



The Hadoop RDBMS

Iwona Milczarek, Małgorzata Marcinkiewicz, Tomasz Staszewski

Poznań, 30.09.2015



The Hadoop RDBMS

Plan

- Geneza
- Architektura
- Cechy
- Instalacja
- Standard SQL
- Transakcje i współbieżność
- Indeksy
- Administracja
- Splice Machine vs. Oracle



The Hadoop RDBMS

Geneza



rozwiązania RDBMS

- wysokie koszty skalowalnych rozwiązań

rozwiązania NoSQL

- alternatywa dla RDBMS
- brak uniwersalnego SQL
- trudności w zapewnieniu spójności danych

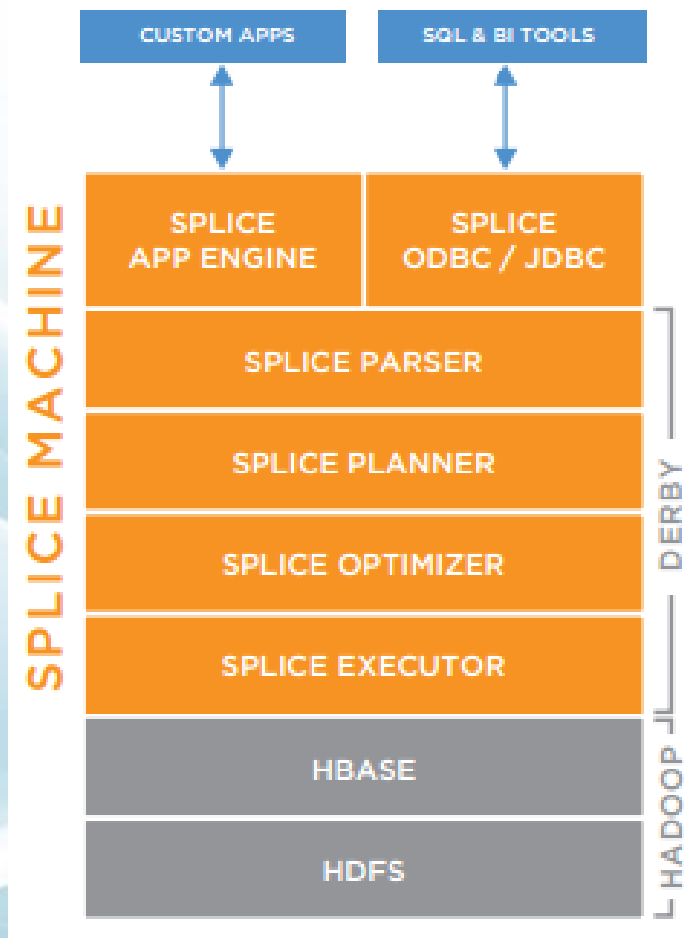
Splice Machine

- standard ANSI SQL- 99
- transakcyjność
- skalowanie poziome (horyzontalne)



