



TECHNOLOGIE ANALIZY DANYCH I CHMUROWE W
ZASTOSOWANIACH BIZNESOWYCH

Poznań, 30 września 2015

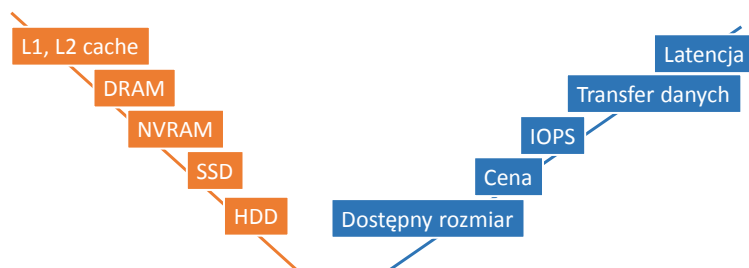
DB2 BLU od środka

Artur Wroński



Baza danych in-memory

Baza danych IN-MEMORY – system zarządzania bazami danych, który wykorzystuje pamięć operacyjną jako główny nośnik do składowania bądź przetwarzania danych.



DB2 BLU

35-73_x

😊 Szybsza analityka, niż
w bazie organizowanej
wierszami

10_x

😊 Oszczędności
przestrzeni dyskowej

Zerø

😊 Strojenia



DB2 BLU – optymalizacja w zakresie

Pamięci

IN-MEMORY

CPU

Single Instruction
Multiple Data

I/O

Kompresja



DB2 BLU – składowanie kolumnowe

```
CREATE TABLE tabela1 (
  k1 BIGINT,
  k2 VARCHAR(30),
  k3 INT,
  k4 DECFLOAT)
```

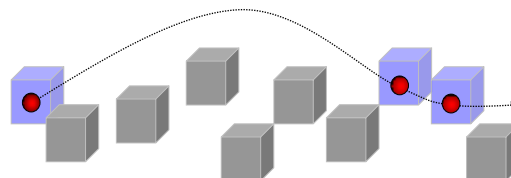
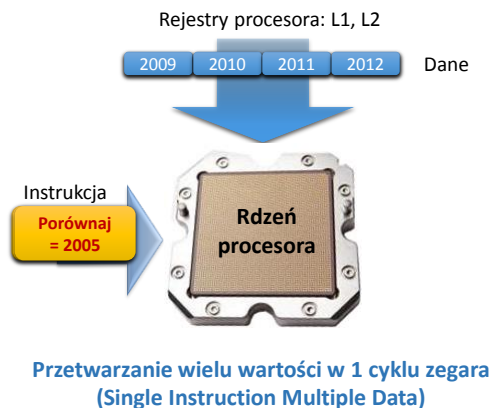
```
PRIMARY KEY (k1)
ORGANIZE BY COLUMN
```



DB2 BLU – kompresja



DB2 BLU – optymalizacja CPU



DB2 BLU - wymiarowanie

? **N** - rozmiar danych przed kompresją

? **%** danych aktywnych (%Kol.|%Wier.)

