

# DB2® Universal Database Version 8.2 Technical Overview



DB2 Worldwide Technical Sales Support  
IBM Toronto Laboratory

## **DB2 UDB Version 8.2 Technical Overview**

This document contains proprietary information of IBM. It is provided under a license agreement and is protected by copyright law. The information contained in this publication does not include any product warranties, and any statements provided in this manual should not be interpreted as such.

You can order IBM publications online or through your local IBM representative.

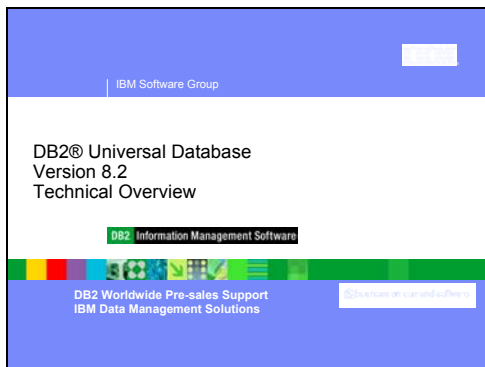
- To order publications online, go to the IBM Publications Center at [www.ibm.com/shop/publications/order](http://www.ibm.com/shop/publications/order)
- To find your local IBM representative, go to the IBM Directory of Worldwide Contacts at [www.ibm.com/planetwide](http://www.ibm.com/planetwide)
- To order DB2 publications from DB2 Marketing and Sales in the United States or Canada, call 1-800-IBM-4YOU (426-4968).

When you send information to IBM, you grant IBM a nonexclusive right to use or distribute the information in any way it believes appropriate without incurring any obligation to you.

© Copyright International Business Machines Corporation 1997 - 2004. All rights reserved. US Government Users Restricted Rights – Use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM Corp.

### **Credits**

Material in this document was prepared by the DB2 Worldwide Pre-sales Support group, along with text developed by the Toronto Information Development group.



## DB2® Universal Database, Version 8.2 Technical Overview

---

DB2® Universal Database (DB2 UDB) Version 8.2 delivers new features to address the ever-increasing demands and requirements on the most important data: our customer's. DB2 UDB Version 8.2 provides new functionality to assist and resolve challenges that you, the customer, experience each day.

Database Administrators can reap immediate benefits from the broadened autonomic (also known as self-managing) computing solutions offered in DB2 UDB Version 8.2. These solutions automate and simplify potentially time-consuming and complex database tasks. For example, this release simplifies performance optimization for specific workloads by introducing the Design Advisor, a self-configuring tool that automates database design.

DBAs in Business Intelligence environments can take advantage of several DB2 UDB Version 8.2 features including runstats enhancements and online import. RUNSTATS now exploits page and row-level sampling for faster statistics gathering. Faster statistics processing means more frequent updates are feasible, resulting in better optimizer plans and faster BI queries. This feature is used by DB2 in recent TPC-H benchmark results. The Online Import Utility exploits fine-grained locking for greater concurrency in online table loads. It can be applied to trickle-feeding a real-time warehouse.

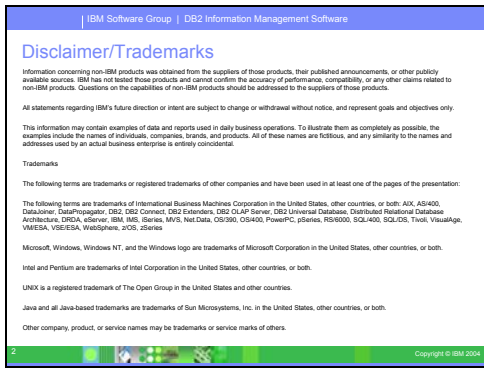
For application developers, DB2 UDB Version 8.2 delivers a significant amount of new capabilities as well as further integration of DB2 tooling into the Microsoft® .NET and WebSphere Studio/Java environments. This simplifies the development and deployment of DB2 applications allowing application developers to take advantage of the openness, performance and scalability of DB2, without regard to the back-end database or the chosen application architecture.

IT managers and DBAs will benefit from the integration of industry-proven high availability disaster recovery (HADR) technology available in DB2 UDB Version 8.2. Line of business managers and the enterprise itself, benefit the most as critical applications face less risk of downtime. HADR technology will help to ensure that the enterprise is business-enabled, around-the-clock without requiring time-consuming application changes.

DBAs and security administrators can provide greater security for their information assets by utilizing the pluggable security architecture. Custom security plug-ins allow DB2 to be plugged into any security scheme that utilizes userids and passwords for authentication, and userids and group membership to implement authorities.

Integrating DB2 into environments utilizing a Windows® security schema has been made easier by improving support for Active Directory.

Regardless of your role or the size of your business, DB2 UDB Version 8.2 has exciting new features that can assist you with your daily challenges.



## Disclaimer and Trademarks

This presentation may contain other proprietary notices and copyright information, the terms of which must be observed and followed. 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.

### Trademarks

The following terms are trademarks or registered trademarks of other companies and have been used in at least one of the pages of the presentation:

The following terms are trademarks of International Business Machines Corporation in the United States, other countries, or both: AIX, AS/400, DataJoiner, DataPropagator, DB2, DB2 Connect, DB2 Extenders, DB2 OLAP Server, DB2 Universal Database, Distributed Relational Database Architecture, DRDA, eServer, IBM, IMS, iSeries, MVS, Net.Data, OS/390, OS/400, PowerPC, pSeries, RS/6000, SQL/400, SQL/DS, Tivoli, VisualAge, VM/ESA, VSE/ESA, WebSphere, z/OS, zSeries

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

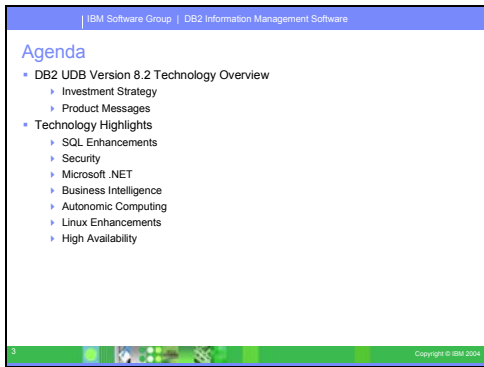
Intel and Pentium are trademarks of Intel Corporation in the United States, other countries, or both.

UNIX is a registered trademark of The Open Group in the United States and other countries.

Java and all Java-based trademarks are trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.

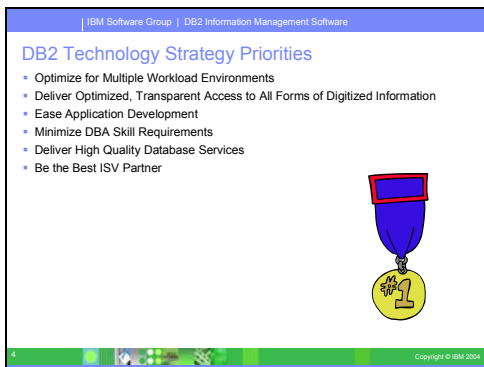
Other company, product, or service names may be trademarks or service marks of others.

© Copyright IBM Corporation 2004. All rights reserved.



## Agenda

This presentation is split up into two sections. The first section deals with the investment and strategic priorities of the DB2 UDB Version 8.2 release, while the second section deals with some of the major features found within the product. Note that the DB2 UDB Version 8.2 has many more features than what are found here. These slides only cover some of the significant features of this release.



## DB2 Technology Strategy Priorities

The DB2 UDB Version 8.2 release was developed around a number of key priorities:

- Optimize for Multiple Workload Environments

The Database needs to give customers excellent performance, whether it is running in an OLTP or decision support environment.

- Deliver Optimized, Transparent Access to All Forms of Digitized Information

Information Integration is a key technology found in the DB2 family of products. Additional technology was developed to improve the performance and scope of the integration capabilities within the product.

- Ease Application Development

With tools like Eclipse and Visual Studio becoming more and more popular with developers, DB2 needs to make sure it exploits these environments and delivers superior ease of user.

