Journal of Platform Technology

ISSN 2289-019X (Online)
http://jpt.ictps.org

Preface
/ R. Young-Chul Kim, Bongen Gu, Dae-Young Kim, Baek-Ki Kim ---------- 1

Design practices of an Integrated Architecture for Knowledge Infrastructure in Science and Technology based on FDP
/ Won Young Lee, Bo Kyung Park, R.Young Chul Kim ---------- 3

EbXML and Adapter Pattern to design integrated architecture of Science and Technology Knowledge Infrastructure
/ Je Seong Hong, R.Young Chul Kim ---------- 10

Keyword-centric item classification for automatic bookkeeping services
/ Daong Bin Choi, Young B. Park, Insu Jo ---------- 16

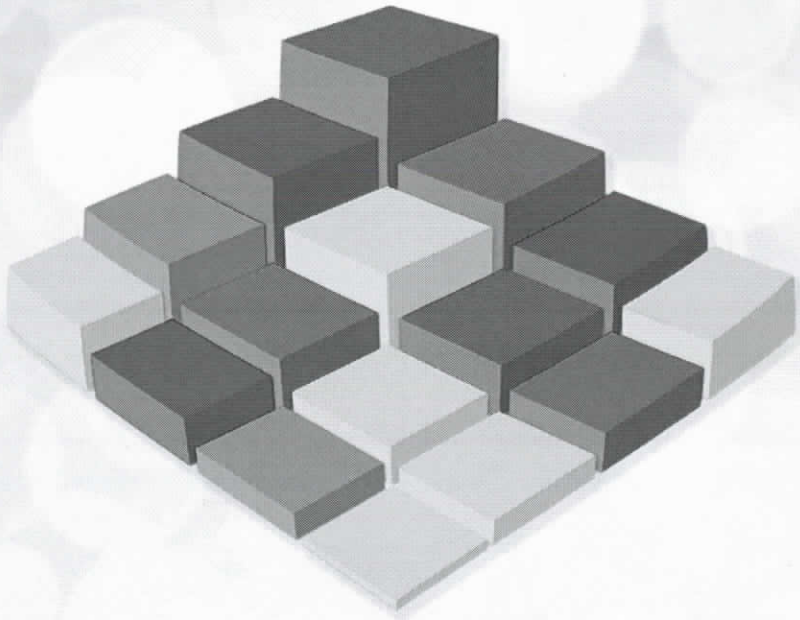
Published by ICT Platform Society

Journal of Platform Technology

Volume 7 Issue 2 June 2019

Journal of Platform Technology

Volume 7, Issue 2
June 2019



ICTPS
ICT Platform Society

FDB를 이용한 과학기술 지식 인프라 통합 아키텍처 설계 연구

[1]이원영, [2]박보경, [*3]김영철
[1, 제1저자]홍익대학교 소프트웨어 공학 연구실, wylee@selab.hongik.ac.kr
[2, 제2저자]홍익대학교 소프트웨어 공학 연구실, park@selab.hongik.ac.kr
[3, 교신저자] 홍익대학교 소프트웨어 공학 연구실, bob@hongik.ac.kr

Design practices of an Integrated Architecture for Knowledge Infrastructure in Science and Technology based on Federal DataBase Paradigm(FDB)

[1]Won Young Lee, [2]Bo Kyung Park, [*3]R. Young Chul Kim
[1, First Author] SELab. Hongik University, leewy@selab.hongik.ac.kr
[2, Second Author] SELab. Hongik University, park@selab.hongik.ac.kr
[*3, Corresponding Author] SELab. Hongik University, bob@hongik.ac.kr

요 약

기존의 과학기술 지식 인프라 서비스는 각 센터별로 분리되어 운영되므로, 각각의 시스템과 서비스의 통합 구축이 필요하다. 이는 각각의 서비스에 대한 사용자의 서비스 접근성을 떨어뜨리는 문제를 야기할 수 있다. 이를 해결하기 위해 이종 시스템 간에 통합을 통한 맞춤형 서비스를 과학기술 지식 인프라 시스템의 통합이 필요하다. 다양한 방법으로 통합을 구축할 수 있지만, 본 연구에서는 과거의 연방 데이터 베이스 Federal DataBase(FDB)방법을 사용하여 상호운용성을 높이는 방안을 제안한다. 이런 제안이 과학기술 지식 인프라 시스템의 서비스 통합 아키텍처 설계에 도움이 될 것으로 기대한다.