The Hadoop RDBMS

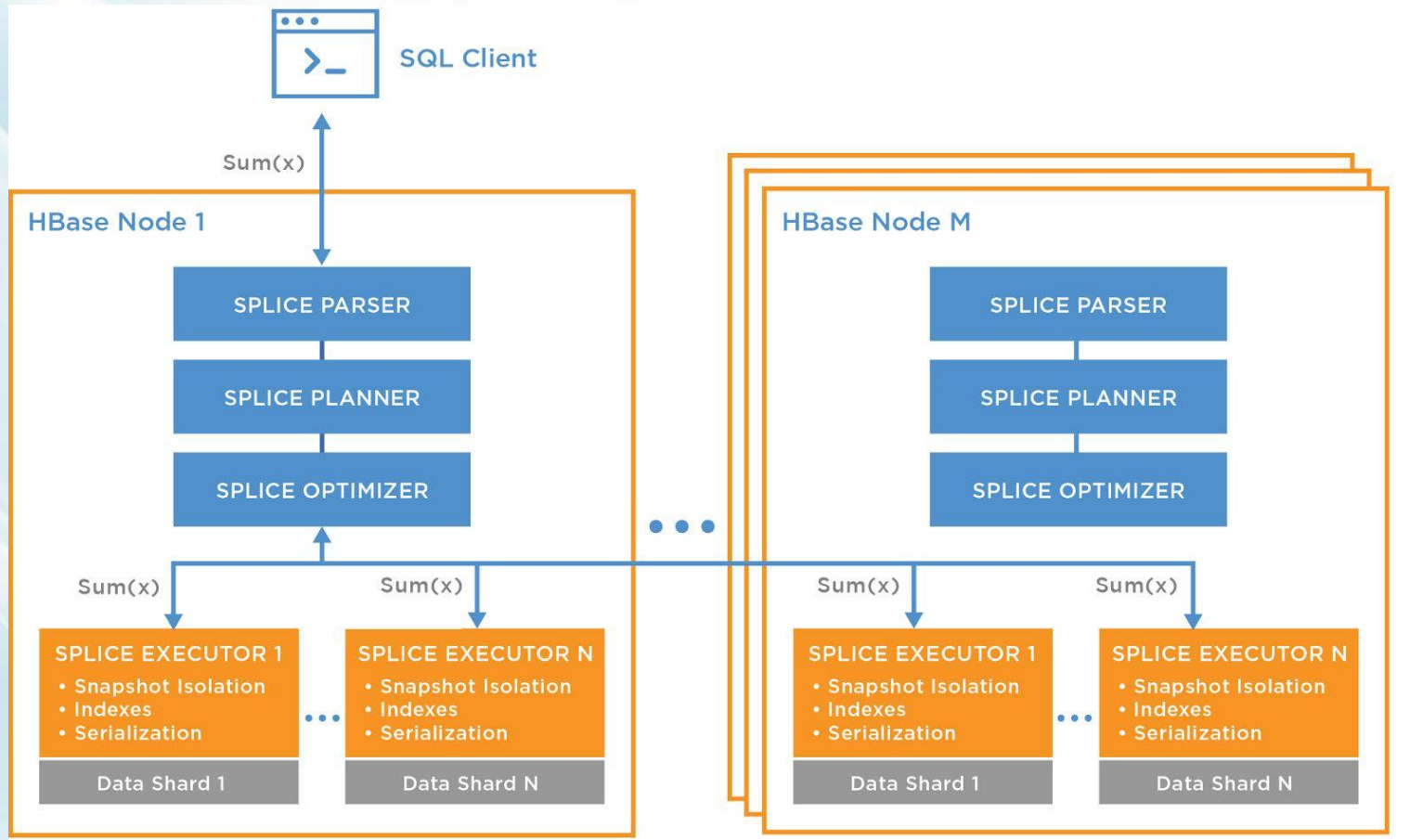
Architektura





The Hadoop RDBMS

Architektura





The Hadoop RDBMS

Instalacja

- system operacyjny – Centos, Red Hat, Ubuntu, MacOS, Windows
- JDK 1.7
- rqlap – do zapamiętywania historii poleceń

konfiguracja środowiska zgodna z dokumentacją

konieczne zwiększenie parametru startowego dla Javy



The Hadoop RDBMS

Instalacja

```
[splice@centos splicemachine]~ ./bin/start-all.sh
Checking for Java...
Starting Splice Machine...
Log file is /home/splice/splicemachine/splice.log
Waiting for Splice...
.Detected zookeeper running@localhost in Mode: standalone
.....
Splice Server is ready
The HBase URI is http://localhost:60010
The JDBC URI is jdbc:splice://localhost:1527/splicedb
Running with debug = false and config file = /home/splice/splicemachine/etc/jetty-
splice-admin.xml
Checking for Java...
Starting Splice Machine Admin...
Log file is /home/splice/splicemachine/splice-admin.log
Jetty Runner jar file is /home/splice/splicemachine/lib/jetty-runner-
8.1.15.v20140411.jar
Admin war file is /home/splice/splicemachine/lib/splice-web-1.0.0.war
Waiting for Splice Admin...
..
Splice Admin Server is ready
The Admin URI is http://centos.localdomain:7020/
```




