

A New Experience in Big Data Analytics

IRIS 지능형 통합 분석 플랫폼

고급 분석 및 다양한 인터페이스 기반으로 사용자 편의 중심의
다양한 분석도구를 제공하는 지능형 통합 분석 플랫폼

제품 설명서



mobigen

Table of Contents

Executive Summary	3
개요	3
Key Features	6
Product Overview	11
System Overview	11
IRIS ETL	12
IRIS 빅데이터 저장소(Big Data Repository).....	14
분석 인프라	18
Contact Information.....	23

Executive Summary

개요

기업들이 운영하고 있는 다양한 종류의 IT 시스템들과 통신 서비스 네트워크, IoT 네트워크 인프라로부터 매일 엄청난 양의 데이터가 발생하고 있습니다. 이들 빅데이터에는 사용자의 트랜잭션, 개별 고객의 서비스 사용 패턴, 장비의 비정상 동작, 보안 위협 등을 파악할 수 있는 중요한 정보들이 포함되어 있습니다. 지난 수 년간 우리 기업들은 폭발적으로 증가하고 있는 빅데이터의 저장과 검색, 분석을 위한 빅데이터 인프라를 구축하기 위하여 많은 투자를 진행하고 있습니다.

특히, 일부 선도 기업들은 장기간 축적된 빅데이터를 기반으로 트렌드를 분석하는 것에 그치지 않고 미래 예측을 시도하는 등 빅데이터로부터 새로운 가치를 창출하기 위한 노력을 계속하고 있습니다.

그러나, 아직도 많은 기업들이 자사의 업무 환경에 맞는 최적의 빅데이터 인프라를 결정하지 못하거나, 기업 내에 흩어져 있는 데이터를 통합적으로 관리하지 못하거나, 빅데이터 인프라를 통해 축적된 데이터를 효율적으로 활용하지 못하는 등 여전히 해결하여야 할 과제들을 안고 있습니다.

많은 기업들이 여전히 빅데이터 활용에 어려움을 겪고 있습니다.

빅데이터 저장소로부터 분석에 필요한 샘플 데이터 추출의 어려움

매일 다양한 형태로 발생하는 수 천만 건(수 TB 분량)의 새로운 데이터가 Data Lake라 불리는 빅데이터 저장소에 저장되는데, 만일 빅데이터 저장소에서 직접 데이터를 분석하려면 성능상의 문제에 직면할 수 있습니다. 따라서, 분석가들은 다양한 검색을 통하여 빅데이터 저장소로부터 분석에 필요한 소량의 데이터를 필터링하고, 이렇게 필터링된 샘플 데이터를 대상으로 의사 결정에 필요한 분석 결과를 도출하게 됩니다. 그런데, 많은 빅데이터 솔루션의 경우, 빅데이터 저장소로부터 데이터를 자유롭게 검색하고 필터링하기 위해서는 SQL, Python, Hive 등 개발 언어에 익숙해 져야 하며, 그렇지 못한 분석가들은 다소 시간이 소요되더라도 개발자에게 요청하여 필요한 데이터를 추출하는 것이 현실입니다. 특히, 개발 언어에 익숙하지 않은 분석가가 데이터 검색과 필터링을 반복할 경우, 개발자에게 필요한 데이터 요청을 반복함으로써 엄청난 시간과 불편함을 감수하여야 합니다.

분석 도구 이용 시 제약

대부분의 분석가들은 자신에게 익숙한 R과 같은 오픈소스 분석 도구 혹은 상용 BI tool을 사용하여 분석 업무를 수행하게 됩니다. 만일, 빅데이터 저장소로부터 어렵게 추출된 샘플

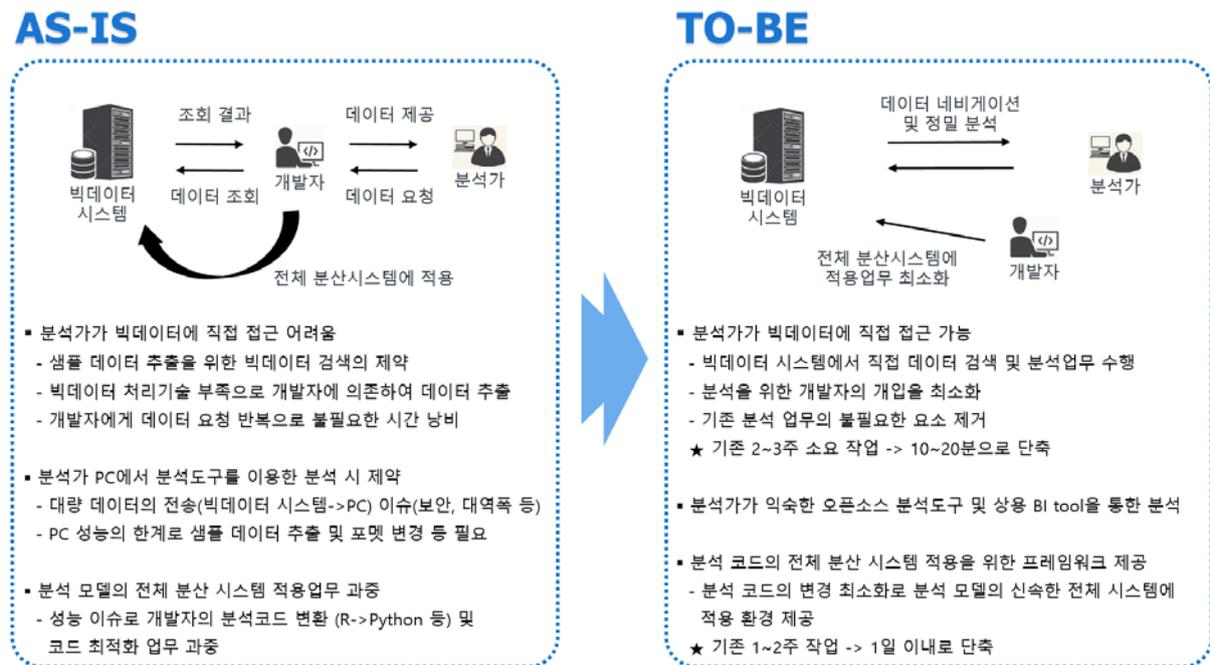
데이터를 분석가가 익숙한 분석 도구에서 사용하기 어렵거나, 기존의 분석 도구가 아닌 새로운 분석 도구를 사용해야 한다면 쉽고 빠른 분석 업무 수행에 제약이 많아 질 수 밖에 없습니다.