Abstract

Conventional service for Knowledge Infrastructure in Science and Technology is operated separately by each center and needs to integrate each heterogeneous services and systems. This can cause problems that reduce service accessibility to users. To solve this problem, we need to integrate user's service for customized service integrated among disparate systems into knowledge infrastructure in science and technology system. In this paper, we propose to enhance interoperability through past Federal DataBase method which create five database schema levels. We expect this proposal to help service architecture design for knowledge infrastructure in science and technology system.

***Keywords**: Federal Database (FDB), Knowledge Infrastructure, Integrated Architecture, Heterogeneous systems, Interoperability*

Ⅰ. 서론

* Corresponding Author
Received: June. 07, 2019, Revised: July. 05, 2019, Accepted: July. 06, 2019

기업에서는 의사결정 지원, 제어, 기업 평가 및 계획을 위해 여러 데이터베이스에 데이터를 액세스하고 공유해야 한다. 이 환경에서는 각기 다른 자율적이고 호환되지 않는 Database Management System(DBMS)를 사용하므로 실시간 정보 연계가 쉽지 않다. 예를 들면 정보 연계의 소스(Source)와 타겟(Target)간 상이한 스키마 구조, 상이한 데이터 생성 규칙, 상이한 데이터베이스 키(Key)체계 등으로 인하여 데이터의 중복, 연계 오류 등의 문제가 발생할 수 있다[1]. 이를 해결하기 위해서는 사용자가 시행착오를 겪지 않는 성공적인 데이터 통합 방안이 제시되어야 한다.

본 논문에서는 Federal DataBase(FDB)방법을 통해 기존 시스템을 구성하는 컴포넌트 및 서비스에 대한 데이터를 통합하여 신규 융합서비스를 효율적으로 개발하기 위한 방법론을 제시한다. 이를 위하여 FDB 의 세가지 특징을 분석하였고, 이를 바탕으로 데이터 통합 오류를 최소화하기 위한 설계 방안을 제시하였다. 본 논문의 구성은 다음과 같다. 2 장 관련연구에서는 FDB 의 구성요소와 특성에 대해 설명한다. 3 장에서는 FDB 를 이용한 데이터 통합 아키텍처 설계 방안을 제시하고 4 장에서는 결론 및 향후 연구를 설명한다.

II. 관련 연구

FDBMS(Federated Database Management System)에서 각 로컬 데이터베이스는 연합(Federation)의 논리적 구성 요소로 간주된다. 이러한 구성요소는 여러 로컬 스키마의 통합을 나타내는 하나 이상의 연합 스키마(Federated Schema)로 묶여 있다. 따라서 연합 스키마는 연합된 각 데이터베이스들이 공유할 수 있는 정보를 나타낸다. 각 데이터베이스들의 제어와 조정을 가능하게 하는 소프트웨어를 FDBMS 이라고 한다. <그림 1>은 연합 데이터베이스 관리 시스템과 구성요소이다.

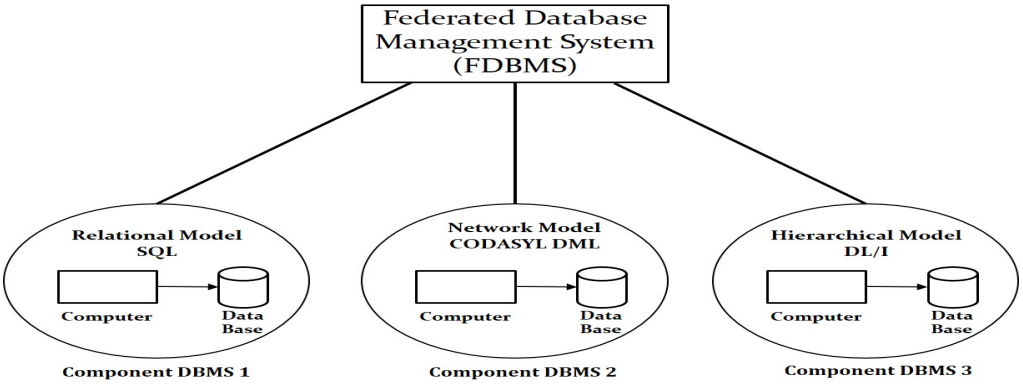


Figure 1 Federated database management system and its components

2.1 자율성(Autonomy)