- Minimize DBA Skill Requirements

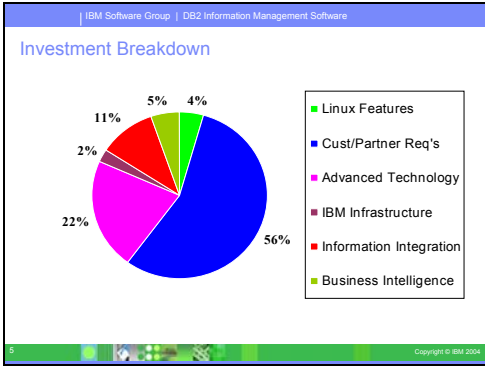
DB2 introduces automation to many of the daily routines that a DBA must take care of this. This will help free DBA resources to work on more design, tuning, and development issues.

- Deliver High Quality Database Services

In conjunction with our DB2 lab services team, and other consulting partners, DB2 continues to provide high-quality services for our customers.

- Be the Best ISV Partner

ISV and customer requirements are key to improving the DB2 product. DB2 UDB Version 8.2 reflects the importance of feedback from our partners and customers.



### Investment Breakdown

You can see from the following pie-chart that the majority of the features found within the DB2 UDB Version 8.2 were based on requirements and feedback from our customers and business partners/ISV's. The types of requirements included support for features that would help in the migration from other database platforms, management (ease-of-use) features, and tools and wizards to make it easier to implement many of the advanced features found within DB2. The investment is broken down as follows:

- 56% in customer and partner requirements
- 22% in advanced technology (like Autonomic computing)
- 11% in Information Integration support
- 5% is business intelligence features (wizards and improved tooling)
- 4% in Linux exploitation (mainly 2.6 Kernel exploitation)
- 2% in cross-product integration (Websphere, DB2, Tivoli)

**SQL Enhancements**

```

with t1(t1.col, avg_sal) as
(
  select t1.col, avg(s.amount) from
  store_sales tablespace system(2)
  group by t1.col
),
t2(t1.col, avg_sal) as
(
  select t1.col, avg(s.amount) from
  store_sales tablespace bernoulli(2)
  group by t1.col
)
select t1.col, t1.avg_sal, t2.avg_sal from
t1, t2
where
t1.avg_sal > t2.avg_sal and
t1.col = t2.col
order by t1.col
    
```

### SQL Enhancements

This section will examine the various SQL enhancements that have been introduced as part of DB2 UDB Version 8.2 product.

**Internal Stored Procedures**

- Internal SQL PL eliminates the current dependency on a C compiler for the creation of SQL procedures
  - ▶ Eliminating this dependency has been the most consistent PSM-related requirement
- New PSM generates a plan and bytecode to execute in the engine
  - ▶ Performance should be comparable to existing implementation
- Bytecode stored in the catalog and loaded at execution time


### Native PSM

Beginning with DB2 UDB, the creation of SQL procedures does not require a C or C++ compiler on the server; therefore C or C++ compiler setup is not required. When you create an SQL procedure its procedural statements are converted to a native representation that is stored in the database catalogs, as is done with other SQL statements. When an SQL procedure is called, the native representation is loaded from the catalogs and the DB2 engine executes the procedure.

IBM Software Group | DB2 Information Management Software

### Support of CALL statement in a Trigger body

- The CALL statement was restricted to external routines and applications
- In DB2 UDB Version 8.2, the CALL statement can be used in inline SQL PL
  - ▶ CALL be used within a Trigger, a SQL Table Function, a SQL Method or in a dynamic compound statement
  - ▶ The result of the RETURN statement of the procedure can be retrieved through the GET DIAGNOSTICS RETURN\_STATUS statement.



Copyright © IBM 2004

## Support of Call Statement in Trigger Body

---


With DB2 UDB Enterprise Edition, you can now invoke procedures from triggers, or any other dynamic compound statement, in single partition environments, by executing a CALL statement that references a procedure within a trigger action. The CALL statement is used to execute SQL and external procedures.

The invocation of a procedure in a trigger-action enables you to encapsulate complex logic in your trigger. The logic can include operations on other tables in the database, or operations external to the database such as sending an e-mail or writing an audit record to a file in the file system of the database server.

IBM Software Group | DB2 Information Management Software

### Support for REOPT Bind option

- Occasionally running dynamic SQL queries using parameter markers will cause performance problems
  - SQL queries could perform poorly during execution if the values used for the input variables (i.e. parameter markers, host variables, and special registers) are outside the predictive range of the default filter factor estimates.
- New REOPT option causes the compilation of the statement is deferred until run-time when the access plan can be optimized using the actual values for the input variables.



Copyright © IBM 2004

### Support for REOPT Bind option

The bind option REOPT enables reoptimization of static and dynamic SQL statements. Specifically, the access path for an SQL statement containing host variables, special registers, or parameter markers is optimized using the values of these variables, rather than the default estimates determined by the compiler. This optimization takes place at query execution time when these values are available.

You can set the bind option REOPT to one of the following three values:

- NONE

The values of any host variables, parameter markers, or special registers in an SQL statement are not used to optimize this statement's access path. The default estimates for these variables are used instead. This is the default behavior.

- ONCE

When the statement is first executed, the access path for a given SQL statement is optimized using the actual values for any host variables, parameter markers, or special registers. This access path will be used in all subsequent executions of the statement.


- ALWAYS

Each time that a statement containing host variables, parameter markers, or special registers is executed, the values of these variables are used to optimize the access path for the statement.

IBM Software Group | DB2 Information Management Software

### Larger SQL Statements

- The 64KB limit on statement size currently limits the total size of statements (e.g. 'CREATE PROCEDURE' or 'CREATE TRIGGER') which therefore limits the size of the object.
- This limit has been increased in order to allow partners who have the majority of their application logic in stored procedures or triggers, or need to use other large statements.
  - Currently partners have to break up the procedures and make them more complex in DB2 than in other RDBMS
  - In some cases, it is not possible to migrate a trigger or psm statement from other RDBMS to DB2 with DB2's 64K maximum statement size
- The limit will be increased to 2MB.



Copyright © IBM 2004

### Larger SQL Statements

DB2 UDB Version 8.2 provides an increased SQL statement size limit of 2 MB, which can limit the size of some created database objects (for example, CREATE PROCEDURE or CREATE TRIGGER statements). The new 2 MB limit allows you to use large statements when your application logic is in stored procedures or triggers.

The increase in the statement size allows you to migrate a trigger or stored procedure statement from another RDBMS to DB2 UDB.





**Nested Savepoints**

- This feature expands the support for external savepoints first introduced in DB2 UDB Version 7 by providing a number of new abilities:
  - The ability to nest savepoints within another savepoint
  - The ability to use savepoint-related statements within atomic compound SQL statements
  - The ability to use atomic compound SQL statements within an active savepoint
  - The ability to create a new savepoint level within a stored procedure invocation
  - The automatic creation of a new savepoint level within each call to a user-defined function (UDF)

```

CREATE TABLE T1 (C1 INT);
SAVEPOINT SP1;
INSERT INTO T1 VALUES (1);
COMMIT WORK;
SAVEPOINT SP2;
INSERT INTO T1 VALUES (2);
COMMIT WORK;
SAVEPOINT SP3;
INSERT INTO T1 VALUES (3);
ROLLBACK TO SAVEPOINT SP1;
INSERT INTO T1 VALUES (4);
    
```

IBM Software Group | DB2 Information Management Software

## Nested Savepoints

DB2 UDB Version 8.2 now supports the nesting of savepoints. This means you can set up a savepoint within another savepoint. DB2 UDB enables you to set as many levels of nested savepoints as you require.

With nested savepoints, an application can have multiple levels of savepoints active simultaneously, and the application can roll back to any active savepoint, as required. A rollback to a particular savepoint statement also releases any active nested savepoints within the savepoint being rolled back.

**Security Enhancements**




IBM Software Group | DB2 Information Management Software

## Security Enhancements

There are a number of security enhancements in DB2 UDB Version 8.2 that improve the integration of DB2 into the Windows environment, as well as add some new security options.