모비젠의 IRIS는

지능형 통합 분석 플랫폼으로, IT 시스템이나 IoT 센서로부터 정형 데이터를 비롯하여, 로그 데이터, 운용 데이터, 머신 데이터 등 다양한 정보를 수집, 처리하여 빅데이터 저장소에 통합 저장하며 저장된 데이터를 다양한 방법으로 실시간 검색할 수 있는 기능을 제공합니다. 또한, IRIS에서 제공하는 분석 UI를 포함하여, 오픈 소스 분석도구나 BI 툴을 통하여 데이터를 분석할 수 있는 분석가 관점의 UX 환경을 제공합니다. IRIS를 도입하는 경우, 빅데이터를 보유한 기업은 이상 탐지, 고장 예측 등 고급 분석을 통하여 의사 결정을 위한 인사이트를 빠르게 도출할 수 있는 통합 빅데이터 분석 환경을 경험할 수 있습니다.

IRIS는 빅데이터 활용 시 겪게 되는 근본적인 문제 해결을 지원합니다.

IRIS는 분석가들이 데이터의 실시간 검색, 필터링, 다양한 분석 도구에서 필터링된 데이터 활용 등 분석 업무를 수행하면서 겪게 되는 근본적인 문제점을 해결하기 위한 목적으로 개발되었습니다. 아래 그림은 분석가가 실제 분석 업무 환경에서 직면하는 문제점을 IRIS를 활용하여 극복한 실제 사례를 보여줍니다.



IRIS는 로그 분석 및 이상 탐지 분야에 특화된 솔루션입니다.

IRIS는 보안, 제조, 통신 서비스, 금융 등 산업분야에서의 전통적인 빅데이터 분석은 물론, 다양한 로그 데이터의 실시간 분석, 머신러닝 기반의 이상 탐지 등의 분야를 중심으로 지원 영역을 확장해 가고 있습니다.

대형 고객사이트에서 검증된 솔루션으로 고객의 추가 구매 지속

IRIS는 SK텔레콤, LGU+, 전력연구원, 재난안전연구원, 국방과학연구소 등 30여 일반 기업 및 관공서에 공급되었으며, 수년간 실제 서비스 환경에서 개선되고 검증된 솔루션입니다.

특히, 기존 고객들은 IRIS의 비용대비 성능의 우수성과 사용자 편의성 및 효율성 등으로 인하여, 획기적인 비용 절감 효과를 경험하고 있습니다.

프로젝트	주요기능	처리용량	비고
전력연구원 빅데이터 플랫폼	전력 검침 데이터를 이용한 개인화된 예측 및 이상 탐지	200만호 매15분 1회 검침 (1,920만건/일) 데이터	
SKT 코어망 이상탐지 시스템	코어망 장비로부터 발생하는 알람 데이터를 이용한 이상탐지 시스템		
일본KDDI 액세스망 품질관리 시스템	액세스망 품질관리 시스템의 데이터 고속 검색을 위한 DB	일일 6GB 유입	
SKT 과금통계 시스템 증설	전사 과금 전수 데이터의 실시간 저장, 통계, 검색	1000억 건/1일 9조 건/90일 2 Peta/90일 보관, 검색, 통계	2013년 구축 후 추가 증설. 기존 운영되던 Oracle Exadata 시스템 대체. 데이터 증가에 따른 용량 증설.

Key Features

빅데이터의 실시간 수집처리 (IRIS ETL)

모비젠은 고객의 요구 시, 빅데이터의 실시간 수집 및 전처리 기능을 제공하는 IRIS 전용의 ETL(Extract, Transform and Load)을 옵션으로 제공합니다. IRIS ETL은 다양한 종류의 외부 시스템으로부터 정형, 반정형, 비정형 데이터 소스를 수집/추출 (Extract)합니다. 또한, 수집된 데이터는 사용자가 요구하는 형태로 변환 (Transform)한 후, IRIS 빅데이터 저장소에 고속으로 적재(Load) 됩니다. IRIS ETL은 특히 데이터를 필터링하거나, 특정 테이블의 join, 다양한 함수를 사용한 데이터 변환 등을 쉽고 빠르게 수행할 수 있는 GUI 환경을 제공함으로써 빅데이터 시스템 구축을 위한 기간과 비용을 대폭 단축할 수 있습니다.

Seamless한 분석 업무 수행을 위한 최적의 환경 제공

오늘날 많은 기업들의 빅데이터 시스템은 매일 새롭게 발생하는 테라바이트 이상 규모의 데이터를 처리하여 빅데이터 저장소에 저장합니다. 이렇게 저장된 데이터는 분석가가 기업에서 필요로 하는 다양한 목적의 분석 업무를 수행하기 위하여 활용됩니다. 분석가는 빅데이터 저장소에서 분석에 필요한 데이터를 쉽고 빨리 탐색하고, 자신에 익숙한 분석 도구를 사용하여 분석 결과를 도출하고자 합니다.

그런데, 적지 않은 기업에서, 분석가가 데이터를 분석하는데 여전히 어려움을 겪고 있습니다. 일부 분석가들은 분석 절차를 진행하면서 많은 부분을 수작업으로 수행하여야 하는 문제를 안고 있으며, 또 다른 분석가들은 자신에게 익숙한 분석 도구를 활용하지 못한 채 시스템에서 제공하는 새로운 분석 도구에 익숙해 져야 하는 불편을 호소하고 있습니다. 이러한 문제점들은 분석가의 분석 효율 저하의 원인이 되며, 분석 결과 도출에 장시간이 소요되는 결과를 초래하게 됩니다.

IRIS는 다양한 빅데이터 프로젝트에서의 경험을 바탕으로, 데이터 탐색 단계에서부터 분석 결과 도출 단계, 분석 결과를 전체 분산환경에 적용하는 단계, 그리고 반복되는 분석 업무에 대한 워크플로우를 개발하여 자동화하는 단계에 이르는 분석 업무의 전체 과정을 seamless하게 수행하고, 각 단계에서 추출된 데이터를 공유할 수 있는 환경을 제공합니다.

획기적이고 실질적인 분석시간의 단축을 위한 다양한 편의기능 제공

모비젠은 분석가가 분석에 필요한 데이터를 신속하게 제공하여 실제 분석 업무에 집중할 수 있도록 다양한 편의기능을 지속적으로 제공하고 있습니다. 또한, 모비젠은 분석가들의 일반적인 분석 업무 과정에서 발생하는 불합리한 요소들을 최소화함으로써, 분석가들이 분석 업무에 소요되는 실질적인 시간을 획기적으로 단축할 수 있도록 IRIS를 획기적으로 개선하였습니다.

쉽고 빠른 데이터 브라우저

IRIS는 빅데이터 저장소에 저장된 엄청난 양의 데이터를 분석가가 자유롭게 사용할 수 있도록, Web GUI 환경의 데이터 브라우저를 제공합니다. IRIS의 데이터 브라우저는 IRIS에 저장된 데이터뿐만 아니라, 오픈 소스에서 일반적으로 사용되는 분산 파일 시스템(HDFS)도 접근이 가능한 UI 환경을 제공합니다. 분석가가 데이터 브라우저에서 특정 데이터 테이블에 접속한 후 SQL문을 작성하여 실행하면, 해당 결과가 차트 형태와 테이블 형태로 출력됩니다. 분석가는 출력된 데이터를 줌인/줌아웃할 수 있으며, 상세 분석을 위한 분석 화면으로 바로 이동할 수도 있습니다. 또한, 출력된 데이터는 CSV 혹은 JSON 형태로 export하여 다른 분석 도구에서 사용할 수 있습니다.

