|  |
| --- |
| ETL 오픈소스 비교자료 |

NICE신용정보 정보기술연구소 2021.01

**1. ETL이란**

- ETL(추출, 변환 및 로드)은 소스 시스템에서 데이터를 동시에 읽고(또는 추출), 쿼리 및 분석을 위해 데이터를 적절한 형식으로 변환(또는 전환)하고, 데이터 웨어하우스, 운영 데이터 저장소 또는 데이터 마트에 로드 하는 데이터 웨어하우징 프로세스

- 일반적으로 별도의 하드웨어에서 호스팅 되고 다른 그룹이나 사용자가 관리하는 여러 애플리케이션 또는 시스템의 데이터를 통합

- 주요 처리내용은 데이터의 임시 하위 집합을 조합하거나, 데이터를 새 데이터베이스로 마이그레이션, 데이터베이스를 새 형식 또는 유형으로 변환하는 데 사용

**2. OpenSource 기반의 ETL 솔루션**

2.1 Pentaho

2.1.1 Pentaho 소개

- Pentaho는 데이터 통합, OLAP 서비스, 보고, 정보 대시보드, 데이터 마이닝 및 ETL(Extract, Transform, Load) 기능을 제공하는 BI(비즈니스 인텔리전스) 소프트웨어

- 데이터의 수집 단계부터 저장, 분석 및 시각화에 이르는 전반적인 빅데이터 작업을 가능하게 하는 데이터 수집 및 분석

- 데이터 수집과 분석을 강력하게 통합함으로써 IT 및 현업 사용자들 모두가 보다 쉽고 빠르게 모든 데이터를 통합/가공 및 분석

2.1.2 주요 기능 및 특징

- 통합 시간이 짧음

- 훌륭한 커뮤니티 지원

- 확장성이 뛰어나고 방대한 데이터 처리 가능

- 다양한 데이터 소스, 시각화

2.1.3 레퍼런스

- JP 모건, 뱅크 오브 아메리카등

2.2 Talend

2.2.1 Talend 소개

- Talend Open Studio는 다양한 ETL 작업을 생성하고 실행할 수 있는 Eclipse 기반 개발자 도구

- 자바 코드가 자동으로 생성되므로 코드를 작성할 필요가 없음

- Talend ETL 툴은 거버넌스와 데이터 통합을 결합하여 안전하고 신뢰할 수 있는 데이터를 쉽게 제공하는 유일한 플랫폼인 Talend Data Fabric으로 구성

2.2.2 기능 및 특징

- 개발 및 구축 시간 단축  
- 비용 효율적이며 우수한 커뮤니티 지원  
- 데이터베이스 통합을 위해 사전 구축된 위젯  
 - 단일 솔루션으로 모든 요구 사항 제공  
 - 통합 프로젝트 지원 제공

2.2.3 레퍼런스

- 존슨 앤 존슨, 바이엘 제약등

**3. 솔루션 별 비교 자료**

3.1 유사성

- 접근이 쉬운 오픈 소스 도구

- Java 코드에 통합 가능

- 포괄적이고 사용자 기반의 IDE 사용

- 우수한 기능 및 커뮤니티 지원 제공

3.2 상세 비교

|  |  |  |
| --- | --- | --- |
| 구분 | Pentaho | Talend |
| 기반 | 상용 오픈 소스 데이터 통합 툴  **\* 일부 기능의 경우 비용발생** | 무료 오픈 소스 데이터 통합 툴 |
| 데이터 품질 | 데이터 품질 조직과의 파트너십, 보안 관련 자제 방화벽 보유 | 클라우드 서비스를 통한 데이터 품질 보장 및 패턴 관리자, 데이터 프로파일러 등의 도구 제공 |
| 데이터 통합 | 데이터베이스에서 애플리케이션으로의 마이그레이션을 포함한 우수한 데이터 통합기능 보유 | 손쉬운 그래픽 개발을 통해 데이터 통합 효율성 향상 |
| 파일 저장 | 파일을 XML 형식으로 저장하고 사용자는 개인 시스템이나 중앙 집중식 데이터베이스에 파일을 저장할 수 있음 | 파일 시스템 수준에서 작동하며, 사용자는 개인 시스템에 파일을 저장할 수 있음 |
| 연결성 | 방대한 데이터베이스에 대한 광범위한 연결 | 동시 데이터베이스로의 연결제한 |
| 성능 | Talend에 비해 속도가 빠름 | Pentaho 대비 속도가 느림 |
| GUI | 현대적이며, 이해하기 쉬움 | 조금 파악하기 어려움 |
| 접근방식 | 메타 기반 다중 스레드 | 단일 스레딩 코드 생성 |
| 반영 | 별도의 서버의 Java 엔진 필요 | Java 및 Perl 파일은 모든 시스템에서 독립적으로 실행 |
| 문서 | 온라인 문서 제공 | PDF 형태의 문서 제공 |
| 플랫폼 지원 | 웹 기반 | 웹 기반 및 아이폰 Apps 지원 |
| 사용고객 | 중소기업 ,대기업 사용 | 중소기업 |

**4. 선정 기준**

- 두 제품의 비교결과, 제품별로 각각의 장단점을 가지고 있음

- 당사의 경우 현재 ETL 솔루션이 없기 때문에 상용기반의 오픈소스인 Pentaho보다는 무료 오픈 소스 기반의 Talend를 선정하여 Docker를 통해 초기 구축, 테스트 및 시범 운영을 진행

- 시범 운영을 통하여 성능 검증 및 요건 확보