

Integrated Data Management: Managing data across its lifecycle

Featuring Optim solutions for Integrated Data Management

Skill Level: Introductory

[Holly Hayes \(hollyann@us.ibm.com\)](mailto:hollyann@us.ibm.com)
Program Director, Data Studio
IBM

08 Jul 2008

Updated 02 Jun 2009

With the June 2009 announcements, IBM is consolidating many of its Data Studio offerings under the Optim name. The Optim portfolio will focus on realizing integrated data management with innovative delivery of application-aware solutions for managing data and data-driven applications across the lifecycle, from requirements to retirement. This overview article explains both the vision and reality of Integrated Data Management and how you -- whether a data architect, developer or tester, DBA, or data steward -- can use IBM solutions today to respond quickly to emerging opportunities, improve quality of service, mitigate risk, and reduce costs.

Welcome to Integrated Data Management by IBM!

IBM has embarked on a strategic initiative to deliver an integrated, yet modular, data management environment to design, develop, deploy, operate, optimize and govern data, databases, and data-driven applications throughout the entire data management lifecycle. We call this *Integrated Data Management*. By focusing across the lifecycle and enabling different roles to collaborate, we believe that you can increase organizational productivity, agility, and effectiveness, while improving the quality of service, cost of ownership, and governance of diverse data, databases, and data-driven applications.

Individual products provide powerful capabilities that target specific data management roles and tasks; more importantly, the components *interoperate*, enabling cross-role collaboration and cross-product synergy. And to deliver a true lifecycle solution, integration extends beyond Optim to include a broad range of IBM offerings.

This article takes a look across phases and roles to articulate how IBM's solutions for Integrated Data Management can help you get more value from your information and help your team be more aligned, productive and effective.

Diverse, distributed, and interrelated environments

Have you noticed that it's hard to find a standalone application anymore? Everything is interrelated. One system feeds another and all must come together to present a common view to the customer – whoever your customer might be. Thus it is increasingly important that we look at how to manage our information assets more holistically and strategically. We need to help organizations inventory and understand what assets they have and how they are related. We need to have a common definition of customer or patient or citizen or supplier across the organization. We need to understand data movement and data lineage. And that's going to mean discovering and sharing the information that we have about our data assets – cross-role, cross-solution, cross-lifecycle.

End-to-end management of the data lifecycle

Today most organizations have a myriad of products in-house from many vendors supporting different roles and tasks. Each focuses on providing rich task-specific value, but puts little emphasis on linkages with the preceding or next phase in the lifecycle. Wouldn't it make life easier to define access or retention policies when the data is first designed and let the tools propagate that information from phase to phase and tool to tool? With IBM software, we can support each phase of a the lifecycle with robust offerings for data-centric tasks and roles, as well as provide support for designing and implementing key cross-phase linkages. This is how we define the key phases in a data-centric software lifecycle:

- **Design** -- Discover, harvest, model, and relate information to drive a common semantic understanding of the business.
- **Develop** -- Code, generate, test, tune, and package data access layers, database routines, and data services.
- **Deploy** -- Install, configure, change, and promote applications, services, and databases into production.
- **Operate** -- Administer databases to meet service level agreements and security requirements while providing responsive service to emergent issues.

- **Optimize** -- Provide pro-active planning and optimization for applications and workloads including trend analysis, capacity and growth planning, and application retirement including executing strategies to meet future requirements
- **Govern** -- Establish, communicate, execute, and audit policies and practices to standardize, protect and retain data in compliance with government, industry, or organizational requirements and regulations. Not limited to a single phase, governance is a practice that must infuse the entire lifecycle.



Cross-organizational collaboration

Maintaining alignment is about communication, collaboration, and clarity across organizational roles. Users and business analysts need to articulate requirements. Architects are responsible for designing the process, application, and data models. Developers must produce effective and efficient code using those models. Administrators must understand security and retention policies established by compliance officers and work with their network and systems administration colleagues to achieve compliance and service objectives. It's critical to the agility, effectiveness and alignment of the organization as a whole that the role-specific capabilities can adapt to multi-role contributors as well as to distributed global teams.

Comprehensive portfolio - emerging integration

Supporting Integrated Data Management is, and will always be, a multi-branded proposition. Today the IBM portfolio encompasses various offerings including Rational, Information Management, Tivoli, and WebSphere offerings. IBM offers broad and deep capabilities for every phase in the lifecycle. But over time what will increasingly differentiate IBM offerings is the value-added integration across the portfolio (either in current product or in roadmaps) with common user interfaces, common components and services, and shared artifacts.

- **Common user interfaces**

Whether Eclipse-based or Web-based, IBM is adopting a standard and integrated approach to user interfaces that makes moving between roles easy and intuitive. The portfolio includes an Eclipse-based user interface for tasks requiring rich object manipulation e.g. design and development. Here, the offerings complement and extend the IBM Rational Software Delivery Platform. The integrated nature of IBM Optim and Rational software simplifies collaboration among business analysts, architects, developers, and administrators. Users can combine products within the same Eclipse instance, providing seamless movement between tasks, or can share objects across geographically distributed teams to make it easier to maintain alignment and work more efficiently.

Operations support requires the ability to monitor and respond from anywhere at any time. The Web-based user interface supports operations-oriented administration. Adopting a common approach with Tivoli software for Web delivered dashboards and portlets provides the greatest flexibility for monitoring, management, and coherent information across the operational stack to improve an organization's ability to meet service level agreements. And sharing all of these capabilities across data servers reduces overall skills requirements and costs. For our z/OS base, existing 3270 interfaces continue to be supported and extended.