**Data Encryption**

- Encrypts user data during client/server communication over the network
- Two new authentication types
  - SQL\_AUTHENTICATION\_DATAENC
    - connections must use the data encryption
  - SQL\_AUTHENTICATION\_DATAENC\_CMP
    - allows for a compatibility mode with down-level products that do not support the new authentication type
    - they will be allowed to connect with SERVER\_ENCRYPT and not encrypt user data
- Encrypts all sensitive data



IBM Software Group | DB2 Information Management Software

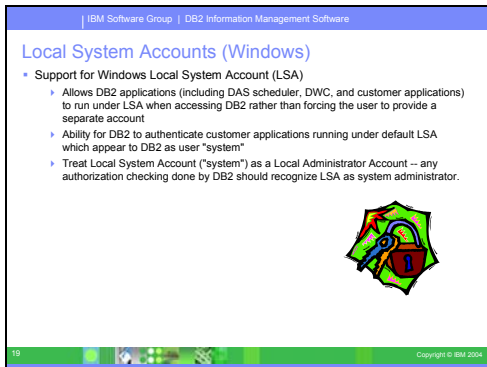
## Data Encryption

To improve the security of user data, there are two new authentication types that allow you to manage the encryption of user data.

One new authentication type, SQL\_AUTHENTICATION\_DATAENC, requires connections to use the data encryption.

The other new authentication type, SQL\_AUTHENTICATION\_DATAENC\_CMP, allows for a compatibility mode with downlevel products that do not support the new authentication type. For such products, the use of this authentication type allows a connection using SERVER\_ENCRYPT where the encryption of user data is not required. Any product that supports SQL\_AUTHENTICATION\_DATAENC will be forced to use it.





## Local System Accounts (Windows)

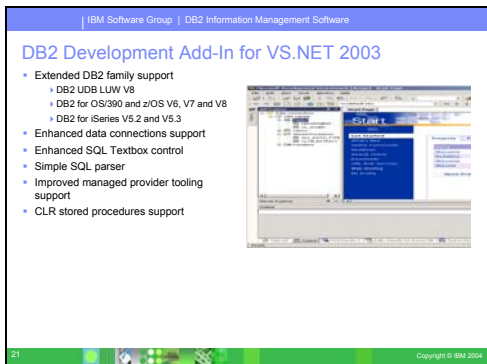
Support for Windows Local System account (LSA) is provided for both the various DB2 services, as well as for applications that access them. The DB2 installation process allows DB2 services to run under SYSTEM (local system account) as an alternative to running under a dedicated user account. You can also perform the DB2 installation itself from a process that is running under the context of the Local System account.

This feature makes it easier to deploy DB2 by avoiding, where appropriate, the complexities associated with handling of userids and passwords. Another enhancement is the ability for DB2 applications that are running in the context of the LSA to access and use the local DB2 server.



## Microsoft Development Enhancements

DB2 includes support for Visual Studio 2003 as well as Eclipse (Java). There are numerous enhancements in this release to help developers in the Microsoft environment work with DB2.



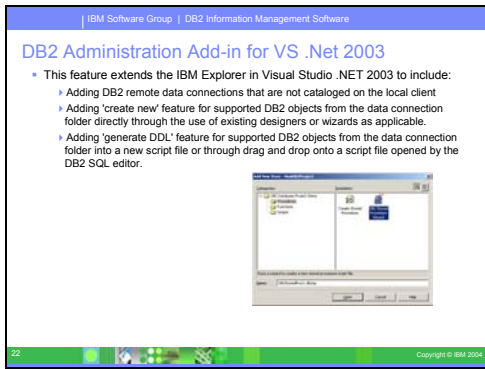
## DB2 Development Add-In for VS.NET 2003

With the DB2 Development Add-In for Microsoft® Visual Studio .NET, you can now create the following DB2 SQL objects using either a wizard or the built-in code editor:

- tables
- indexes
- views
- triggers

In addition, you can now develop DB2 common language runtime (CLR) procedures from existing methods in a .NET managed language, such as C# and Visual Basic. A DB2 .NET CLR procedure is a user-created DB2 external routine that has its procedure- body implemented by a method in a .NET assembly compiled from source code written in a .NET managed language.

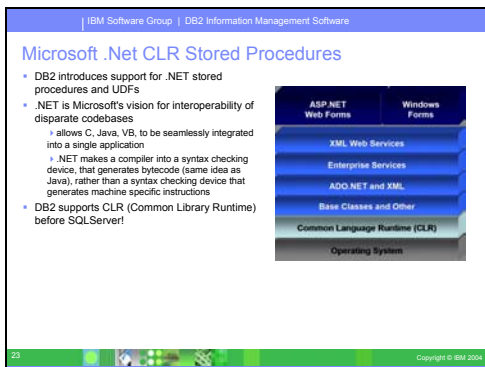
Lastly, you can perform source-level debugging of SQL procedures using the DB2 database project.



## DB2 Administration Add-in for VS .Net 2003

The DB2 .NET Data Provider is an extension of the ADO.NET interface that allows .NET applications to access a DB2 database through a secure connection, execute commands, and retrieve results. DB2 UDB Version 8.2 features the following improvements to the DB2 .NET Data Provider:

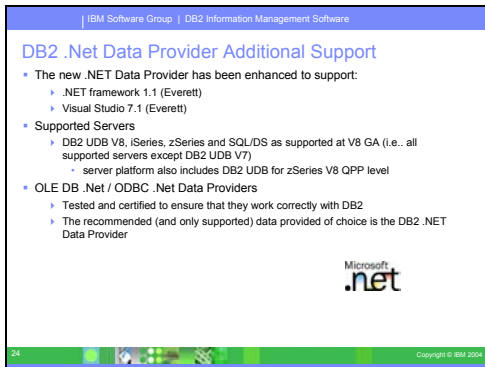
- Performance is faster.
- Support for the Microsoft® .NET Framework, Version 1.1. This includes support for the following new System.Data features: the DB2DataReader.HasRows and DB2Connection.EnlistDistributedTransaction methods.
- A new ConnectionString keyword to specify the isolation level.
- Four new DB2Connection properties enable DBAs to attribute workload to a particular source. The new properties are:
  - DB2Connection.ClientUser - client user ID
  - DB2Connection.ClientWorkStation - client workstation name
  - DB2Connection.ClientAccountingInformation - client accounting string
  - DB2Connection.ClientApplicationInformation - client application name
- .NET applications can now access the following database management systems through the DB2 .NET Data Provider:
  - DB2 Universal Database™ Version 5, Release 1 (or later) for AS/400® and iSeries™, through DB2 Connect™ development enhancements
  - DB2 Universal Database Version 7.3 (or later) for VSE & VM, through DB2 Connect



## Microsoft .Net CLR Stored Procedures

DB2 Universal Database allows you to create routines (procedures, user- defined functions (UDFs), and methods) in SQL and in other programming languages such as C, C++, Java™, and OLE. Now you can create procedures and UDFs using any .NET common language runtime (CLR) compatible programming language, including C#, Visual Basic, managed C++, and all other CLR compatible languages.

You can create CLR routines in the same way as other external (non-SQL) routines are created: by executing a CREATE statement that associates a database routine signature with a .NET assembly residing on the database server. You can use the routines to encapsulate commonly used database operations and logic, to extend the functionality of SQL, and to improve the performance of client applications.

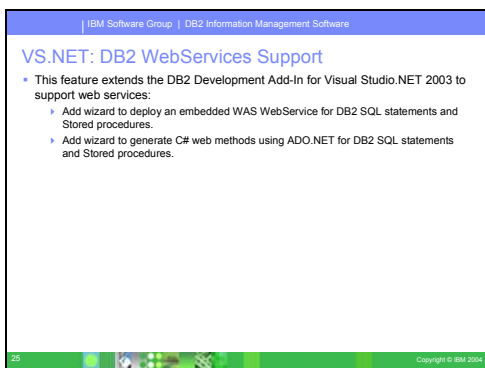


## DB2 .Net Data Provider Additional Support

In addition to the .NET Data Provider, DB2 includes an OLE DB .NET Data Provider and a ODBC .NET Data Provider.

The OLE DB .NET Data Provider uses the IBM® DB2 OLE DB Driver. Therefore, the connection string keywords supported by the OLE DB .NET Data Provider are the same as those supported by the IBM OLE DB Provider for DB2. Also, the OLE DB .NET Data Provider has the same restrictions as the IBM DB2 OLE DB Provider.

The ODBC .NET Data Provider makes ODBC calls to the DB2 data source using the IBM DB2 CLI Driver. Therefore, the connection string keywords supported by the OLE DB .NET Data Provider are the same as those supported by the IBM DB2 CLI Driver.

