

Leonard GRINKE

Akademia Techniczno-Humanistyczna w Bielsku-Białej
Katedra Mechaniki i Inżynierskich Metod Komputerowych

INFRASTRUKTURA SIECIOWA A BAZY DANYCH W SYSTEMIE WIRTUALNYCH PRZEDSIĘBIORSTW PRODUKCYJNYCH

Streszczenie. W referacie omówiono problem doboru topologii sieci komputerowej oraz zastosowania narzędzi i metod do przesyłania danych z systemów produkcyjnych do baz danych. Rozpatrzono dwa przypadki: dużej intensywności napływu małej ilości danych poprzez sieć lokalną oraz replikacji dużej ilości danych poprzez sieć rozległą do zdalnego centrum.

Słowa kluczowe: infrastruktura sieciowa, baza danych, replikacja danych.

NETWORK INFRASTRUCTURE AND DATABASES USED IN VIRTUAL PRODUCTION ENTERPRISES

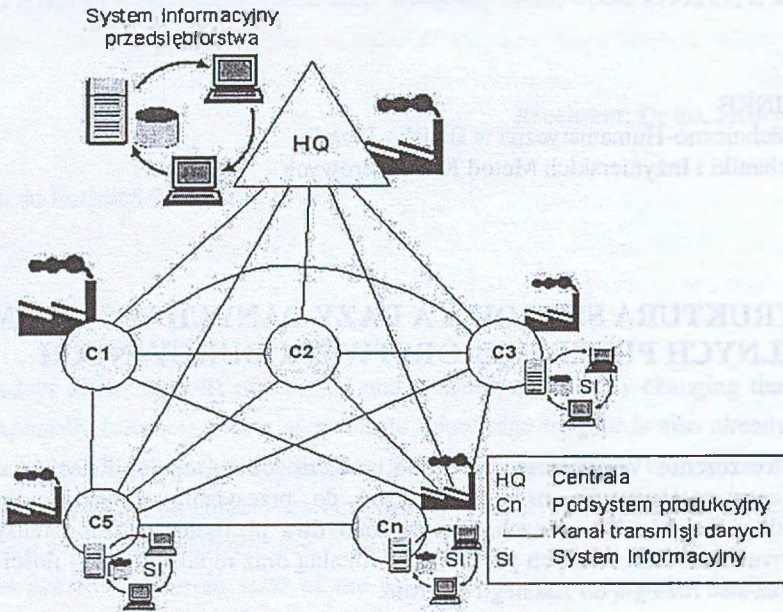
Summary. In the paper the problems of choosing network topology, tools and utilities used to transfers data to heterogeneous databases are presented. Two cases: request of high rate of data insert transaction against database using Local Area Network, and asynchronous replication of high volume data to a remote Central Site are discussed.

Keywords: network infrastructure, database, data replication.

1. Wprowadzenie

Scentralizowany nadzór nad wykonywaną produkcją związany z planowaniem i monitorowaniem jej przebiegu w skali całej korporacji wiąże się z napływem danych od wszystkich jej podsystemów produkcyjnych. Sytuacja taka coraz częściej występuje w warunkach wirtualizacji procesów produkcyjnych zachodzących w sieci kooperujących pomiędzy sobą małych i średnich przedsiębiorstw, zwłaszcza w warunkach jednostkowej i małoseryjnej produkcji oraz świadczenia usług między nimi.

Topologia połączeń pomiędzy firmami tworząca graf jest następująca (rys.1):



Rys. 1. Sieć powiązanych ze sobą podsystemów produkcyjnych (przedsiębiorstwo wirtualne)

Fig. 1. Network of cooperate manufacturing systems (Virtual Enterprise)

