Oracle 12.2 Database Sharding

Maciej Zakrzewicz

XVII Konferencja PLOUG, 8.06.2017, Warszawa

PLAN PREZENTACJI

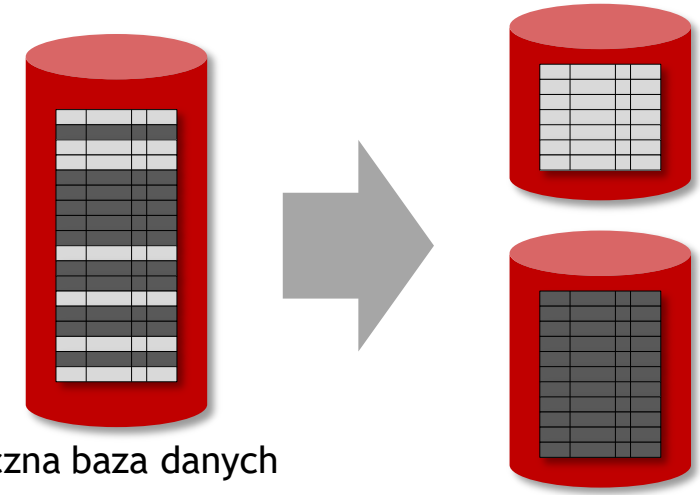
Oracle 12.2 Database Sharding

- Wprowadzenie do Database Sharding
- Mechanizmy Database Sharding w Oracle Database 12.2
- Budowa dzielonej bazy danych
  - narzędzia
  - przygotowanie środowiska
  - implementacja struktur danych
- Realizacja zapytań do tabel w dzielonej bazie danych
- Połączenia aplikacji z dzieloną bazą danych
- Podsumowanie

DATABASE SHARDING

Wprowadzenie

- Technika skalowania baz danych oparta na poziomym partycjonowaniu danych pomiędzy wiele fizycznych lokalizacji (*baz cząstkowych*)
  - Shared Nothing
- Każda baza cząstkowa (*shard*) przechowuje poziomy fragment tabeli



logiczna baza danych

fizyczne cząstkowe bazy danych

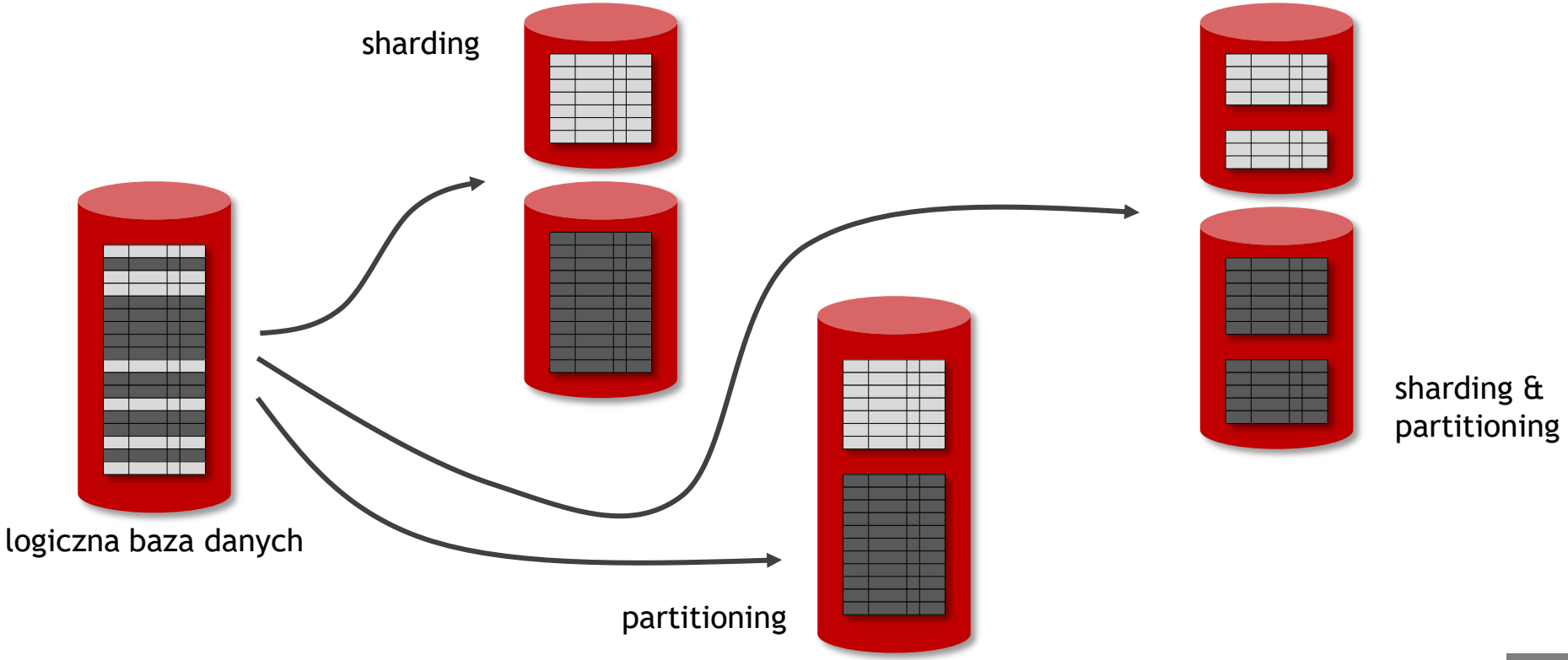
Wcześniejsze rozwiązania Database Sharding

