

그래프 데이터베이스의 전력시스템 적용 방안 연구

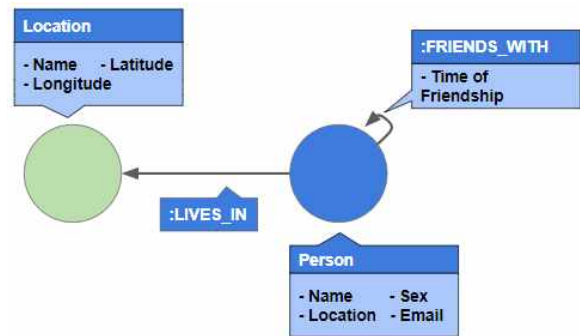
정남준*, 김동욱*, 홍수빈*, 이동혁*
한국전력공사 전력연구원*

A Study on the Application of Graph Database to the Power System

Nam-Joon Jung*, Dong-Wook Kim*, Su-Bin Hong*, Dong-Hyuk Lee*
KEPCO Research Institute*

Abstract - 최근 4차 산업 혁명과 디지털변환의 업무 환경에서 데이터의 폭발적인 증가와 더불어 데이터 모델의 복잡성으로 인하여 데이터 처리시간이 이슈가 되고 있다. 데이터의 저장과 처리를 위하여 자주 활용되었던 기존의 관계형 데이터베이스(RDB)의 기술적 한계를 드러내면서 IDC, 금융, 보험, 소셜 미디어 데이터 처리 등 다양한 분야에서 비정형 데이터에 대한 새로운 분석방안으로 그래프 데이터베이스(GDB) 기술에 관심이 높아지고 있다. 본 논문에서는 관계형 데이터베이스와 그래프 기반 데이터베이스 성능 비교 사례들을 분석해보고 전력시스템에서 그래프 데이터베이스를 활용하여 시스템을 구축한 사례들을 활용하여 한전의 전력시스템에 활용하는 방안을 연구하였다.

자율주행과 같은 다양한 분야에 활용된다.



<그림 1> 소셜 네트워크 표현 그래프 개념

1. 서 론

우리는 일상생활에서 무수히 많은 데이터를 생산하고 있다. 시장 조사 기관인 IDC(International Data Corporation)에 따르면 매년 전세계 데이터의 총량은 2018년에 33ZB(Zettabytes)이고, 2025년에는 175ZB에 이를 것 이라고 한다[1], 데이터의 저장과 처리를 위하여 과거에는 관계형 데이터베이스(RDB)의 기술을 활용하였지만, 최근에는 그래프 데이터베이스를 활용하여 처리 성능을 향상하고 있다. 그래프 데이터베이스는 2015년 파나마 페이퍼즈(Panama Papers) 사건을 계기로 데이터간 연관관계 분석에 특화된 데이터베이스로 자리 매김을 하게 되었다[2]. S Srivastava, L Zhuhadar 등은 기존 RDB로 분석이 어려운 파나마 사건(약 총 2.6 테라바이트)의 방대한 자료를 GDB 분석으로 조세회피 근거를 추적할 수 있음을 연구로 증명했다. GDB와 같은 그래프 마이닝 기술은 RDB 보다 높은 쿼리(Query) 성능과 비정형 데이터의 새로운 데이터 분석 방법을 제공한다[3]. 본 연구는 다방면에 활용되고 있는 그래프 데이터베이스에 대하여 분석하고 전력시스템에서 그래프 데이터베이스를 적용한 사례들을 기반으로 한전의 업무 시스템에 그래프 데이터베이스를 적용하는 방안을 설명하고자 한다.