- **Common components and services**

Sharing components and services across offerings help organizations achieve cost, productivity and consistency objectives. When the products share components, such as the Data Source Explorer, learning new tools becomes easier. For example, sharing a common connections repository saves times for team members. Shared services, such as data privacy policies, mean personal identification numbers will be handled consistently whether creating test data or sharing research data.

- **Shared policies, models, and metadata**

This is the glue that truly holds everything together. The ability to express policies for machine interpretation, to associate policies with data models or data workloads, and communicate both through shared metadata is the

crux of the challenge as well as the critical lynchpin for greatest value. For example, shared configuration information between database administrators and application server administrators can vastly reduce deployment costs while improving quality of service. Shared privacy policies together with the services that implement them can improve security and compliance.

Heterogeneous flexibility

Recognizing the heterogeneity of most organizations, the vision spans IBM and non-IBM databases. While we will deliver first on DB2® and Informix® Dynamic Server databases, we are building out the portfolio across a range of heterogeneous databases. Design, development, optimization and governance phases are already supported by solutions that span IBM and non-IBM databases. And our roadmap includes expanding deployment, and operations offerings in that direction.

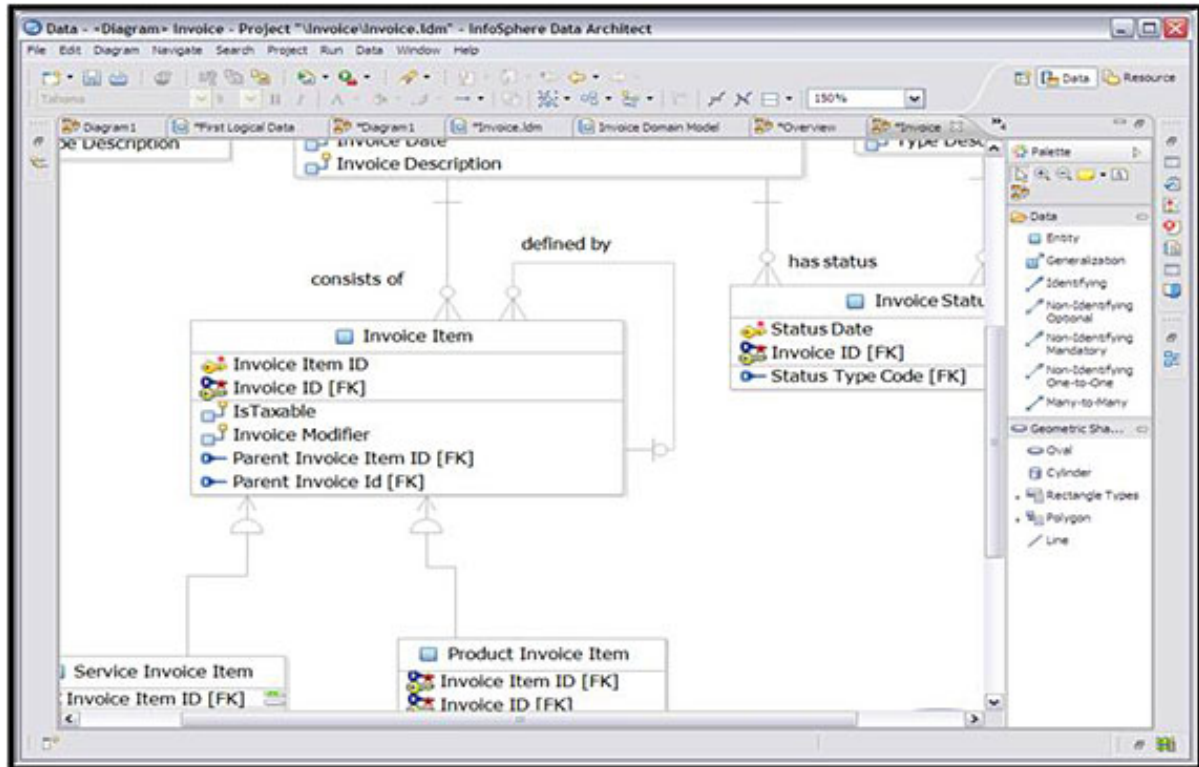
Data-centric roles

Turning our attention now to various key roles that Integrated Data Management supports, let's look at some of the key offerings and value we expect them to deliver.

Data architect – Better data quality and enterprise consistency

The data architect's key tool is InfoSphere Data Architect (formerly Rational Data Architect), used for discovering, modeling, relating, and standardizing data. Like any good data modeling offering, it supports logical and physical modeling and automation features for diverse databases that simplify tasks such as reverse engineering from existing databases, generating physical models from logical models, generating DDL from physical models, and visualizing the impact of changes.

Figure 1. InfoSphere Data Architect for modeling an image



Resources

- Product site: [InfoSphere Data Architect](#)
- Product site: [Optim Database Relationship Analyzer](#)
- Demo: [Introduction to InfoSphere Data Architect](#)
- developerWorks product page: [InfoSphere Data Architect](#)
- Download: [InfoSphere Data Architect](#)

But beyond core data modeling, InfoSphere Data Architect also helps data architects:

- Integrate information by discovering and identifying mappings between models; Data Architect's metadata-driven discovery is complemented by query-driven and data-driven capabilities in Optim Data Relationship Analyzer and InfoSphere Information Analyzer. Data models can then be delivered to to InfoSphere Information Server or InfoSphere Warehouse.
- Implement best practices based on naming standards enforcement, business glossary integration, and industry model integration.
- Achieve architectural alignment across process, service, application and data models with built-in transformation between models and clear linkage to business requirements. Built-in integration with the Rational

portfolio offerings simplify model interchange and alignment.