Pochodzące głównie z obszaru systemów NoSQL

- Apache HBase
- Azure SQL Elastic Database
- Couchbase
- CUBRID
- Elasticsearch
- eXtreme Scale
- IBM Informix
- Kdb+
- MonetDB
- MongoDB
- MySQL Cluster
- Oracle NoSQL Database
- OrientDB
- Grails/Grails Sharding Plugin
- Octopus/Ruby Active Record
- PostgreSQL/Citus
- Shard Query
- Solr Search Server
- Spanner
- Teradata

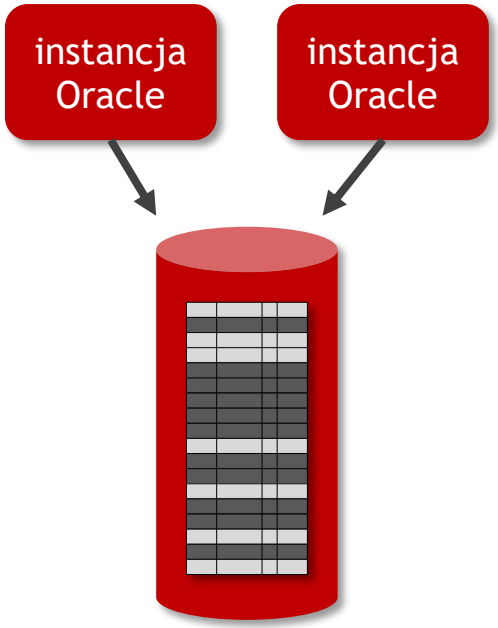
SHARDING VS. PARTITIONING

Porównanie metod partycjonowania poziomego

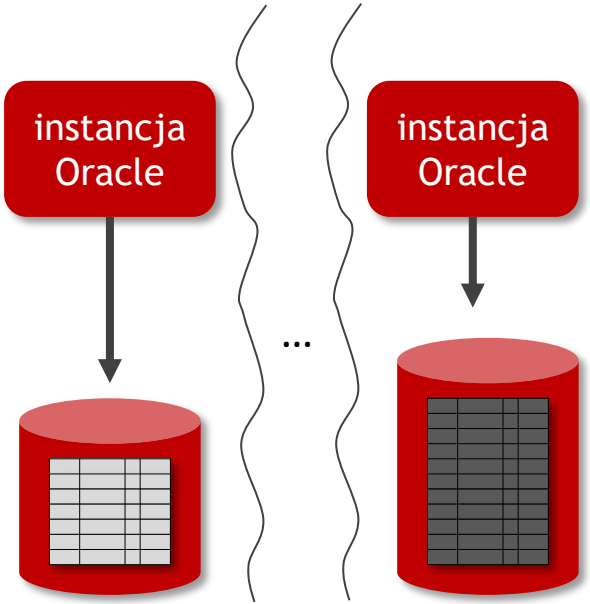


SHARDING VS. REAL APPLICATION CLUSTERS