2.1.2 관계형 데이터베이스의 한계 및 성능 분석 사례

기존 업계에서 사용 비중이 높은 RDB는 정형화된 스키마 구조의 유연성 부족과 비정형 데이터에 대한 성능 저하 등 문제가 존재한다. 테이블(Table) 간에 외부 참조키를 통해 주키 열(Column)에 연결하는 방법은 실제로 많은 계산과 메모리가 필요하다. GDB는 노드(Node)와 관계(Relation)로 연결된 노드의 n depth로 떨어져 있는 다른 노드 검색 성능이 RDB에 비해 훨씬 효율적이다. <표 1>은 기존 RDB와 NoSQL 계열의 GDB의 성능 비교 분석을 나타낸다. 인터넷 기업 라쿠텐에 의하면 MySQL를 NewSQL 계열 데이터베이스로 전환하면 기존 서버 대수를 90%가량 줄일 수 있고, 분산 처리에 샤딩(Sharding)이 필요할 경우 애플리케이션마다 12개월이 걸리는 개발을 한 달 내로 줄일 수 있다고 한다[5].

2. 본 론

2.1 관련 연구

2.1.1 그래프 데이터베이스

그래프 데이터베이스는 정점(Node)과 간선(Edge)의 집합으로 구성된다. 보다 쉽게 표현하면 정점들을 연결하는 간선은 정점들 간의 관계를 표현한다. 그래프는 개체를 정점으로 나타내며 이러한 개체들 간의 관계를 실제 세계와 직접적으로 표현한다. 이러한 범용적 표현 구조는 다양한 시나리오를 모델링 하는데 유용하다. 그래프 데이터베이스의 그래프는 특정 엣지의 유형 또는 전체 그래프 전체를 트래버스 할 수 있다[4]. 그래프 데이터베이스는 맵핑테이블이 필요하지 않은(Schemaless) 특성 때문에 RDB 대비 월등한 연산성능을 보여주며, 비정형데이터를 보다 유연하게 관리되고, 직관적인 데이터구조를 제공한다. 그래프 데이터베이스는 데이터 간의 관계를 만들고 이러한 관계를 신속하게 쿼리해야 하는 소셜 네트워킹, 추천 엔진, 이상 거래 탐지,

<표 1> RDB와 GDB(Neo4j) 성능 비교 사례

Depth	RDBMS execution time(s)	Neo4j execution time(s)	Records returned
2	0.016	0.01	~2500
3	30.267	0.168	~110,000
4	1543.505	1.359	~600,000
5	Unfinished	2.132	~800,000

2.1.3 에너지 섹터의 그래프 데이터베이스 적용 방법론 사례

스웨덴 왕립공과대학교(KTH)는 핀란드에 본사를 둔 에너지 회사 Fortum을 대상으로 그래프 데이터 기반 알고리즘이 전력시스템에 적용한 사례를 분석하는 연구를 하였다[6].

그 연구에 의하면 그래프 이론이 적용된 전력 분야의 유스 케이스는 9개분야에 총 310개의 유스 케이스로 분석되었다.

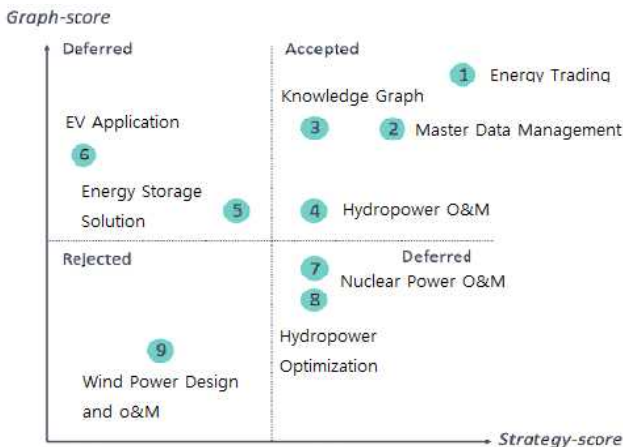
〈표 1〉 Number of past use-cases of Fortum