FDBS 의 각 구성 요소들은 로컬 데이터베이스이다. 로컬 데이터베이스들은 일반적으로 한 시스템에 여러 데이터베이스를 관리하며. 각 로컬 데이터베이스들은 자율성을 갖는다. FDBS 에서는 로컬 데이터베이스가 독립적으로 작동할 수 있는 방법을 정의해야 한다. 각각의 데이터베이스들에 대한 자율성 측면을 이해하고 데이터베이스들이 FDBS 에 연결하기 위한 처리 방법을 이해하는 것이 중요하다. FDBS 에 참여하는 데이터베이스 시스템들은 통신 자율성, 실행 자율성, 연계 자율성으로 총 세가지 유형으로 나타낼 수 있다. 통신 자율성(Communication Autonomy)은 DBMS 의 일반적인 작동을 말하며, 다른 DBMS 와 통신하는지 여부를 나타낸다. 두번째는 실행 자율성(Execution Autonomy)은 DBMS 구성요소가 로컬 및 외부 조작에 의해 요청된 조작을 제어할 수 있게 한다. 연계 자율성(Association Autonomy)은 DBS 구성요소가 연합과의 관계를 끊을 수 있는 권한을 부여한다. 즉, FDBS 가 단일 DBS 와 독립적으로 작동할 수 있음을 의미한다.

2.2 이질성(Heterogeneity)

각 로컬 DBMS는 데이터 모델, 의미 및 시스템 작업을 표현하는 다양한 방법이 존재한다. 데이터 모델은 구조 또는 쿼리 언어 등의 고려가 필요하다. 또한 모든 프로그래밍 언어는 구조를 다르게 표현하며, 쿼리 언어는 시스템마다 조금씩 다르다. 따라서 FDBS에는 상호운용성을 고려해야 한다. 데이터베이스의 이질성에서 의미론에 대한 고려도 필요하다. DBMS가 지원하는 쿼리 언어는 다른 DBMS 간의 이질성 문제를 발생시킨다. 예를 들어, 동일한 데이터 모델이나 쿼리 언어가 다른 경우 이질성 문제가 발생할 수 있다. 의미론적 이질성은 데이터의 의미, 해석, 의도된 사용 방법에 대한 의견 차이가 있을 때 발생한다. <표 1>은 스키마와 데이터 수준에서 발생할 수 있는 이질성의 분류이다.

Table 1 The heterogeneity at the schema and data level

classification of heterogeneity	Example
Collision Naming	A database that uses a different name to represent the same concept.
Collision Domain or Collision Data Expression	A database that uses different values to represent the same concept
Precision Collision	A database that uses the same data values in domains of different cardinality for the same data
Collision Metadata	Same concept appears at schema level and instance level
Collision Data	Missing attributes
Collision Schema	Table-to-table conflicts including Collision Naming and Collision Data

2.3 분산성(Distribution)

FDBS 에서의 데이터 분산은 FDBS 가 구축되기 전에 다중 데이터베이스 시스템이 존재하기 때문에 발생한다. 데이터는 단일 컴퓨터 또는 여러 컴퓨터에 저장될 수 있는 여러 데이터베이스에 분산될 수 있다. 또한 이 컴퓨터는 동일한 위치에 있을 수도 있고, 지리적으로 다른 위치에 있을 수 있다. FDBS 는 네트워크를 통한 상호 연결이 가능하므로 단축된 액세스 시간, 향상된 가용성 및 신뢰성을 가진다. 분산 데이터베이스 시스템에서 의도적으로 데이터가 분산될 수 있다. 그러나 FDBS 에서는 데이터가 일반적으로 이미 분산되고 각 로컬 노드가 이미 가지고 있는 데이터에 의해 관리된다.

