

빅데이터 기반, 스마트 모니터링 및 분석 시스템

이정규 tbs 교통방송 TV기술국

tbs 교통방송은 HD 제작시스템 업그레이드와 상암동 신사옥 이전 계획에 따라 이전까지 한 번도 시도되지 않았던 방송/미디어 빅데이터를 활용한 스마트 모니터링 및 분석 시스템을 시범운영 중에 있다. 원고 청탁을 받고 아직 시작단계이지만 실제 구축과 운영을 해본 경험을 이 분야에 관심 많은 방송기술인을 위해 공유하여 다가올 DT(Data Technology) 시대에 필요한 기술과 대응방안을 함께 마련하고자 한다.

빅데이터 이해하기

우리는 매일 데이터 속에서 하루를 시작하고, 하루를 마감하고 있다. 아침에 일어나면 제일 먼저 스마트폰을 확인한다. 지난밤 사이 수신된 문자메시지나 이메일이 없는지, 오늘 할 일은 무엇인지, 집을 나와서 출근길 버스와 지하철에서 주요 뉴스에 대한 기사를 읽고 회사에 출근하여 지문인식기 출근체크 후 정해진 업무의 일상으로 하루를 시작한다. 미리 정해진 업무 외에도 업무 중에 수시로 발생하는 회의를 통해서 쏟아지는 수많은 데이터들 속에서 심신이 지칠 때 즈음에 출근 때와 마찬가지로 지문인식기 퇴근체크를 하고 또다시 버스와 지하철에서 포털 검색이나 SNS 서핑을 하면서 집에 와서 하루를 마감한다.

빅데이터란, 사전적 의미로 해석하면 “디지털 환경에서 생성되는 데이터로 그 규모가 방대하고, 생성 주기도 짧고, 형태도 수치 데이터뿐 아니라 문자와 영상 데이터를 포함하는 대규모 데이터를 말한다. 빅데이터 환경은 과거에 비해 데이터의 양이 폭증했다는 점과 함께 데이터의 종류도 다양해져 사람들의 행동은 물론 위치정보와 SNS를 통해 생각과 의견까지 분석하고 예측할 수 있다.” 라고 정의하고 있다. [네이버 지식백과] 빅데이터 정의 (빅데이터, 2013. 2. 25., 커뮤니케이션북스)



인터넷이라는 새로운 소통의 도구가 시작된 지도 벌써 20년이나 되었다. 구글, 아마존, 알리바바 등은 굳이 이들이 누군지, 무엇을 하는지를 설명하지 않아도 세계 최고의 기업가치를 인정받는 기업이라는 것을 알 수 있다. 고객이 누군지, 무엇을 구매했는지, 어떤 걸 좋아하는지, 매출은 얼마인지, 무엇을 해야 하는지 등 BI(Business Intelligence) 분석을 통한 결과에 대한 과거의 데이터 분석에

서 현재는 어떤 고객이 왔는지, 무엇을 하고 있는지, 어떤 행동을 했는지, 무엇을 구매했는지, 다음에 또 방문했는지, 매출 손실의 원인이 무엇인지 등 O(Operaton Intelligence) 분석을 통한 좀 더 세밀하고 정확한 실시간 트렌드에 대한 데이터 분석에 중점을 두고 있다. 구글, 아마존, 알리바바 등 그들은 이미 실시간 트렌드 데이터 분석이 고객 이탈과 신규고객에 결정적인 지표를 제공하리라고 예견했을 것이다.