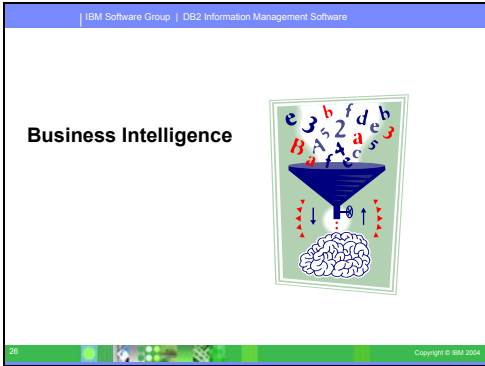


## VS.NET: DB2 WebServices Support

DB2 UDB Version 8.2 is a significant release for application developers with significant new SQL features, better support for open standard APIs, web services and seamless integration with the key development environments for both Java™ and Microsoft® .NET.

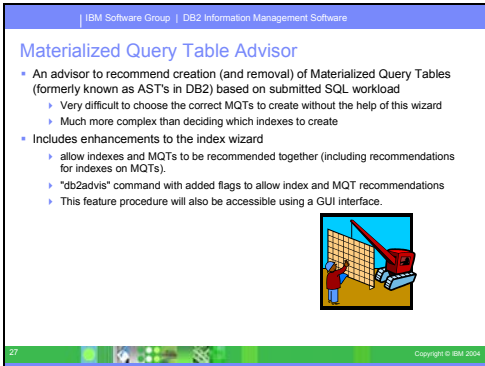
It builds on the strong momentum we've gained since the release of DB2 UDB Version 8 for .NET support in both driver enablement and tooling. The DB2 .NET Data Provider has much improved performance and now supports DB2 UDB server on iSeries™. DB2 UDB servers also support stored procedures written in Microsoft CLR (Common Language Runtime) compliant languages such as Visual Basic .NET and C#.

The DB2 tools add-ins to Microsoft Visual Studio .NET are further augmented with schema operation capabilities, wizards for developing and deploying CLR stored procedures as well as WORF (Web services Object Runtime Framework) web services.



## Business Intelligence

DB2 already has a significant number of features that support business intelligence workloads. In DB2 UDB Version 8.2, a number of wizards have been introduced that will make it easier for the DBA or developer to implement many of these advanced BI features.



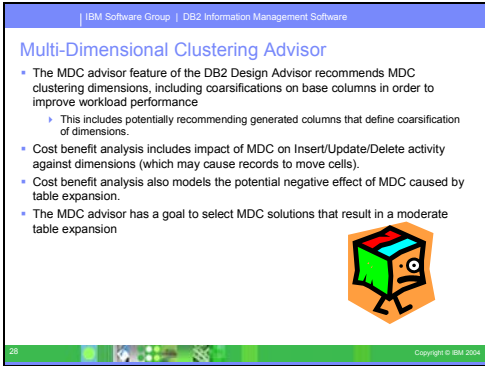
## Materialized Query Table Advisor

The DB2 Design advisor is a tool that can help you significantly improve your workload performance. The Design advisor identifies all of the objects needed to improve the performance of your workload. Given a set of SQL statements in a workload, the Design advisor will generate recommendations for new materialized query tables.

While designing your database, use the Design advisor to generate design alternatives. After your database is set up, you can use the Design advisor to help you meet the following tuning goals:

- Improve performance of a particular statement or workload.
- Improve general database performance, using the performance of a sample workload as a gauge.
- Improve performance of the most frequently executed queries, for example, as identified by the Activity Monitor.
- Determine how to optimize the performance of a new key query.
- Respond to Health Center recommendations regarding shared memory utility or sort heap problems in a sort-intensive workload.
- Find objects that are not used in a workload.





## Multi-Dimensional Clustering Advisor

The DB2 Design Advisor (db2advis), formerly known as the Index Advisor, has an MDC feature. This feature recommends clustering dimensions for use in an MDC table, including coarsifications on base columns in order to improve workload performance. The term coarsification refers to a mathematic expression to reduce the cardinality (the number of distinct values) of a clustering dimension. A common example of a coarsification is the date where coarsification could be by date, week of the date, month of the date, or quarter of the year.

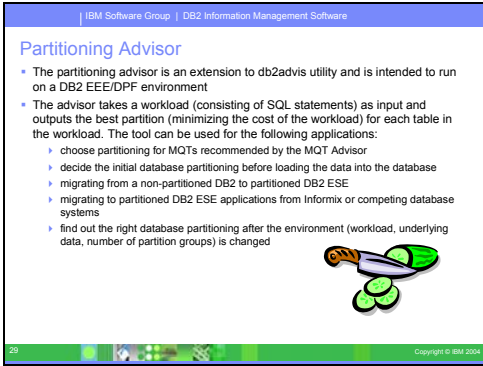
The recommendation includes identifying potential generated columns that define coarsification of dimensions. The recommendation does not include possible block sizes. The extent size of the table space is used when making recommendations for MDC tables. The assumption is that the recommended MDC table will be created in the same table space as the existing table, and will therefore have the same extent size. The recommendations for MDC dimensions would change depending on the extent size of the table space since the extent size impacts the number of records that can fit into a block or cell. This directly affects the density of the cells.

Only single-column dimensions, and not composite-column dimensions, are considered, although single or multiple dimensions may be recommended for the table. The MDC feature will recommend coarsifications for most supported data types with the goal of reducing the cardinality of cells in the resulting MDC solution. The data type exceptions include: CHAR, VARCHAR, GRAPHIC, and VARGRAPH data types. All supported data types are cast to INTEGER and are coarsified through a generated expression.

The goal of the MDC feature of the DB2 Design Advisor is to select MDC solutions that result in improved performance. A secondary goal is to keep the storage expansion of the database constrained to a modest level. A statistical method is used to determine the maximum storage expansion on each table.

The analysis operation within the advisor includes not only the benefits of block index access but also the impact of MDC on insert, update, and delete operations against dimensions of the table. These actions on the table have the potential to cause records to be moved between cells. The analysis operation also models the potential performance impact of any table expansion resulting from the organization of data along particular MDC dimensions.



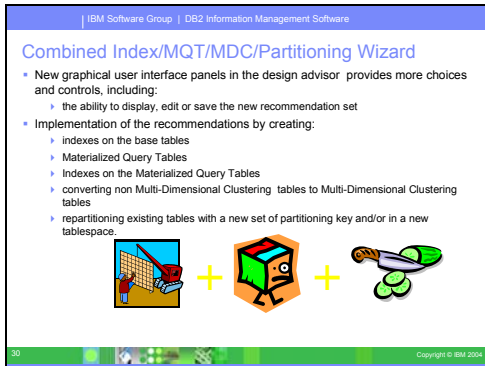


## Partitioning Advisor

The DB2 Design advisor is a tool that can help you select which partitions to create for a complex workload. Given a set of SQL statements in a workload, the Design advisor can generate recommendations for:

- Determining initial database partitioning before loading data into a database.
- Migrating from a non-partitioned DB2 database to a partitioned DB2 database.
- Migrating to DB2 in a partitioned environment from another database product.
- Evaluating partitions that have been generated manually.
- Repartitioning of tables for better performance.

You can choose to have the Design advisor GUI tool implement some or all of these recommendations immediately or schedule them for a later time.



## Combined Index/MQT/MDC/Partitioning Wizard

The DB2 Design advisor is a tool that can help you significantly improve your workload performance. The task of selecting which indexes, clustering dimensions, or partitions to create for a complex workload can be quite daunting. The Design advisor identifies all of the objects needed to improve the performance of your workload. Given a set of SQL statements in a workload, the Design advisor will generate recommendations for:

- new indexes
- new materialized query tables
- conversion to multidimensional clustering tables
- repartitioning of tables
- deletion of objects unused by the specified workload


You can choose to have the Design advisor GUI tool implement some or all of these recommendations immediately or schedule them for a later time. Using either the Design advisor GUI or the command-line tool, the Design advisor can help simplify the following tasks:

- Planning for or setting up a new database or partitioning structure
  - Generate design alternatives in a test environment for partitioning, indexes, MQTs, and MDC tables.
  - Determine initial database partitioning before loading data into a database.
  - Assist in migrating from a non-partitioned DB2 database to a partitioned DB2 database.
  - Assist in migrating to DB2 in a partitioned environment from another database product.
  - Evaluate indexes, MQTs or partitions that have been generated manually.
- Workload performance tuning
  - Improve performance of a particular statement or workload.
  - Improve general database performance, using the performance of a sample workload as a gauge.
  - Improve performance of the most frequently executed queries, for example, as identified by the Activity Monitor.
  - Determine how to optimize the performance of a new key query.
  - Respond to Health Center recommendations regarding shared memory utility or sort heap problems in a sort-intensive workload.
  - Find objects that are not used in a workload.