The Hadoop RDBMS

Instalacja

- dostęp do bazy danych : JDBC, ODBC
- sqlshell.sh

Running Splice Machine SQL shell

For help: "splice> help;"

```
SPLICE* - jdbc:splice://localhost:1527/splicedb
```

* = current connection

```
splice>
```




The Hadoop RDBMS

Instalacja

- mapowanie tabel Hbase na tabele w splicedb
(Hbase wykorzystywany jako magazyn danych,
zapis informacji o mapowaniu w tabelach systemowych)

```
splice> show tables;
TABLE_SCHEM      |TABLE_NAME                                     |CONGLOM_ID|REMARKS
-----|-----|-----|-----
SYS              |SYSALIASES                                   |336       |
SYS              |SYSCHECKS                                   |272       |
SYS              |SYSCOLPERMS                                 |784       |
SYS              |SYSCOLUMNS                                 |80        |
SYS              |SYSCONGLOMERATES                            |32        |
SYS              |SYSCONSTRAINTS                              |288       |
SYS              |SYSDEPENDS                                  |256       |
SYS              |SYSFILES                                     |384       |
SYS              |SYSFOREIGNKEYS                              |368       |
SYS              |SYSKEYS                                      |240       |
SYS              |SYSOPERATIONHISTORY                         |1200      |
SYS              |SYSPERMS                                    |848       |
SYS              |SYSPRIMARYKEYS                              |1168      |
SYS              |SYSROLES                                     |816       |
SYS              |SYSROUTINEPERMS                             |800       |
SYS              |SYSSCHEMAS                                  |48        |
SYS              |SYSSEQUENCES                                |832       |
SYS              |SYSSTATEMENTHISTORY                         |1184      |
SYS              |SYSSTATEMENTS                               |304       |
SYS              |SYSSTATISTICS                               |656       |
SYS              |SYSTABLEPERMS                              |768       |
SYS              |SYSTABLES                                   |64        |
SYS              |SYSTASKHISTORY                              |1216      |
SYS              |SYSTRIGGERS                                 |320       |
SYS              |SYSUSERS                                    |864       |
SYS              |SYSVIEWS                                    |352       |
```

Instalacja

192.168.90.3:60010/master-status?filter=operation ewa blasik

Tables

Catalog Table	Description
-ROOT-	The -ROOT- table holds references to all .META. regions.
.META.	The .META. table holds references to all User Table regions

102 table(s) in set. [\[Details\]](#)