원시데이터의 Summary 기능

분석가가 데이터를 분석할 때, 원시 데이터에서 바로 분석할 수도 있지만, 특정 컬럼에 대해서 1분 주기, 5분 주기, 또는 1시간 주기로 summary하여 요약된 데이터를 바로 얻을 수 있다면 분석가 입장에서는 유용한 툴이 될 수 있습니다. 따라서, IRIS는 분석가가 임의로 summary 쿼리문을 작성하여 실행함으로써, summary 결과를 바로 확인할 수 있으며, summary 결과를 저장하여 재사용할 수 있는 기능을 제공합니다.

Full Text 검색

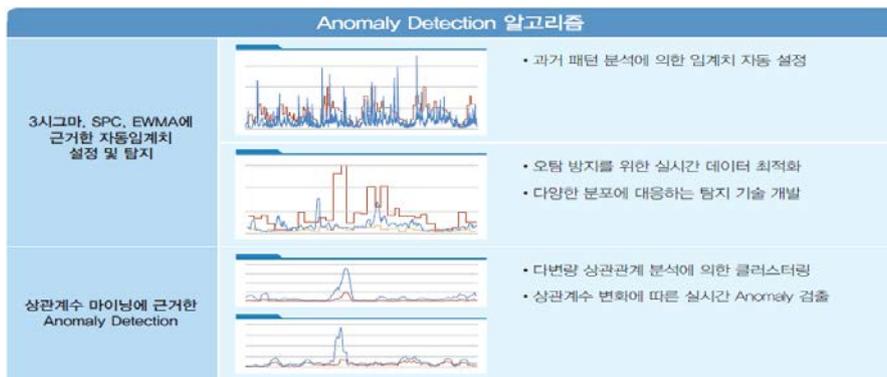
IRIS의 FTS(Full Text Search)를 사용하면 전체 컬럼 데이터에 대하여 고속으로 문자열을 검색할 수 있습니다. 자체 테스트 결과, 사용자는 일반적인 SQL의 Like문 보다 최소 수십 ~ 수 백배 이상 빠르게 텍스트를 검색할 수 있습니다. 또한 사용자는 검색된 결과를 CSV 혹은 JSON 형태로 저장할 수 있습니다.

데이터 분석을 위한 전처리 기능

IRIS는 원시 데이터의 특정 컬럼을 분리하거나, 2개 컬럼 값을 하나로 통합하거나, 컬럼 join 등을 통해 원시 데이터를 가공하고, 가공된 데이터를 저장하여 분석 업무에 사용할 수 있도록 분석 데이터 전처리 기능을 제공합니다.

다양한 고급분석 틀 제공

모비젠은 빅데이터의 수집, 저장, 분석 등 전통적인 빅데이터 시스템의 기능은 물론, 머신러닝 기반의 이상 징후 탐지(Anomaly Detection), 시계열 데이터의 분석을 통한 예측(Time Series Prediction) 등 고급 분석 기술에 대한 지속적인 연구를 진행하고 있습니다. 이를 기반으로, 기업에 통합 분석 환경과 인사이트를 제공하기 위하여 다양한 기술이 적용된 고급 분석 엔진들을 개발하여 IRIS에 탑재하고 있습니다.



사용자가 검색 화면에서 검색된 데이터의 특정 영역을 드래그하고 anomaly detection 명령을 실행하면, IRIS는 고급분석 알고리즘을 반영하여 학습된 모델을 기반으로 anomaly 이벤트를 차트 혹은 표 형태로 보여주게 됩니다. 따라서, 사용자는 단지 클릭 한번으로 anomaly 이벤트를 찾아낼 수 있습니다.

데이터 브라우저를 통한 데이터 검색

IRIS는 빅데이터 저장소에 저장된 각종 데이터로부터 특정 키워드를 검색하거나 통계 처리를 통하여 분석할 수 있는 환경을 제공합니다. 데이터 브라우저는 시스템의 고장 탐지는 물론, 보안 모니터링, 고객 패턴 분석 등 다양한 분야에서 활용될 수 있습니다. IRIS 데이터 브라우저가 제공하는 기본 기능은 다음과 같습니다.

기능	설명
검색	IRIS에 저장된 대량의 데이터 중, 특정 키워드(숫자, 문자, 단어, 문장 등)가 포함된 이벤트를 고속으로 검색하며, 통계 및 차트를 통해 비정상 이벤트를 검색할 수 있는 환경을 제공합니다.
피벗	분석하고자 하는 데이터를 사용자가 임의대로 정렬하고 필터링 함으로써, 보다 빠르고 쉽게 데이터를 분석할 수 있는 피벗 기능을 제공합니다.

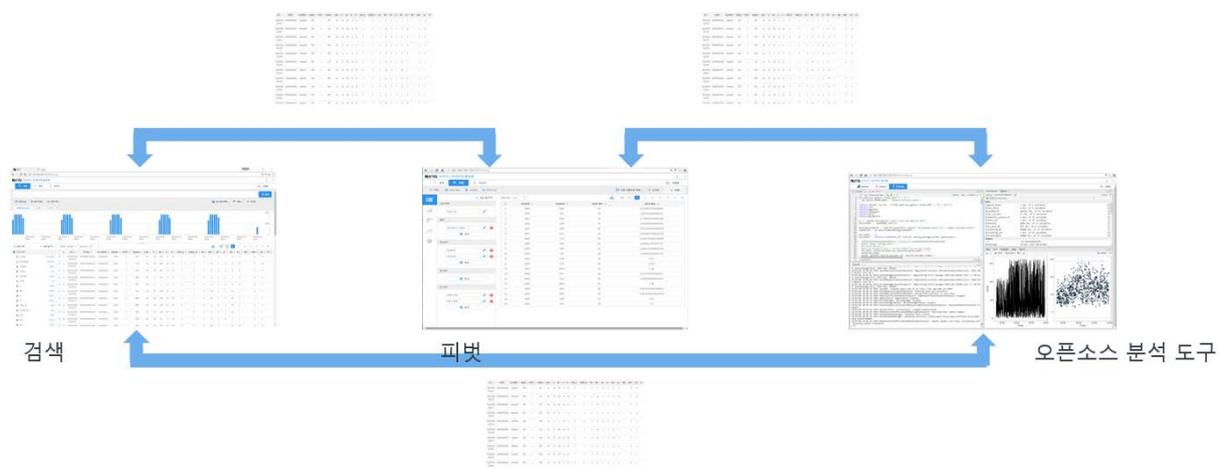
데이터 유형에 따라 시각화를 위한 다양한 차트 지원

IRIS는 분석가가 분석 결과를 쉽게 이해할 수 있도록, 데이터의 시각화를 위한 다양한 차트를 지원합니다. 분석가는 분석하고자 하는 특정 필드를 차트 영역으로 드래그함으로써 쉽고 편리하게 차트를 생성할 수 있습니다. 분석가는 시계열, 분포, 관계형 등 데이터의 유형에 따라 적절한 차트를 사용하여 데이터를 시각화함으로써 분석 결과를 직관적으로 해석할 수 있습니다.