IBM Software Group | DB2 Information Management Software

### RUNSTATS Enhancements

- Sampling Support
  - ▶ Boosts the performance of RUNSTATS by exploiting page level and row level sampling
  - ▶ Page level sampling can improve performance because it reduces the number of table pages that need to be accessed to collect the statistics
- New RUNSTATS options flag will be added to indicate the type of sampling
  - RUNSTATS ON TABLE <table> WITH DISTRIBUTION TABLESAMPLE BERNOULLI (30)
  - RUNSTATS ON TABLE <table> WITH DISTRIBUTION TABLESAMPLE BERNOULLI (30) REPEATABLE
  - RUNSTATS ON TABLE <table> WITH DISTRIBUTION TABLESAMPLE SYSTEM (1.5)
- Support for utilities using a statistics profile
  - ▶ There are some utilities such as LOAD and REORGCHK which can also specify that statistics are to be gathered.



31 Copyright © IBM 2004

## RUNSTATS Enhancements


Table statistics are used by the query optimizer in selecting the best access plan for any given query, so it is important that statistics remain current to accurately reflect the state of a table at any given time. As the activity against a table increases, so should the frequency of statistics collection. With the increasing size of databases, it is becoming more important to find efficient ways to collect statistics. Random sampling of table data on which to collect statistics can reduce the amount of time that it takes to collect statistics. For I/O bound or CPU bound systems, the performance benefits can be enormous. The smaller the sample, the faster statistics collection completes.

Starting in DB2 UDB Version 8.2, the RUNSTATS command provides the option to collect statistics on a sample of the data in the table by using the TABLESAMPLE option. This feature can increase the efficiency of statistics collection since sampling uses only a subset of the data. At the same time, the sampling methods ensure a high level of accuracy.

IBM Software Group | DB2 Information Management Software

### Spatial Geodetic feature

- The DB2 Geodetic Supplement for Spatial Extender for DB2 UDB Version 8.2 will provide DB2 Customers with the ability to treat the earth as one seamless globe
- The DB2 Geodetic Supplement enables the existing Spatial Types to work with the understanding that the Earth is a closed sphere.
  - ▶ This enables the creation of trans-polar and trans-dateline seamless spatial values and enables trans-polar and trans-dateline seamless queries
- The DB2 Geodetic Supplement only adds new functionality to complement the current functionality in the Spatial Extender



32 Copyright © IBM 2004

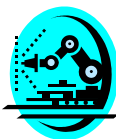
## Spatial Geodetic feature

DB2 Geodetic Extender uses the same spatial data types and functions as DB2 Spatial Extender to store and manipulate geographic data in a DB2 database. DB2 Geodetic Extender treats the Earth as a globe so that you can run seamless queries on geographic coordinates around the poles and coordinates that cross the date line. Distance and area calculations are precise, regardless of location on the surface of the Earth.

DB2 Geodetic Extender is a separately- priced feature, and requires that you purchase a separate license key.

IBM Software Group | DB2 Information Management Software

### Autonomic Computing Enhancements

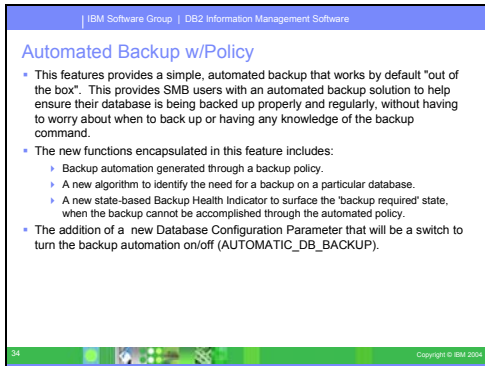


33 Copyright © IBM 2004

## Autonomic Computing Enhancements

If you want to improve the productivity and effectiveness of your database administrators, check out the new manageability features in DB2 UDB Version 8.2.

An exciting example of these features is the Design advisor, which assists DBAs in making optimal and comprehensive database design decisions. Many DBAs would agree that the decisions made with respect to the design of a database are some of the most challenging, time consuming, and critical to make. This self-configuring tool greatly simplifies the design process by using workload, database, and hardware information to recommend specific performance acceleration options for routine design tasks.



## Automated Backup w/Policy

A database may become unusable due to a wide variety of hardware or software failures. Automatic database backup simplifies database backup management tasks for the DBA by always ensuring that a recent full backup of the database is performed as needed. It determines the need to perform a backup operation based on one or more of the following measures:

- You have never completed a full database backup
- The time elapsed since the last full backup is more than a specified number of hours
- The transaction log space consumed since the last backup is more than a specified number of 4 KB pages (in archive logging mode only).

Protect your data by planning and implementing a disaster recovery strategy for your system. If suitable to your needs, you may incorporate the automatic database backup feature as part of your backup and recovery strategy.

If the database is enabled for roll-forward recovery (archive logging), then automatic database backup can be enabled for either online or offline backup. Otherwise, only offline backup is available. Automatic database backup supports disk, tape, Tivoli® Storage Manager (TSM), and vendor DLL media types.

Through the Configure Automatic Maintenance wizard in the Control Center or Health Center, you can configure:

- The requested time or number of log pages between backups
- The backup media
- Whether it will be an online or offline backup.


If backup to disk is selected, the automatic backup feature will regularly delete backup images from the directory specified in the Configure Automatic Maintenance wizard. Only the most recent backup image is guaranteed to be available at any given time. It is recommended that this directory be kept exclusively for the automatic backup feature and not be used to store other backup images.

The automatic database backup feature can be enabled or disabled by using the `auto_db_backup` and `auto_maint` database configuration parameters. In a multiple database partitioned environment, the automatic database backup runs on each partition if the database configuration parameters are enabled on that partition.

IBM Software Group | DB2 Information Management Software

### Self-tuning Backup and Restore

- Backup and Restore will automatically choose an optimal value for, # buffers, Parallelism, Buffer size whenever these values are not explicitly specified
- Any one or combination of these parameters can be explicitly specified
- For those customers using CLP, if a value for BUFFERS, BUFFER and PARALLELISM is not explicitly specified then a value of 0 will be assumed and the utility will choose optimal values.
- The BACKBUFSZ and RESTBUFSZ in the DBM CFG will be ignored. If a customer wished to use the values they must explicitly specify the value on the utility invocation



35 Copyright © IBM 2004

## Self-tuning Backup and Restore

DB2 will now automatically choose the number of buffers, the buffer size and the parallelism settings for both backup and restore operations. The values chosen are based on the amount of memory available, the number of processors available and the database configuration. The objective is to minimize the amount of time required for backup and restore operations to complete. The BACKUP DATABASE and RESTORE DATABASE commands will automatically choose an optimal value for the following parameters whenever they are not explicitly specified:

- WITH num-buffers BUFFERS
- PARALLELISM n
- BUFFER buffer-size


For restore database operations, a multiple of the buffer size used for the backup operation will always be used.

The values specified by the database manager configuration parameters BACKBUFSZ and RESTBUFSZ are ignored. If you want to use these values, you must explicitly specify them when you issue the BACKUP DATABASE or RESTORE DATABASE command.

IBM Software Group | DB2 Information Management Software

### Inclusion of log files in online backup images

- Current architecture of the backup and restore utilities requires that the backup image and the required log files be shipped as separate objects
  - Possibility of losing the log files required for recovery
- Online backup images of databases and tablespaces will include the set of log files necessary to restore and recover the backup image to a consistent point in time.
  - Require to have a single object to ship to disaster recovery sites that would allow the customer to bring the database back to a consistent point in time.
- New INCLUDE LOGS/EXCLUDE LOGS on ONLINE backup



36 Copyright © IBM 2004

## Inclusion of log files in online backup images


When creating an online backup image, you can include the log files required to restore and recover a database in the image. This means that if you need to ship backup images to a disaster recovery site, you do not have to send the log files separately or package them together yourself. Further, you do not have to decide which log files are required to guarantee the consistency of an online backup, and you will not be able to delete the log files that pertain to a particular backup image.