User Table	Online Regions	Description
1009	1	{NAME => '1009', FAMILIES => [{(NAME => 'V', BLOOMFILTER => 'ROW', VERSIONS => '2147483647')}]}
1025	1	{NAME => '1025', FAMILIES => [{(NAME => 'V', BLOOMFILTER => 'ROW', VERSIONS => '2147483647')}]}
1041	1	{NAME => '1041', FAMILIES => [{(NAME => 'V', BLOOMFILTER => 'ROW', VERSIONS => '2147483647')}]}
1057	1	{NAME => '1057', FAMILIES => [{(NAME => 'V', BLOOMFILTER => 'ROW', VERSIONS => '2147483647')}]}
1073	1	{NAME => '1073', FAMILIES => [{(NAME => 'V', BLOOMFILTER => 'ROW', VERSIONS => '2147483647')}]}
1089	1	{NAME => '1089', FAMILIES => [{(NAME => 'V', BLOOMFILTER => 'ROW', VERSIONS => '2147483647')}]}
1105	1	{NAME => '1105', FAMILIES => [{(NAME => 'V', BLOOMFILTER => 'ROW', VERSIONS => '2147483647')}]}
1121	1	{NAME => '1121', FAMILIES => [{(NAME => 'V', BLOOMFILTER => 'ROW', VERSIONS => '2147483647')}]}
113	1	{NAME => '113', FAMILIES => [{(NAME => 'V', BLOOMFILTER => 'ROW', VERSIONS => '2147483647')}]}
1137	1	{NAME => '1137', FAMILIES => [{(NAME => 'V', BLOOMFILTER => 'ROW', VERSIONS => '2147483647')}]}
1153	1	{NAME => '1153', FAMILIES => [{(NAME => 'V', BLOOMFILTER => 'ROW', VERSIONS => '2147483647')}]}
1168	1	{NAME => '1168', FAMILIES => [{(NAME => 'V', BLOOMFILTER => 'ROW', VERSIONS => '2147483647')}]}
1184	1	{NAME => '1184', FAMILIES => [{(NAME => 'V', BLOOMFILTER => 'ROW', VERSIONS => '2147483647')}]}
1200	1	{NAME => '1200', FAMILIES => [{(NAME => 'V', BLOOMFILTER => 'ROW', VERSIONS => '2147483647')}]}
1216	1	{NAME => '1216', FAMILIES => [{(NAME => 'V', BLOOMFILTER => 'ROW', VERSIONS => '2147483647')}]}
129	1	{NAME => '129', FAMILIES => [{(NAME => 'V', BLOOMFILTER => 'ROW', VERSIONS => '2147483647')}]}
1312	1	{NAME => '1312', FAMILIES => [{(NAME => 'V', BLOOMFILTER => 'ROW', VERSIONS => '2147483647')}]}
1328	1	{NAME => '1328', FAMILIES => [{(NAME => 'V', BLOOMFILTER => 'ROW', VERSIONS => '2147483647')}]}
1344	1	{NAME => '1344', FAMILIES => [{(NAME => 'V', BLOOMFILTER => 'ROW', VERSIONS => '2147483647')}]}
1360	1	{NAME => '1360', FAMILIES => [{(NAME => 'V', BLOOMFILTER => 'ROW', VERSIONS => '2147483647')}]}
1376	8	{NAME => '1376', FAMILIES => [{(NAME => 'V', BLOOMFILTER => 'ROW', VERSIONS => '2147483647')}]}
1392	1	{NAME => '1392', FAMILIES => [{(NAME => 'V', BLOOMFILTER => 'ROW', VERSIONS => '2147483647')}]}



The Hadoop RDBMS

Standard SQL

- umożliwia tworzenie, usuwanie i modyfikację:
schematów, kluczy, indeksów, widoków, warunków integralności i
praw dostępu
- język zapytań oparty jest na algebrze relacji obejmującej polecenia
dodające, usuwające i aktualizujące dane w bazie danych (SELECT,
INSERT, DELETE, UPDATE)
- zapewnia kontrolę transakcji - posiada polecenia blokowania danych
do współbieżnych operacji (COMMIT, ROLLBACK)



The Hadoop RDBMS

Standard SQL

- podzapytania

zapytanie do tabeli po słowie FROM, WHERE, z EXISTS, z IN, z ORDER BY i

FETCH FIRST, podzapytania skalarne

```
SELECT *  
  FROM TPCD.SUPPLIER  
 WHERE EXISTS  
   (SELECT *  
     FROM TPCD.SUPPLIER  
     WHERE dest_CITY = 'RUSSIA' AND S_REGION = 'EUROPE');
```

- optymalizacja zapytań SQL

```
select * from TPCD.CUSTOMER --splice-properties index=CUST_INDX;
```



The Hadoop RDBMS

Standard SQL

DDL	CREATE TABLE, CREATE SCHEMA, ALTER TABLE, DELETE, UPDATE
DML	INSERT, DELETE, UPDATE, SELECT
Funkcje agregujące	AVG, MAX, COUNT
Funkcje okna	DENSE_RANK, RANK, ROW_NUMBER
Zapytania	SELECT DISTINCT, GROUP BY, HAVING
Sekwencje	TAK
Podzapytania	TAK
Synonimy	TAK
Wyzwalacze	NIE