Porównanie metod rozdzielania zasobów



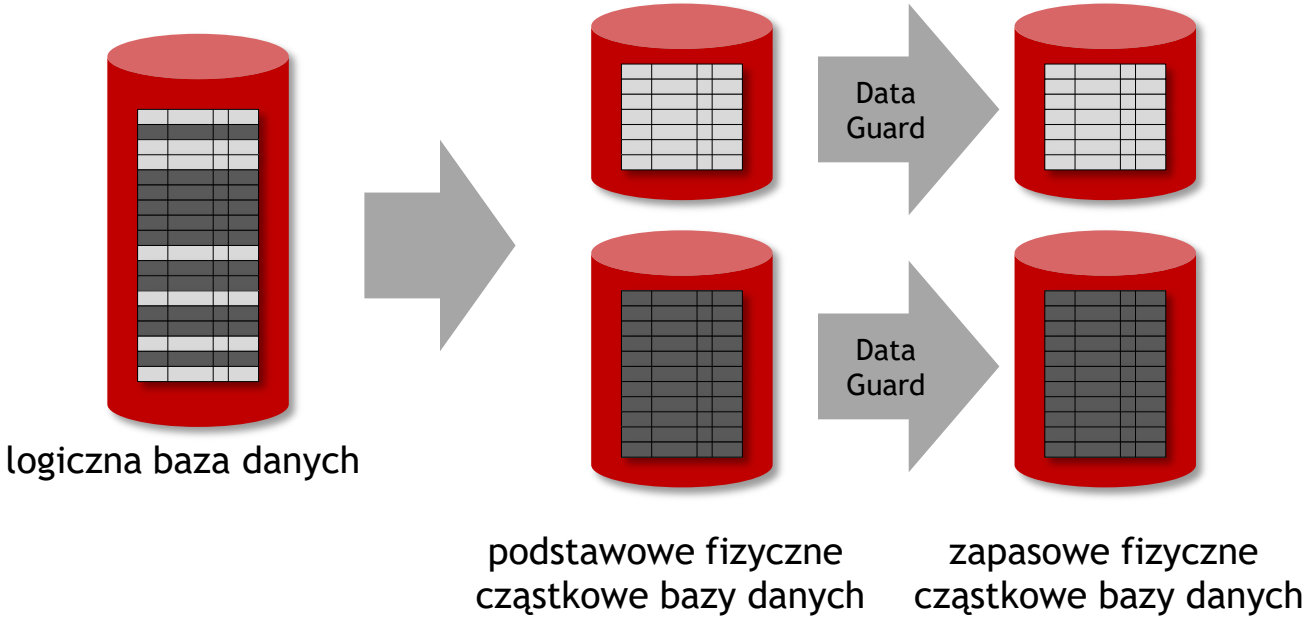
RAC



Database Sharding

NIEZAWODNOŚĆ

Za pomocą replikacji baz danych - Oracle Data Guard



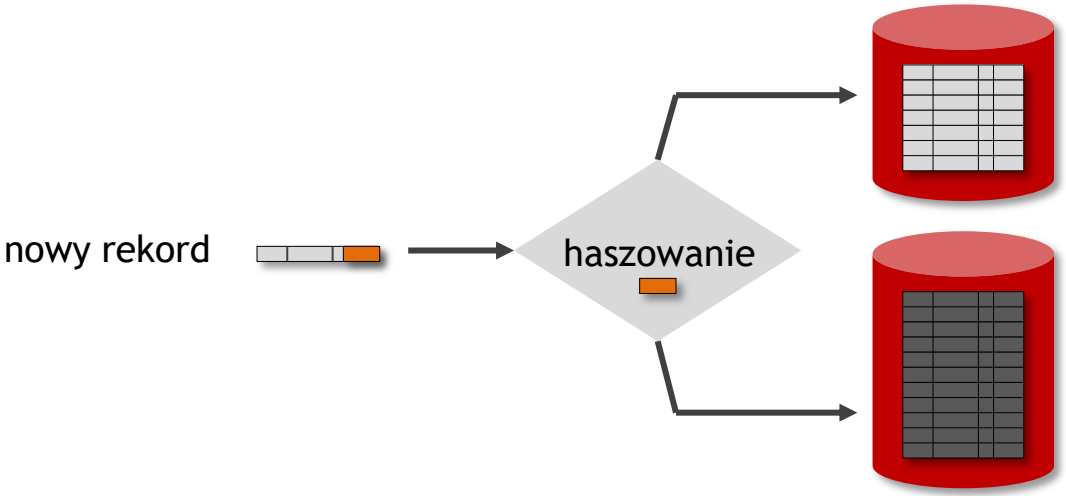
Mechanizmy Database sharding

W Oracle Database 12.2

- **System-managed sharding**
  - przydział rekordu do bazy cząstkowej na podstawie wartości funkcji haszowej
- **Composite sharding**
  - przydział rekordu do grupy baz cząstkowych (*shard space*) na wartości klucza, a następnie do bazy cząstkowej na podstawie wartości funkcji haszowej

