

May 6-10, 2007
San Jose Convention Center
San Jose, California, USA

Session: H03

Text search on DB2 for z/OS data



IDUG® 2007
North America

Peggy Zagelow
IBM

May 07, 2007 01:40 p.m. – 02:40 p.m.

Platform: DB2 for z/OS

GoFurther



If you have text data in DB2 for z/OS character, varchar, and CLOB fields, how can you search that data? The DB2 for z/OS and IBM search teams have been working on solutions to this problem, both for current and future DB2 releases. Come hear what you can do now, and what we are looking into for our future releases.

Drivers for text search solution

- Customer demand
- Problems with prior text extender offerings
- OmniFind index and search technology

IDUG® 2007 North America

Google Web Images Video News Maps more...
 enterprise database search Search Advanced Search Preferences

Web Results 1 - 10 of about 73,400,000 for enterprise database search (0.86 seconds)

Enterprise Systems
 Home | About ES | Subscribe | Contact Us | RSS XML

Business Intelligence

Oracle Searches for Extra-Relational Data
 Now that Oracle's on board, all three market-leading database vendors have articulated enterprise search strategies.

By Stephen Swoy 3/8/2006
 Listen to some fo Also Sprach Zora
 It can't be that ga Oracle Corp.'s an software, the three strategies.
 Although technical well in advance. A double use in its which—along with its traditional data

More Business Intelligence News
 The Appliance Tipping Point: Commoditization

c|net NEWS.com Search:

Today on CNET Reviews News Downloads Tips & Tricks CNET TV Compare Prices NEW ON CNET Do-
 Today on News Business Tech Cutting Edge Access Threats Media 2.0 Markets Digital Life Blogs Ext

Microsoft to Google: Hands off enterprise search

By Mike Ricciuti
 Staff Writer, CNET News.com
 Published: July 13, 2006, 1:33 PM PDT
 TalkBack E-mail Print del.icio.us Digg this

BOSTON--Microsoft is digging in for a fight with Google in the

Brimey Spears
 Keira Knightley

There is much interest in the industry about Enterprise search – and DB2 for z/OS contains enterprise data.

Customer search scenarios

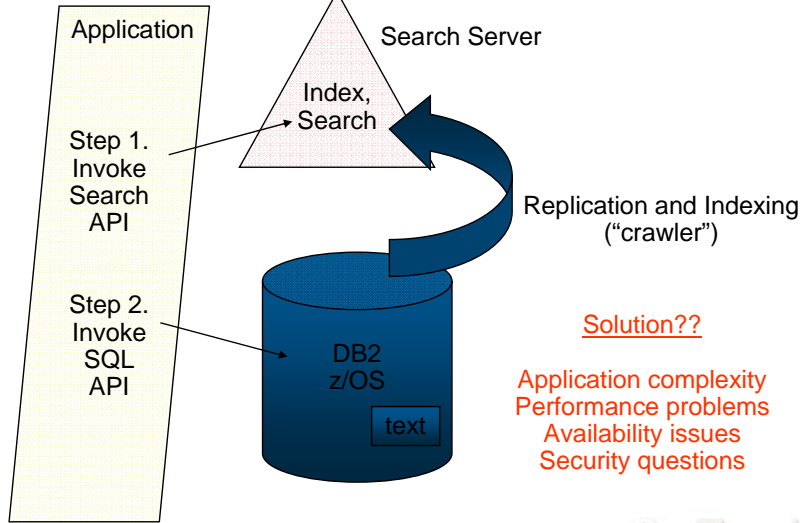
- DB2 for z/OS table with catalog item descriptions
 - VARCHAR column, average 256 bytes
 - Online web searching with familiar interface
- DB2 for z/OS table with insurance agent notes
 - CLOB column, average 1K bytes
 - Agent remembers a claim but not who made it!
- DB2 for z/OS table with item names
 - CHAR column, padded to 80 bytes, 400K+ rows
 - Find items with keyword ordered by a customer

4

GoFurther

These are representative of some examples we've heard from our customers on their search needs.

Current Customer Search scenario



Existing search applications are limited by product capability.



One example of some press that our OmniFind search technology received.