FDBS 는 협력적이고 자율적인 데이터베이스 관리 시스템이다. 각 DBMS 는 FDBS 에 참여하고 FDBS 는 안전하고 통제된 방식으로 기본 데이터베이스 시스템에 응용 프로그램 액세스 소프트웨어를 제공한다. FDBS 에 대한 중앙 집중식 제어는 없으며, 모든 작업에는 로컬 DBMS 또는 전체 연합이 관련된다. <그림 2>은 여러 유형의 로컬 데이터베이스를 연합해서 함께 그룹화할 수 있는 방법을 나타낸 것이다.

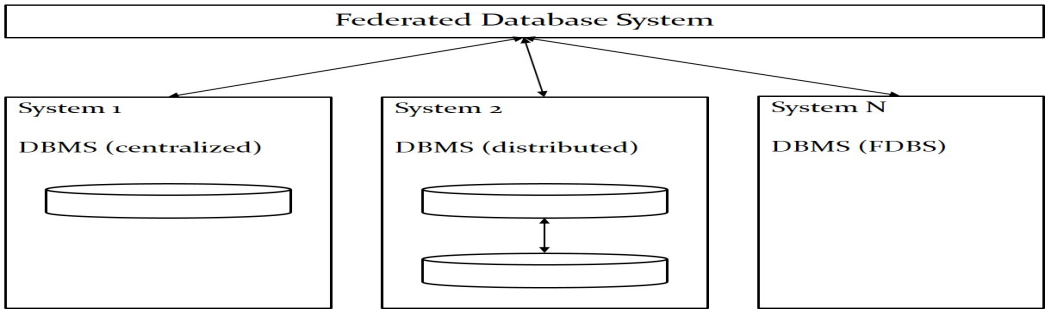


Figure 2 The grouping method for different type local DBMS

III. FDB 를 이용한 데이터 통합 아키텍처 설계 방안

새로운 FDBS 개발은 주로 기존 구성 요소 데이터베이스를 통합하는 것으로 구성된다. 이러한 목적으로 상향식 FDBS 개발 프로세스를 수행 할 수 있다. 이 프로세스는 FDBS 에 새로운 컴포넌트 데이터베이스를 추가할 수도 있다. 기존 FDBS 를 사용하여 새 응용 프로그램을 개발할 경우 응용 프로그램의 데이터 요구 사항이 연합 스키마에서 지원되는지 여부를 확인해야 한다. 일치하지 않으면 연합 스키마를 확장하거나 새로운 연합 스키마를 작성하고 기존 컴포넌트 데이터베이스를 확장하거나 작성해야한다. 이 프로세스를 하향식 FDBS 개발 프로세스라고 한다. 이는 전통적인 분산 데이터베이스 설계 프로세스를 확장한 것이다. 상향식 및 하향식 프로세스의 요소는 FDBS 를 개발하는 데 사용된다. FDB 개발 시, 데이터 사전/디렉토리 (DD/D)는 필수 정보를 저장하여 다양한 활동을 조정하는 데 중요한 역할을 한다. DD/D 는 FDBS 에 의해 관리되는 데이터 정보를 나타내는 모든 스키마를 저장한다. 또한 스키마 간에 매핑, 스키마 및 데이터베이스에 대한 정보, 스키마 독립 정보 및 다양한 유형의 시스템 정보(예: 각 DBMS 의 기능, DBMS 를 호스팅하는 각 시스템의 네트워크 주소 및 지정된 시스템과 통신하는 데 사용되는 통신 기능)등을 저장한다.

3.1 상향식 개발 방법

상향식 FDBS 개발 프로세스는 여러 기존 데이터베이스를 통합하여 FDBS 를 개발하는 데 사용된다. <그림 3>은 상향식 프로세스를 보인 것이다.