- Facilitate governance practices regarding privacy standards for test data generation by capturing privacy policies and business objects for downstream tasks. Share privacy policies with developers and publish extract scripts for the Optim Test Data Management Solution and the Optim Data Privacy Solution.

InfoSphere Data Architect represents a key integration point among the Rational, InfoSphere, and Optim portfolios. There is a rich product roadmap that plans to extend governance policies to include retention attributes. And additional integration is planned with the InfoSphere portfolio, including enriching the data model based on metadata discovery from Information Analyzer as well as model extensions and exchange with Metadata Server.

Developer – Better productivity and better application performance

Resources

- Product site: [Optim Development Studio](#)
- Product site: [Optim pureQuery Runtime](#)
- Product site: [Optim Query Tuner](#)
- Platform page: [pureQuery platform](#)
- developerWorks product page: [Optim Development Studio and pureQuery Runtime](#)
- Demo: [New features in 2.1 are highlighted](#)
- Technical article: ["Migrate from EJB 2 container-managed persistence to pureQuery for IBM Master Data Management Server, Part 1: Evaluating pureQuery technology"](#)
- Download: [Data Studio Developer and pureQuery Runtime \(new Optim release will be available by end of June\)](#)

Optim Development Studio, Optim Query Tuner, and Optim pureQuery Runtime offerings target data-centric developers and application DBAs.

Optim Development Studio (formerly Data Studio Developer) provides an Eclipse-based integrated development environment, to speed data-centric development targeting DB2, Informix, and Oracle databases. Customers and partners have reported a 25% to 50% productivity improvement with the product. And the data-centric development capability seamlessly extends functionality within

the Rational Software Delivery Platform, such as Rational Application Developer for WebSphere Software. In particular, Optim Development Studio delivers SQL content assist integrated with the Java editor, stored procedure development (both SQL/PL and PL/SQL), data access layer generation, Web services tooling, SQL hot spot analysis, impact analysis, and much more.

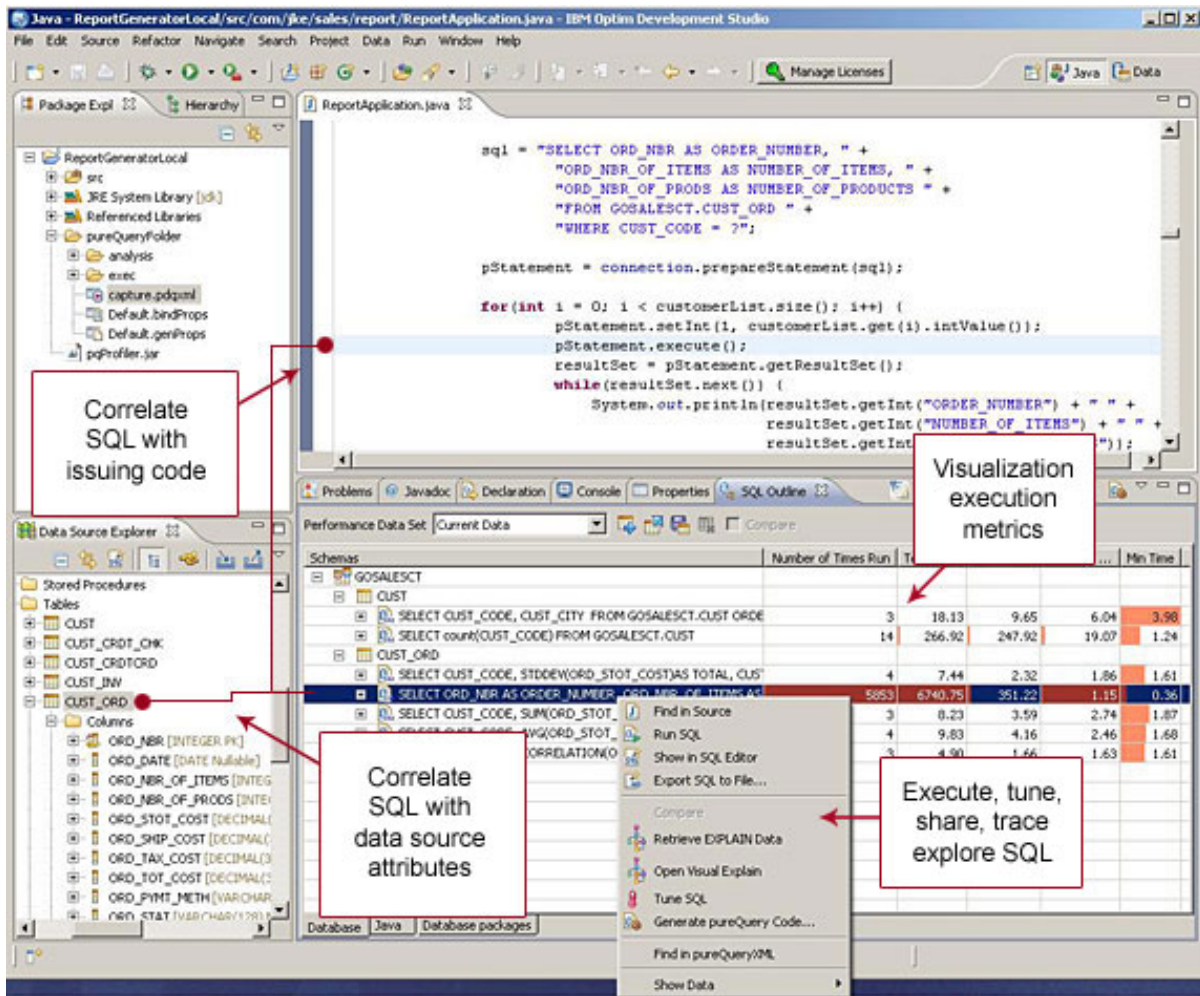
The data access layer generation includes support for the standard Data Access Object (DAO) pattern and leverages the pureQuery API, an intuitive and simple API that balances the productivity boost from object-relational mapping with the control of customized SQL generation. It also simplifies the use of best practices for enhanced database performance. Optim pureQuery Runtime (formerly Data Studio pureQuery Runtime) is used with pureQuery data access layers.