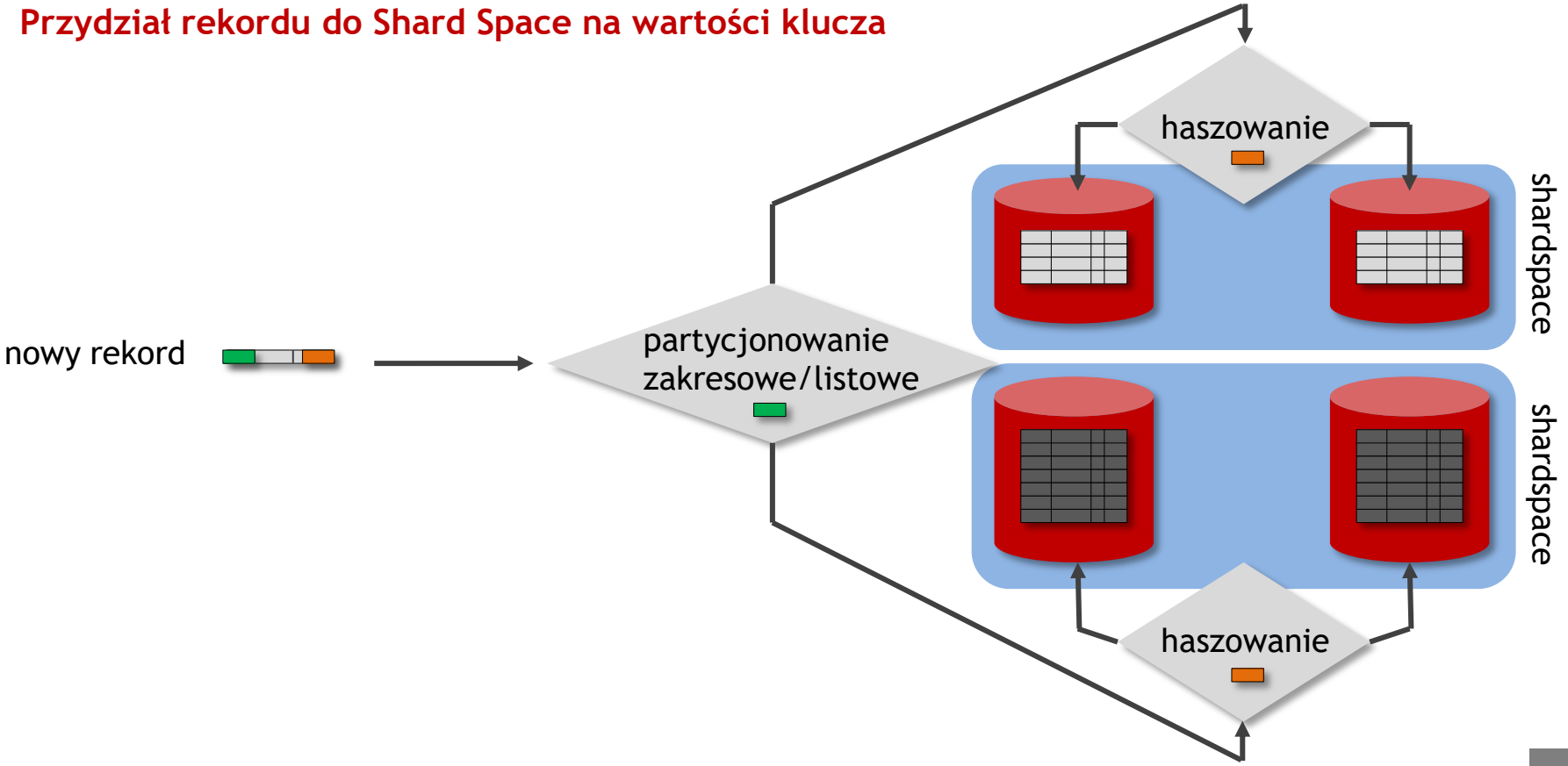
SYSTEM-MANAGED SHARDING

Przydział rekordu do bazy cząstkowej na podstawie funkcji haszowej



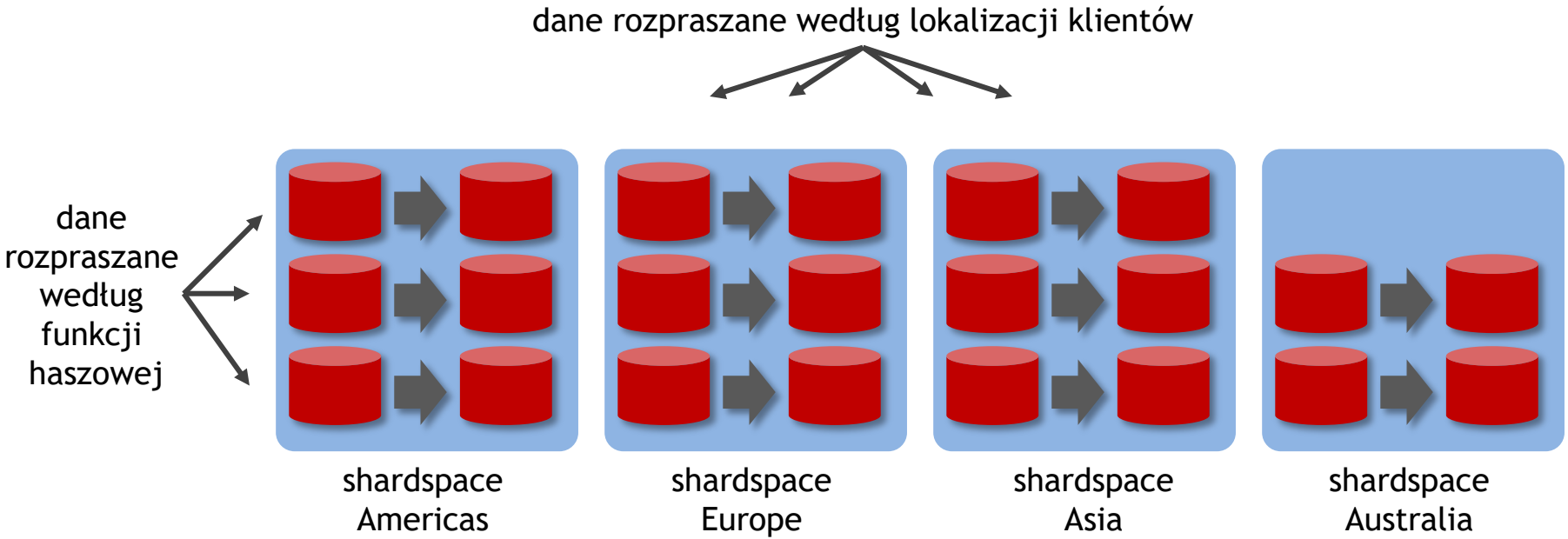
COMPOSITE SHARDING

Przydział rekordu do Shard Space na wartości klucza



PRZYKŁADOWA ARCHITEKTURA

Composite Sharding w większej skali

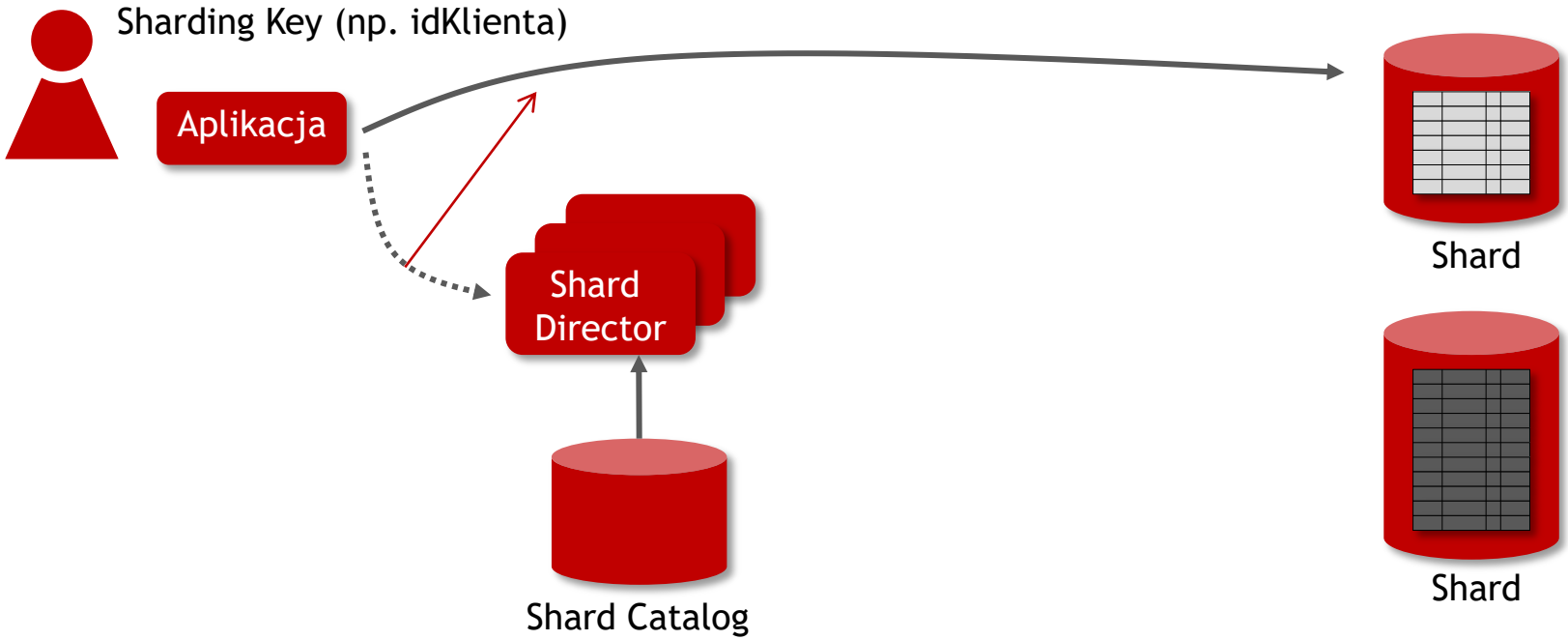


Zalety
Database Shardingu

- Prawie liniowa skalowalność rozmiaru bazy danych
- Izolacja usterek, eliminacja pojedynczego punktu awarii
- Możliwość geograficznej dyslokacji danych w pobliżu ich konsumentów
- Kroczące aktualizacje (*rolling upgrades*) nie wstrzymujące pracy systemu
- Łatwość wdrożenia w środowiskach Cloud

