

Super Analytics, Super Easy



Introducing IBM DB2 10.5 with BLU Acceleration

By [Sam Lightstone](#), [Guy Lohman](#) and [Berni Schiefer](#) | Published April 3, 2013

Imagine a database technology that gives you 10-20 times faster performance right out of the box, requires dramatically less storage, and nearly eliminates the need for tuning. Too good to be true?

Not anymore.

IBM® DB2® 10.5 with BLU Acceleration changes everything. This revolutionary technology for complex analytic queries originated in the Blink project at IBM Research for in-memory, hardware-optimized analytics. It was then perfected and seamlessly integrated with DB2 through a collaboration between DB2 product development, the IBM Systems Optimization Competency Center, and IBM Research—adding columnar processing, broader SQL support, I/O and CPU efficiencies, and integration with the DB2 SQL compiler, query optimizer, and storage layer. BLU Acceleration is all about reducing costs and improving time-to-value by making complex analytics faster, easier, and more resource-friendly.

Making complex analytics faster

While the speed improvements that BLU Acceleration delivers vary by server, workload, and data characteristics, 10-20 times performance improvements are common. What's really exciting is that no tuning is required on DB2 10.5 to achieve these results.

Workload	Speedup using DB2 10.5 with BLU Acceleration
Financial services (Linux)	46.8x
European analytic ISV (IBM AIX®)	37.4x
North American analytic ISV (Linux)	13.0x
Banking (AIX)	6.1x
Food manufacturer (Linux)	5.6x
Analytics benchmark (AIX)	3.0x

Table 1. Performance increases for a set of workloads measured on DB2 10.5 with BLU Acceleration compared against traditional row-organized processing with optimal indexing on the same server. Based on customer testing.

Overall performance is often skewed by outliers—those few queries that seem to run much longer than others. BLU Acceleration actually provides the biggest speed boost to those queries having the longest execution times. IBM has found that performance in DB2 10.5 is often at least three times less variable than traditional business intelligence systems because access plans for column-organized tables are simplified.

Although BLU Acceleration is in-memory optimized, it is not main memory-limited. BLU Acceleration is highly optimized for accessing data in RAM, but performance won't suffer as data size grows beyond RAM. These remarkable benefits are achieved by combining columnar and vector processing, operating on compressed data, carefully exploiting modern microprocessor designs, and accessing memory efficiently. The result is a system that simultaneously looks and feels like DB2 while being in-memory optimized, CPU-optimized, and I/O-optimized.

Making complex analytics easy

When developing DB2 with BLU Acceleration, IBM's mantra has been “super analytics, super easy.” This means the focus was on making DB2 easy and intuitive to use, similar to IBM's work involving query speeds.

The goal was to allow users to create and load their tables and then immediately start running queries—which is achieved using a new single registry setting, DB2_WORKLOAD=ANALYTICS. Once set—ideally before creating the database—DB2 automatically adapts resources, configuration, workload management, and storage parameters to optimize resource consumption for the target server. It also enables BLU Acceleration by default, creating all new user tables in column-organized format. Subsequently, users simply load data to run their queries, without the need for tuning.

It's not just the setup that's easier, either. There's less ongoing maintenance to worry about. There are no indexes or materialized query tables (MQTs) to define or tune. Storage is automatically freed and returned to the system for reuse as data is deleted over time. Even the compression algorithms will automatically adapt to changing data patterns.

You can easily convert tables to the column-organized format using the new db2convert command-line utility or by leveraging similar tooling in IBM Data Studio 4.1, which can convert any number of tables from row to column organization.