To use this feature, specify the INCLUDE LOGS option of the BACKUP DATABASE command. When you specify this option, the backup utility will truncate the currently active log file and copy the necessary set of log extents into the backup image.

IBM Software Group | DB2 Information Management Software

### Integrated and automated log file management

- DB2 introduces autonomic log management issues
- Allows DB2 to manage the use, archive, retrieval, and eventual deletion of log files with minimal user interaction
  - Many of the requirements came from ISVs
- New log manager introduced in DB2 UDB Version 8.2
- New History file entries
  - Indicates when log file are archived, and where they are archived
  - Can determine which database backup, and exactly which log files are required to restore to any given point in time

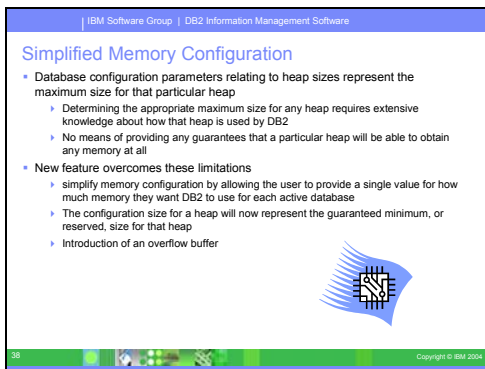


37 Copyright © IBM 2004

## Integrated and automated log file management

Other manageability enhancements in DB2 UDB Version 8.2 include self-healing features such as the Health Center Recommendation advisor and automated log file management; and self-tuning features for backup and restore operations as well as the ability to throttle backup operations and statistics collection.

This facilities allows DB2 to manage the use, archive, retrieval, and eventual deletion of log files with minimal user interaction. The history file entries indicate when log file are archived, and where they are archived. With this information, DB2 can determine which database backup, and exactly which log files are required to restore to any given point in time.

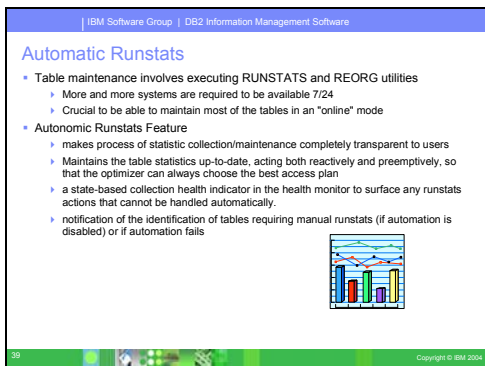


## Simplified Memory Configuration

There are several disadvantages to configuring database heap sizes as maximum hard limits on memory usage:

- Determining the appropriate maximum size for a particular database heap requires extensive knowledge about how the heap will be used by .
- Setting a limit on the amount of memory that can be allocated to a heap does not guarantee that memory will be available to that heap when it is required.
- The maximum size of a heap represents a hard limit, so that memory allocation for a heap cannot exceed this limit, even for a short period of time.

To address these limitations, database shared memory heaps and instance shared memory heaps will be configured and interpreted differently. To simplify configuration of these heaps, you will now be able to specify a single value for how much memory DB2 uses for each active database and specify guaranteed minimums reserved for individual heaps. Interpreting configuration parameter values for heap sizes as guaranteed minimums will reduce the number of memory errors experienced during memory usage peaks.



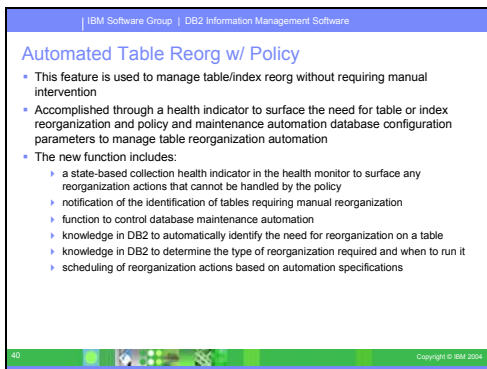
## Automatic Runstats

Another example of a new time saving feature for a critical DBA task is the ability to automate database maintenance activities such as backup, table defragmentation (reorg), and table statistics gathering (runstats). DB2 Universal Database™ version 8.2 provides the ability to easily schedule these tasks to run automatically based on factors determined by the DBA such as available maintenance windows, percentage of log files used, and so on.

The RUNSTATS utility provides an option to register and use a statistics profile, which is a set of options that specify which statistics are to be collected on a particular table, for example, table statistics, index statistics, or distribution statistics.

Statistics profiles can now be generated automatically by the DB2 automatic statistics profiling feature. When this feature is enabled, information about database activity is collected and stored in a query feedback warehouse. Based on this data, a statistics profile is generated. Enabling this feature can alleviate the problem of uncertainty about which statistics are relevant to a particular workload and permits the collection of the minimal set of statistics to provide optimal database workload performance.

This feature can be used together with the automatic statistics collection feature to decide which statistics to collect on a particular table based on past table activity.

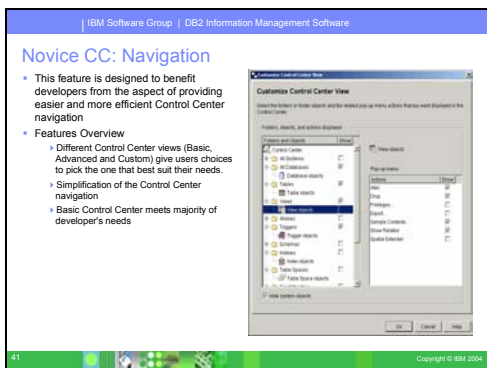


## Automated Table Reorg w/ Policy

After many changes to table data, logically sequential data may be on non- sequential physical pages so the database manager has to perform additional read operations to access data.

Among other information, the statistical information collected by RUNSTATS show the data distribution within a table. In particular, analysis of these statistics can indicate when and what kind of reorganization is necessary. Automatic reorganization determines the need for reorganization on tables by using the REORGCHK formulas. It periodically evaluates tables that have had their statistics updated to see if reorganization is required. If so, it internally schedules a classic reorganization for the table. This requires that your applications function without write access to the tables being reorganized.

The automatic reorganization feature can be enabled or disabled by using the auto\_reorg, auto\_tbl\_maint, and auto\_maint database configuration parameters.



## Novice CC: Navigation

The DB2 UDB Version 8.2 Control Center is available in three different views:

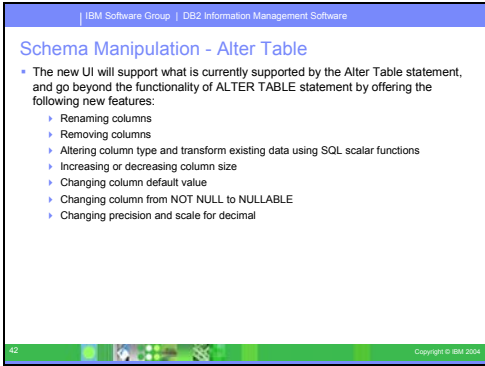
- Basic. This view provides core DB2 UDB functionality, which includes the essential objects, such as databases, tables, and stored procedures.
- Advanced. This view displays all objects and actions available in the Control Center. This is the view that you should select if you are working in an enterprise environment and want to connect to DB2 for z/OS™ or IMS™.
- Custom. This view gives you the ability to tailor the object tree and the object actions to your specific needs.

You can change your view if your needs change. Tailoring the Control Center to display the folders and folder objects that you work with regularly (the objects within a folder are called folder objects).

The Control Center now has a third pane called the Object details pane. It displays information about the folder or folder object that you have selected in the Control Center's object tree or contents pane. It also lists some of the actions that you can perform on the selected folder or folder object.

The Control Center now supports additional types of IMSplex members and commands. You can now query or update databases, areas, data groups, or online reorganizations. In addition, IMS command windows make it easier to issue commands.