ARCHITEKTURA TECHNICZNA ROZWIĄZANIA

Oracle Database Sharding



Proces instalacji środowiska Sharded Database

- Maszyna **Shard Catalog**
  - Instalacja oprogramowania Oracle Database 12.2 i utworzenie bazy danych, która będzie pełnić rolę Shard Catalog
    - non-CDB, OMF, FRA, `open_links>16`, `open_links_per_instance>16`
    - odblokowane konto `gsmcatuser`
    - utworzenie konta właściciela Shard Catalogu
- Maszyny **Shard** (*maks. 1000*)
  - Instalacja oprogramowania Oracle Database 12.2
  - Uruchomienie Remote Scheduler Agenta (`schagent -start`)
  - Zarejestrowanie planowanej bazy danych w Shard Catalogu (`schagent -registerDatabase`)
- Maszyny **Shard Director**
  - Instalacja oprogramowania Shard Director (*Oracle Database 12.2 Global Service Manager*)
  - Za pomocą narzędzia GDSCTL (`gdsctl`):
    - połączenie z maszyną Shard Catalog i utworzenie Shard Catalogu (`create shardcatalog`)
    - utworzenie i uruchomienie procesu Shard Director (`add gsm, start gsm`)
    - zdefiniowanie i zdalne utworzenie cząstkowych baz danych na maszynach Shard (`create shard, deploy`)

NOWE NARZĘDZIE KONFIGURACYJNE: GDSCTL

```
gdsctl> create shardcatalog -database shardcat:1521:shardcat -chunks 12 -
user mysdbadmin/oracle -sdb shardcat -region region1, region2  -agent_port
8000 -agent_password oracle

gdsctl> add gsm -gsm sharddirector01 -listener 9000 -pwd oracle -catalog
shardcat:1521:shardcat -region region1

gdsctl> start gsm -gsm sharddirector01

gdsctl> add credential -credential oraclecred -osaccount oracle -ospassword
oracle

gdsctl> set gsm -gsm sharddirector01

gdsctl> connect mysdbadmin/oracle

gdsctl> add invitednode shard1

gdsctl> create shard -shardgroup shgrp1 -destination shard1 -credential
oraclecred -sys_password oracle

...
```

PRZYGOTOWANIE SCHEMATÓW DANYCH

- Zbiór przestrzeni tabel (*tablespace set*)
  - logiczny rozproszony kontener przechowujący rozproszone tabele w cząstkowych bazach danych
  - tworzony z poziomu bazy danych Shard Catalog
  - polecenie `create tablespace set`
- Tabele dzielone (*sharded*) i duplikowane (*duplicated*) są definiowane z poziomu bazy danych Shard Catalog
  - polecenie `alter session enable shard ddl`
  - polecenie `create sharded table`
  - polecenie `create duplicated table`

SYSTEM-MANAGED SHARDING

Tworzenie tabeli dzielonej

```
create tablespace set ts1;

alter session enable shard ddl;

create sharded table customers
(cust_id number(8) primary key,
 name varchar2(50),
 location varchar2(20),
)
partition by consistent hash(cust_id)
partitions auto tablespace set ts1;
```

- Zbiór przestrzeni tabel rozproszonych pomiędzy bazy cząstkowe
- Tabela rozdzielona pomiędzy bazy cząstkowe
  - rekordy rozpraszane według wartości funkcji haszowej `hash(cust_id)`