Use-case cluster	Number of past use-cases
Hydropower: optimisation	13
Hydropower: operation and maintenance	12
Nuclear power: operation and maintenance	47
Wind power: design, operation and maintenance	65
Electric vehicle applications	42
Energy trading	110
Master data management	16
Energy storage	21
Knowledge graphs	20

KTH 연구에서는 Fortum의 전문가들을 활용하여 위에서 분석된 각 클러스터의 유스 케이스들을 Graph Applicability, Technical Feasibility, Economic Potential, Strategic Alignment 관점에서 각 항목에 평가 점수를 부여하여 각 유스케이스가 그래프 이론을 활용한 시스템 개발에 유용했는지에 대한 평가를 진행했다. 각 클러스터의 유스케이스를 점수화한 내용은 그림 2와 같다.

Criteria nr	Criteria	C1	C2	C3	C4	C5	C6	C7	C8	C9
1.	Graph applicability (/8)	3	6	7	7	8	7	5	8	8
1.a.	Underlying graph structure	1	2	2	2	2	2	2	2	2
1.b.	Richness of relationships	1	1	2	2	2	2	2	2	2
1.c.	Identified concepts and algorithms	1	2	1	2	2	2	1	2	2
1.d.	Availability of supporting use-cases	0	1	2	1	2	1	0	2	2
2.	Technical feasibility (/8)	4	3	1	1	7	5	3	7	5
2.a.	Simplicity of model setup	0	1	0	0	1	1	0	2	0
2.b.	Homogeneity of expertise	2	1	1	0	2	2	1	1	1
2.c.	Low computational constraint	1	0	0	0	2	1	2	2	2
2.d.	Risk-free	1	1	0	1	2	1	0	2	2
3.	Economic potential (/8)	3	5	4	4	3	5	4	4	3
3.a.	Sector size	2	2	2	1	0	2	0	1	1
3.b.	Sector growth	0	0	0	2	2	0	1	1	1
3.c.	Substitute tools	0	1	0	0	1	1	2	0	1
3.d.	Scalability	1	2	2	1	0	2	1	2	0
4.	Workability (/8)	5	4	4	1	2	6	6	2	5
4.a.	Relevance	2	2	2	1	0	2	2	1	2
4.b.	Data alignment	1	1	1	0	0	2	1	0	2
4.c.	Human alignment	2	1	1	0	1	1	2	1	1
4.d.	Integrability and maintainability	0	0	0	0	1	1	1	0	0
5.	Strategic alignment (/20)	8	8	8	4	2	12	6	10	8
5.a.	Strategy point 1	2	2	2	0	0	2	1	2	2
5.b.	Strategy point 2	0	0	0	0	0	1	0	0	0
5.c.	Strategy point 3	0	0	0	2	0	0	0	0	0
5.d.	Strategy point 4	0	0	0	0	2	1	2	2	0
	Graph-score (/10)	4,7	5,6	5,0	4,1	6,3	7,2	5,6	6,6	6,6
	Strategic-score (/10)	4	4	4	2	1	6	3	5	4
	Final score (/20)	8,7	9,6	9,0	6,1	7,3	13,2	8,6	11,6	10,6

〈그림 2〉 각 분야별 유스케이스의 평가 점수화



〈그림 3〉 유스케이스별 그래프 및 추진 전략 점수

각 유스케이스들을 점수화하기 위한 점수 항목과 점수의 범위는 유틸리티 회사의 상황에 따라 달라질 수는 있지만, 그래프 이론을 적용하여 업무의 효율성을 높이는 차원에서 본다면 이