오픈소스 분석도구 및 상용 BI tool 통합

모비젠은 분석가가 각종 오픈소스 분석 도구 또는 상용 BI tool을 이용하여 IRIS에 저장되어 있는 데이터를 기반으로 분석 업무를 수행할 수 있도록, IRIS에 오픈소스 분석 도구와 주요 BI tool들을 통합하였습니다. 따라서, 분석가들은 IRIS의 GUI 화면에서 오픈소스 분석 도구 혹은 BI tool을 바로 사용하여 IRIS에 저장된 빅데이터를 분석할 수 있습니다. 일부 오픈소스 혹은 상용 분석 도구 연동만을 지원하는 타 빅데이터 시스템과 달리, IRIS는 다양한 분석 도구를 IRIS에 통합함으로써, IRIS와 오픈소스 분석 도구 혹은 BI tool 간에 데이터를 자유롭게 공유할 수 있습니다.

또한, 분석가는 외부의 HDFS로부터 얻은 데이터 파일을 오픈 소스 및 상용 BI tool의 분석 화면에서 바로 import하여 분석할 수 있습니다. 분석가는 IRIS의 데이터 브라우저에서 분석을 위한 데이터를 샘플링한 후, 해당 데이터를 Zeppelin, Jupyter, R-Studio 혹은 상용 BI tool 등 자신이 익숙한 오픈소스 분석 화면에서 바로 분석할 수 있으므로, 분석에 소요되는 시간을 대폭 단축할 수 있습니다.



자동화

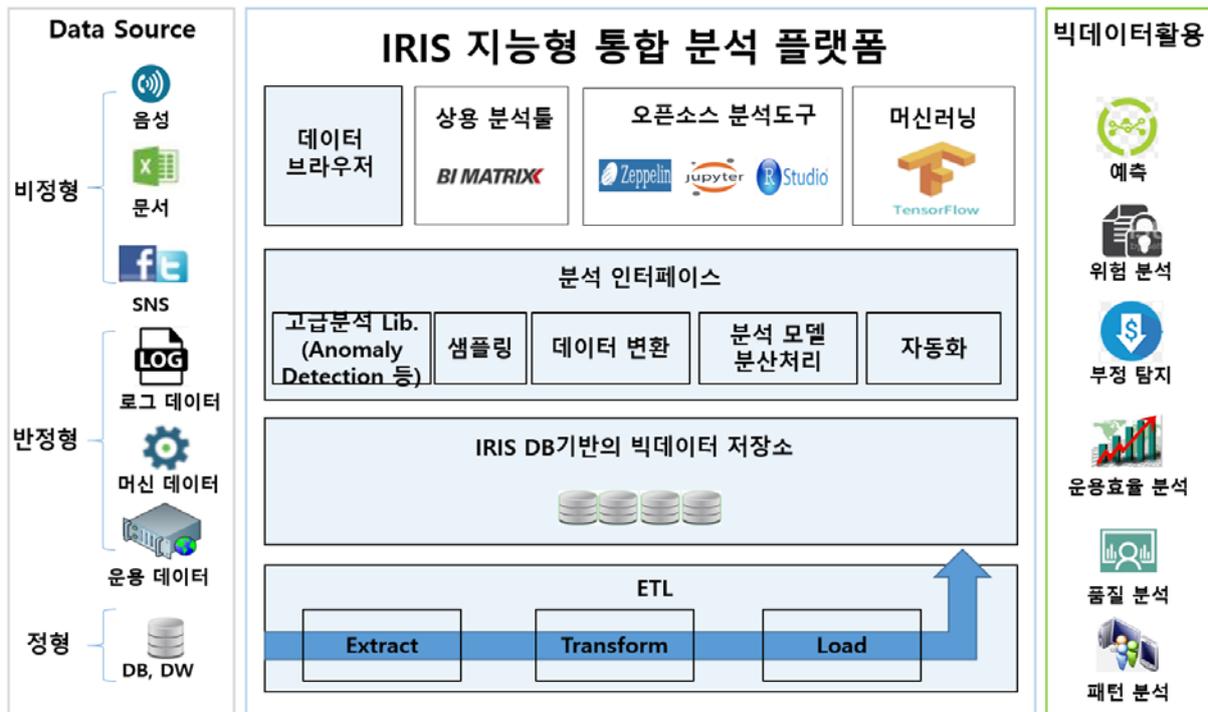
동일한 조건의 분석 업무를 반복적으로 수행하여 보고서를 생성하여야 할 경우나 일정 시간이 소요되는 통계 생성 업무의 경우, 해당 job을 등록하여 자동으로 실행할 수 있습니다. 등록된 job은 사용자가 간단한 설정을 통하여 정의한 스케줄에 따라 자동으로 실행되기 때문에 업무 효율을 향상시킬 수 있습니다.

Product Overview

System Overview

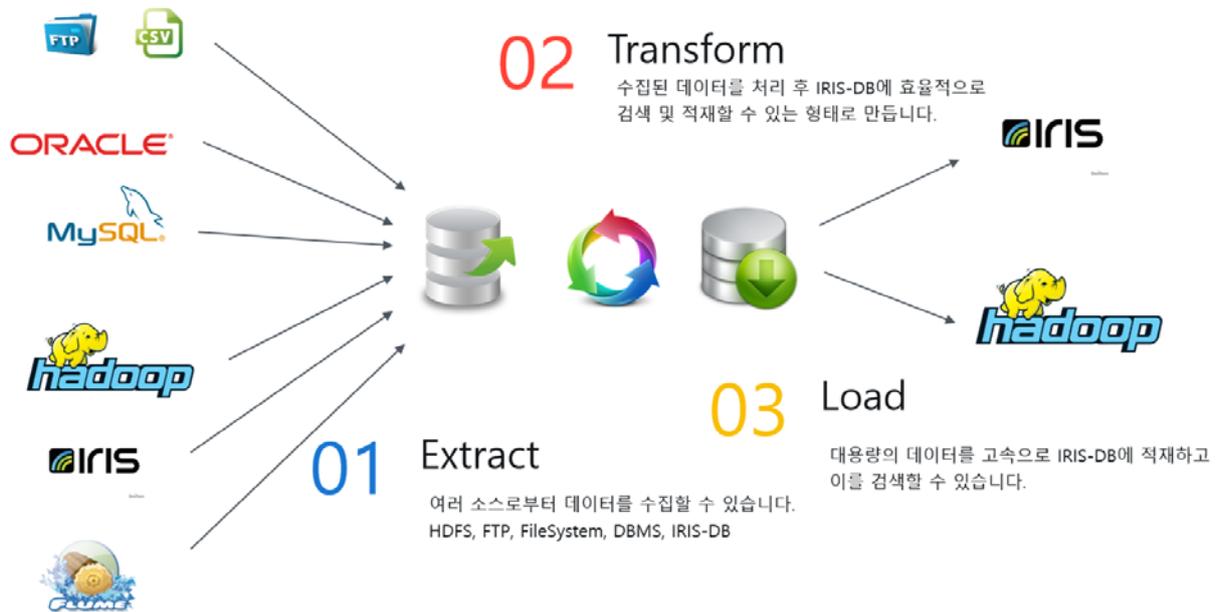
IRIS는 단일 플랫폼에서 빅데이터의 수집, 처리, 저장, 분석이 가능한 통합 분석 플랫폼입니다.