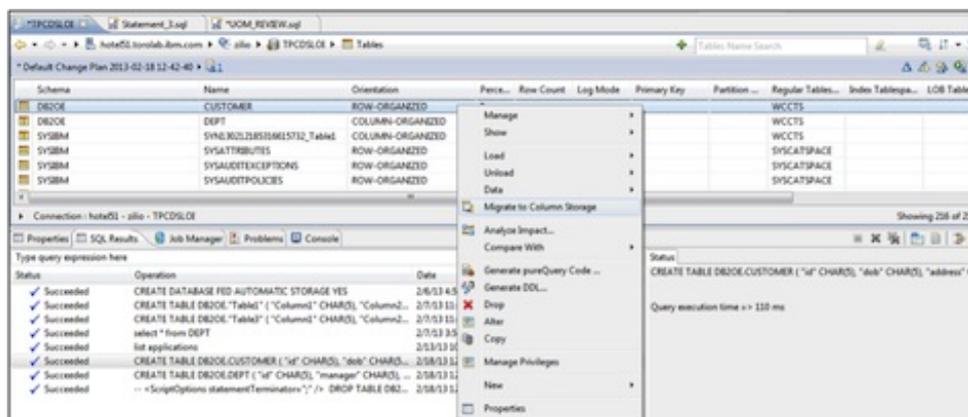


Figure 1. IBM Data Studio 4.1 lets DBAs convert tables from row to column organization quickly and easily.

You can do most of the typical tasks you're used to in DB2. There's no need to change the SQL in existing applications because BLU Acceleration reuses the same SQL compiler and optimizer. Most utilities—including LOAD, INGEST, EXPORT, BACKUP and RESTORE, ROLLFORWARD, and many others—work as usual. In addition, DB2 10.5 introduces an exciting new feature: the ability to mix row-organized and column-organized tables in the same storage (that is, tablespace), bufferpool, schema, and even within the same SQL statement. However, testing shows that the performance of any analytics query is best if all the tables referenced in that query are column-organized.

Making complex analytics more resource-friendly

DB2 10.5 with BLU Acceleration introduces automatic workload management when `DB2_WORKLOAD=ANALYTICS` is set. This feature ensures that while any number of queries may be submitted by applications, only a controlled number are allowed to consume resources simultaneously. By providing more resources per query, queries can zip through the system without competing with each other for memory, locks, CPU, and I/O bandwidth. As a result, all queries run faster—even under heavy load.

BLU Acceleration also reduces storage requirements in a number of important ways. First, since no secondary indexes or MQTs are needed on column-organized tables, you save storage space. Second, BLU Acceleration exploits multiple compression techniques on each column, including separate order-preserving, frequency-encoded dictionaries, and offsets (or "deltas") from dictionary elements to compress each value to just a few bits. This approach allows multiple values to fit into a machine word.

These patented techniques permit DB2 to not only store the data more efficiently, but also to better process it while it is still compressed. BLU Acceleration applies predicates, performs joins, and does grouping, all on the compressed values of column-organized tables. This combination brings together all resources—I/O bandwidth, bufferpools, memory bandwidth, processor caches, and even machine cycles—through single-instruction, multiple data (SIMD) operations.