빅데이터 처리 기술

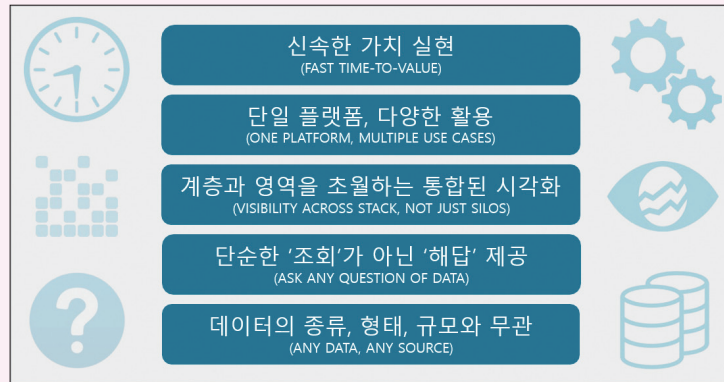
빅데이터 하면, 대부분은 구글부터 생각할 것이다. 구글은 실시간 검색, 번역, 이메일, 클라우드 서비스 등 우리가 필요로 하는 모든 요소의 정보와 서비스를 제공해 주고 있다. 빅데이터 오픈소스 소프트웨어(open-source software)의 대표적인 하둡(hadoop)을 비롯하여 지금까지 소개되고 있는 대부분의 빅데이터 플랫폼 기술이 구글이 보유한 방대한 데이터를 활용한 실시간 서비스로부터 착안되었다고 해도 과언이 아닐 정도로 구글은 우리가 상상할 수 없을 정도로 많은 데이터와 연관된 플랫폼을 보유하고 있다.

구글의 검색엔진 기술은 대량의 정보를 효과적으로 저장하기 위한 분산파일 시스템(GFS, Google File System), 대용량 데이터의 읽기와 쓰기를 위한 분산 스토리지 시스템인 빅테이블(Bigtable), 분산 데이터 처리를 위한 맵리듀스(MapReduce)로 요약할 수 있다. 구글에 대한 하둡의 빅데이터 기술은 하둡 분산파일 시스템(HDFS, Hadoop Distributed File System), 하둡 맵리듀스(Hadoop MapReduce), 에이치베이스(Hbase)가 각각 담당하고 있다.

데이터의 종류와 양이 폭증하면서 분석된 결과에 대한 종합적인 제안을 위해서 좀 더 고도화된 데이터 마이닝 기술을 필요로 하고 있다. 데이터 마이닝 표준 처리 과정(CRISP-DM, Cross Industry Standard Process for Data Mining)은 비즈니스 이해(Business Understanding), 데이터 이해(Data Understanding), 데이터 준비(Data Preparation), 모형(Modeling), 평가(Evaluation), 적용(Deployment)의 6단계로 구성되어 있다. 수치 데이터와 문자, 영상에 이르기까지 정형화된 데이터부터 비정형 데이터를 의미 있는 데이터로 추출(selection)해서 사전 처리(preprocessing)와 변환 과정(transformation)을 거쳐 분석(data mining)하고 결과를 해석하는 전 과정을 데이터 마이닝 기술로 요약할 수 있다. (Chapman, et al., 2000).

온라인 쇼핑몰 사업을 하는 기업에서는 실시간 트렌드 분석 결과가 매출에 지대한 영향을 끼친다는 것을 잘 알고 있다. 사업 특성상 수많은 경쟁사와의 치열한 경쟁에서 이기기 위해 24시간 뒤의 결과값만으로 대응하기에는 소비자는 기다려 주지 않는다. 빅데이터 플랫폼 운영에서 흩어진 데이터를 빠르게 수집해서 분석하는 것만큼 중요한 것은 분석된 데이터를 사업에 즉시 반영될 수 있도록 하는 의사결정에 절대적으로 필요한 시각화 작업이다.

운전을 하는 대부분의 사람들은 길안내 서비스 도구인 내비게이션을 사용하거나 사용경험을 가지고 있을 것이다. 가고자 하는 목적지를 기준으로 거리, 시간, 경로, 주변시설, 교통법규 제한 등 다양한 데이터를 일목요연하게 운전자 가 잘 보고 들을 수 있도록 안내하는 기술을 데이터 시각화라고 이해한다면 데이터 시각화가 그리 어렵지 않게 이해될 것이다. 데이터 시각화(data visualization)는 데이터 분석 결과를



쉽게 이해할 수 있도록 시각적으로 표현하고 전달하는 과정을 말한다. 데이터 시각화의 목적은 도표(graph)라는 수단을 통해 정보를 명확하고 효과적으로 전달하는 것이다(Friedman, 2008).