IRIS는 ETL, 실시간 빅데이터 저장소, 분석 인프라, 데이터 브라우저, 오픈소스/상용 분석 도구 등의 컴포넌트들로 구성되어, 빅데이터 시스템을 구축하고 빅데이터로부터 인사이트를 얻고자 하는 기업에게 토탈 솔루션을 제공합니다. 특히, 분석가가 분석에 필요한 데이터를 추출하기 위하여, 또는 분석 도구와의 연동 등을 목적으로 추가 개발할 필요 없이, 다양한 분석 도구를 사용하여 분석 업무를 수행할 수 있도록 오픈 소스 분석도구 및 상용 분석 툴을 점진적으로 통합하고 있습니다.



IRIS ETL

IRIS ETL은 다양한 종류의 외부 시스템으로부터 정형, 반정형, 비정형 데이터를 수집/추출(Extract)할 수 있습니다. 또한, 수집된 데이터는 사용자가 요구하는 형태로 변환(Transform)하거나, 통합, 요약 등의 작업을 거친 후, IRIS 스토리지에 고속으로 적재(Load) 됩니다. IRIS ETL은 고객의 요구사항을 수용하기 위하여 개발 용역 업무를 수반하던 기존의 방식을 개선하여 생산성과 안정성을 높였습니다.



IRIS ETL 내부의 주요 기능 모듈은 다음과 같습니다.

기능 모듈	설명
Extract	<ul style="list-style-type: none"> 대상 시스템 연동을 통한 다양한 데이터 수집 지원 대상: <ul style="list-style-type: none"> - 상용 데이터베이스 (Oracle, MySQL) - 하둡기반 분산파일 시스템(아파치 배포판, 호트웍스 배포판) - NoSQL 데이터베이스 - 파일시스템 (Excel, CSV 등) 다양한 운영체제 지원 (Linux, NT, UNIX 등)
Transform	<ul style="list-style-type: none"> 수집된 원시 데이터에 대하여, GUI를 통한 자유로운 변환기능 제공 <ul style="list-style-type: none"> - 데이터 필터링 - 룰기반 파서 - 이기종 DB table간의 join - Regular expression을 통한 데이터 변환

	- 다양한 데이터 변환 함수 제공
iLoader (Fast Loader)	<ul style="list-style-type: none"> • ETL에서 수집 및 처리된 데이터를 IRIS 스토리지에 고속으로 로드
관리기능	<ul style="list-style-type: none"> • 스케줄링: 작업 스케줄에 따른 job 실행 혹은 일괄 배치 처리 가능 • 모니터링: job, 데이터 수집현황, 시스템 자원에 대한 모니터링 및 실시간 수집 데이터의 통계 정보 제공 • 장애관리: <ul style="list-style-type: none"> - 장애 발생 시 메일 통보 기능 - 장애 로그 생성 - 자원 모니터링 - 장애 발생 데이터에 대한 복구 프로세스에 의한 재처리

IRIS 빅데이터 저장소(Big Data Repository)

IRIS의 빅데이터 저장소인 IRIS Enterprise DB(이하 IRIS DB)는 하이브리드(In-memory and on-disk) 방식의 분산 병렬 DB 클러스터로 구성됩니다. 단일 노드 DBMS와 달리 데이터를 각 노드에 분산하여 저장하고, 분산된 각 노드에서 개별적으로 SQL 쿼리를 실행함으로써, 일반 DBMS에서 수용할 수 없는 빅데이터의 실시간 처리를 가능하게 합니다.

IRIS DB의 이중화된 마스터 노드는 분산과 조합을 통제하며, 스토리지, 메모리, CPU를 별도로 보유하는 각 데이터 노드는 SQL 연산을 병렬로 처리함으로써 부하 분산과 고속 연산을 가능하게 합니다. 분산 노드간의 데이터 중복저장, 부하분산, 장애에 대한 내성 보장(fault tolerance), 데이터 압축 지원 등을 제공함으로써, 오픈소스 빅데이터 플랫폼을 이용한 빅데이터 시스템 구성에 비하여 비용과 성능 면에서 유리합니다.

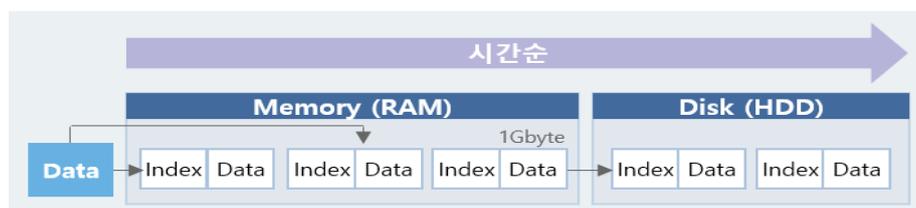
IRIS DB는 또한 확장성 (Scalability) 측면에서 수평적 증설 (Scale-out)을 지원하기 때문에, 용량을 확장할 경우에는 서비스 운용중인 상태에서도 손쉽게 데이터 노드를 추가 증설할 수 있도록 설계되었습니다.

In-memory 기반의 대용량 데이터 실시간 색인 생성

IRIS DB는 실시간으로 입력되는 데이터에 대하여, 각 분산 데이터 노드의 메모리 상에서 일정한 크기의 블록파일 단위로 색인을 생성하여 저장합니다. 대량의 데이터를 IRIS DB에 로딩할 때, 디스크 I/O 병목이 가장 심한 색인 생성을 메모리상에서 수행함으로써, 실시간으로 데이터를 처리할 수 있습니다.