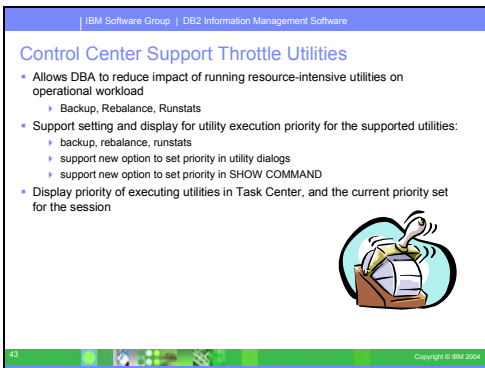


## Schema Manipulation - Alter Table

When altering a table from within the Alter Table notebook, you can now make more changes to table columns than you could previously. These changes are possible because the Alter Table notebook now drops and re-creates the table on an as-needed basis. You can make the following additional changes when altering a table:

- Rename a column
- Drop a column
- Change the data type of a column
- Change the length, scope, or precision values for a column
- Change whether a column is nullable

If the table is dropped and re-created, attempts will be made to restore any dependent objects and to transform the existing data into the target data type of each remaining column.

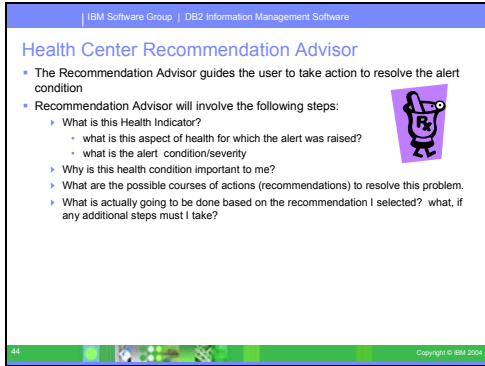


## Control Center Support Throttle Utilities

Executing resource-intensive utilities like RUNSTATS can negatively impact overall database performance. However, in order to maintain efficient database operation, statistics must be collected regularly, leaving database administrators with the task of identifying periods of time when the impact of utility execution is most tolerated. In many environments, there are no regular windows of reduced database activity.

Throttling of the RUNSTATS utility limits the amount of resources consumed by the utility, based on the current level of database activity. When database activity is low, the utility runs more aggressively; when database activity increases, the resources allocated to executing RUNSTATS are reduced.





## Health Center Recommendation Advisor

---

The Health Center for DB2 UDB Version 8.2 has a number of new features:

- Health Indicator Configuration launchpad

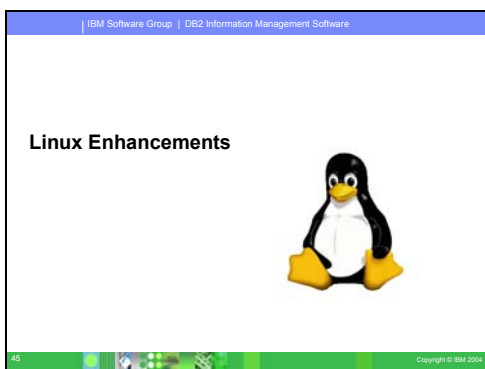
The Health Indicator Configuration launchpad provides easy access to various levels of health indicator settings. This launchpad also helps you to understand how these configurations apply to health monitoring, allowing you to make changes appropriate to your database system environment.

- Recommendation advisor

The Recommendation advisor helps you to resolve health alerts on DB2 objects. This advisor provides recommendations that can correct the issue causing a health alert. In addition, the Recommendation advisor helps you to implement the recommendation that you select, whether this requires launching a tool, running a script, or adjusting the configuration parameter settings on an instance or a database.

- Troubleshoot Health Alert Notification wizard

The Troubleshoot Health Alert Notification wizard helps you to troubleshoot various problems related to health alert notification, which ensures that notifications are received properly by the correct contacts.



## Linux Enhancements

---

DB2 UDB Version 8.2 includes a number of features that exploit the capability of the Linux 2.6 Kernel.

IBM Software Group | DB2 Information Management Software

### DIRECT I/O for Linux

- Currently, the file system caching policy of all I/O operations in DB2 are performed in buffered mode by default
- While this caching policy is extremely effective when the cache hit ratio is high, it has an overhead of making an extra copy of the buffer from the disk to file cache (in case of read) or from file cache to disk (in case of write).
  - Since the buffer is already cached in DB2's bufferpool layer, this dual level of caching proves to be unnecessary in situations where the cache hit ratio is low and a lot of I/O is performed.
- Direct I/O (DIO) is an alternate caching policy that reduces CPU utilization for reads and writes by eliminating the copy from file cache to user buffer.
- This feature extends DIO support on Linux distribution which are based on a 2.6 kernel (this was not ported back to previous kernel levels)
- Lab tests indicate a 12% performance improvement (OLTP workload)
- Specified by using the NO FILE SYSTEM CACHING parameter on the CREATE/ALTER TABLESPACE or CREATE DATABASE statement

46 Copyright © IBM 2004

IBM Software Group | DB2 Information Management Software

### CPU pinning on Linux

- NUMA technology allows for multiple cpus to run workloads in Linux
  - Processes can move from cpu to cpu
  - Penalty for switching CPUs in the middle of a task can be quite high (L1, L2 cached data)
- This feature allows db2 processes to be pinned to a CPU once assigned by the OS to a specific CPU.
  - This is not the same as having to decide which CPU to pin a process to, but rather whether or not to pin once a process is on a certain CPU.
  - In this case, the OS still decides which CPU to run the task on, but DB2 decides if the task should stay on that CPU until the task is finished.
- Enabled by specifying an XML format file (called the resource affinity configuration file)
- Can cause performance degradation
- Properly tuned can result in a 1%-2% improvement

47 Copyright © IBM 2004

IBM Software Group | DB2 Information Management Software

### Asynchronous I/O on Linux

- Allows DB2 to take advantage of the kernel-based asynchronous I/O capabilities of the 2.6 Linux kernel
- AIO is not enabled by default in DB2 UDB Version 8.2
- use db2set and the DB2NOLIOAIO variable to enable AIO
  - If false, AIO will be enabled in DB2
  - If true, AIO will be disabled.
- DB2 uses the Linux AIO interface for it's page cleaners
- AIO can noticeably improve performance on Red Hat AS 2.1, SuSE SLES 8 (2.4 kernel) also supports it, so customers can manually enable and run AIO with DB2, even with existing code.
- Lab tests have shown a 14% performance improvement when running OLTP workload on a file system. 5% on Raw Devices

48 Copyright © IBM 2004

IBM Software Group | DB2 Information Management Software

### Large Page Support on Linux

- The Linux 2.6 kernel allows for large page sizes. Reduces overhead of mapping large memory in the system.
  - This support has also been back ported to the 2.4 kernel in certain distributions (Red Hat AS 2.1 and SuSE SLES 8)
  - This feature is necessary for improving performance on fast CPUs today that require lots of memory to be fully utilized.
- Must enable both Linux and DB2!!
- Internal lab tests indicate potential 9% improvement when DB requires large memory

49 Copyright © IBM 2004

## Direct I/O for Linux

Direct I/O is now supported on both file systems and block devices for distributions of Linux with a 2.6 kernel. Direct I/O on block devices is an alternative way to specify device containers for direct disk access or for raw I/O.

The performance of Direct I/O is equivalent to the raw character device method. DB2 UDB enables Direct I/O while opening the table space when the CREATE TABLESPACE statement specifies a block device name for the container path. Previously, the same performance was achieved using the raw I/O method, which required binding the block device to a character device using the raw utility.

## CPU Pinning

DB2 UDB now supports precise configuration of DB2 UDB use of some operating system resources. For example, you can pin each engine EDU to a single logical processor. Such advanced configurations can improve performance in some scenarios. However, these configurations should be attempted only by advanced users who are knowledgeable about their operating environment and its workload. Misconfiguration can result in degraded performance. A resource policy is defined to describe how DB2 UDB interacts with the operating system. The resource policy typically limits what resources DB2 UDB uses, or it contains rules for assigning specific operating system resources to specific DB2 UDB objects. The extent of resource control varies depending on the operating system.