? **X** współczynnik kompresji

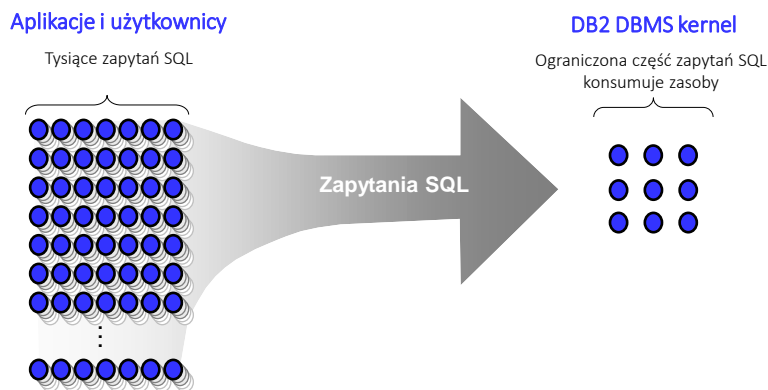
40% RAM – bufor danych

60% RAM – obsługa sesji

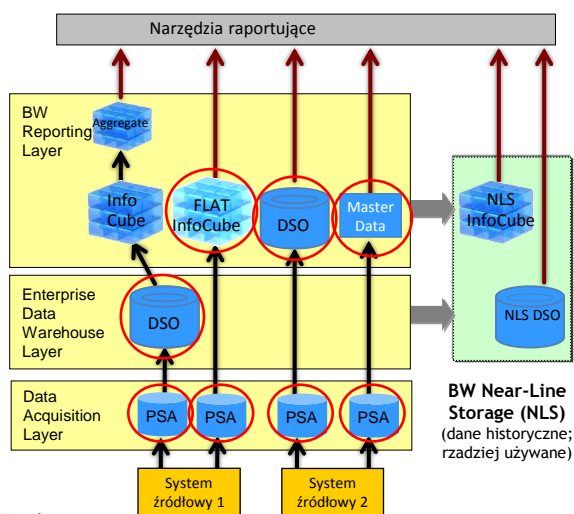
	Małe	Średnie	Duże
Dane surowe	1 TB	5 TB	10 TB
Minimalne wymagania dla wydajności			
Rdzenie	8	16	32
Pamięć	64 GB	256 GB	512 GB
Wysokie wymagania dla wydajności			
Rdzenie	16	32	64
Pamięć	128-256 GB	384-512 GB	1024-2048 GB



DB2 BLU - zarządzanie obciążeniem



DB2 BLU – przykład użycia: SAP BW



😊 DB2 BLU jest **certyfikowane** do użycia w systemie SAP BW i zintegrowane z narzędziami SAP

😊 **Wszystkie** obiekty SAP BW mogą być składowane kolumnowo i akcelerowane w pamięci (od aktualizacji FP4).



DB2 BLU – wdrożenie w SAP BW

BALLUFF
sensors worldwide



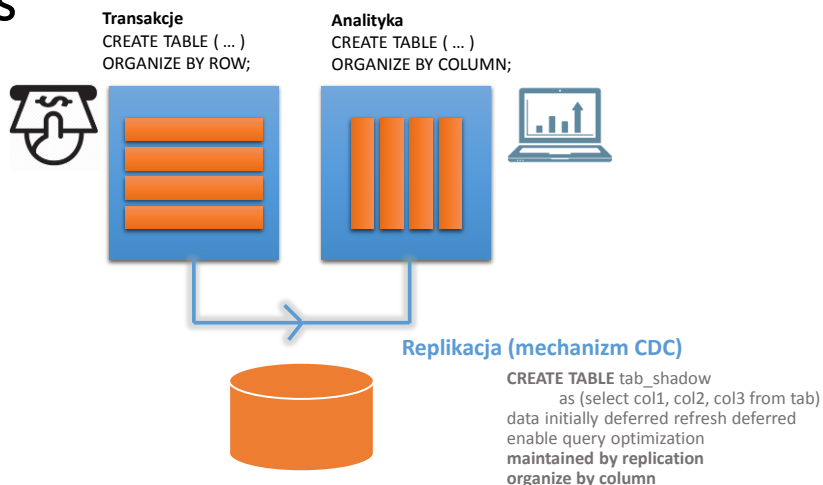
"Since we moved to IBM DB2 with BLU Acceleration, our sales team can react much quicker to fast-changing market demands. For example, we cut database run time of complex reports by up to 98 percent. Across all business reports, we achieved savings of 30 percent on average."

- Bernd Dummel, Team Manager SAP BW and Application Development at Balluff



„Shadow Tables”

- Połączenie przetwarzania transakcyjnego i analitycznego na tych samych danych
- „Shadow table” jest rodzajem agregatu kolumnowego (MQT)
- Przezroczyste dla aplikacji
- Pozwala zrezygnować z indeksów (do analityki) na tabeli wierszowej.