Transakcje i współbieżność

Anomalia	Czy występuje w Splice Machine?
brudny odczyt (ang. dirty read)	nie
utracona modyfikacja (ang. lost update)	nie
niepowtarzalny odczyt (ang. non-repeatable read)	tak
fantomy (ang. phantoms)	nie

Transakcje i współbieżność

Poziom izolacji	Czy występuje w Splice Machine?
READ UNCOMMITTED (poziom izolacji 0)	nie
READ COMMITTED (poziom izolacji 1)	tak
REPEATABLE READ (poziom izolacji 2)	nie
SERIALIZABLE (poziom izolacji 3)	tak

Indeksy

- klucz główny i unikalny generuje indeksy
- nazwy indeksów są unikalne w ramach schematu
- indeks B- tree

```
splice> show indexes from sm_tab1;
```

TABLE_NAME	INDEX_NAME	COLUMN_NAME	ORDINAL	NON_UNIQUE	TYPE	ASC	CONGLOM_NO
SM_TAB1	TAB1_I1	WOJ	1	true	BTREE	A	1425
SM_TAB1	TAB1_I1	POW	2	true	BTREE	A	1425
SM_TAB1	TAB1_I1	R_GHI	3	true	BTREE	A	1425
SM_TAB1	TAB1_I1	GMI	4	true	BTREE	A	1425
SM_TAB1	TAB1_I1	OBR	5	true	BTREE	A	1425
SM_TAB1	TAB1_I2	ARK_MAP	1	false	BTREE	A	1441
SM_TAB1	TAB1_I2	NR_DZ	2	false	BTREE	A	1441
SM_TAB1	TAB1_I2	WOJ	3	false	BTREE	A	1441
SM_TAB1	TAB1_I2	POW	4	false	BTREE	A	1441
SM_TAB1	TAB1_I2	GMI	5	false	BTREE	A	1441
SM_TAB1	TAB1_I2	R_GHI	6	false	BTREE	A	1441
SM_TAB1	TAB1_I2	OBR	7	false	BTREE	A	1441
SM_TAB1	TAB1_I2	LP	8	false	BTREE	A	1441

13 wierszy wybranych



The Hadoop RDBMS

Administracja

system zarządzania bazą danych posiadający mechanizmy:

- zapewniające integralność i bezpieczeństwo danych
- pozwalające na odtworzenie zawartości bazy danych po awarii
- dostęp do danych poprzez język zapytań,
- wielodostępność danych (poprzez transakcje),
- mechanizmy pozwalające na autoryzację dostępu do danych
- optymalizujące czas dostępu do danych (indeksy)

Administracja

- czynności administracyjne wykonywane są poprzez procedury systemowe/funkcje

```
splice> CALL SYSCS_UTIL.SYSCS_CREATE_USER('Fred', 'fredpassword');
```

```
splice> call SYSCS_UTIL.SYSCS_BACKUP_DATABASE('BACKUP');
```

- tabele systemowe w schemacie użytkownika SYS

```
splice@centos:~/splicemachine
splice> show tables;
```

TABLE_SCHEM	TABLE_NAME	CONGLOM_ID	REMARKS
SYS	SYSALIASES	336	
SYS	SYSCHECKS	272	
SYS	SYSCOLPERMS	784	
SYS	SYSCOLUMNS	80	
SYS	SYSCONGLOMERATES	32	
SYS	SYSCONSTRAINTS	288	
SYS	SYSDEPENDS	256	
SYS	SYSFILES	384	
SYS	SYSFOREIGNKEYS	368	
SYS	SYSKEYS	240	
SYS	SYSOPERATIONHISTORY	1200	