빅데이터 접근하기

최근 들어 국가정보화 3.0의 핵심정책 방향은 新국가정보화 추진 전략을 마련하고, 공공정보를 최대한 개방해서 국민의 삶을 편리하게, 창업 일자리를 창출하는 것이다. 공공데이터 제공 및 이용활성화에 관한 법률 제정으로 법 시행 전 5,000여 개에 불과하던 개방 데이터가 현재는 1만4,000여 개에 이르고, 공공 데이터 이용은 25배, 민간 앱 개발은 42개에서 570개로 13배가 늘어났다.

빅데이터 플랫폼 확보를 위해서는 수집, 분석, 처리, 저장, 시각화 등 데이터 관리에도 중점을 두어야 하지만 안정적인 운영과 고가 용성을 위해 클러스터링 기반의 스케일아웃 아키텍처 컴퓨팅 인프라 기술에도 역점을 두어야 한다. 또한 흩어져 있는 셀프 서비스 BI들을 통합관리하여 히스토리컬한 데이터를 쌓고 분석될 수 있도록 하여야 한다.

누구나 다 좋아하고 콘텐츠를 골라 잘 보이는 곳에 위치시키는 방식으로 운영하던 기존의 방식에서 특정 시간, 특정 디바이스, 특정 콘텐츠 등 소비자들의 반응에 관심을 가지고 고객 맞춤형 서비스를 통해서 15%~20%에서 25%~30%로 높아진 CRT(노출대비 클릭수) 기대효과를 경험한 다음 뉴스 서비스에 빅데이터 분석을 적용한 사례와 고객 단말기 사용현황(기간과 자사제품 유무), 신제품에 대한 반응 등에 대한 분석을 통해서 신제품 출시 시점, 출시 이후 마케팅 전략을 적용한 갤럭시S6에 빅데이터 분석을 적용한 사례는 수많은 빅데이터 속에서 살아가는 우리에게 시사하는 바가 크다.

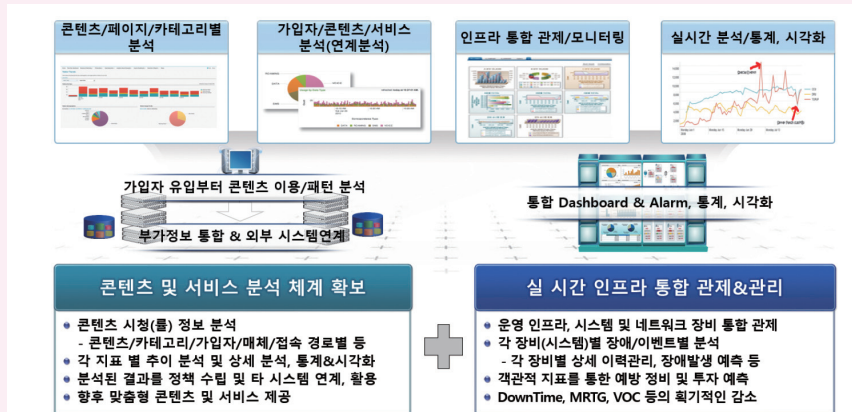
빅데이터, 접근하기 쉽지 않은 영역임에는 분명하다. 지금 이 순간에도 우리 기업 내에서 의미 있는 수많은 데이터가 쏟아져 나오고 방치되고 있다는 것을 알아야 할 때이다. 데이터는 곧 미래를 준비할 수 있는 강력한 도구이자 해결방안을 제시해줄 중요한 척도라는 사실을 이해하고 고객 대응 역량 향상과 기업경쟁력을 위해 빅데이터 통찰력을 높이는 방안을 마련할 수 있도록 하여야 한다.