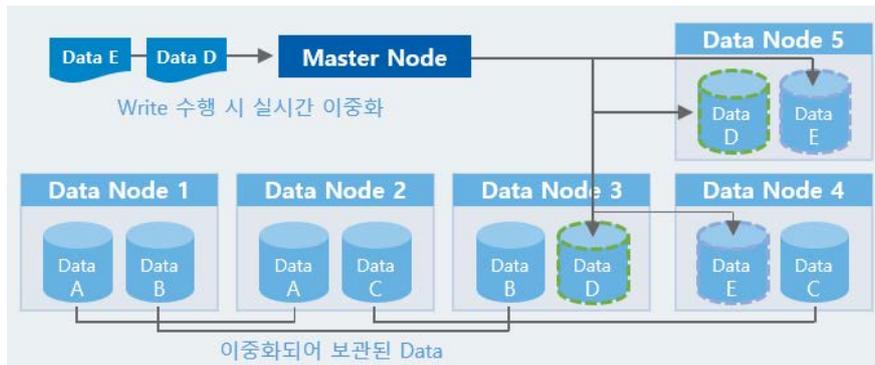
IRIS DB에 적용된 실시간 색인 생성 기술은 지난 10여 년의 대량 데이터 고속 처리 경험을 기반으로 자체 개발한 기술로서, 대형 고객 사이트에서 초당 1백만 ~ 1천만 레코드에 대한 실시간 색인을 성능 저하 없이 처리함으로써 충분히 검증되고 안정화되었습니다. 또한, 색인과 함께 메모리에 저장된 블록파일은 가용 자원의 수용 용량의 범위 내에서 데이터를 일정 기간 동안 메모리에 유지한 후, 디스크로 이전하여 저장됩니다.

IRIS DB는 특히, 시간 순으로 수집되는 로그 데이터 등 time-stamped 데이터를 실시간으로 처리하며, 특정 로그를 실시간으로 조회할 때 탁월한 성능을 경험할 수 있습니다.



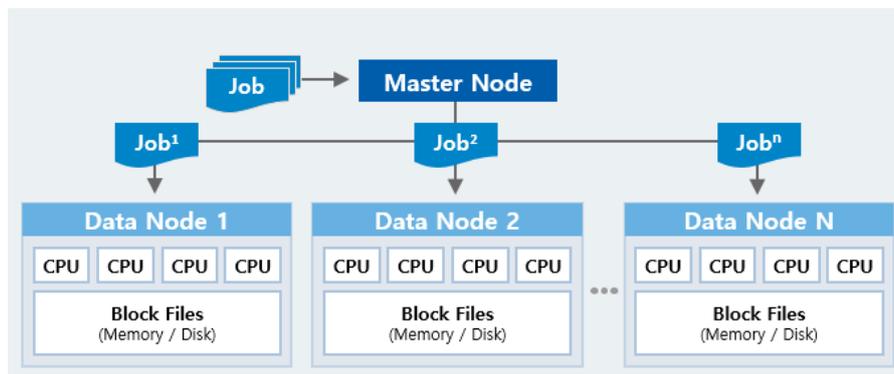
데이터 다중화를 통한 안정성 확보 및 부하 분산

IRIS DB로 입력되는 데이터는 마스터 노드의 관리하에, 물리적으로 분리된 데이터 노드에 저장됩니다. 따라서, 특정 데이터 노드에 장애가 발생한 경우에도 데이터 유실을 방지할 수 있으며, 데이터 처리를 위한 부하가 각 노드에 분산되므로 성능 저하를 방지할 수 있습니다.



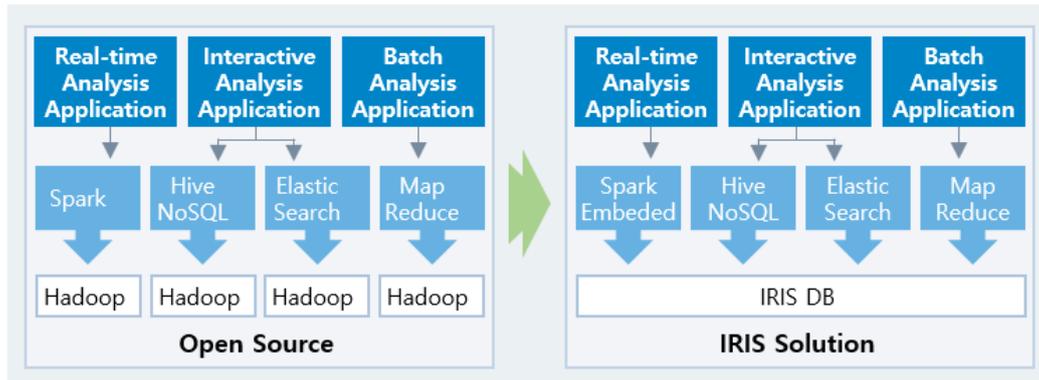
성능과 속도 향상을 위한 데이터의 분산처리

IRIS DB는 sharding 기술을 채택하여, 입력된 데이터를 다수의 데이터 노드에 분산하여 저장합니다. 또한, IRIS DB는 각 노드의 메모리와 디스크를 활용하여, 사용자의 쿼리 요구를 병렬로 처리하기 때문에, 대량의 데이터를 고속으로 처리할 수 있습니다. 특히, 데이터 처리 부하가 각 데이터 노드로 분산되므로 성능 저하로 인하여 발생할 수 있는 문제를 예방할 수 있습니다.



IRIS DB 기반의 오픈소스 빅데이터 플랫폼으로의 진화

IRIS DB는 오픈소스 프레임워크에 대하여 단일 데이터 저장소 역할을 수행하며, Spark을 포함한 Hadoop 에코시스템 등의 오픈소스 프레임워크와 통합하여 오픈소스 빅데이터 플랫폼으로 활용할 수 있습니다. 특히, 분 단위 미만의 짧은 반응 속도를 필요로 하는 작업의 경우, IRIS DB는 탁월한 투자비용 절감 효과를 제공합니다.



사용자 편의를 고려한 클러스터 통합 관리

빅데이터 클러스터를 운영하는 데 있어서 문제점 중의 하나는 분산 노드로 구성된 시스템을 업무량과 비용을 고려하여 효율적으로 관리할 수 있는 솔루션을 찾는 것입니다. 만일, 이런 솔루션이 없다면 독립된 개별 서버들을 관리하는 데 발생하는 수준의 업무량과 비용이 발생할 수 있습니다. IRIS DB는 관리자가 IRIS DB 클러스터를 통합 환경에서 관리할 수 있는 Web GUI 환경을 제공합니다. IRIS DB에서 제공하는 주요 클러스터 관리기능은 다음과 같습니다.

관리 기능	설명
통합 모니터링	<ul style="list-style-type: none"> • 각 노드의 자원(CPU, 메모리, 디스크 등) 상태 감시 • 쿼리 에러 발생 노드 감시 • 세션 상태 감시
DB테이블 관리	<ul style="list-style-type: none"> • DB 테이블의 정보 조회 및 신규 테이블 생성 • 쿼리 실행
프로세스 관리	<ul style="list-style-type: none"> • 각 노드에서 실행되는 프로세스의 동작 관리
리소스 상태 관리	<ul style="list-style-type: none"> • 각 노드별 주요 자원의 상태 관리 (CPU, 메모리, 디스크, 테이블 등) • 각 노드의 자원 사용율을 통계화하여 자원 사용율 히스토리 관리
세션 관리	<ul style="list-style-type: none"> • 쿼리 세션 현황 관리
사용자 관리	<ul style="list-style-type: none"> • IRIS DB에 접속중인 사용자의 접근 권한 등 사용자 정보 관리

다양한 고객사에 공급하여 검증되고 안정화된 빅데이터 데이터베이스

모비젠은 풍부한 경험의 빅데이터 전문 인력 및 검증된 빅데이터 솔루션을 기반으로, 대규모 IRIS DB 공급 프로젝트 (1일 약 1,000억건, 2 PB 규모), 대형 하둡 클러스터 구축 및 운용(약 800대 하둡 서버에서 6PB 데이터 처리) 등 다양한 산업 분야에서 대형 프로젝트들을 수행하고 있습니다.