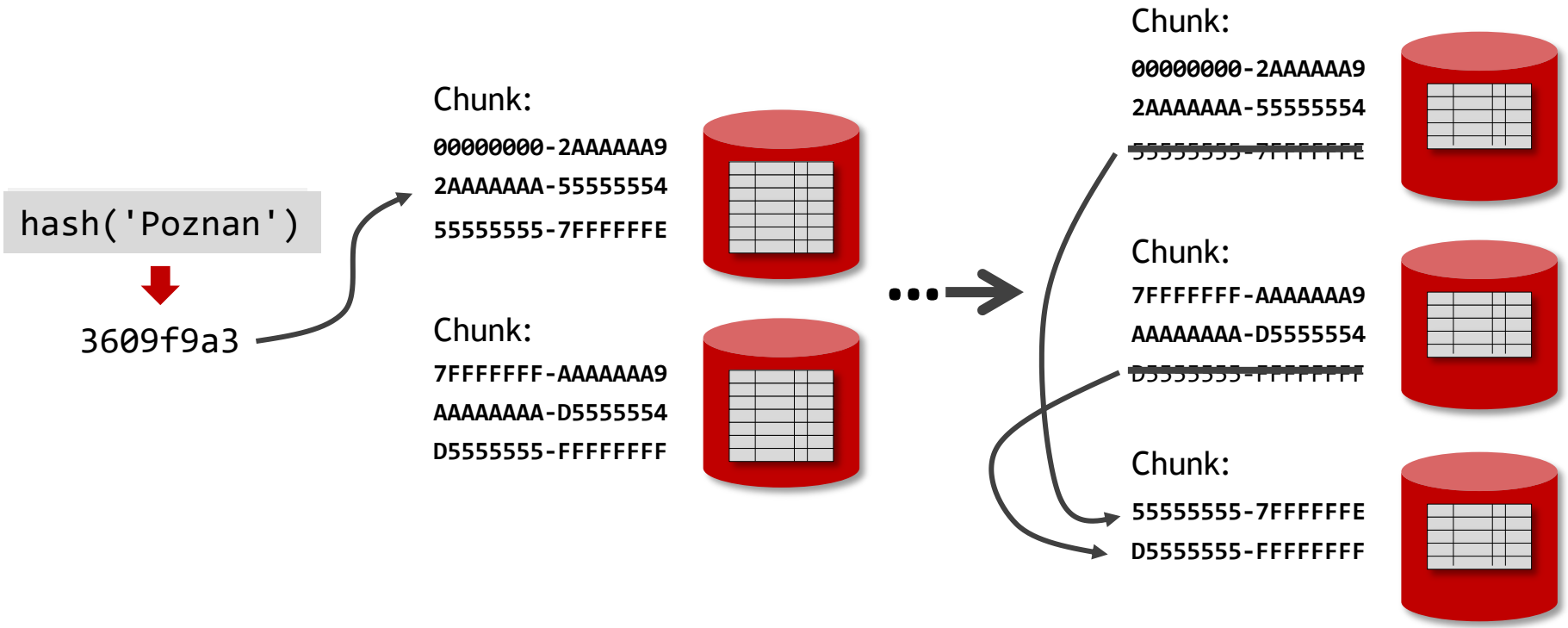
Consistent Hashing

System-Managed Sharding

- W jaki sposób rekordy tabeli są przydzielane do konkretnej cząstkowej bazy danych?
  - klasyczne haszowanie: `hash(klucz) mod N`
    - gdy zmienia się `N`, konieczne remapowanie wszystkich kluczy!
  - haszowanie spójne (*consistent hashing*)
    - gdy zmienia się `N`, remapowaniu podlega tylko `1/N` kluczy
    - przestrzeń wartości funkcji `hash(klucz)` dzielona jest na `N` równych przedziałów (*chunks*)
    - przedziały są przypisywane do cząstkowych baz danych (wiele-do-jednego)
    - gdy zmienia się liczba cząstkowych baz danych, przenoszone są tylko niektóre przedziały, a zawartość przedziałów nie ulega zmianie

Consistent Hashing

Przykład uproszczony - dodanie nowej bazy cząstkowej



RODZINA TABEL DZIELONYCH

Sharded table family

- Gdy z tabelą dzieloną powiązane są inne tabele, ich rekordy powinny być rozpraszane w taki sam sposób
- Rodzina tabel dzielonych (*sharded table family*) obejmuje tabele, których rekordy są rozpraszane (przydzielane do baz cząstkowych) w jednakowy sposób
  - podczas tworzenia tabel należy zastosować klauzulę partycjonowania referencyjnego względem tabeli głównej
  - `partition by reference (fk)`

SYSTEM-MANAGED SHARDING

Tworzenie tabeli duplikowanej

```
create tablespace tbsp1 datafile size 1G;

create duplicated table products
  (
    product_id  number(8) primary key,
    name        varchar2(128),
    price  number(19,4)
  ) tablespace tbsp1;
```

- Tabela tworzona w Shard Catalogu
- W bazach cząstkowych automatycznie utworzone zostaną perspektywy materializowane (*read only*)
  - częstotliwość odświeżania SHRD_DUPL_TABLE_REFRESH_RATE (*domyślnie 60s*)

PLAN WYKONANIA ZAPYTANIA DO TABELI DZIELONEJ

Z poziomu Shard Catalogu

```
SQL> select * from customers where name like '%one%';


-------------------------------------------------------------------
| Id | Operation        | Name | Cost (%CPU)| Inst   |IN-OUT|
-------------------------------------------------------------------
| 0  | SELECT STATEMENT |      |       0 (0)|        |      |
| 1  |  SHARD ITERATOR  |      |            |        |      |
| 2  |   REMOTE         |      |            | ORA_S~| R->S |
-------------------------------------------------------------------
```