Developers can visualize SQL hot spots within the application during development. Adding Optim Query Tuner helps developers tune SQL for DB2 based on expert guidance to build skills and reduce query tuning needs in production where risks and costs are much higher.

Portfolio integration helps developers be cognizant of sensitive data. Developers can readily identify sensitive data based on the privacy metadata capture in InfoSphere Data Architect. They can provision test databases directly from fictional data or can generate extract definitions for Optim Test Data Management and Data Privacy to create customized test databases.

Developers can spend considerable time isolating performance issues: first to a specific SQL statement, then to the source application, then to the originating code. Three-tier architectures and popular frameworks make this isolation more difficult as the developer may never see the SQL generated by the framework. Optim Development Studio makes it easier to isolate problems by providing an outline that traces SQL statements back to the originating line in the source application, even when using Java frameworks like Hibernate, OpenJPA, Spring, and others.

Figure 2. Outline view in Optim Development Studio



As more and more abstraction is introduced into the application architecture, developers and DBAs have become increasingly isolated from one another. And developers have less and less involvement, or even control, over the SQL that is executed to manage database access and persistence. Optim Development Studio supports collaboration between the developer and DBA, giving them an easy way to capture, share, review, optimize, and restrict SQL that will be put into production.

Tester – Better quality test data without revealing sensitive information

Resources

- Product site: [Optim Test Data Management Solution](#)
- Product site: [Optim Data Privacy Solution](#)
- Demo: [A Day in the Life of a DBA](#)

The key role of the tester is to assure application quality. Historically, testers have used clones or extracts of live customer data to attempt to provide contextually accurate data, but a simple extract may not be sufficient and full clones can quickly break the budget. The test data needs to be reflective of application processing constraints as well as error and boundary conditions. IT staff are also challenged to protect confidential data and personally identifiable information ("PII") like bank account numbers and national identifiers.

The Optim Test Data Management Solution together with the Optim Data Privacy Solution create a right-sized, "production-like" test environment that accurately reflects end-to-end business processes while de-identifying sensitive information providing the perfect option for test data creation. It offers built-in knowledge of packaged application business objects.

The Optim solutions support an iterative testing model that simplifies specification of error and boundary conditions and comparison of test results to baseline data. Determination of errors is difficult, especially when you don't know if, who or how the data has changed. The Optim Test Data Management Solution enables comparison between data pre and post test to determine data inconsistencies and identify errors earlier in the lifecycle. It offers built-in knowledge of packaged application business objects and pre-defined masking algorithms for common sensitive information. Privacy attributes can be defined and managed consistently in InfoSphere Data Architect and used to generate test definitions directly from the Data Architect's desktop or from Optim Development Studio helping organizations to assure compliance.

Database Administrator – More control, efficient problem isolation

The portfolio of products supporting the DBA are too numerous to mention individually, but you can find more information at [Tools for z/OS](#) and [Tools for DB2 for Linux®, UNIX®, and Windows®](#) . So rather than individual offerings, let's focus on the strategic priorities and look at examples of particular tools that illustrate those priorities.

Give the DBA more control

Resources

- Product site: [Optim pureQuery Runtime](#)
- Overview article on static SQL and pureQuery: "[No Excuses database programming for Java](#)"
- Tutorial: "[Optimize existing JDBC applications with pureQuery](#)"

- Tutorial: ["Optimize existing .NET applications with pureQuery"](#)
- Performance study with Java: ["pureQuery Runtime for DB2 for z/OS Performance"](#)
- Performance study with .NET: [".NET and pureQuery Runtime for DB2 for z/OS Performance"](#)
- Download: [Data Studio Developer and pureQuery Runtime \(new Optim release will be available by end of June\)](#)

Over time, the DBA's ability to control database performance has eroded, or at least become much harder, as additional layers emerge in the application stack. SQL is generated by frameworks not programmers, database connections are managed by systems administrators not DBAs, and dynamic SQL complicates security management.

We think DBAs like the added control they can gain from using static SQL, and now it is possible to gain that control easily over existing Java and .NET applications by using client optimization technology delivered in Optim pureQuery Runtime. This is an innovative approach to performance optimization that focuses on how to optimize database access from the database client rather than only looking within the database engine. Client optimization captures SQL from executing applications and enables administrators to bind the SQL to DB2 for static execution without changing a single line of application code. All of the gain of static SQL - making response time stable, reducing security risks, increasing throughput, improving manageability – and none of the pain. What's more, pureQuery can alleviate novice programming errors, for example, by consolidating common SQL statements that use literals and converting them to parameter markers or enabling DBAs to replace poor SQL generation by frameworks with optimized SQL. Now frameworks are a little less scary for conservative DBAs.

Future enhancements include plans to give DBAs control over performance knobs in the application server and to make client configuration manageable – finally.