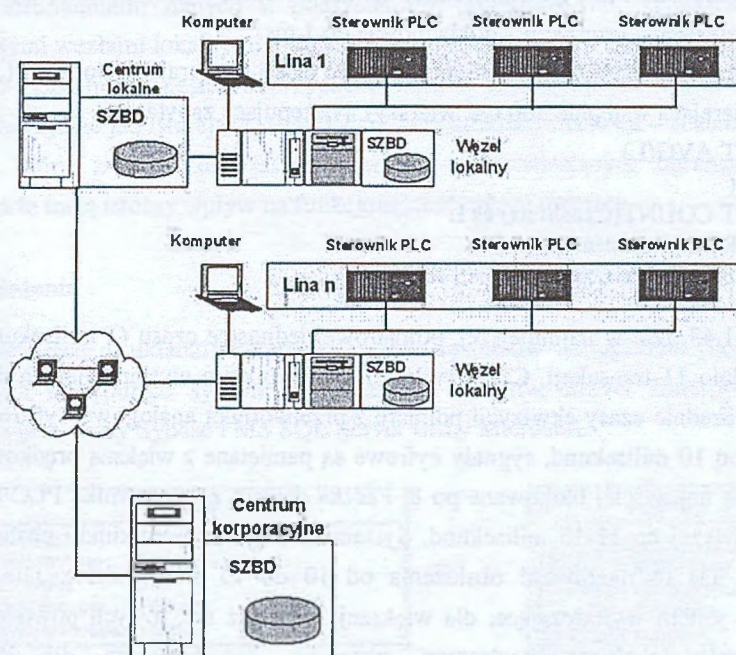
Topologię tę najczęściej upraszcza się do struktury drzewa jako lepiej przystającej do obecnej realizacji wirtualnej organizacji produkcyjnej w istniejącej rzeczywistości gospodarczej. Nie wynika to z przyczyn technologicznych, a raczej organizacyjnych, musi istnieć centrum narzucające standardy, koordynujące prace i najczęściej reprezentujące wirtualną organizację w środowisku zewnętrznym. Reprezentacja ta jest konieczna dla spełnienia wymogów prawno-społeczno-gospodarczych. Struktura przedsiębiorstwa wirtualnego jest zmienna, przedsiębiorstwa wchodzi w jej skład na okres realizacji zamówienia produkcyjnego.

W trakcie gromadzenia i wykorzystania danych mogą występować następujące problemy:

- A. Wydajności bazy danych w przypadku dużego napływu danych z systemów wytwarzania poprzez sieć przemysłową pracującą w czasie rzeczywistym.
- B. Obsługi dużego przepływu danych pomiędzy systemami heterogenicznymi podsystemów produkcyjnych z wykorzystaniem sieci rozległej.

2. Duży napływ danych do bazy z wykorzystaniem sieci lokalnej

Podstawowym problemem jest określenie infrastruktury transportu danych do poszczególnych węzłów układu. Jest on szczególnie interesujący, jeżeli założy się dużą intensywność napływu danych.



Rys. 2. Infrastruktura lokalnego podsystemu produkcyjnego

Fig. 2. Infrastructure of local manufacturing center

Lokalne centra produkcyjne przechowują dane otrzymywane ze stacji pomiarowych znajdujących się na liniach produkcyjnych. Dane transportowane są z wykorzystaniem sieci lokalnej. Najczęściej na tym poziomie można nie traktować przepustowości sieci jako ograniczenia. Sieć lokalna może zostać zbudowana od podstaw, dobór wykorzystanej do budowy sieci technologii jest związany z wymaganiami systemów z niej korzystających. Problemem najczęściej jest obsługa przez SZBD dużych strumieni danych. Z tego powodu w wielu rozwiązaniach do pamiętania danych stosuje się pliki płaskie. Metoda ta ma poważną wadę, którą jest rezygnacja z potężnych mechanizmów manipulacji danymi. Stosowanie plików płaskich wymusza każdorazowo tworzenie dodatkowego kodu i wyklucza stosowanie jakiegokolwiek optymalizacji. Konieczność tworzenia zaawansowanych raportów i zestawień nie tylko z wykorzystaniem narzędzi oferowanych przez SZBD, ale firm trzecich np. MS Excel jest oczywista.

Ograniczenia wydajnościowe istnieją po stronie bazy danych, która może nie obsłużyć wszystkich napływających do niej żądań. Do sprawdzenia wydajności bazy (Adaptive Server Anywhere firmy Sybase, komputer Intel Pentium 4 1.2 MHz, 0,5 GB RAM, karta sieciowa 100Mb/s i system operacyjny Windows 2000 Server) użyto ośmiu aplikacji generujących dane. W celu określenia maksymalnej wydajności bazy danych każdy z wierszy danych posiada kolumnę CzasStart typu timestamp (znacznik czasowy). Wartością tej kolumny są data i czas rozpoczęcia transakcji z dokładnością do 1 ms.