방송/미디어에서 빅데이터 활용 방안

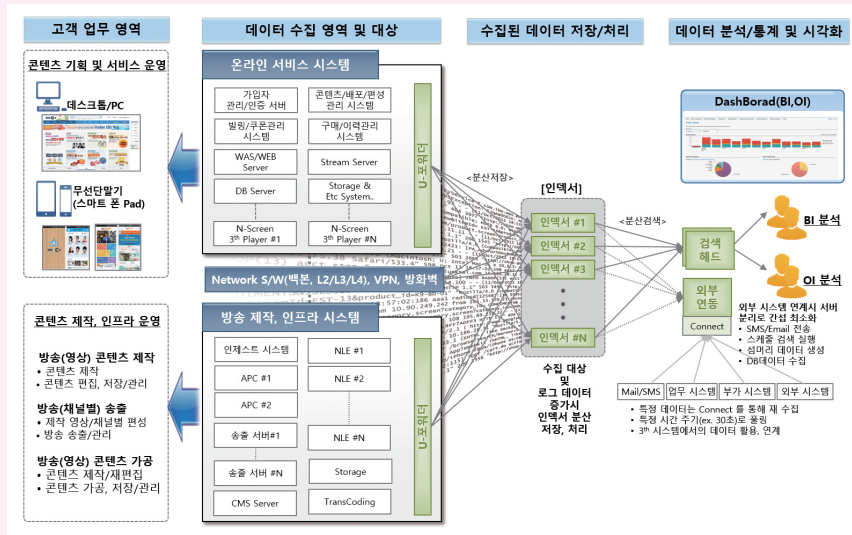
최근 몇 년 사이 디지털 방송과 다양한 디바이스를 통한 서비스가 다양해지면서 방송국에서는 파일기반 제작시스템에 많은 예산과 시간을 투자해 왔다. 파일기반 제작시스템을 도입하면 가장 어려움을 겪는 부분이 모든 방송 시스템들이 네트워크 스위치에 연결되어 동작되고 있는 탓에 어느 한 분야의 IT 기술만 집중해서는 관리, 운영할 수 없다는 것이다.

파일기반 제작시스템의 가장 중심에 있는 NPS(Network Production System)와 그 외 보도정보시스템, 편성정보시스템, 문자발생기, 비디오 스위처, 오디오 믹서, APC, 키어, 압축다중화 시스템, UPS, 항온항습기, 기타 방송시스템 등 방송 운영 전반에 걸쳐 수많은 시스템과 장비들은 대부분 네트워크 스위치에 연결되어 X86 기반의 서버에 Windows, Linux/Unix 등과 같은 OS를 사용하고 있으며, 여기에 솔루션 업체가 개발한 Application에 의해서 구동되고 있다.

방송제작을 마치면 자사가 운영하는 웹/앱 서비스와 3rd Player(IPTV, 케이블TV, 유튜브, 네이버, 다음, 티빙 등)을 위해 콘텐츠 배



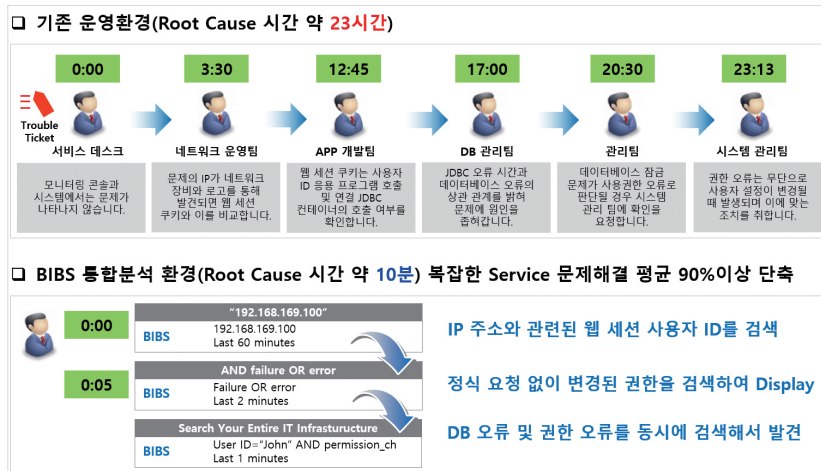
포와 서비스 연계를 진행한다. 아날로그 방송 때부터 전통적으로 해오던 외부 리서치 기관에 의한 시청률 조사 데이터만으로는 시청자 반응을 확인하여 프로그램 제작/기획, 프로그램 및 광고 편성 등에 대한 정책을 마련한다는 것은 다소 무리가 있다.