분석 인프라

일반적으로, 분석가가 빅데이터로부터 원하는 분석 결과를 도출하려면, 빅데이터 저장소에서 분석에 필요한 데이터를 필터링하고, 분석 용도에 따라 일부 데이터를 변환하는 등 일종의 분석 데이터 처리 과정이 필요하게 됩니다.

예를 들어, 엄청나게 많은 데이터 중에서 분석에 유의미한 데이터 영역을 탐색하거나, 여러 종류의 필드값을 하나로 통합하거나, 또는 특정 값을 일정 범위로 그룹핑하여 레이블을 지정할 필요가 있습니다. 이 경우, 분석가 스스로 혹은 개발자에게 요청하여 해당 결과를 얻기 위한 프로그램을 개발하여야 합니다. 물론, 가시화 툴이나 BI 툴이 이런 문제를 일정 부분 해결해 줄 수는 있겠지만, 많은 분석가들은 여전히 분석 데이터 처리를 위한 프로그램 개발이나 수작업으로 인하여 불필요한 시간과 인력을 낭비하고 있습니다.

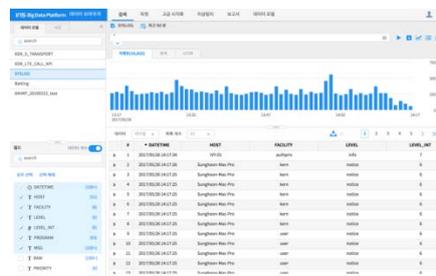
IRIS는 분석가가 자신이 익숙한 분석 도구를 사용하여 IRIS에 저장된 빅데이터를 쉽고 빠르게 분석할 수 있도록 고급분석 기능을 비롯하여 다양한 분석지원 기능을 제공하고 있습니다.

쉽고 편리한 데이터 브라우저

데이터 브라우저는 IRIS 저장소에 저장된 각종 데이터로부터 특정 키워드 혹은 문장을 검색하거나, IRIS가 제공하는 다양한 함수를 사용하여 데이터를 분석할 수 있는 환경을 제공합니다. 또한, 데이터 브라우저는 GUI를 통하여 오픈소스 분산 파일 시스템(HDFS)에서 데이터를 분석할 수 있도록 지원합니다. 데이터 브라우저에서 제공되는 기본 기능은 다음과 같습니다.

데이터의 고속 검색

IRIS에 저장된 대량의 데이터 중, 특정 키워드(숫자, 문자, 단어, 문장 등)가 포함된 이벤트를 고속으로 검색하며, 통계 및 차트를 통해 비정상 상황을 분석할 수 있는 기능을 제공합니다.

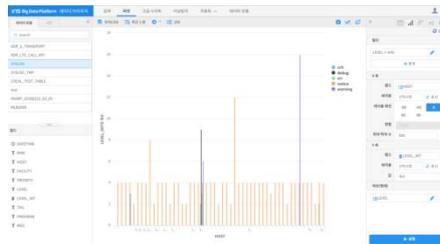


- **데이터 검색:** 특정 검색어 또는 검색 조건을 입력하여 이벤트를 검색하고, 해당 이벤트의 정렬, 연산, 재 검색 등을 수행합니다.

- **통계:** 검색된 이벤트에 대하여, 특정 필드의 개수, 평균, 최대값/최소값/중간값, 표준편차, 합계, 분산 등 통계를 활용하여 분석 업무를 수행할 수 있습니다.
- **차트:** 상기 통계에 대하여 꺾은 선형, 영역형, 세로 막대형, 막대형, 원형 등 차트 형태로 분석 업무를 수행할 수 있습니다.

피벗 분석

분석하고자 하는 데이터 기준을 사용자가 임의대로 정렬하고 필터링 함으로써, 보다 빠르고 쉽게 데이터를 분석할 수 있는 피벗 기능을 제공합니다.



피벗 기능은 사용자가 검색 결과의 각 필드, 행, 열 등을 GUI 화면상에서 적절히 조합함으로써 원하는 분석 결과를 쉽고 빠르게 도출하기 위하여 사용되는 강력한 분석 툴입니다. IRIS 데이터 브라우저는 사용자가 피벗 기능을 쉽게 사용할 수 있도록 직관적인 사용자 인터페이스를 제공합니다. 사용자는 데이터 검색과 분석을 위한 명령어에 대한 교육을 받지 않더라도 화면의 메뉴를 클릭함으로써 쉽게 피벗 기능을 사용할 수 있습니다.

자동화

반복적인 분석 보고서 생성 혹은 주기적인 통계 생성 등 동일한 업무를 반복적으로 수행하는 경우, 해당 job을 등록하여 자동으로 실행할 수 있는 자동화 기능을 제공합니다. 자동화 기능은 다음과 같이 사용자가 job의 등록에서부터 현황 파악까지 job의 전반적인 관리를 할 수 있도록 지원합니다.

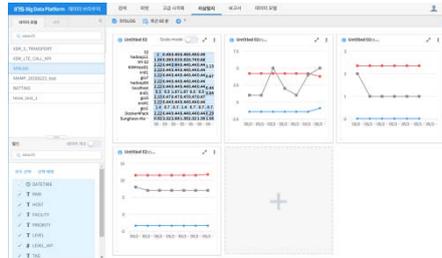
- 템플릿 기반의 job 등록 UI 제공
- 등록된 job의 현황 및 현재 실행중인 job의 상태 파악을 위한 운용 현황 정보 제공
- 등록된 job의 실행 스케줄 관리
- 기 실행된 job에 대한 이력 관리

머신러닝 기반의 이상 탐지(Anomaly Detection)

IRIS에서 제공하는 고급 분석 기능인 Anomaly Detection (이하 AD) 기능은 머신러닝 기반의 과거 데이터 학습을 통하여 IRIS ETL에서 수집 및 처리한 다양한 형태의 대량 데이터를 자체적으로 분석하고, 사용자가 비정상적인 이벤트를 쉽고 빠르게 파악할 수 있도록 지원합니다. AD 기능의

특징은 다음과 같습니다.

- GUI상에서 특정 필드의 드래그만으로 이상 탐지 가능
- Heat map, 시계열 차트 등을 통한 직관적 분석
- 세밀한 이상탐지를 위한 다양한 옵션 제공



특히, 모비젠의 전문 분석가 그룹에서 수 년 동안 대량의 데이터를 기반으로 최적화된 이상탐지 모델을 적용하여 탐지 오류를 최소화하였습니다.