Po zakończeniu pomiaru wyświetlono zawartość tabeli zawierającej pomiary („Rejestracja działań” zawierająca wstępnie 300 tys. wierszy). Następujące zapytanie:

```
SELECT AVG(L)
FROM (
SELECT COUNT(CzasStart) as L
FROM RejestracjaDzialan
GROUP BY CzasStart ) AS K; (1)
```

dało wynik 11.43, tzn. w najmniejszej, pomiarowej jednostce czasu (1 milisekunda) system wykonał średnio 11 transakcji. Czy wynik ten jest wynikiem wystarczającym do realizacji wymagania? Średnie czasy akwizycji pomiaru z przetwornika analogowo-cyfrowego rzadko są mniejsze od 10 milisekund, sygnały cyfrowe są pamiętane z większą prędkością, ale ich przysłanie jest najczęściej blokowane po 8. Paczka danych ze sterownika PLC będzie więc wysyłana najwyżej co 12-15 milisekund. System w ciągu 1 milisekundy obsłużył średnio 10 transakcji. Da to możliwość obsłużenia od 10 do 15 sterowników. Dla jednej linii produkcyjnej jest to wystarczające, dla większej ilości już nie. Z tych powodów przyjęto istnienie węzła lokalnego z własnym systemem bazodanowym dla każdej linii produkcyjnej (rys. 2). Wynik ten osiągnięto rezygnując z następujących mechanizmów wymuszających spójność danych w bazie: sprawdzenia wartości wprowadzanych danych do bazy, związków referencyjnych, wyzwalaczy, stosowania zatwierdzeń (COMMIT) po każdej transakcji.

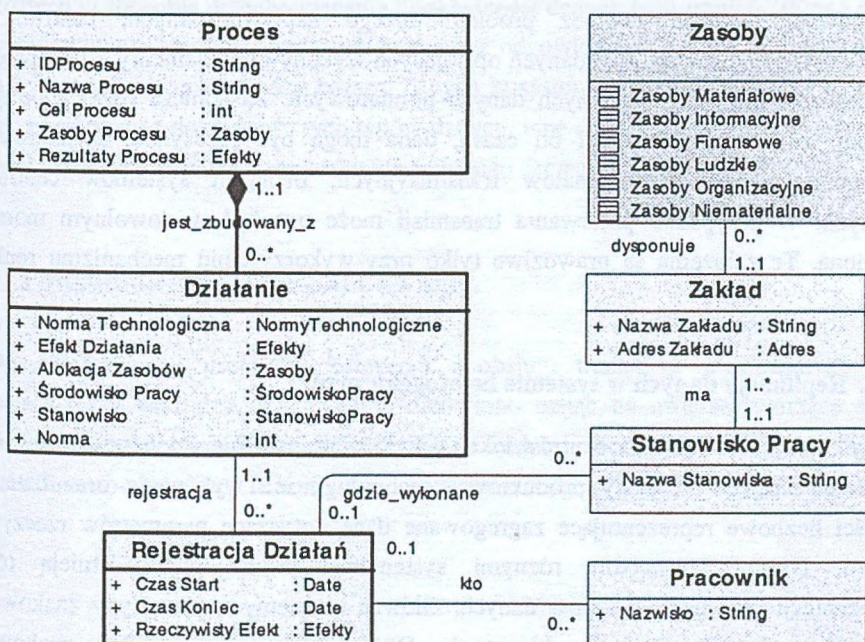
Czy można zrezygnować z tych bardzo istotnych mechanizmów wymuszenia poprawności danych? Systemy bazodanowe zastosowane do przechowywania danych z linii produkcyjnych najczęściej dane otrzymują od innych komputerów (sterowników programowalnych). Sprawdzony i przetestowany mechanizm wprowadzania danych, brak operacji wycofywania transakcji itp. pozwalają na rezygnację z wielu czasochłonnych operacji wykonywanych przez SZBD. W takich warunkach system może pracować w trybie zbliżonym do pracy w czasie rzeczywistym. Baza danych nie musi więc zawierać mechanizmów sprawdzających złożone warunki biznesowe.