PLAN WYKONANIA ZAPYTANIA DO TABELI DUPLIKOWANEJ

Z poziomu bazy cząstkowej

```
SQL> select * from products where product_id>20;


-------------------------------------------------------------------------------
| Id| Operation                                | Name     | Rows | Bytes | Cost|
-------------------------------------------------------------------------------
| 0 | SELECT STATEMENT                         |          | 1    | 112 | 0 (0)|
| 1 |  MAT_VIEW ACCESS BY INDEX ROWID BATCHED| PRODUCTS  | 1    | 112 | 0 (0)|
|* 2|   INDEX RANGE SCAN                       |SYS_C007361| 1    |     | 0 (0)|
-------------------------------------------------------------------------------
```

COMPOSITE SHARDING

Tworzenie tabeli dzielonej

```
create tablespace set ts1 in shardspace americas;
create tablespace set ts2 in shardspace europe;

alter session enable shard ddl;

create sharded table customers
(cust_id number(8) primary key,
 name varchar2(50),
 location varchar2(20),
) partitionset by list(location)
partition by consistent hash(cust_id)
partitions auto
(partitionset americas values ('AMERICA') tablespace set tsp1,
 partitionset europe values ('EUROPE') tablespace set ts2);
```

- Oddzielny zbiór przestrzeni tabel w każdej grupie baz cząstkowych (*shard space*)
- Tabela rozdzielona pomiędzy bazy cząstkowe
  - rekordy rozpraszane pomiędzy grupy baz cząstkowych na podstawie wartości kolumny `location`
  - następnie rozpraszane pomiędzy bazy cząstkowe według wartości funkcji haszowej `hash(cust_id)`

POŁĄCZENIA APLIKACJI Z DZIELONĄ BAZĄ DANYCH

W celu dostępu do tabel dzielonych i duplikowanych

- Bezpośrednie z bazą cząstkową
  - na podstawie wartości klucza podanego przez aplikację klienta – Sharding Key
  - wspierane przez JDBC, OCI, ODP.NET, itp.
- Za pośrednictwem Proxy Routing
  - dla zapytań globalnych
  - wymagają połączenia się z bazą Shard Catalog, z usługą `GDS$CATALOG`

Połączenie bezpośrednie - przykład

Przekierowanie do właściwej bazy cząstkowej na podstawie wartości klucza

```
PoolDataSource pds = PoolDataSourceFactory.getPoolDataSource();

...

OracleShardingKey shardingKey = pds.createShardingKeyBuilder()
                    .subkey(12345, OracleType.NUMBER).build();

Connection conn = pds.createConnectionBuilder()
                    .shardingKey(shardingKey).build(); ...
```

PROXY ROUTING - RESTRYKCJE

Nieobsługiwane operacje

- Nieobsługiwane zapytania połączeniowe do tabel dzielonych, z wyjątkiem:
  - połączeń równościowych w oparciu o klucz partycjonowania
  - użycia dodatkowego filtra w zapytaniu, prowadzącego do tylko jednej cząstkowej bazy danych
- Nieobsługiwane zapytania do tabel dzielonych wykorzystujące `EXISTS` lub `NOT EXISTS`
- Nieobsługiwane funkcje grupowe `AVG, COUNT, SUM, MIN, MAX` w zapytaniach do tabel dzielonych
- Polecenia DML operujące na wielu bazach cząstkowych

DATABASE SHARDING - DOSTĘPNOŚĆ

W różnych modelach licencyjnych

Table 1-2 Feature, Option, and Management Pack Availability by Oracle Database Offering

Feature/Option/Pack	SE2	EE	EE-Exa	DBCS SE	DBCS EE	DBCS EE-HP	DBCS EE-EP	ExaCS	Notes
Oracle Sharding	N	Y	Y	N	Y	Y	Y	Y	**EE** and **EE-Exa**: No limit on the number of either primary shards or standby shards if every shard has an Active Data Guard, Golden Gate, or Oracle RAC license. Without an Active Data Guard, Golden Gate, or Oracle RAC license, use is limited to three primary shards, with basic Data Guard standbys. **DBCS EE** and **DBCS EE-HP**: Use is limited to three primary shards; there is no limit on the number of standby shards. **DBCS EE-EP** and **ExaCS**: No limit on the number of either primary shards or standby shards.

Posumowanie

Oracle 12.2 Database Sharding

- Nowy poziom horyzontalnego partycjonowania (skalowania) bazy danych
- Możliwość łatwej dyslokacji danych w pobliże konsumentów danych
- Istotne ograniczenia dla zapytań globalnych
- Konieczność modyfikacji aplikacji i modeli danych
- W porównaniu z rozwiązaniami NoSQL:
  - relacyjny model danych
  - ACID
  - SQL i PL/SQL
  - kompresja i szyfrowanie danych
  - niezawodność poprzez replikację on-line
  - zaawansowane kopie bezpieczeństwa i odtwarzanie po awarii