는 그 영역의 업무의 특성을 잘 평가하고 점수화했다는 점에서 의미가 크다고 할 수 있다. 가장 높은 점수를 얻은 Energy Trading 분야 그래프 적용 가능성 평가의 경우 날씨나 수요 공급 조건, 다양한 변수 세트를 가지는 특성으로 그래프 유형에 적합할 것으로 평가되었고, 기술적 가능성 측면에서는 중간 정도의 평가를 받았다. 많은 매개 변수들을 고려하는 시간이 많이 걸리는 모델 구축을 고려해야 할 것으로 평가 했다. 경제적인 잠재성 관점에서는 유틸리티 기업들이 경쟁력을 높이기 위하여 내부 도구를 개발해야 한다는 관점에서 5점을 얻었다. 작업성 관점에서는 다양한 데이터 세트를 활용하여 작업이 가능하다는 측면에서 6점 높은 점수를 받았다. 또한 마지막으로 전략적 제휴 측면에서는 전력시장의 자유화 이후 가격 변동성의 증가를 고려할 때 에너지 거래는 모든 참여자의 핵심 활동이 될 것이라는 측면에서 높은 점수를 받았다.

2.2 분산 트랜잭션 가시화 업무 그래프 데이터베이스 적용 분석

2.2.1 분산 트랜잭션 가시화 업무 개요

저자들은 다양한 사내외 사용자들이 사용하는 업무시스템의 트랜잭션을 분석하는 연구를 추진 중에 있다. 그 연구에서 수행하고자 하는 주요 내용은 트랜잭션 장애가 있기 전에 시스템 부하 현황 및 자원의 상태 그리고 어플리케이션의 운영 상태를 인공지능으로 분석하여 시스템 담당자에게 알람을 주고 분석한 결과를 제공하고자 한다. 그렇게 하기 위해서는 복잡한 시스템 운영 환경을 한눈에 볼 수 있도록 가시화하여 제공하는 내용도 중요한 영역 중에 하나이다. 그 문제를 해결하기 위하여 우리는 그래프 데이터베이스를 활용하는 방안을 고려하고자 한다.

2.2.2 KTH 연구 템플릿을 이용한 적합도 검토

앞에서 소개한 스웨덴 왕립공과대학교의 연구 템플릿을 활용한다면 저자들이 연구 중인 업무의 그래프 데이터베이스 기술 응용 적합도에 관한 평가를 쉽게 할 수 있을 것으로 판단되며 기존 연구의 결과가 의미 있어 보인다. 특히, Master Data Management 분야처럼 다양한 정보들의 트랜잭션들이 실시간으로 발생하는 상황이기 때문에 시계열 데이터의 특성을 반영하는 아키텍처 모델의 설계가 가능할 것으로 판단된다.

3. 결 론

본 연구에서는 그래프 데이터베이스의 필요성 및 성능에 대하여 간단하게 분석하고, 에너지 섹터에서 그래프 데이터베이스를 활용한 유스케이스 그리고 그 유스케이스에 대한 평가 방법에 대하여 연구해 보았다. 본 연구를 기반으로 현재 저자들이 연구 중인 업무에 그래프 이론을 적용하고 아키텍처를 설계하는 연구를 향후 진행 예정이다.

〔참 고 문 헌〕

[1] Datanami, "https://www.datanami.com/2018/11/27/global-datasphere-to-hit-175-zettabytes-by-2025-idx-says/", 2018
 [2] S. Srivastava & A. K. Singh, "Graph Based Analysis of Panama Papers", Distributed and Grid Computing(PDGC), IEEE, 822-827, 2018
 [3] 배석민, 김진형, 유재민, 양성열, 정재진, "관계형데이터를 이용한 그래프 데이터베이스의 모델별 구조 분석과 쿼리 성능 비교 연구", Journal of Korea Multimedia Society, Vol. 22, 1036-1046, 2019
 [4] Native Multi-model Can Compete with Pure Document and Graph Databases, "https://www.arangodb.com/2015/06/multi-modelbenchmark", 2019
 [5] Ryutaro Yada, "How Rakuten Reduced Database Management Spending by 90% through Clustrix implementation", Rakuten, tech showcase(Online). 2012
 [6] Royal Institute of Technology, "Graph theory applications in the energy sector", 2020