Bring the information together

Resources

- Product site: [DB2 Performance Expert for Linux, Unix, and Windows](#)
- Product site: [IBM Tivoli OMEGAMON XE for DB2 Performance Expert on z/OS](#)
- Overview article: [End-to-end database monitoring with DB2 Performance Expert Extended Insight Feature](#)
- Demo: [A Day in the Life of a DBA](#)

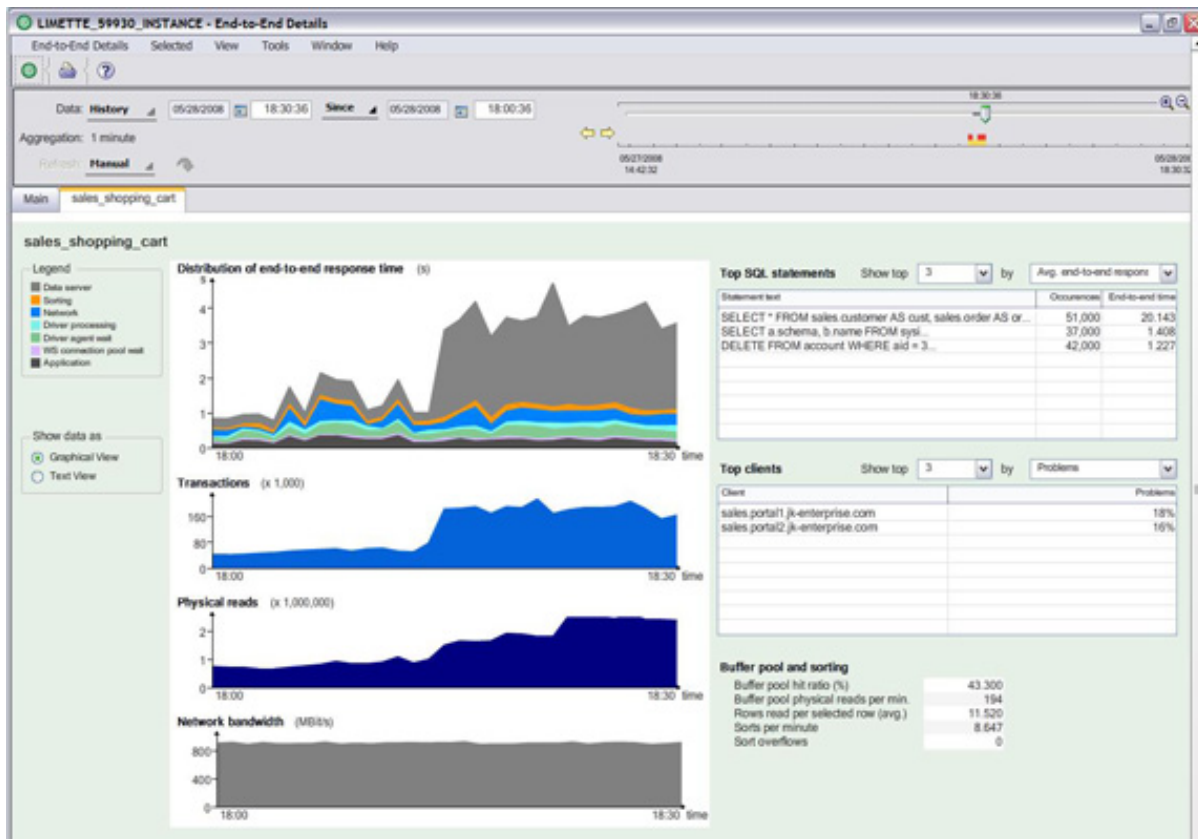
Ever spend 3 or 4 days just isolating a performance problem to a particular query, and then spend another few days isolating it to the application? Performance issues are particularly difficult to isolate given that the problem could be in the application, the application server, the database client, the network, the database server, or the operating system. Each of these layers has performance information, but none have the information in aggregate. A key objective is to give administrators the ability to aggregate and correlate information enabling fast problem isolation not only to the offending SQL statement, but also to the originating application source.

The performance monitors DB2 Performance Expert and Tivoli OMEGAMON XE for DB2 Performance Expert on z/OS offer a wealth of information regarding the performance of their respective database servers. Tivoli software gathers information about the application servers, networks, and hardware devices. The Optim Development Studio outline view adds a missing piece that correlates Java code, SQL statements, and table information. But the trick is putting them all together.

An extension to DB2 Performance Expert 3.2, DB2 Performance Expert Extended Insight Feature, extends database monitoring across the database client, the application server, and the network, giving DBAs immediate insight into where database workloads, transactions, and SQL requests are spending their time. The DBA can now readily identify the SQL statement and the application. The developer can trace the SQL statement to the source code using the SQL outline.

Planned enhancements to the performance monitors, and integration with Tivoli software, will provide comprehensive views across the application stack to include the development metadata, to further streamline problem isolation.

Figure 3. Distribution of end-to-end response time in DB2 Performance Expert Extended Insight Feature