Administracja

- plany wykonania zapytań – polecenie EXPLAIN

```
splice> explain select C_CUSTKEY from TPCD.CUSTOMER3 WHERE C_CUSTKEY BETWEEN 17000 AND 25000;
Plan
-----
ScrollInsensitiveResultSetNode ((n=3, estRowCount=9765, regions=1, cost=2441.25, estSingleScanCount=9765.0))
  ProjectRestrictNode ((n=2, estRowCount=9765, regions=1, cost=2441.25, estSingleScanCount=9765.0))
    ProjectRestrictNode ((quals=[true], n=1, estRowCount=9765, regions=1, cost=2441.25, estSingleScanCount=9765.0))
      FromBaseTable ((n=0, estRowCount=9765, table=CUSTOMER3,1728, using-index=CUSTOM3_INX4,1761, regions=1, cost=2441.25, estS&
4 rows selected
splice>
```

- wykonanie zapytania

select C_CUSTKEY from TPCD.CUSTOMER3 WHERE C_CUSTKEY BETWEEN 17000 AND 25000

▼ BulkTableScan [In: 0 rows, Out: 0 rows, Time: 0ms]
 RegionScan [In: 0 rows, Out: 8,001 rows, Time: 23ms]

RegionScan Details		
Summary	Input	Output
<ul style="list-style-type: none"> • Host: centos.localdomain • Iterations: 1 • Total Wall Time: 23ms • Region: 1761... 	<ul style="list-style-type: none"> • Local Scan Rows: 8,001 • Local Scan Wall Time: 4ms 	<ul style="list-style-type: none"> • Output Rows: 8,001



The Hadoop RDBMS

Splice Machine vs. Oracle

przyjęte założenia:

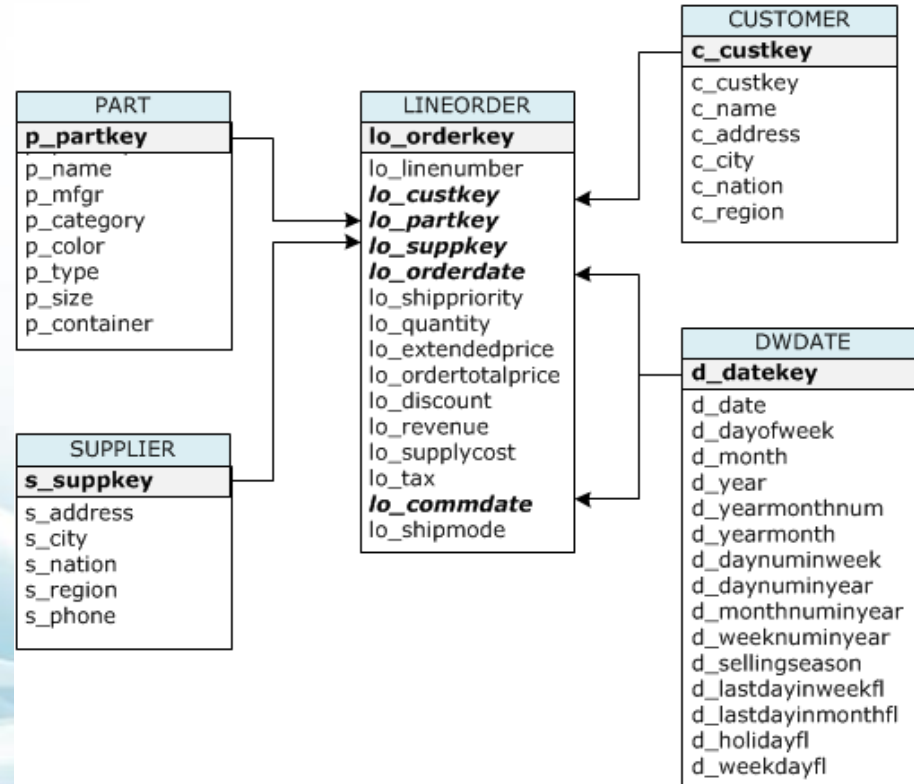
- wykorzystanie SSB do weryfikacji czy Splice Machine nadaje się do zapytań analitycznych
- sprawdzenie jak te same zapytania wyglądają w bazie Oracle