다양한 이기종의 방송시스템으로부터 발생하는 머신 데이터는 시스템 성능, 용량, 사이버보안 위협, 애플리케이션 동작방식, 콘텐츠 이동경로, 서비스 수준, 사기성 활동과 사용자 경험 등 다양한 이력을 담고 있지만 과거의 방법론으로 분석하는 것이 쉽지 않다.

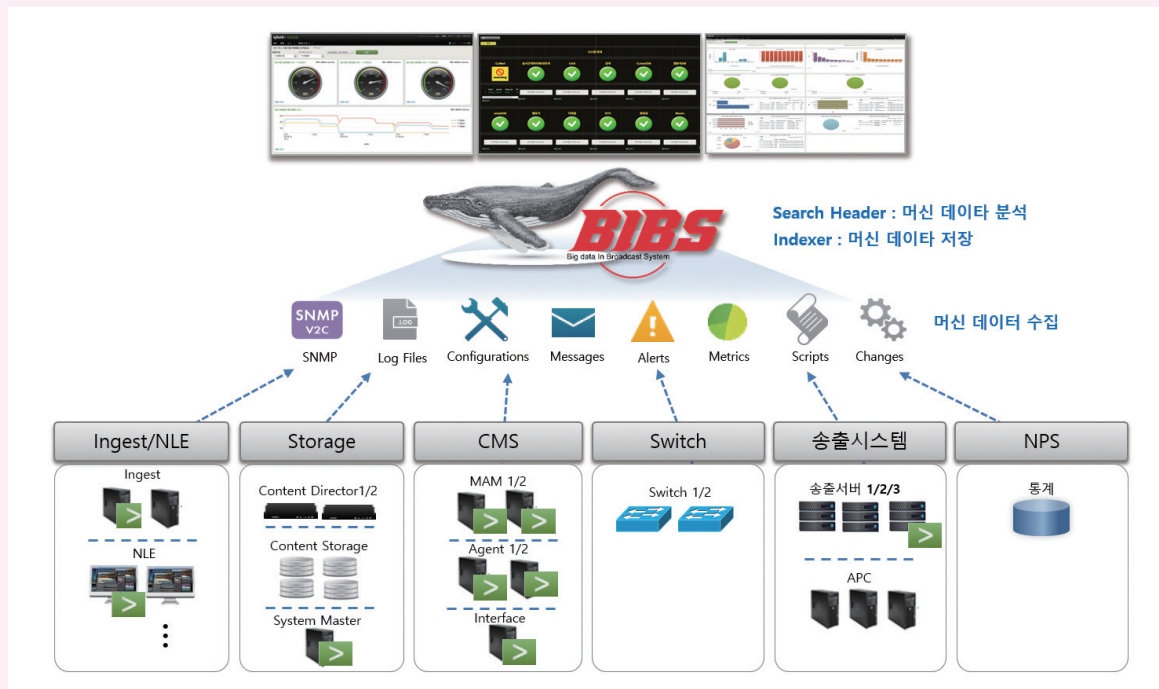
품질 좋은 방송과 우수한 콘텐츠 제작에 앞서 그러한 기대에 대응할 수 있는 관리와 운영을 하고