dashDB – DB2 BLU+ w chmurze



dashDB – DB2 BLU+ w chmurze

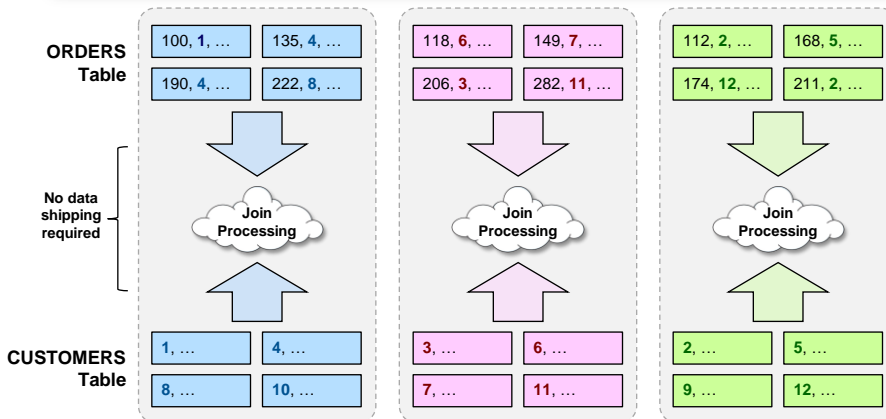
Plan	CPU rdzenie	Pamięć	Dyski	Cena
Freemium	współdzielone	współdzielona	< 1 GB	\$ 0
Entry	współdzielone	współdzielona	20 GB	\$50 USD (za miesiąc)
Enterprise 4 TB	32	256 GB	4 TB HDD	\$4560 USD (za miesiąc)
Enterprise 12 TB	32	256 GB	12 TB HDD	\$7160 USD (za miesiąc)
Enterprise MPP	24 (nowsze CPU) (per węzeł)	256 GB (per węzeł)	4 TB SSD (per węzeł)	\$5260 USD (za miesiąc)



dashDB MPP – klaster DB2 BLU

```
CREATE TABLE ORDERS (ORDER_NO, CUST_NO, ...) DISTRIBUTE BY HASH (CUST_NO)
CREATE TABLE CUSTOMERS (CUST_NO, ...) DISTRIBUTE BY HASH (CUST_NO)
SELECT ... FROM ORDERS O, CUSTOMERS C WHERE O.CUST_NO = C.CUST_NO
```

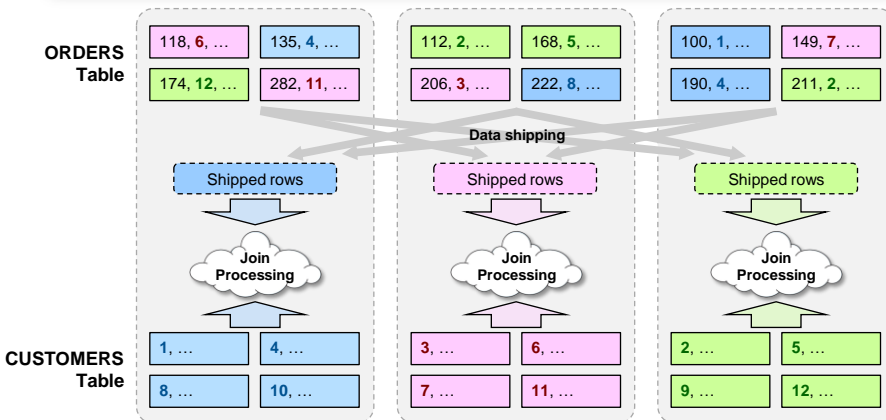
Te same klucze dystrybucji.



dashDB MPP – klaster DB2 BLU

```
CREATE TABLE ORDERS (ORDER_NO, CUST_NO, ...) DISTRIBUTE BY HASH (ORDER_NO)
CREATE TABLE CUSTOMERS (CUST_NO, ...) DISTRIBUTE BY HASH (CUST_NO)
SELECT ... FROM ORDERS O, CUSTOMERS C WHERE O.CUST_NO = C.CUST_NO
```

Różne klucze dystrybucji



www.ibm.com/software/data/dashdb/



IBM dashDB is here.