3. Replikacja danych z węzłów lokalnych do węzła centralnego ze szczególnym uwzględnieniem heterogeniczności systemu

W Centrali (HQ) powinny znaleźć się dane ze wszystkich węzłów i centrów lokalnych. Rozpatrując rozwiązanie realizowane w najprostszy sposób, Centrum zostanie zalane potężnymi strumieniami danych z podsystemów produkcyjnych. Odległości pomiędzy poszczególnymi węzłami lokalnymi a Centrum są najczęściej duże. Budowa infrastruktury od podstaw, co może mieć miejsce w przypadku węzłów lokalnych, jest całkowicie nierealna i należy wykorzystać już istniejącą. W układzie infrastruktura sieciowa – baza danych należy rozpatrzyć, który z elementów jest elementem wprowadzającym ograniczenia i czy ograniczenia te mają istotny wpływ na funkcjonalność całego systemu.

3.1. Założenia

Założono pracę w układzie heterogenicznych Systemów Zarządzania Bazami Danych. Wykorzystano następujące systemy bazodanowe: Adaptive Server Enterprise, Adaptive Server Anywhere firmy Sybase i MS SQL Server firmy Microsoft.



Rys. 3. Diagram klas pomiaru działań podsystemów produkcyjnych
 Fig. 3. Class diagram of local manufacturing

Na rysunku 3 podano diagram klas dla rozpatrywanych systemów bazodanowych. Pokazuje on tylko podstawowe atrybuty klas i istniejące pomiędzy nimi asocjacje. Diagram

nie zawiera dokładnego opisu klas, a jego celem jest pokazanie złożoności i wielości struktur danych. Klasa „Działanie” zawiera opis takich złożonych danych, jak: zasoby, środowisko pracy i inne. Diagram ten może zostać zastosowany dla każdego z istniejących Relacyjnych Systemów Zarządzania Bazą Danych. Zastosowanie diagramu klas zamiast diagramu związków encji (DZE) zostało podyktowane lepszą adekwatnością modelu obiektowego i łatwością wyboru technologii docelowej. Szczegółowy opis schematu bazy danych wykracza poza ramy niniejszego opracowania.

Dane pomiarowe, które podlegają rejestracji z linii produkcyjnej:

- czas rozpoczęcia działania,
- czas zakończenia działania,
- efekty,
- działanie,
- pracownik,
- stanowisko.

Lokalne podsystemy produkcyjne przesyłają dane do centrum. Problemy spotykane na tym poziomie to głównie przepustowość i jakość połączeń z wykorzystaniem sieci rozległych (WAN) oraz heterogeniczność zastosowanych platform sprzętowych i systemów bazodanowych. Istnieje również problem dużego napływu danych, centrum żąda szczegółowych (elementarnych) danych opisujących wykonywane elementy, opisu procesów ich wykonania oraz zagregowanych danych pomiarowych. Zasadnicza różnica polega na rezygnacji ze ścisłej zależności od czasu, dane mogą być przesyłane asynchronicznie w zależności od obciążeń kanałów transmisyjnych, obciążeń systemów centralnego i lokalnych. W przypadku przerwania transmisji może ona być w dowolnym momencie wznowiona. Te założenia są prawdziwe tylko przy wykorzystaniu mechanizmu replikacji danych.

3.2. Replikacja danych w systemie heterogenicznym

Przy kopiowaniu danych z centrów lokalnych do centrum korporacyjnego podstawowymi elementami danych są: opisy produktów i technologii ich wykonania oraz data, czas i wartości liczbowe reprezentujące zagregowane dane dotyczące parametrów rzeczywistej realizacji. Niestety pomiędzy różnymi systemami bazodanowymi istnieją różnice w reprezentacji wewnętrznej typów danych. Główne problemy wnoszą typy znakowe, dat i czasu oraz dużych obiektów binarnych. Duże obiekty binarne lub znakowe są wykorzystywane do pamiętania danych o produkcie i technologii jego wykonania (pliki CAD/CAM, STEP itp.). Problem użycia różnych stron kodowych do reprezentacji znaków łącznie ze znakami narodowymi w tym przypadku można pominąć.