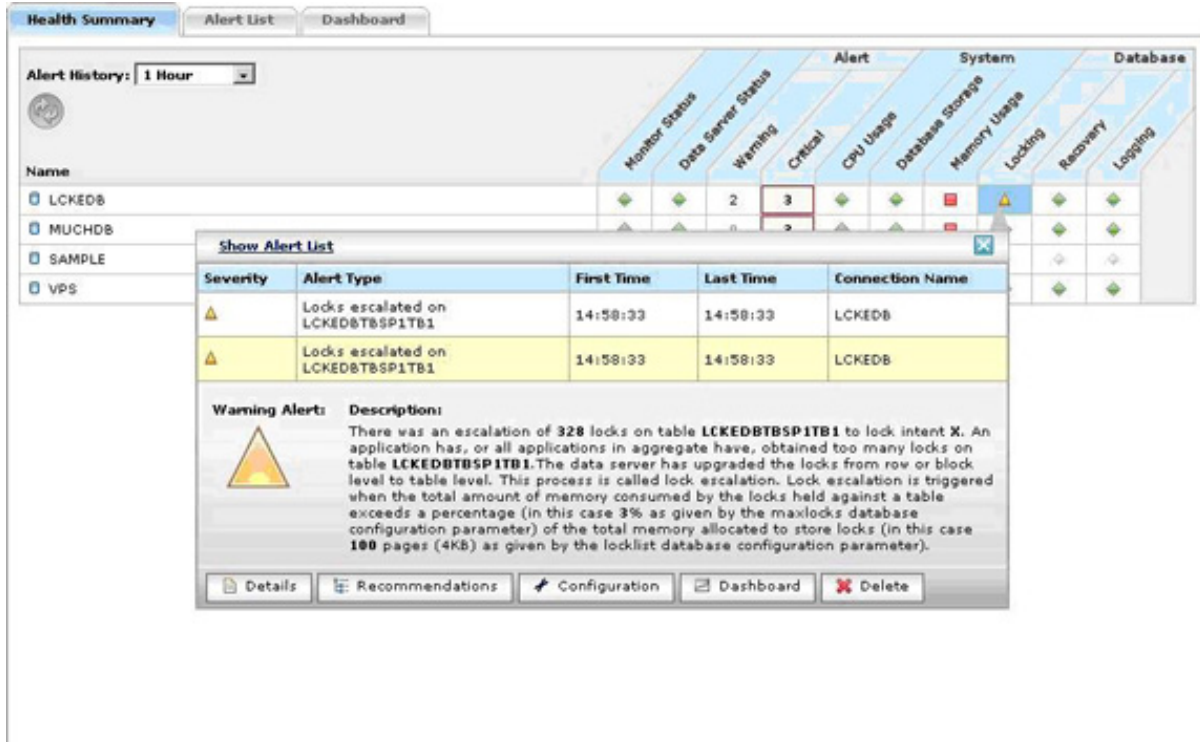
Provide task-specific flows and context

Resources

- Tutorial: ["Using Data Studio administration console for health monitoring"](#)
- Download: [Data Studio administration console](#)

The Data Studio administration console, provided at no charge with DB2 databases, provides an example of task-specific flows, as well as a glimpse into the future of the operations-based user interface. Administrators need to be able to establish objectives, then leave the system to alert them when something is awry and provide relevant context to manage the alerted condition. The health monitor will flag a problem on the dashboard when it detects a threshold condition. Built into the console are decision trees that lead you through root cause analysis. Plus it automatically displays relevant configuration parameters and performance indicators along with recommendations for problem resolution, thus streamlining problem resolution.