Keeping data warehouse infrastructure out of your way

dashDB is a fully managed data warehouse service in the cloud,
it's an analytics powerhouse at your finger tips

Get started with dashDB

Sign-in at Cloudant.com

SIGN UP FOR A FREE TRIAL

For anyone who wants a data warehouse in the cloud

For Cloudant users who want to leverage analytics



www.ibm.com/software/data/dashdb/

The screenshot shows the IBM dashDB user interface. At the top right, the URL www.ibm.com/software/data/dashdb/ is visible. The user is logged in as 'awronsk1'. A dark sidebar on the left contains a navigation menu with the following items: Home, Analyze, Manage, Monitor, Set Up, Documentation, and Help. The main content area features a blue header with the text 'Welcome. Your database is ready.' and a close button (X). Below this, a message reads: 'Welcome to IBM dashDB. You can load your data, run SQL queries against the data, and use tools to explore predictive analysis and in-database analytics. Sample data is already loaded. Start exploring the power of dashDB!'. Two buttons are present: 'Load your data' and 'Go to your tables'. At the bottom, a large white banner with a background image of a desk says 'Work Faster'.



IBM Bluemix

Big Data

Harness the power

- BigInsights for Hadoop (IBM)
- dashDB (IBM)**
- Geospatial Analytics (IBM)
- IBM Analytics for Hadoop (IBM BETA)
- IBM Insights for Twitter (IBM BETA)
- Time Series Database (IBM)



IBM dashDB™

- Home
- Analyze
- Manage
- Monitor
- Set Up**
 - Downloads
 - Connect Applications
- Documentation
- Help

Driver

You might need to download and install a driver package:

- Driver package for Windows 64-bit (92.6 MB)
- Driver package for Linux 64-bit (66.6 MB)
- Driver package for Mac OS X 64-bit (53.8 MB)

Connection without SSL

Settings to connect to the database are shown in v

Individual database settings:

Host name:	bluemix05.blufc
Port number:	50000
Database name:	BLUDB
User ID:	dash012858

Database Administration - BLUDB - DB2 UDB V10.5 - IBM Data Studio

Administration Explorer

- All Databases
 - 10.0.203.107
 - bluemix05.bluforcloud.com
 - db2inst1
 - BLUDB (DB2 for Linux, UNIX, and Windows)
 - Change Plans
 - Tables
 - Views
 - Indexes
 - Constraints
 - Triggers
 - MQTS
 - Sequences
 - Aliases
 - Schemas
 - Temporary Tables
 - Storage Groups
 - Table Spaces
 - Buffer Pools
 - Partition Groups
 - Fine Grained Access Controls
 - Application Objects
 - Users and Groups
 - Federated Database Objects
 - XML Schemas
 - Working Sets

Task Launcher: bluemix05.bluforcloud.com | db2inst1 | BLUDB | Tables

Schema	Name	Per...	Row Count	Primary Key	Partitio...
SAMPLES	SHOWCASE_SYSUS...	-1	5000000		
SAMPLES	PERSON	-1	1048575	SQL150212...	
SAMPLES	GEO_CUSTOMER	-1	1000000	SQL150212...	
SYSIBM	SYSCOLDIST	-1	600270		
GOSALES	ORDER_DETAILS	-1	446023	SQL140620...	
GOSALES...	SLS_SALES_FACT	-1	446023	SQL140620...	
GOSALES...	SLS_SALES_ORDER...	-1	446023	SQL140620...	
GOSALES	SALES_TARGET	-1	233625	SQL140620...	
GOSALES...	SLS_SALES_TARG_F...	-1	233625	SQL140620...	
GOSALES...	MRK_PRODUCT_SU...	-1	165074	SQL140620...	
GOSALEMR	PRODUCT_SURVEY...	-1	165074	SQL140620...	
GYSALFSD	FIN_FINANCE_FACT	-1	164132	SQL140620...	