Dla przykładu omówiono problem wewnętrznej reprezentacji daty i czasu. Dla wymienionych SZBD reprezentacja wymienionych typów przedstawia się następująco:

- **Microsoft SQL Server** posiada dwa typy danych do przechowywania jednocześnie daty i czasu: *datetime* i *smalldatetime*. Cechy typu *smalldatetime*: dokładność do minuty, wartości od 1 stycznia 1900 do 6 czerwca 2079, cechy typu *datetime*: zakres wartości od 1 stycznia 1753 do 31 grudnia 9999, z zaokrągleniem do 0.00, 0.03 lub 0.07 sekundy.
- **Adaptive Server Anywhere (ASA)** posiada dwa typy danych przechowujące czas i datę: *datetime* i *smalldatetime*. Obydwa typy mają tę samą dokładność i zakres przyjmowanych wartości. Obydwa typy danych mogą przyjmować wartości od 1600-02-28 23:59:59 do 7911-01-01 00:00:00. Dane są przechowywane z dokładnością do milisekundy.
- **Adaptive Server Enterprise (ASE)** również dysponuje typami danych *datetime* i *smalldatetime*. W kolumnach typu *smalldatetime* można przechowywać daty z zakresu od 1990-01-01 do 2079-06-06, w kolumnach typu *datetime* wartości z zakresu od 1753-01-01 do 9999-12-31 z dokładnością do setnej części sekundy.

Różnice w sposobie przechowywania i dokładności danych typu *smalldatetime* i *datetime* są powodem wielu błędów. Dane tych typów bez odpowiednich konwersji i zaokrągleń nie mogą być wymieniane pomiędzy bazami danych w systemie heterogenicznym. Każda baza danych posiada inną dokładność tych typów danych, inne dopuszczalne zakresy poprawnych wartości, choć typy te służą do przechowywania tego samego rodzaju danych.

4. Podsumowanie, wnioski i uwagi

Do automatyzacji procesów konwersji i obsługi transmisji wykorzystano podane w tabelach 1 i 2 narzędzia. Zagadnienie omówiono mając na uwadze tworzące się sieci wirtualnych przedsiębiorstw, najczęściej małej i średniej wielkości. Z racji ograniczonych środków finansowych firmy te korzystają z rozwiązań technicznych podanych w tabeli 1.

Tabela 1

Podstawowe narzędzia replikacji danych

Narzędzie	Uniwersalność, prostota konfiguracji, niezawodność i wydajność
SQL Remote	<p>Obsługuje replikację tylko pomiędzy bazami danych ASE i ASA. SQL Remote jest łatwy w konfiguracji, nie obciąża baz danych, nie korzysta z wyzwalaczy ani reguł, może pracować przy okresowych połączeniach z siecią. Cechuje go łatwość administrowania.</p> <p>Wymaga jednak bardzo dokładnego określenia zakresu danych, które będą replikowane i planowania na poziomie obsługi procesu replikacji. Wykorzystanie współdzielonych plików jako mechanizmu przenoszenia danych wymaga udostępnienia odpowiednich zasobów dyskowych na zdalnych i lokalnych komputerach połączonych sieciami typu LAN lub WAN. Usunięcie udostępnienia przez niepowołanego użytkownika lub nawet administratora uniemożliwia dalszą replikację.</p>
Mechanizmy replikacji w MS SQL Server 7.0	<p>Microsoft SQL Server pozwala na replikację do źródeł heterogenicznych, które dostarczają trzydziestodwubitowych sterowników ODBC lub OLE DB. Replikacja jest prosta do konfiguracji, można używać kreatorów, konfigurować replikację z poziomu Enterprise Manager lub poprzez wykonanie odpowiednich skryptów.</p>
DTS	<p>Obsługuje ogromną ilość źródeł i odbiorców danych, jest prosty w konfiguracji dzięki dostępnym narzędziom graficznym i kreatorom pakietów.</p> <p>Przenosząc dane pomiędzy różnymi systemami baz danych, DTS nie zawsze wybiera w bazie docelowej odpowiedni typ danych. Często jest wybierany typ o innym zakresie lub precyzji przechowywanych wartości. DTS nie uwzględnia ograniczeń zdefiniowanych w bazie danych.</p>
UltraLite i MobiLink	<p>Aplikacje UltraLite mogą być pisane dla pager'ów, palmtopów i innych urządzeń przenośnych. Aplikacja MobiLink współpracuje ze źródłami danych ODBC, np. z Sybase Adaptive Server Anywhere, Sybase Adaptive Server Enterprise, Oracle i Microsoft SQL Server.</p> <p>UltraLite posiada mechanizmy zabezpieczające przed awariami poszczególnych komponentów systemu synchronizacji danych. Błędy mogą być obsługiwane automatycznie lub z wykorzystaniem stworzonych skryptów. Na podstawie ustalonych priorytetów potrafi rozwiązywać konflikty występujące w czasie modyfikacji danych. Budowa aplikacji UltraLite wymaga poznania odpowiednich bibliotek programistycznych, a konfiguracja synchronizacji MobiLink wymaga napisania skryptów synchronizacyjnych, co nie jest proste.</p>

W przypadku przedsiębiorstw globalnych (dużych firm) w przepływie informacji stosowane są inne narzędzia. Problemy finansowe schodzą na dalszy plan (tabela 2).