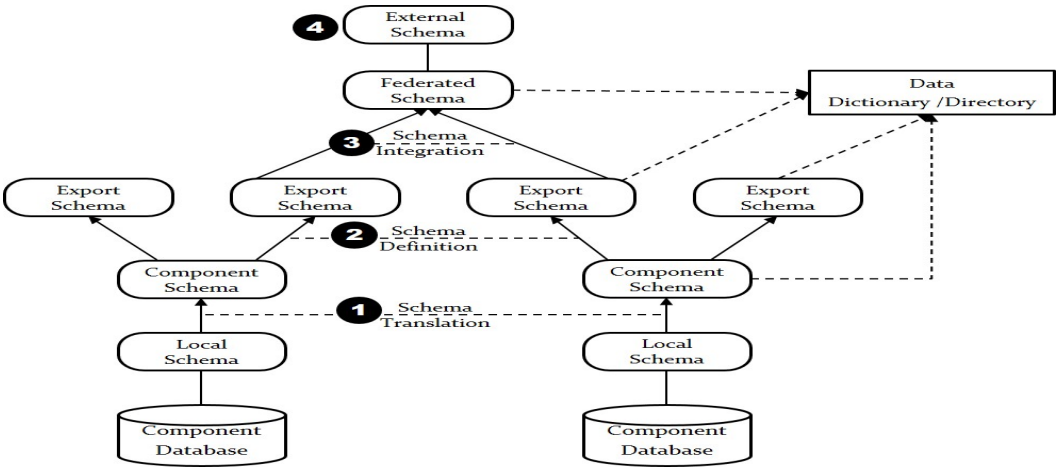


Figure 3 Bottom-up FDBS Developing Process

① Translate schemas: Component 데이터베이스의 개념 스키마인 Local 스키마를 공통 데이터 모델을 참고하여 Component 스키마로 변환한다. 그런 다음 두 스키마 사이의 맵핑 작업을 수행한다. Component 스키마에 표시된 명령을 해당 Local 스키마에 표시된 명령으로 변환 할 수 있는 프로세서(Transforming Processor)를 개발한다.

② Define export schemas: Component 스키마에서 export 스키마를 정의한다. 이 단계는 각 Component 데이터베이스들의 관리자가 연합 DBA 와의 협상을 기반으로 FDBS 에 포함될 데이터베이스의 일부를 인증한다. 적절한 필터링 프로세서(Filtering Processor)를 개발한다.

③ Integrate schemas: 통합과 관련된 export 스키마들을 선택하고 통합한다. 각각의 export 스키마들을 통합하면 하나의 federated 스키마가 생성된다. 그런 다음, Constructing 프로세서를 개발한다. 이 프로세서는 Federated 스키마의 명령을 해당 export 스키마의 명령으로 변환한다. 여기에는 적절한 분산(Distribution) 정보가 포함된 매핑 생성 과정이 포함된다. 이 단계는 관련된 export 스키마들과 해당 federated 스키마에 대해 한 번 반복된다.

④ Define external schemas: 필요한 경우 각 연합 사용자에 대한 external 스키마를 정의한다. 그런 다음 필요한 필터링 및 변환 프로세서를 구축하거나 식별한다. 변환 스키마는 external 스키마의 데이터 모델이 CDM 과 다른 경우 스키마 변환을 수행한다.

3.2 하향식 개발 방법

하향식 DBS 개발 프로세스는 FDBS 가 이미 존재하고 추가 사용자 요구 사항 (예: 새로운 응용 프로그램 지원)이 배치 될 때 사용된다. <그림 4>는 아래에 요약 된 하향식 프로세스이다.