Figure 4. Data Studio administration console



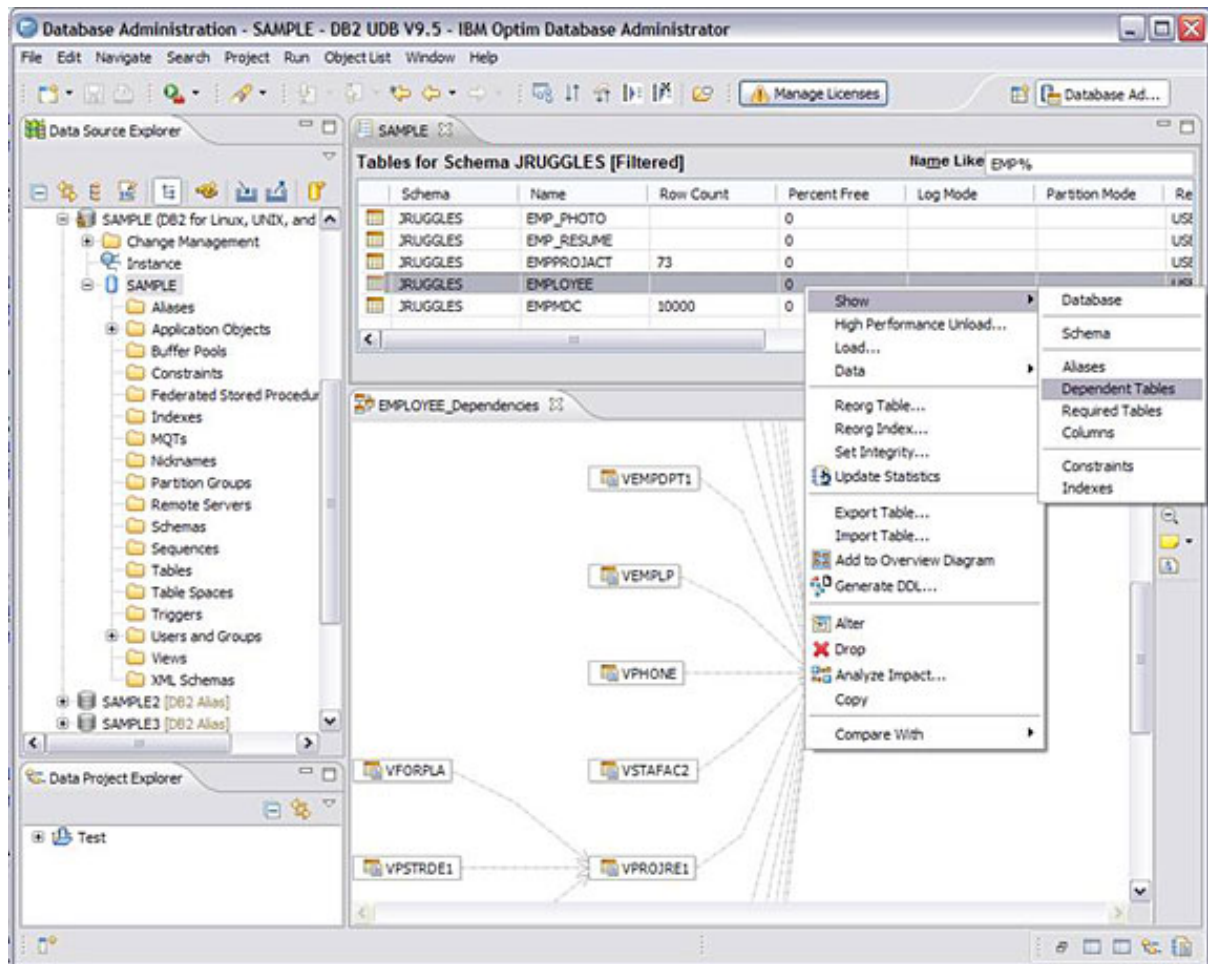
Resources

- Product site: [Optim Database Administrator](#)
- Demo: [A Day in the Life of a DBA](#)
- Technical article: "[Restoring your database](#)"
- Technical article: "[Data Migration and Change Management](#)"
- Technical article: "[Fast and efficient data and schema migration](#)"
- Technical article: "[Database version control](#)"
- Download: [Data Studio Administrator \(new Optim release will be available by end of June\)](#)

Make tools smarter

We are continuing on our journey toward autonomic operations integrating best practices and advisory functions into the products. Optm Database Administrator increases productivity and reduces application outages through task automation. It facilitates impact and dependency analysis to mitigate risk, generates customizable deployment scripts to automate and accelerate changes, and supports object, data, and authorization migration in support of database migration scenarios.

Figure 5. Identifying dependencies with Optm Database Administrator

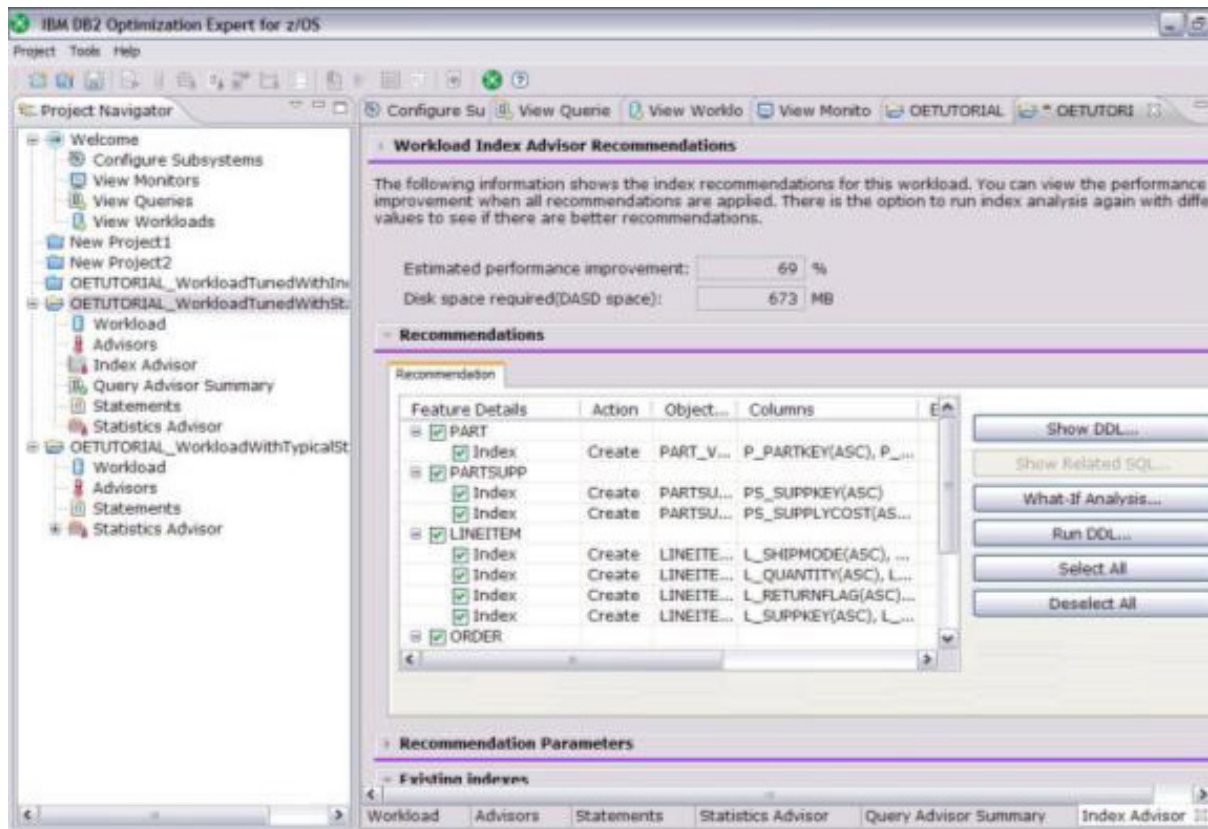


Resources

- Product site: [DB2 Optimization Expert for z/OS](#)
- Redbook: [DB2 for z/OS V9: New tools for query optimization](#)

Another such example is DB2 Optimization Expert. DB2 Optimization Expert offers a comprehensive set of tools and expert advisors that can help identify and improve problematic queries for DB2 for z/OS. It provides support for single query tuning as well as workload tuning. The advisors provide a rich set of recommendations for the type of statistics needed to improve performance, new indexes to improve query response time, as well as query and access path recommendations. It can shell share with Optim Development Studio, providing a single workspace for DBAs to optimize and revise application SQL without changing the application.