Customer demand for text search

- Common Scenario:
 - Text data in DB2 for z/OS: catalog, reports
 - Need online capability for users to search
 - Data extracted from DB2 to enable search
 - Search and retrieve are separate operations
- Customers want to use DB2 SQL API for a single search and retrieve
 - Security, optimization

7

GoFurther

What our customers have asked for.

Problems with Prior search solutions

- Early attempt was “Text Extender”, with many technical and usability problems.
- Newer “Net Search Extender” was not ported to z/OS, no data sharing solution
- Not integrated with DB2 for z/OS – a “wart” on top
 - Difficult for customers to manage, use of file system, etc.
 - Performance limited by external functions
- Index size scalability issues
- Lack of linguistic processing
- Used z/OS MIPS with associated MLC woes

Go Further

What we tried in the past to meet the need of our customers.

DB2 for z/OS requirements

- Not NSE extension - DB2 for z/OS and Omnifind feature
- All administration from invoking administrative stored procedure
 - No Unix System Services commands
 - Admin SP defined by DB2 installation JCL
 - No new authorizations, ids, password files, etc.
- No HFS file system access from DB2 engine or admin SP
 - All data in DB2 tables or static data in programs

,

GoFurther

These are the requirements with which we began the DB2 for z/OS and Omnifind project.

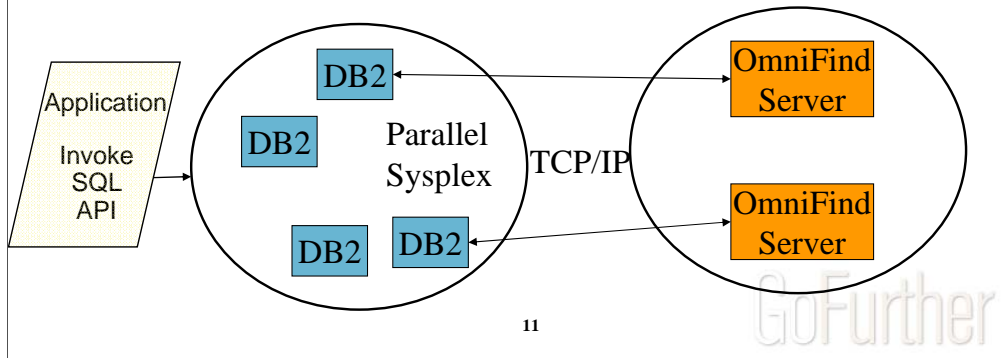
DB2 for z/OS requirements

- Must support data sharing group data
 - All data and indexes must be accessible from all members
- Backup of text search server index data must be able to be coordinated with DB2 data
- Must be able to support failover to a different text search server

Most important “-ilities” we need to provide with our solution.

OmniFind Text Search Support

- Provides text search for CHAR/VARCHAR/LOB/XML columns
- OmniFind provides a text index server
- Efficient communication interactions with DB2 for z/OS
- OmniFind text indexes are persisted into DB2 tables for backup/recovery purposes



This is the basic architecture of our solution.

Key Tasks for enabling text search

- Obtain and install the OmniFind text search server on a windows or Linux machine
- Run the installation JCL in DB2 for z/OS to define the stored procedures and create the tables
- Populate the administration table with the OmniFind server information
 - Location, user/pw, DB2 user/pw
- Install and configure the type-4 Universal Java Driver on the Omnifind text search server machine
- Execute the START() stored procedure

What the DB2 administrators will need to do.

Text Search Administration Stored Procedures

- Start – DB2 connects to OF server, Allows contains() score() to be invoked
- Create Index
- Update Index
- Alter index, Drop index, Clear events
- Stop OF – contains(), score() disabled

What we are providing as an administration interface.

Key tasks for administering text search

- Choose the text columns to enable indexing
 - CHAR, VARCHAR, CLOB, BLOB, XML
 - Table must have a ROWID column
- Invoke the SYSPROC.SYSTS_CREATE() stored procedure to create the index for each column
 - language may be specified
- Invoke the SYSPROC.SYSTS_UPDATE() stored procedure to index the documents
- Schedule the SYSPROC.SYSTS_UPDATE() stored procedure to run periodically

More tasks for administering text search with DB2 for z/OS.