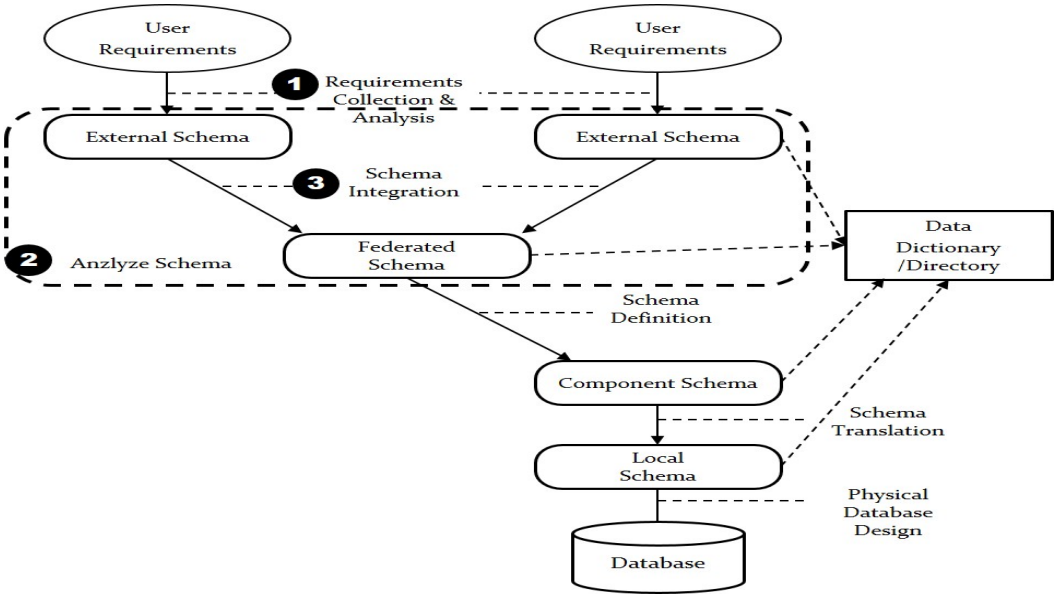


Figure 4 Top-down FDBS developing process

① Define or modify external schemas: 연합 사용자 요구 사항을 수집하고 분석하여 기존 external 스키마에 대한 새로운 external 스키마 또는 확장을 정의한다.

② Analyze schemas: 관련된 federated 스키마를 external 스키마와 비교하여 FDBS 에서 지원되는 external 스키마 부분을 식별한다. external 스키마의 일부가 FDBS 에서 아직 지원되지 않은 경우 federated 스키마를 확장하거나 개발하여 해당 부분을 포함해야한다. 이 부분을 임시 스키마(통합 프로세스가 끝날 때 삭제됨)라고 한다. 하나 이상의 component 데이터베이스가 임시 스키마를 지원해야 한다.

③ Integrate schemas: 임시 스키마를 관련된 연합 스키마와 통합하고 임시 스키마를 삭제한다.

3.3 통합 방법론

5 Step 통합 방법론은 이기종 환경에서 스키마 통합을 위한 방법이다. 5 Step 통합 방법론의 목적은 데이터베이스를 통합하려는 사람들에게 체계적인 접근법을 제공하고, 성공적인 통합을 위한 것이다. 다음은 이기종 환경에서 스키마 통합을 위한 방법이다.

Step 1: Formulation of an Integration Policy

통합 정책은 이기종 환경에서의 스키마 통합이 이루어지기 전에 공식화해야 한다. 이 정책에는 export 스키마와 각 사이트의 맞춤형 통합 전역 뷰를 결정하는 것이 포함된다. 이러한 정책 결정은 각 사이트의 데이터베이스 관리자와 긴밀한 상호 작용을 통해 높은 수준의 조직에서 수행된다.

Step 2: Schema Transformation

통합 정책이 결정되고 각 사이트에 대한 export 스키마가 합의되면 각 로컬 스키마는 중간의 공통 데이터 모델에서 동일한 스키마로 변환된다. 이 스키마는 공통 모델 로컬 스키마라고 한다. 이후 export 스키마는 공통 모델 로컬 스키마의 하위 스키마로 지정된다.

Step 3: Conflict Identification

이 단계에서, 개별 스키마를 분석하고 비교하여 가능한 충돌을 식별한다. 이 단계에서 스키마 간 관계를 식별할 수 있다. 충돌의 종류는 1. 개요의 가. 개발 필요요소 및 고려사항에서 언급하였다.

Step 4: Conflict Resolution

충돌이 확인되면 문제를 해결해야 한다. 이 단계에서 각 스키마의 의미를 명확히 하기 위해 사용자 피드백이 중요하다. 병합할 스키마의 구조적 및 의미적 차이로 인해 스키마 통합에 문제가 발생하곤 한다. 따라서 충돌 식별과 해결은 스키마 통합에 매우 중요하다.

Step 5: 스키마 병합

이 단계에서 개별 사이트의 export 스키마를 전역 스키마로 병합한다.

이 방법론의 개발 프로세스는 <그림 5>과 같이 다중 레벨 아키텍처로 표현된다.