Connection: bluemix05 - db2inst1 - BLUDB

Properties | SQL Results | Console | Data Project Explorer

<Column-organized table> SHOWCASE_SYSUSAGE

Privileges	Name	Primar...	Domain	Data Type	Length
Table Spaces	SID	<input type="checkbox"/>		INTEGER	
Versioning	DATE	<input type="checkbox"/>		TIMESTAMP	6
Relationships	USERS	<input type="checkbox"/>		INTEGER	
Statistics	MEMUSED	<input type="checkbox"/>		INTEGER	
Documentation	AI_FRT	<input type="checkbox"/>		VARCHAR	1

Temporal attributes: System time period Business time period



Run SQL Scripts

Enter one or more SQL statements, which you can validate and run against the database. Use the Syntax Assistant to help with a basic SELECT statement for

You can view the results of the query, which you can sort or download as a CSV file. You can also run the query directly from Excel by downloading an ODC file

Run Validate Syntax Validation: Not Validated Syntax Assist Save Open

```

1 -- Example SQL queries that you can run against the sample data in dashDB.
2 ;
3
4 SELECT ORGANIZATION_KEY, COUNT(*) AS "Row Count" FROM GOSALESDW.EMP_EXPENSE_FACT GROUP BY ORGANIZATION_KEY ORDER BY ORGANIZATION_KEY;
5
6 SELECT CLOSE_DAY_KEY, EMPLOYEE_KEY, GROSS_MARGIN, GROSS_PROFIT, ORDER_DAY_KEY,
7 ORDER_METHOD_KEY, ORGANIZATION_KEY, PRODUCT_KEY, PROMOTION_KEY, QUANTITY, RETAILER_KEY
8 RETAILER_SITE_KEY, SALES_ORDER_KEY, SALE_TOTAL, SHIP_DAY_KEY, UNIT_COST, UNIT_PRICE, UNIT_SALE_PRICE
    
```

Status	Run time (seconds)	Statement
✓ Succeeded - BLUDB	1.026	
✓ Succeeded	0.379	SELECT ORGANIZATION_KEY, COUNT(*) AS "Row Count" FROM G...

Run Query in Excel via ODC file Log

ORGANIZATION_KEY	Row Count
11101	1022
11102	1364



Develop R Scripts

Click Submit to generate a plot. Errors, warnings, or messages are sent to the Console Output page.

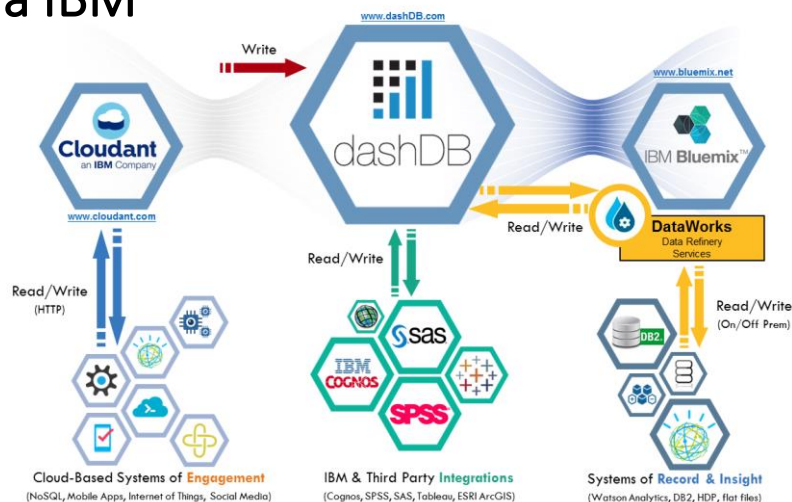
Script Console Output Plots

Histogram Of Education Level by Gender

- Associate's degree
- Bachelor's degree
- Doctorate degree
- Master's degree
- One or more years of college, no degree
- Professional school degree
- Some college, but less than 1 year



Chmura IBM



SoftLayer Infrastructure as a Service

Kontakt:

<http://pl.linkedin.com/in/arturwronski/en>
artur.wronski@pl.ibm.com

603-886649

DB2 BLU

- Ponadprzeciętna wydajność (35-73x)
- Efektywna kompresja (10x)
- Prostota użycia: Załaduj dane i działaj
- dashDB w chmurze:
 - Usługa w IBM BlueMix
 - Integracja z NOSQL, Cloudant
 - R oraz analityka na danych przestrzennych
- „Shadow tables” – raportowanie na tabelach OLTP
- Zgodność z bazami Oracle @98%+
- Klaster geograficzny HADR