AD 기능은 네트워크 운용, 보안 등 다양한 분야에서 활용함으로써, 신속한 이상 탐지가 가능하고 사용자의 업무 효율이 크게 개선되고 있습니다. AD 기능의 주요 활용 사례는 다음과 같습니다.

- ICT 장비 혹은 공장 설비의 비정상 동작에 대한 모니터링 및 예측 정비
- 사이버 공격 등 네트워크의 이상징후에 대한 조기 감지
- 이상금융거래 탐지

원시 데이터의 summary용 SQL 지원

IRIS에서는 분석가가 원시 데이터를 대상으로 summary SQL문을 작성하여 실행하고, 그 결과를 차트 및 테이블 형태로 확인할 수 있습니다. 분석가는 원시 데이터로부터 요약된 정보를 분석하여 상세 분석을 위한 유의미한 데이터 영역을 찾아낼 수 있을 것입니다. Summary SQL 지원 기능을 요약하면 다음과 같습니다.

- 분석가가 지정한 초/분/시간/일/주/월 단위로 원시 데이터의 요약
- Aggregation 함수 설정: 합계, 도수, 평균, 최대, 최소, 분산, 표준 편차, 중간값, IQR 등
- 누락된 데이터의 사용자 지정 보정
- Summary SQL 실행 결과의 HDFS 저장

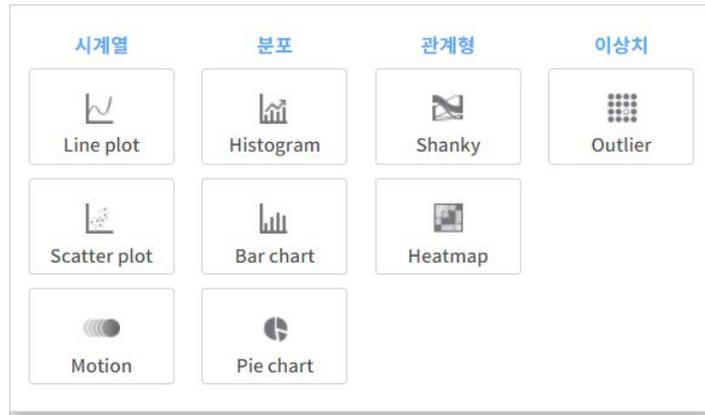
분석 전처리 (분석용 데이터의 변환)

분석가들이 데이터 분석을 위하여, 원시 데이터로부터 특정 값들을 조합, 분리, 변환하는 등 가공하여 분석에 사용하는 경우가 빈번하게 발생합니다. 이때, 경우에 따라서는 분석을 위한 데이터의 가공을 위해 적지 않은 시간과 노력이 수반됩니다. IRIS는 분석가가 불필요한 시간을 줄이고 데이터 분석에 전념할 수 있도록 실제 업무 환경에서 필요로 하는 다양한 분석 지원 툴들을 제공합니다.

상세 기능	설명
컬럼 split	특정 컬럼의 값을 분리하는 기능
컬럼 concatenation	사용자가 선택한 다수 개 컬럼의 값을 하나의 컬럼으로 합치는 기능
반올림/내림	특정 컬럼의 수치형 값에 대하여 반올림, 내림, 버림하는 기능
범위 지정	특정 컬럼의 수치형 값에 대하여 사용자가 지정한 범위로 그룹을 지정하는 기능 (예, 컬럼 값이 0 ~ 40 : A 그룹, 41 ~ 60 : B 그룹 등으로 지정)
컬럼 Join	특정 테이블의 컬럼들을 join하여 새로운 컬럼을 추가하거나, 기존 컬럼을 대체하는 기능
데이터 변환	<ul style="list-style-type: none"> Time format 변환: 원시 데이터 중 날짜/시간의 포맷을 사용자 지정 포맷으로 변환 (예, YYYYMMDDHHMMSS -> YYYY-MM-DD HH:MM:SS) 단위 변환: 온도, 길이, 데이터량 등 계량/계측 단위를 변환
데이터 가공	정규 표현식을 이용하여 특정 컬럼의 값을 가공 (예, 45005* -> 9999*)
필터	특정 컬럼의 값을 필터링 (<, >, =, !=,)
가공 데이터 export	가공된 데이터를 export한 후, 다른 화면에서 import하여 재사용할 수 있는 기능

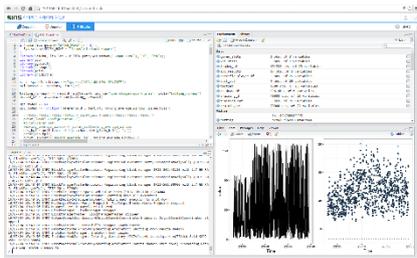
시각화 차트 지원

IRIS의 시각화 기능은 빅데이터의 분석 결과를 쉽게 이해하고 빠른 의사 결정을 위하여, 데이터의 유형에 따라 다양한 차트를 생성할 수 있도록 지원합니다. 분석가는 데이터의 유형에 따라 적절한 차트를 선택한 후, 특정 데이터 필드를 해당 차트 영역에 드래그하기만 하면 됩니다. IRIS에서 기본적으로 제공하는 차트는 다음과 같으며, 지속적으로 차트의 종류를 확장하고 있습니다.



오픈소스 분석도구 및 상용 BI tool 통합

모비젠은 분석가들이 익숙한 분석 도구를 사용하여 쉽고 빠르게 분석 업무를 수행할 수 있도록, R-Studio, Zeppelin, Jupyter 등 각종 오픈소스 분석 도구는 물론, BI Matrix, Metatron, Tableau 등



상용 BI tool들을 IRIS에 통합하였습니다. 기존의 일부 빅데이터 플랫폼들도 오픈소스 분석 도구와의 연동을 지원하지는 하지만, 빅데이터 플랫폼과 오픈소스 분석 도구 간 데이터 공유의 제약이나 사용의 불편함 등으로 인하여 분석가들이 불만을 제기하는 경우가 자주 발생되고 있습니다. IRIS는 데이터 브라우저에서 검색되고 필터링된 데이터를

오픈소스 분석도구나 BI tool에서 바로 사용하여 분석하거나, 오픈소스 분석도구 혹은 BI tool에서 사용된 데이터를 IRIS 화면에서 재활용하는 등 IRIS와 오픈소스 분석도구 간에 제한 없이 데이터를 공유할 수 있습니다.

A New Experience in Big Data Analytics

IRIS



www.mobigen.com

Contact Information

본사

(06187) 서울 강남구 역삼로 552 케이엠빌딩 3층

TEL : 02-538-9360

FAX : 02-538-9369

Email: sales@mobigen.com

연구소

(06187) 서울 강남구 역삼로 552 케이엠빌딩 5층

분당지점

(13595) 경기도 성남시 분당구 수내로 46번길 12, 401호

TEL : 070-8666-3892

FAX : 070-4009-6924