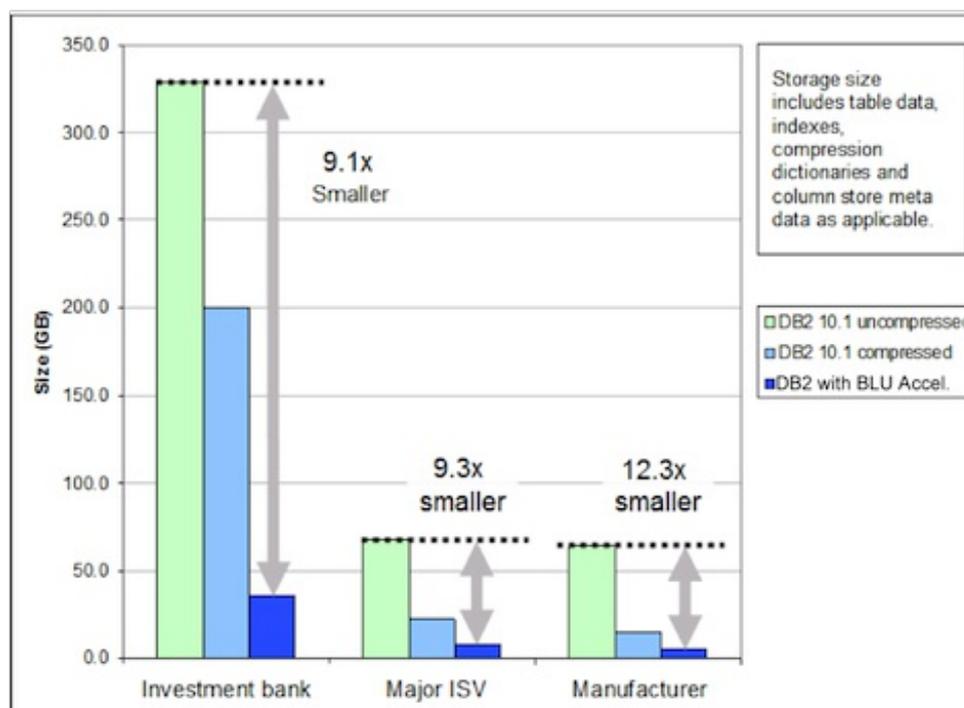


Figure 2. BLU Acceleration dramatically reduces storage requirements for analytics databases, typically by 10 times compared to uncompressed data in traditional databases.

When to use BLU Acceleration

If you have a workload that exclusively executes deep analytic queries, then the decision is easy: use BLU Acceleration. If your workload is somewhat mixed, then the Workload Table Organization Advisor in IBM Data Studio 4.1 can analyze your workload and recommend which tables should take advantage of this new technology.

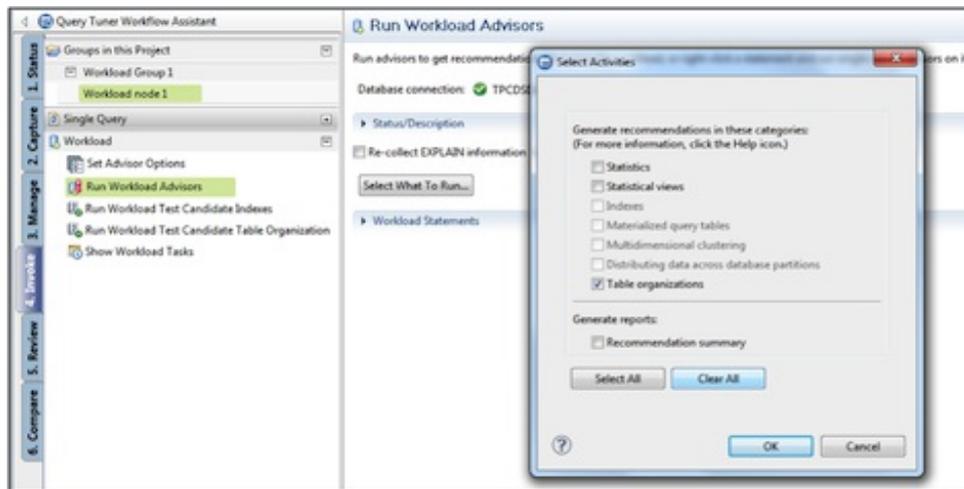


Figure 3. Optim Query Workload Tuner 4.1 can help you analyze workloads and decide when to use BLU Acceleration.

You'll have to try DB2 10.5 with BLU Acceleration to fully appreciate just how fast your queries will run, how simple it is to administer, and how small your column-organized tables can be compressed. We are confident you will be pleased with your improved time-to-value.

Resources

To learn more, visit the [DB2 Technology Preview](#).

Related Topics

1. [Hyper-Speed Analytics with IBM DB2 for z/OS](#)
2. [Gear up with IDAA: IBM DB2 Analytics Accelerator](#)
3. [IBM DB2 Analytics Accelerator V2](#)
4. [Exploring the Origins of IBM DB2 Analytics Accelerator](#)
5. [Making the Move to IBM DB2 on Intel Xeon Processors](#)