The Hadoop RDBMS

Star Schema Benchmark

	SF=1	SF=5	SF=10
CUSTOMER	30 000	150 000	300 000
SUPPLIER	2000	10 000	20 000
DWDATE	2556	2556	2556
PART	200 000	600 000	800 000
LINEORDER	6 001 171	29 999 810	59 986 214





The Hadoop RDBMS

Splice Machine vs. Oracle

Ładowanie danych

- Splice Machine – funkcja SYSCS_UTIL.IMPORT_DATA

```
splice> call SYSCS_UTIL.IMPORT_DATA('TPCD','SUPPLIER',null,  
'demodata/ssbdata/supplier.tbl','|',null,null,null,null,0,'demodata/ssbdata/');
```

- Oracle – SQL Loader z opcją DIRECT=TRUE

```
sqlldr CONTROL=ctrl_customer10ctl userid=tpcd10/tpcd10, DIRECT=true;
```

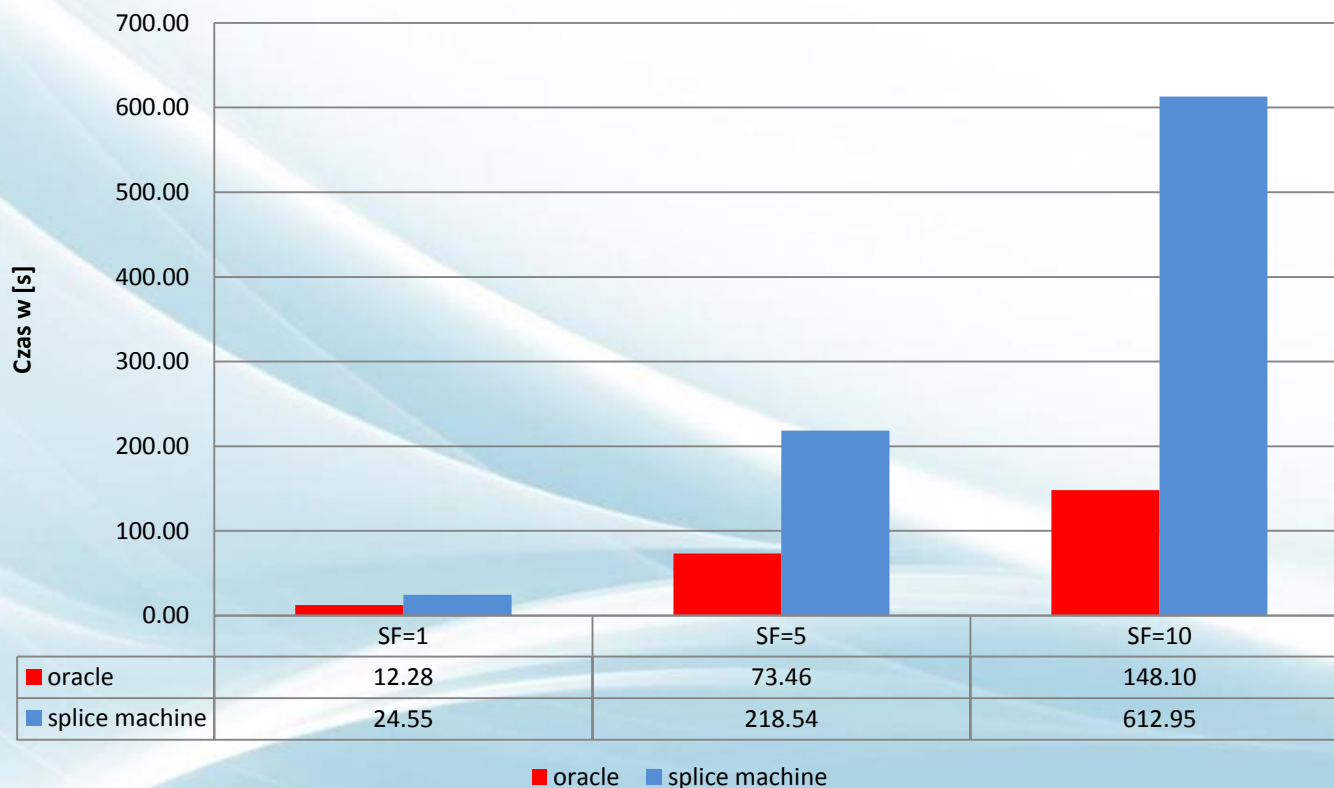
plik ctrl_customer10ctl

```
load data infile '/u01/ssbdata/1G/customer.tbl' into table customer fields  
terminated by "|"
```

```
(c_custkey,c_name,c_address,c_city,c_nation,c_region,c_phone,c_mktsegment)
```

