

빅데이터 플랫폼 요구사항 분석 - 신소연

1. 빅데이터란 무엇인가? (빅데이터의 정의 및 특징, 분석효과에 대해서 설명하시오.)

통신 기술이 발달하면서 비정형 데이터(SNS, LOG, Streaming, Video)가 폭발적으로 증가하면서 발생하게 되었다.

분석하기 어려운 데이터지만 IT가 발전하면서 다룰 수 있게 된 방대한 양의 데이터다. 비정형 데이터 중심이며 이를 분석하여 인사이트(시사점), 가치를 부여하는 새로운 기술이다.

3가지 특징 혹은 4가지 특징이 있다. 이를 3V 혹은 4V라 하고,
3V에는

규모(Volume) : 대량의 데이터

다양성(Variety) : 새로운 종류의 데이터와 분석, 비전통적 형태의 미디어, 소셜 미디어 데이터

속도(Velocity) : 실시간 정보, 신기술에서 유입되는 데이터

가 있고, 4V가 되려면 이 3가지에 가치(Value) 혹은 정확성(Veracity)를 추가하기도 한다.

정확성(Veracity)는 오류와 부정확한 데이터의 처리나 분석 목적과 다른 데이터를 처리하는 실수를 범하지 않아야 한다는 뜻이다.

분석효과에는 이전에 얻을 수 없었던 지식, 통찰, 식견과 시사점을 얻을 수 있고,
비즈니스 측면과 공공 서비스 측면 등 분야별로 서로 다른 가치를 제공할 수 있다.

2. 빅데이터의 비즈니스 적인 가치를 실제 사례를 들어 설명하시오.

빅데이터를 통해 고객 이탈 방지하는 사례이다.

고객 '홍길동'은 오랜 기간 A업체의 인터넷과 IPTV서비스를 이용했지만, 해당 IPTV에 골프 채널이 안 나오는 게 항상 불만이였다. 이에 홍길동은 콜센터에 불만을 표시하고, 이 빅데이터 아키텍처를 갖춘 기업은 음성 응답 정보를 수집하여 분석 후 서비스 팀과 협의해서 서비스 부서가 골프 채널을 무료로 시청하도록 제시했다. 홍길동은 새로운 서비스를 받는 것에 만족했고 이를 통해 홍길동의 서비스 해지를 방지했다.

3. 빅데이터 구축을 위한 방법(단계)에 대해 설명하시오.

구축방법론

수요 인식 및 목표 설정 -> 데이터 역량 검토 -> 도입 검토 및 사업 계획 -> 빅데이터 도입

수요 인식 및 목표 설정에서

수요 인식은 유행에 따라 우리도 한 번 해보자는 식의 접근은 피하고, 기업에서 기존 시스템을 수행하던 중, 새로운 지식이 필요할 경우 빅데이터의 수요를 인식 해야하고

목표 설정은 부서간 통합된 문제 제기과 협의 후 방향을 설정하고 방향 설정이 어려울 경우 전문 컨설팅 업체의 도움을 받는다.

데이터 역량 검토 시에는
문제 해결을 위한 해결 방안을 찾는 통찰력이 있는 데이터 분석가를 양성해 조직의 데이터 역량을 키워야 한다.

도입 검토 및 사업 계획시에는
기술적인 부문과 비용적인 부문으로 나누어지는데 오픈소스, 상용 패키지, 오픈소스 기반의 서드파티나 여러 빅데이터 기술을 다룰 수 있는 고급 인력이 존재하는지, 해당 기술을 도입할 수 있는 비용을 감당 할 수 있는지 확인하고, 데이터 수집할 때의 비용 또한 고려 후 전체 사업 계획을 세워야 한다.

빅데이터 도입
조직이 생각하는 방향과 실제 시스템이 일치하는지 파악하고, PoC를 도입하고자 할 때에는 최대한 위험을 회피하고 조정할 필요가 있다.

4. 데이터 유형과 수집방법에 대해서 설명하시오.

데이터의 유형에는 정형 데이터, 반정형 데이터, 비정형 데이터가 있고

정형 데이터 수집에는 RDB Aggregator, 관계형 데이터 베이스에서 정형데이터를 수집한다.

반정형 데이터 수집에는 Crawling, OpenAPI, RSS, Log Aggregator를 이용한다.

Crawling은 인터넷 상의 웹 문서와 정보를 수집하는 것으로 HTML의 형태이다.

Open API는 개방된 API로 수집하며 실시간 데이터의 형태이다.

RSS는 웹 기반 최신 정보 공유 서비스이며 XML의 형태이다.

Log Aggregator는 각종 로그 데이터를 수집하는 것이며 Log데이터 형태이다.

비정형 데이터 수집에는 FTP, STREAMING으로

FTP는 각종 파일 송수신으로 수집하고 FILE데이터이다.

STREAMING은 음성, 오디오, 비디오를 실시간으로 수집하는 것이며 WMV, MP3 형태이다.

인터넷에서 수집한다.

5. 빅데이터 구축의 핵심요소에 대해서 설명하시오.

빅데이터 구축의 핵심요소에는 자원, 기술, 인력이 있으며

자원은 조직 목표에 맞는 데이터 활용을 위한 데이터 발견이며

기술에는 빅데이터 처리를 위한 수집, 저장, 변환, 분석, 시각화 기술이 있고

인력에는 빅데이터에서 통찰과 인사이트를 찾는 데이터 과학자나 전문가를 말한다.