IDUG® 2007 North America

DB2 for z/OS | TCP/IP network | OmniFind Server (Windows/Linux)

When a text search index is created with the SYSTS_CREATE SP,

A staging table is created and admin data written

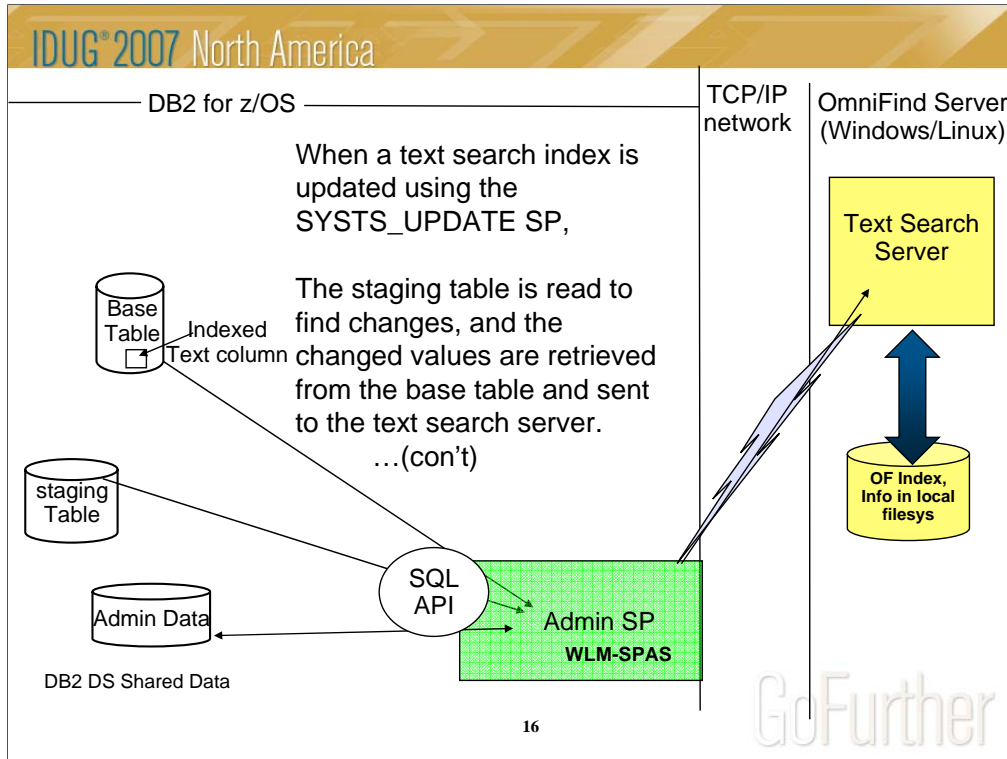
A trigger is created on the base table to propagate changes to a staging table

An index collection is created on the text search server (but not populated yet).