Figure 6. DB2 Optimization Expert for z/OS



Planning for strategic growth

Resources

- Product site: [Optim Data Growth Solution](#)
- Technical article: "[Access Optim archived files using Cognos](#)"

Overgrown databases can impair the performance of your mission-critical ERP, CRM and custom applications. Optim Data Growth Solutions solve the data growth problem at the source - by managing your enterprise application data. Optim enables you to archive historical transaction records, storing them securely and cost-effectively. With less data to sift through, you speed reporting and improve responsiveness of mission-critical business processes.

But data archiving isn't just about performance and cost improvements. Data archiving also facilitates application upgrades, consolidations, and retirement. Why consolidate all the data when only 20% is actively used? Archiving before an upgrade or consolidation speeds up the process, reduces risk, and reduces cost. If you find that some of the archived data should have been active, you can easily restore it to active status.

Have you been afraid to retire an application because you fear you may someday need the underlying data? Optim makes application retirement easier and safer by providing the capability to archive data from decommissioned applications while providing ongoing access to the data for query and reporting. You reduce risk and cut cost, without jeopardizing data retention compliance.

Data Stewards (or those by any other name) – Greater consistency for reduced risk

The role of data stewardship is often a role in the line of business reporting directly to senior executives, but the implementation of data steward functions typically come down to a Security Administrator, Compliance Administrator, or Database Administrator.

Data governance has many facets: availability, security, privacy, quality, audit, and retention to name a few. These tasks are split across many roles with few offerings that really aggregate the compliance story. IBM has a portfolio of robust data governance offerings that span the facets mentioned above. Key portfolio goals here are:

Compliance-savvy tools

More than the brute force to implement compliance initiatives, we believe the tools themselves should be providing intelligence regarding how best to comply with specific regulatory requirements. An example is Optim Data Privacy Solution that comes with prepackaged intelligent data masking routines to transform complex data elements such as credit card numbers, email addresses and national identifiers required to comply with HIPAA, GLBA, DDP, PIPEDA, Safe Harbour, PCI DSS and others.

Resources

- Product site: [Optim Data Growth Solution](#)
- Product site: [Optim Data Privacy Solution](#)

Cross life-cycle consistency

We want to deliver the ability to define governance policies once and have them implemented across the portfolio stack where appropriate. The first step in this direction is the model-driven governance, mentioned above. With the data model as a key architectural hub, privacy and retention attributes should be able to be propagated to other model-based tools such as Optim Data Privacy Solution or Optim Data Growth Solution.

Protection from threats

Resources

- Product site: [IBM Database Encryption Expert](#)
- Product site: [IBM Database Encryption for IMS and DB2 Databases](#)

The use of advanced access control techniques within databases, such as Label Based Access Control, Multilevel Security, and Trusted Context, are fundamental to protecting the data from misuse within the database. However, there are an increasing number of attacks on sensitive data from outside the database. These can be attacks from outsiders, internal privileged users, as well as inadvertent data loss. To ensure the sensitive data is protected, an accepted best practice is to encrypt all sensitive data. IBM Database Encryption Expert, and IBM Database Encryption for IMS and DB2 for z/OS, provide robust and application transparent encryption to ensure data is secure and enables compliance with many industry and government regulations governing the protection of sensitive data.

Coherent auditability

Resources

- Product site: [Tivoli Security Information and Event Manager](#)
- Product site: [DB2 Audit Management Expert for z/OS](#)

Gathering audit data is largely a manual process across most enterprises. We aspire to make this information both easily scoped and accessible to auditors. DB2 and IDS databases offer comprehensive audit facilities to capture all the information your auditors might need to ensure compliance with business controls. In addition, DB2 Audit Management Expert has enhanced analysis and reporting, custom built for auditors, to let them answer the who, what, when, where, and how regarding the database objects and users without giving them unfettered access to the audited databases. This audit information is also brought to an enterprise-wide view through the Tivoli Security Information and Event Manager, which provides end-to-end auditing that spans the database, operating system, application, and network.

Something for everyone - but more together

Whether a data architect, developer, tester, administrator, or steward, the Integrated Data Management portfolio by IBM has capabilities that that can help you be more effective and efficient. But more importantly, the portfolio and roadmap are delivering a collaborative environment that will deliver organization productivity and efficiency

to make organization more responsive to opportunities, improve the quality of service, mitigate risk and reduce costs for diverse data, databases, and data-driven applications. We hope after reading these examples, you'll agree.

Resources

- [Participate in the discussion forum for this content.](#)
- [Optim family page on developerWorks](#)
- Check out [Integrated Data Management experts team blog](#). Find links to discussion forums, articles, demos, blogs and more at the [Integrated Data Management community space](#).
- Find and download [Integrated Data Management](#) software.

About the author

Holly Hayes

Holly Hayes is a Program Director for the Data Studio Enablement Team. A 27-year IBM veteran, she has held development, strategy, marketing, and management positions working with operating systems microcode, replication technology, data warehouse infrastructure, database management, and information integration technology. She has been a prominent speaker at industry events and customer briefings and a frequent contributor to industry articles, analyst research, and other publications. She holds a U.S. patent in replication technology.

Trademarks

IBM, DB2, Informix, Rational, Tivoli, and WebSphere are registered trademarks of International Business Machines Corporation in the United States, other countries or both

Windows is a trademark of Microsoft Corporation in the United States, other countries or both.

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

Linux is a registered trademark of Linus Torvalds in the United States, other countries or both.

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