share and deliver information
- Performance and flexibility were provided by
 - *Hardware accelerators*
 - *Network Attached Storage provided*
 - *TeraText DBS*

System has scaled two orders of magnitude over the last four years

Contact Information

Steve Rizzi

Bill Wolf

Bill Kovalick

SAIC

<http://www.saic.com>

410-266-0993

rizzis@saic.com

Kim Kingsford

TeraText Solutions

<http://www.teratext.com>

301-371-3283

kingsfordk@saic.com