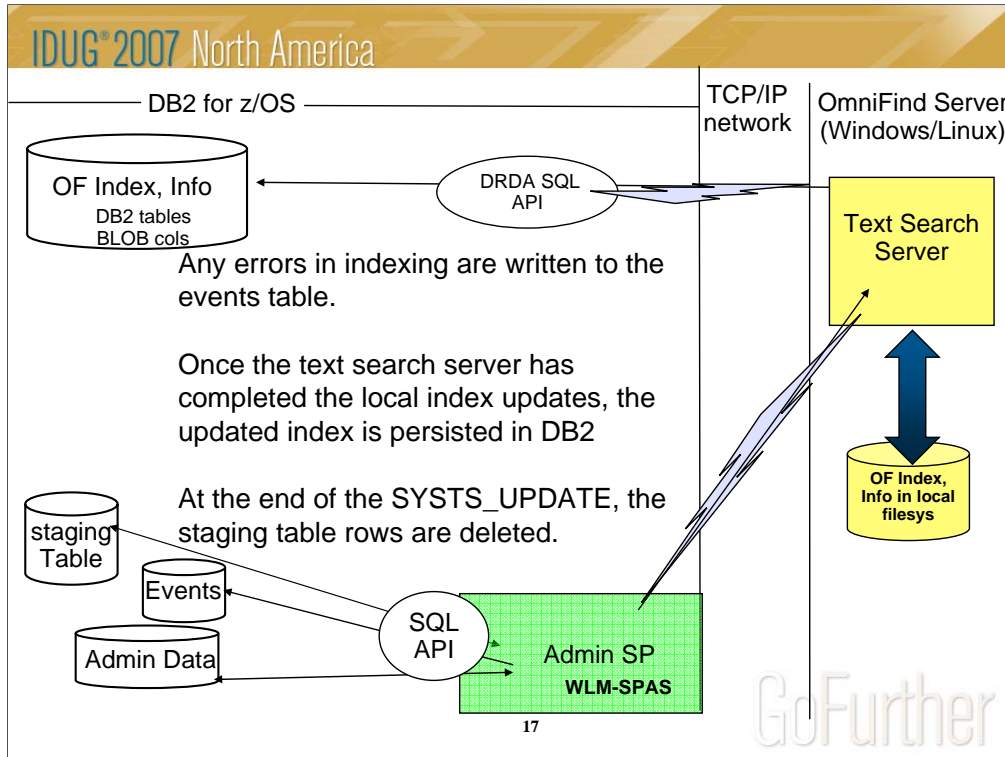
15

GoFurther

Create index processing



The first part of update processing.



The completion of index update processing. The text search server makes a separate connection to DB2 using the type4 Universal Java Driver to persist a binary copy of the index to DB2. The text search index copy in DB2 is the master copy, and can be backed up / restored with other DB2 data.

Using a text index

- Built-in functions CONTAINS() and SCORE ()
- Built-in functions, optimizer does selectivity evaluations on them
 - For example, a more complex WHERE clause may be able to use a DB2 index more efficiently – for example narrowing by state.
- `SELECT CUSTOMER FROM CLAIM_TABLE WHERE CONTAINS(REPORT,"mentos diet coke ");`

Sample query returns all the customers who made a claim for diet coke and mentos accidents!

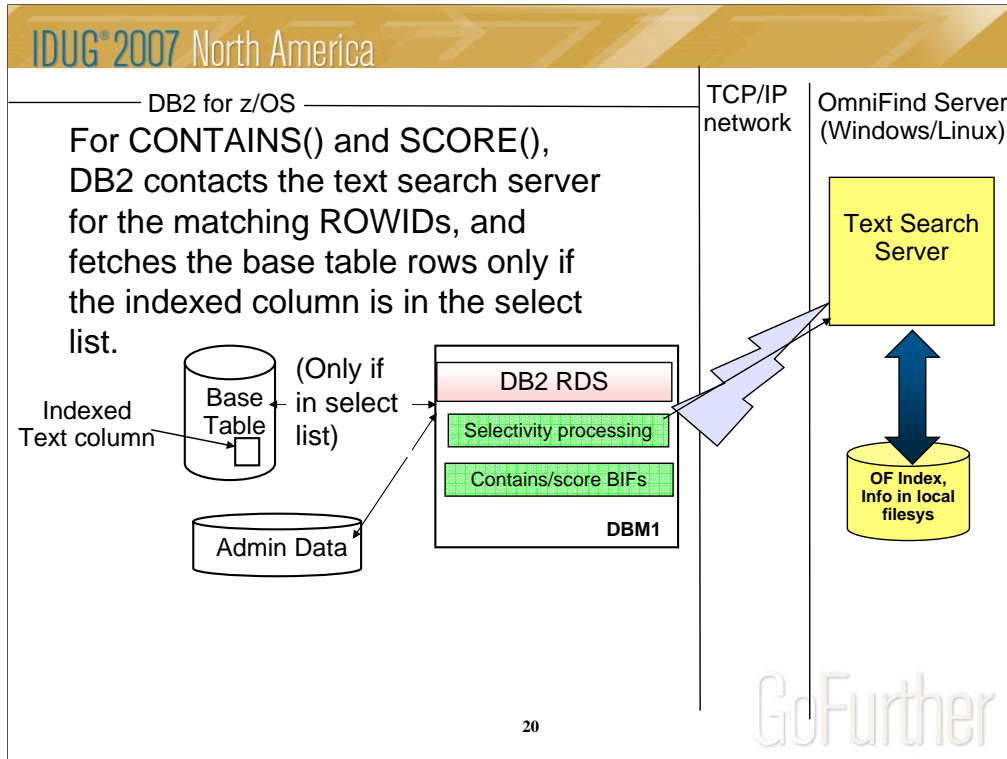
Huh?

Why would an accident
Report contain both
Mentos (candy) and
Diet Coke?