있는지가 우선이 되어야 한다. 현재 대부분의 방송운영 실무에서는 방송시스템에 대한 장애감시를 NMS(Network Management System)를 통해서 인지하고 고장수리를 해당 업체에 의존하고 있다. 이러한 현상 가운데 특이점은 일부 장애에 대해서는 원인도 모르고 넘어가는 경우가 많다. 다행히 NMS에서 감시된 장애는 그나마 고장수리를 통해서 정상화 가능하지만 감시되지 않은 내재된 장애 이벤트로 인해 누적될 경우 심각한 장애로 이어질 수 있다는 것이다.



tbs 교통방송 빅데이터 기반 “스마트 모니터링 및 분석 시스템”

tbs 교통방송 스마트 모니터링 및 분석 시스템은 콘텐츠 생성, 저장, 전송, 송출 등 워크플로우 관점에서 시스템 성능 분석, 사전 장애감지, 다양한 패턴 분석을 통해서 문제점을 신속히 파악하여 VOC를 절감하고, 안정적인 방송 운영을 지원하기 위해 인제스트, NLE, 스토리지, CMS, 송출서버, 네트워크, DB 등 파일기반 제작시스템의 가장 중심에 있는 NPS를 구성하는 주요 방송시스템에 구축되어 있다. NPS를 구성하는 주요 방송시스템은 X86 기반의 서버 하드웨어와 Windows, Linux/Unix 등과 같은 OS, 운영 프로그램인 Application 등으로 구성되어 있다. 머신 데이터 분석을 위해 가장 먼저 시스템 성능 분석을 위한 SNMP Trip(MIB) 데이터와 콘텐츠 데이터 이력관리에 필요한 Syslog 데이터, 사전 장애감지를 하기 위한 Event log 데이터를 수집하고, 수집된 데이터 분석을 통해서 전체적인 서비스 상태와 작업통계를 일/주/월 단위로 세밀하게 확인할 수 있었다.