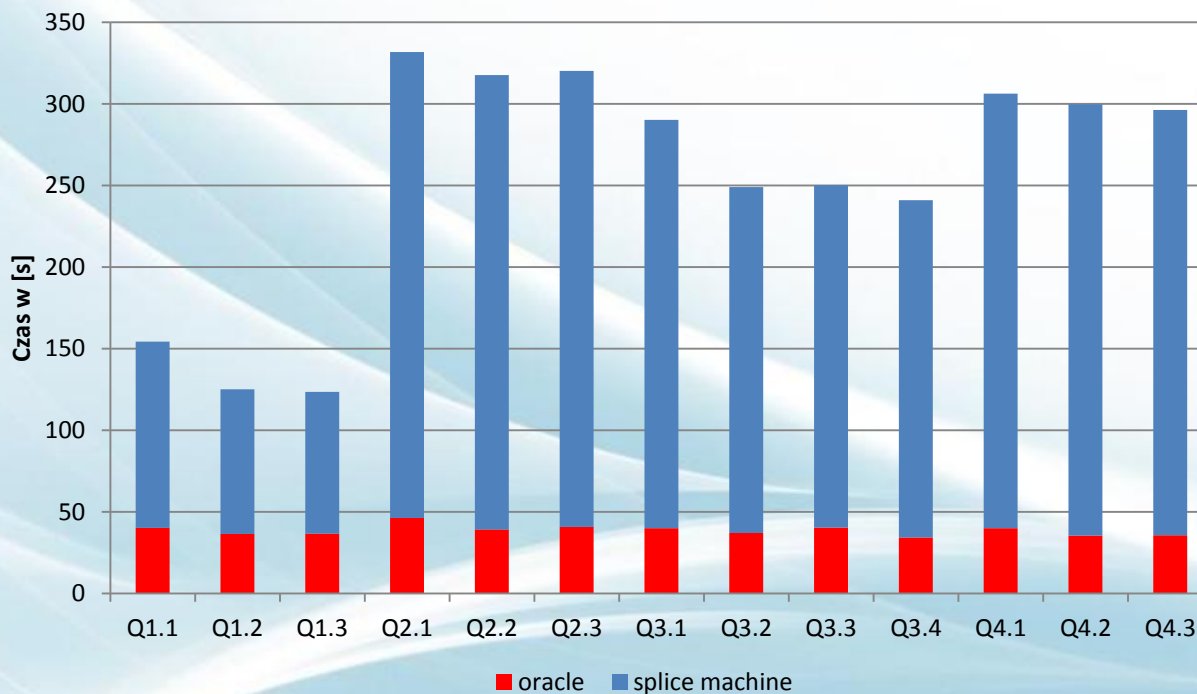

Star Schema Benchmark – czas ładowania danych do tabeli

Tabela lineorder



Star Schema Benchmark – czas wykonania zapytań

SSB - czasy wykonania zapytań dla SF=10





The Hadoop RDBMS

Podsumowanie



standard ANSI - SQL-99

ACID

skalowalność

przetwarzanie w czasie rzeczywistym

domyślnie nie korzysta z MapReduce

OLTP + OLAP

bazy NoSQL

brak transakcji

brak złączeń

ograniczony SQL

duży nakład pracy na zmianę aplikacji w przypadku przejścia na bazę NoSQL

bazy relacyjne

duże koszty skalowania rozwiązań

konieczność ręcznego partycjonowania

tabel