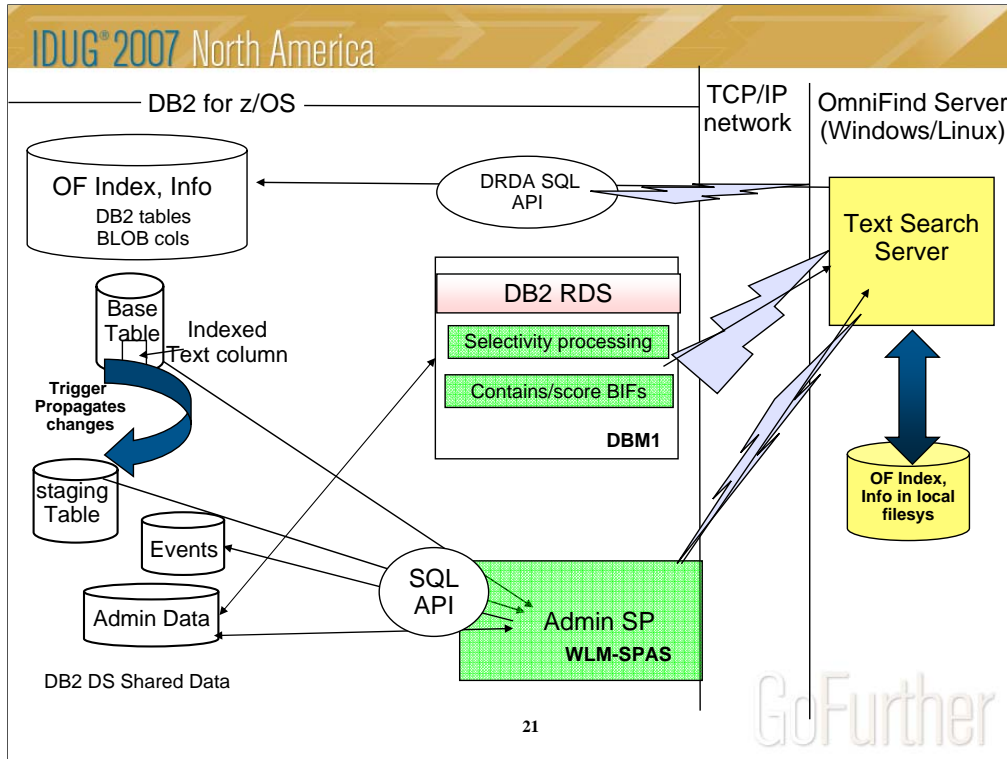


19

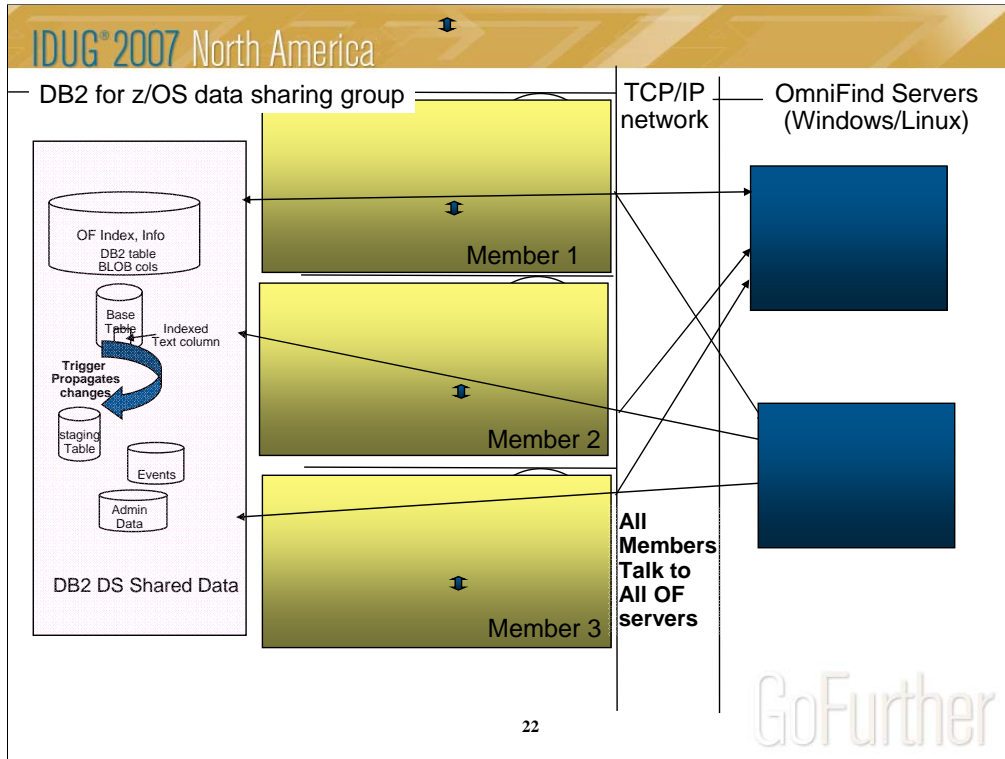
GoFurther



CONTAINS() processing – the most optimized, for queries. The text search server uses the copy it has in its local file system.



A view of the complete text search processing system.



A view of the text search processing system with datasharing and multiple text search servers.

Error Handling

- Server errors exposed in both SQLCA and operator console messages
- SYSPROC.SYSTS_TAKEOVER stored procedure to try again and move the index handling to another available server
- SYSPROC.SYSTS_RESTORE for planned server outages or recovering DB2 to point in time

Serviceability – how OmniFind server errors are exposed to the client, and some administration SPs to aid in availability.

Points to know about text search

1. Requires separately obtained and installed OmniFind server
2. Number of indexes on one server is limited
3. Text indexes are not automatically updated by DB2
 - Base table changes may not immediately be reflected
 - Must invoke SYSPROC.SYSTS_UPDATE
 - administrative enhancements for auto-SP are coming!
4. Text indexes are not dropped when table is dropped

24

GoFuriner

Important points to be considered about the solution.

Sounds Great! When can I have it?

- Our intention is to provide this support for DB2 9 for z/OS, available after General Availability of the product.
- Beta program first – will require DB2 9 for z/OS and an OmniFind server on Linux.

Disclaimer and Trademarks

Information contained in this material has not been submitted to any formal IBM review and is distributed on "as is" basis without any warranty either expressed or implied. Measurements data have been obtained in laboratory environment. **Information in this presentation about IBM's future plans reflect current thinking and is subject to change at IBM's business discretion. You should not rely on such information to make business plans. The use of this information is a customer responsibility.**