Tabela 2

Zaawansowane narzędzia replikacji danych

Narzędzie	Uniwersalność, prostota konfiguracji, niezawodność i wydajność
Replication Server	Obsługuje wiele różnych SZBD, między innymi bazy Sybase i Oracle, wydajny, pozwala na rozłożenie obciążenia w sieci, trudny do konfiguracji, wymaga specjalistycznych szkoleń, zwykle każdy Replication Server wymaga obsługi administratora.
CIS i OmniConnect	Component Integration Services w połączeniu z OmniCONNECT jest używany w celu realizacji równoczesnego dostępu do wielu źródeł danych lub przesyłania danych z jednego do drugiego serwera. Pozwala na tworzenie związków pomiędzy tabelami w systemie heterogenicznych baz danych oraz migrację dużej liczby obiektów różnego rodzaju.
Enterprise CONNECT	Pozwala klientom Sybase na dostęp do źródeł danych takich jak: DB2 i DL/1, kolejki TD i TS, pliki VSAM i inne SZBD; jest to oprogramowanie wspomagające pracę stacji roboczych z komputerami typu mainframe.
Adaptive Server Distributed Transaction Management	Integralność transakcji jest zachowywana tylko wtedy, gdy wszystkie serwery, na których jest wykonywana, obsługują Transaction Coordination Services lub następuje współpraca z zewnętrznym menadżerem transakcji.
Microsoft Distributed Transaction Coordinator	Obsługuje interfejs XA, przez co może współpracować z bazami danych takimi jak Oracle czy DB2; wspomaga wykonywania transakcji obejmujących wiele menadżerów transakcji komunikujących się przy użyciu Transaction Internet Protocol; pozwala na wykorzystanie COMTI. Używa 2PC do wykonywania rozproszonych transakcji. Zawiera wbudowane mechanizmy zabezpieczające przed awarią.

LITERATURA

1. Matuszek J.: Inżynieria kosztów przedsiębiorstwa wirtualnego. Inżynieria Maszyn, Wrocław 2002, Vol 7, zeszyt 2, 2002.
2. Grudzewski W., Hejduk I.: Przedsiębiorstwo wirtualne, Centrum Doradztwa i Informacji Difin sp. z o.o., Warszawa 2002.
3. Eriksson H., Penker M.: Business Modeling with UML Business Patterns ant Work, Wiley Computer Publishing, New York 2000.
4. www.zifa.com: Web Site of The Zachman Institute for Framework Advancement, luty 2003.

Recenzent: Prof. dr hab. inż. Stanisław Kozielski

Wpłynęło do Redakcji 2 kwietnia 2003 r.

Abstract

The development of information technology powered by Internet's concepts and ideas will lead to the development of new forms of economic undertakings. New economic actors join the market introducing a concept of Virtual Enterprise. Virtual Enterprise is a set of loosely coupled firms, which associate their strengths to provide a specific service traditionally provided by a single enterprise. Network structure of Virtual Enterprise is shown in Fig.1. All members of Virtual Enterprise are connected using a Wide Area Network. Each of them has own application systems, different databases, procedures and culture. The ability of interacting raises several problems that must be considered. In this paper the problems of choosing network topology, tools and utilities used to transfers data to heterogeneous databases are presented. In general there are two cases:

- Collecting data of activity (see Fig.2 and Fig. 3) from manufacturing lines in database. Database can be a bottleneck if there is many insert data transactions. One of the solution is to remove all data constrains and set up a local data center.
- Transferring large amount of data from one manufacturing system to another. Interacting systems are heterogeneous in nature. Class diagram (Fig.2) shows the complexity of replicated data. In short, tools for data replications are described in Table 1 and Table 2.

Adres

Leonard GRINKE: Akademia Techniczno-Humanistyczna, Katedra Mechaniki i Inżynierskich Metod Komputerowych, ul. Willowa 2, 43-309 Bielsko-Biała, Polska, lgrinke@ath.bielsko.pl.