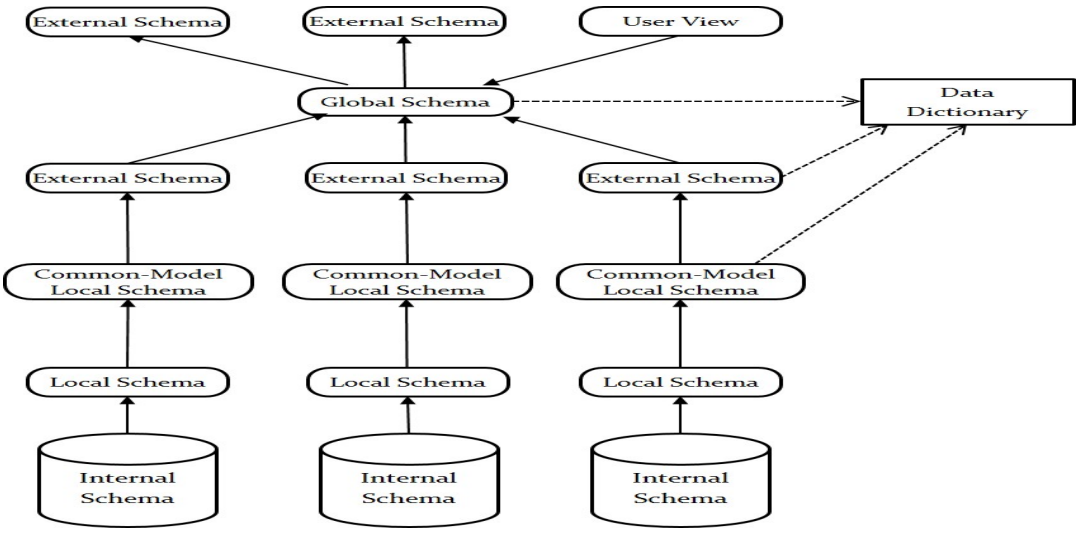


Figure 5 The architecture method for FDBS development

IV. 결론

과학기술 지식인프라의 데이터 처리 및 분석하여 과학기술 지식인프라의 효율적인 데이터 사용을 위한 시스템 개발에 대한 요구가 증가하고 있다. 하지만, 기업에서는 각기 다른 자율적이고 호환되지 않는 DBMS를 사용하므로 실시간 정보 연계가 쉽지 않다. 본 연구에서는 FDB 기반의 데이터 및 서비스 연계 설정을 위한 데이터 통합 아키텍처 설계 방법을 제시하였다. 본 연구를 통하여 이질적인 정보 구조를 가진 정보 간에 실시간 데이터 통합체계 구축과 효율적인 연계체계 운영을 위한 지침이 될 것으로 기대된다.

V. 감사의 글

본 논문은 한국과학기술정보연구원이 지원하는 2018 년도 위탁연구과제(No. P18014)와 2017 년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(NRF-2017R1D1A3B03035421)의 연구수행으로 인한 결과물 임을 밝힙니다.

VI. 참고문헌

[1] B. Park, H. Choi, J. Yang, B. Lee, J. Kim, “Information associated approach between heterogeneous information”, KISTI 2009 Conference Vol. 36.
[2] Heimbigner, Dennis, and Dennis McLeod. "A federated architecture for information management." ACM Transactions on Information Systems (TOIS) 3.3 (1985): 253-278.
[3] Sheth, Amit P., and James A. Larson. "Federated database systems for managing distributed, heterogeneous, and autonomous databases." ACM Computing Surveys (CSUR) 22.3 (1990): 183-236.
[4] Kamel, Magdi N., and Nabil N. Kamel. "Federated database management system: Requirements, issues and solutions." (1992).
[5] W. Kim, “Research directions for integrating heterogeneous databases” Position papers of 1989 Workshop on Heterogeneous Databases, Evanston, IL, USA (December 1989).
[6] Y. Kang, Y. Kim, “Design Methodology of Integrated Architecture for S&T Knowledge Infrastructure”, Journal of Platform Technology 2018 Vol. 6 No. 3 31-37.
[7] Jeffrey Robert Spies. 2013. The Open Science Framework: Improving Science by Making it Open and Accessible. Ph.D. Dissertation. University of Virginia, Charlottesville, VA, USA. Advisor(s) Brian Nosek. AAI3570855.
[8] Northrop L, Feiler P, Gabriel RP, Goodenough J, Linger R, Kazman R, et al. Ultra-large-scale systems-the software challenge of the future. Technical report Software Engineering Institute Carnegie Mellon University ISBN. 2006.