IBM MAY HAVE PATENTS OR PENDING PATENT APPLICATIONS COVERING SUBJECT MATTER IN THIS DOCUMENT. THE FURNISHING OF THIS DOCUMENT DOES NOT IMPLY GIVING LICENSE TO THESE PATENTS.

TRADEMARKS: THE FOLLOWING TERMS ARE TRADEMARKS OR ® REGISTERED TRADEMARKS OF THE IBM CORPORATION IN THE UNITED STATES AND/OR OTHER COUNTRIES: AIX, AS/400, DATABASE 2, DB2, e-business logo, Enterprise Storage Server, ESCON, FICON, OS/390, OS/400, ES/9000, MVS/ESA, Netfinity, RISC, RISC SYSTEM/6000, iSeries, pSeries, xSeries, SYSTEM/390, IBM, Lotus, NOTES, WebSphere, z/Architecture, z/OS, zSeries, OmniFind



The FOLLOWING TERMS ARE TRADEMARKS OR REGISTERED TRADEMARKS OF THE MICROSOFT CORPORATION IN THE UNITED STATES AND/OR OTHER COUNTRIES: MICROSOFT, WINDOWS, WINDOWS NT, ODBC, WINDOWS 95

For additional information see ibm.com/legal/copytrade.phtml

1

26

GoFurther

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice. Any references in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.

The licensed program described in this information and all licensed material available for it are provided by IBM under terms of the IBM Customer Agreement, IBM International Program License Agreement, or any equivalent agreement between us.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in the operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

All statements regarding IBM's future direction or intent are subject to change or withdrawal without notice, and represent goals and objectives only. This information may contain examples of data and reports used in daily business operations. To

illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

Session: H03

Text search on DB2 for z/OS data

Summary:

Our intent is to provide needed text search for DB2 text data with all the “ilities” in a cost-effective manner.

Peggy Zagelow

IBM

(e-mail will be provided in session)