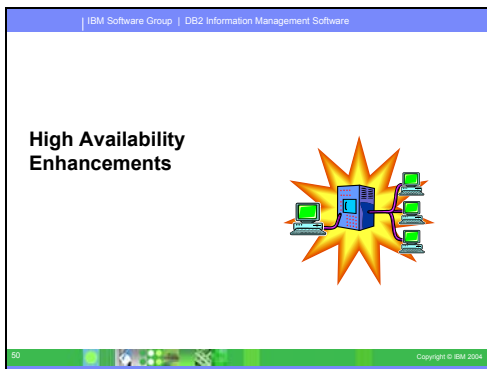
## Asynchronous I/O

Asynchronous I/O (AIO) support is now available on Linux (2.6 and some 2.4 kernels) for raw devices and O\_DIRECT file systems. AIO improves page cleaner performance. You can enable or disable AIO on Linux by issuing the db2set command.

To use AIO, users must install libaio-0.3.98 or later and have a kernel that supports AIO. Users must also run the db2set DB2LINUXAIO=true command and restart DB2 UDB.

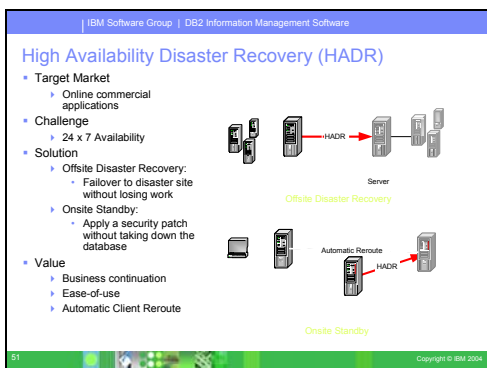
## Large Page Support

Large page support in buffer pools is supported on DB2 UDB for Linux in addition to DB2 UDB for AIX® 64-bit. Buffer pools are kept in memory and are not swapped out when using large pages. This improves performance of the DB2 UDB for Linux databases and applications.



## High Availability Enhancements

DB2 UDB Version 8.2 includes the new HADR feature. This feature will be of particular interest to customers who want to maintain a highly available DB2 database, or want to be able to maintain a database without the need to stop production.



## High Availability Disaster Recovery (HADR)

DB2 high availability disaster recovery (HADR) is a data replication feature that provides a high availability solution for both partial and complete site failures. HADR protects against data loss by replicating data changes from a source database, called the primary, to a target database, called the standby.

A partial site failure can be caused by a hardware, network, or software (DB2 or operating system) failure. Without HADR, the database management system (DBMS) server or the machine where the database resides has to be rebooted. The length of time it takes to restart the database and the machine where it resides is unpredictable. It can take several minutes before the database is brought back to a consistent state and made available. With HADR, the standby database can take over in seconds. Further, you can redirect the clients that were using the old primary database to the standby database (new primary database) by using automatic client reroute or retry logic in the application.

A complete site failure can occur when a disaster, such as a fire, causes the entire site to be destroyed. Since HADR uses TCP/IP for communication between the primary and standby databases, the databases can be situated in different locations. If a disaster occurs at the primary site, data availability is maintained by having the remote standby database take over as the primary database with full DB2 functionality. After a takeover operation occurs, you can bring the original primary database back up and return it to its status of primary database; this is known as failback.

IBM Software Group | DB2 Information Management Software

### High Availability Disaster Recovery

- DB2 High Availability Disaster Recovery (HADR) is an easy to use data replication feature that provides a high availability (HA) solution for both partial and complete site failures
  - HADR is intended to be the primary HA solution for complete site failures as well as the choice for applications demanding ultra-fast failover for partial site failures.
- HADR replicates data changes from a source database (called the primary) to a target database (called the standby)
  - Using synchronous mode, HADR can guarantee that any transaction committed on the primary is also committed on the standby
  - HADR allows failover and failback between the two systems
  - HADR requires the same hardware, OS and DB2 software on the two systems (except some minor differences during rolling upgrade).
- HADR propagates data changes by shipping database log records from the primary to the standby
  - It is tightly coupled with DB2 logging and recovery.

52 Copyright © IBM 2004

## What is HADR?

With HADR, you can choose the level of protection you want from potential loss of data by specifying one of three synchronization modes: synchronous, near synchronous, or asynchronous.

HADR allows the standby database to take over as the primary database with full DB2 functionality. It is also possible for the original primary database to be brought back up and returned to its status of primary database.


When a failure occurs on the primary database, you can initiate a takeover operation on the standby database which then becomes the new primary. Since the standby database is already online, failover can be accomplished very quickly resulting in minimal down time.

Once the failed old primary is repaired, it can rejoin the HADR pair as a standby database if the two copies of the database can be made consistent. After the original primary database is reintegrated into the HADR pair as the standby database, a failback operation can be performed so that the original primary database is once again the primary database.

IBM Software Group | DB2 Information Management Software

### Client Reroute

- Whenever a server crashes, each client that is connected to that server gets a communication error which terminates the connection leading to an application error.
  - The DB2 UDB client code attempts to re-establish the connection to either the original server or to a new server.
  - When the connection is re-established, the application will receive an error that informs it of the transaction failure, but the application can continue with the next transaction.
- The Automatic Client Reroute feature could be used in following configuration environments:
  - Enterprise Server Edition (ESE) / Data Partitioning Feature (DPF)
  - Dpropr-style Replication
  - High Availability Cluster Multiprocessor (HACMP)
  - High Availability Disaster Recovery (HADR) environment
- Also available with LDAP Clients

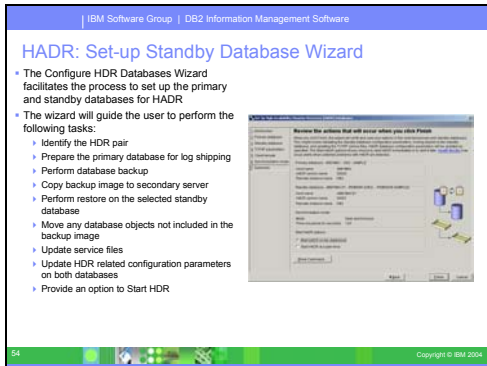


53 Copyright © IBM 2004

## Client Reroute

The automatic client reroute feature allows client applications to recover from a loss of communication with the server so that they can continue to work with minimal interruption. After a loss of communication, the client application attempts to reconnect to the server. If this fails, the client is then rerouted to a different server. You can specify an alternate location through the command line processor (CLP), by invoking an application programming interface (API), or when adding a database using the Control Center or the advanced view of the Configuration Assistant.

You can use automatic client reroute with high availability disaster recovery (HADR) to make client applications connect to the new primary database after a takeover operation.

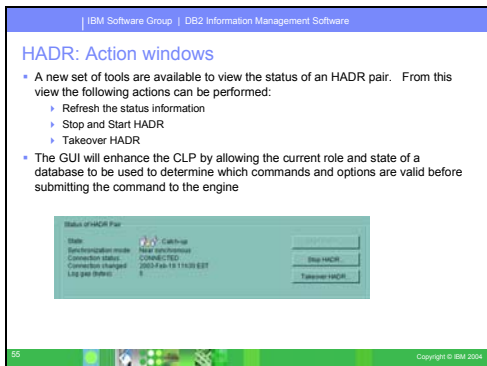


## HADR: Set-up Standby Database Wizard

Use the Set Up High Availability Disaster Recovery (HADR) Databases wizard to set up and configure primary and standby databases for high availability disaster recovery.

The wizard guides you through the following tasks:

- Identifying the HADR pair, specifically the primary and standby databases and instances.
- Setting the logging-related configuration parameters to prepare the primary database for log shipping.
- Updating the HADR-related database configuration parameters on the primary and standby databases.
- Taking a backup of the primary database, copying the backup image to the standby database, and restoring the backup image to a new or existing standby database.
- Moving database objects that are not included in the backup image.
- Updating the services file with the service name and port number pair used by the HADR agent.
- Configuring databases for automatic client reroute.
- Starting HADR as soon as the configuration is completed, or postponing starting HADR to a later time.



## HADR: Action windows

Use the Manage High Availability Disaster Recovery window to configure and to check the status of your HADR system. This window will notify you if there are any problems with the HADR configuration. From this window you can perform the following tasks:

- Start HADR on one or both of the databases.
- Stop HADR on one or both of the databases.
- Initialize a takeover operation.
- Check or alter database configuration parameters related to HADR.
- Catalog an HADR database (if required).