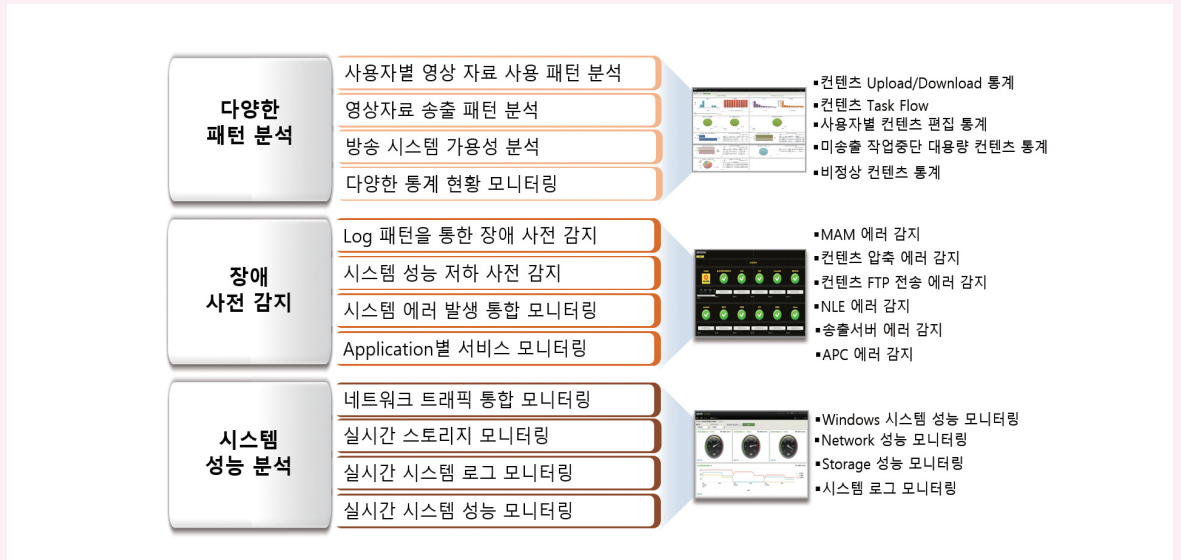


빅데이터 기반 스마트 모니터링 및 분석 시스템의 활용가치

대부분의 NPS 업무를 주관하는 CMS는 주요 관심 대상에서 최우선 순위로 관리되어야 한다. 제작된 영상을 인제스트 서버와 NLE를 통해서 편집된 완성본의 콘텐츠는 CMS의 Agent가 스토리지에서 송출서버로 전송하도록 제어한다. 이때 네트워크 스위치를 통해서 대용량의 콘텐츠 데이터는 FTP로 전달되는데, 스토리지와 송출서버 사이에서 안정되게 콘텐츠가 전송될 수 있도록 하는 중요한 역할을 한다. HD 제작 시스템 업그레이드 이전에는 경험하지 못한 다양한 장애를 유발하고 그 원인을 찾는 데 많은 시간을 소비하고, 운영 애플리케이션 소프트웨어 디버깅을 통해서 교정 작업을 수 차례 반복하였다.

수집된 머신 데이터를 토대로 장애유발에 영향을 미치는 의미 있는 데이터를 추출하여 인덱싱 후 실시간 모니터링 및 상관관계 분석을 한 결과, 원인분석과 해결방안 마련까지 그리 어렵지 않게 작업을 할 수 있었다.

tbs 교통방송 스마트 모니터링 및 분석 시스템은 먼저 적용된 NPS 외에도 보도정보시스템, 편성정보시스템, 문자발생기, 비디오 스위처, 오디오 믹서, APC, 키어, 압축다중화 시스템, UPS, 항온습습기, 기타 방송 시스템 등 방송 운영 전반에 걸쳐 수 많은 TV/라디오 시스템과 장비들에 적용하기 위한 시스템 구조로 설계하였다. 또한 서울지하철 구간에 서비스되고 있는 IPTV 시설 사전장애 분석과 웹/앱 서비스 관련된 시스템, 콘텐츠 소비 패턴분석 등 각 분야별로 별도구성 또는 통합 구성될 수 있도록 클러스터링 및 스케일 아웃 아키텍처 컴퓨팅 인프라 기술에도 역점을 두고 미리 설계된 시스템이다.



스마트 모니터링 및 분석 시스템 구축을 통해 개선된 점으로는, 어디가 문제인지 해당 시스템 로그를 즉시 확인할 수 있어서 고장수리 시간을 단축할 수 있었으며, 새로운 모니터링 및 분석이 필요한 운영 이슈에 따라 사용자가 직접 대시보드를 생성할 수 있을 정도로 쉽게 운영할 수 있다는 것이다. 잠재 장애 감지, 원인분석과 사전정비로 인해 KPI 향상과 월/분기/연간 사용패턴 트렌드 분석을 통해 추가 증설 지점을 파악할 수 있어서 ROI 개선에 많은 도움을 받을 수 있는 다양한 지표로의 활용 가능성을 기대할 수 있었다.

방송/미디어 관점에서 빅데이터를 활용하여 무언가를 하고자 하거나 정의하기 시작하면 접근방법 자체가 어려울뿐더러 시작도 하기 전에 쉬이 포기하고 말 것이다. 모든 데이터는 기업의 워크플로우에서부터 시작된다는 관점에서 대중화된 상용 빅데이터 통합 솔루션 도입을 통해서 우선 가법게 시작할 수 있는 부문부터 시작한다는 생각으로 방송/미디어 빅데이터 활용 방안을 찾아보는 편이 더 이해하기 쉽고 편할 것이다.

그동안 수많은 기업들이 보유한 데이터와 공개된 데이터를 활용하여 기업경쟁력을 높이기에 앞다투어 빅데이터에 대한 투자와 인력을 양성하고 있다. 그럼에도 불구하고 빅데이터를 활용한 사례를 통해서 우리 기업에 적용하기란 매우 어려운 현실이다. 이번 스마트 모니터링 및 분석 시스템 구축을 통해 구축을 고려하거나 관심 있는 모든 이들에게 작은 안내가 되었으면 하는 바람이다. 📖

