

“성공적인 빅데이터 구축과 활용을 위하여”

빅데이터의 일반적인 이슈들

장동인

한국 테라데이터

기업 내에서 빅데이터를 바라보는 관점이 다르다

- 빅데이터는 누르면 나오는거 아냐?
- 아랫사람 시키면 되지.
- CIO가 TF팀 맡으세요.



경영자



- 전문가 없어요...
- 진짜 성공사례도 없어요...
- 우리 회사는 데이터도 크지 않아요...
- 예산도 없어요....
- 그런데 빅데이터가 뭐예요?
- 우리 회사의 어떤 분야에 적용해야 하죠?

현업담당자



- 빅데이터가 뭐길래 이렇게 떠들지?
- IT에서 알아서 하겠지. 내일 아니니까.
- 분석 같은 것은 outsourcing 하면 되지

...

IT 담당자



- TF팀은 말았는데 어떻게 하지?
- Vendor불러서 들어보자.
- BMT, POC 시키자.
- Hadoop 이 뭐야? 꼭 배워야해?

관점의 차이는 커다란 문제를 가져온다

관점의
차이



- 일단 검토와 공부는 하는데, 부서간에 대화가 안된다.
- 따라서, 빅데이터 사업에 대한 진행이 어렵다
- 서로 모르는 상황에서 목소리 큰사람이 이긴다
- 한번 도입하면 끝나는 One time 프로젝트라고 생각한다.
- TF팀을 맡은 IT 부서는 머리를 싸매고 고민만 한다
- 경영진은 도입 지시를 하면 일이 끝나는 줄 안다
- 어렵게 프로젝트에 들어가도 부서간 협조가 안된다
- 특히 현업 부서는 남의 일이라고 생각한다
- 중간 보고 할 때 쫄 되면, 실망하기 시작하고 마지막에는 '이건 아니다' 라고 생각한다
- 빅데이터 도입 완료는 했으나 쓰는 부서가 없다
- 빅데이터에 대한 회의론이 일기 시작한다

이 상황에서 빅데이터 프로젝트 실패 가능성
90% 이상!

경영진의 생각



- 빅데이터를 너무 쉽게 생각한다. (TV, 매체 영향)
- 빅데이터에 대한 공부를 안한다. (시키면 되니까)
- 의사결정은 자신의 감에 의해서 한다. 중요한 것은 실적이고 숫자이다! (지금까지 그래왔으니까)

➔ 당연히 실무자들의 관심사는 오직 실적!

➔ 문제가 발생하면 왜 그런지 분석하라는 요구 없음

- 경영진 중에서 빅데이터를 제일 많이 아는 것 같은 임원 = CIO.
- 따라서 빅데이터 TF팀장은 당연히 CIO.
- 그리고 부탁. "돈 많이 안드는 방향으로 하세요"

무엇이 문제 일까요?

경영진에게 빅데이터는.... Fact based decision making !!



- 빅데이터를 도입하려고 하는 기업은... 최소한 데이터에 근거한 의사결정을 하는 것이 정착되어야 한다.
 - ✓ 실적이 나오지 않는 근본적인 이유를 데이터에서 찾는다.
- “시키니까 한다” 광역버스 입석 금지(2014.7.16)
 - ✓ 기본적인 분석도 없이 그대로 시행. 대혼란
 - ✓ 사후에 분석한 결과. (2014.8.11)

“이날 오전 6~9시 서울 방향 광역버스를 분석한 결과, 오는 9월부터 584회가 운행(404대 증차 기준) 돼야 출근 대란을 피하고 좌석제가 원활히 추진될 것으로 예측됐다”고 밝혔다. 이에 따른 소요 예산은 6개월간 216억원(대당 3천900만원)이 필요할 것으로 내다봤다. 이번 예측은 도가 버스운송관리시스템(BMS)과 교통카드를 분석하고 9월 승객 증가량을 적용해 산출했다(2014.8.11 보도)

경영진은 직원에게 끊임없이 원인분석을 지시



- 근본 원인 분석
- 미래에 대한 예측
- 예측치에 대한 대책 강구
- 효과 체험
- 좀 더 정교한 근본 원인 분석과 예측 모델 업그레이드



Fact based 경영의 선순환 구조

이 상황에서 빅데이터 프로젝트 성공 가능성
90% 이상!

경영진의 마인드가 빅데이터 성공의 결정적 영향!



지시



신규TF



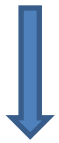
- Fact based 경영이 정착화 되기 전에는 빅데이터는 성공할 수 없다.
- 빅데이터 프로젝트를 시작하기전에 fact based decision making 분위기 조성이 먼저.
- 통계, 원인 분석, 대처 방안이 없는 보고서는 던져라
- 직원들에게 통계와 원인분석, 예측 방법을 교육 시켜라
- 경영진 스스로도 기본 통계와 원인분석, 예측 방법을 배워야 한다
- 사람은 오고 갈 수 있으나, 결국 "원인분석-예측모델"이 기업의 최고의 자산으로 남는다.

여기 까지 오는데 드는 비용 ZERO !!

빅데이터를 도입하는데 처음부터 빅머니가 필요하지는 않다



지시



신규TF



- 경영자의 고민. 빅데이터 = 빅머니
- 실제로 초기에는 직원 및 임원 교육비만 필요하다
- 데이터를 분석하는 인프라는 기존 것을 활용할 수도 있고, 클라우드를 활용할 수도 있다.
- 통계분석, 예측을 위한 소프트웨어는 Excel도 가능하고 오픈 소프트웨어인 R 은 매우 좋다.
- 소셜미디어 분석에는 네버트렌드, 소셜매트릭스 등 무료툴을 사용한다.
- 직원의 분석 능력을 신속하게 증진시키는 방법을 강구한다.

직원의 분석능력을 신속하게 증진시키는 방법
은 분석 워크샾을 하는 것이다

현업 담당자들의 생각



- 빅데이터는 내일이 아니다. TF를 맡은 IT팀의 일이다.
- 데이터 분석은 내 업무가 아니다.
- 빅데이터 분석? 그건 뭐지?
- 정 필요하면 아웃소싱하면 된다
- 내 업무는 숫자와 성과를 챙기는 것이다.
- 빅데이터 TF팀에 있기는 하나, 뭘 해야하는지 잘 모르겠다.
- 빅데이터 테마? 내가 빅데이터를 모르는데 어떻게 빅데이터 테마를 선정할 수 있나?
- 빅데이터도 얼마간 떠들다가 시간이 지나면 사라질거야.

현업 담당자들의 생각을 바꾸는 것!

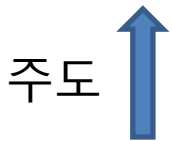


신규TF



- 현업 담당자들의 생각을 바꾸는 유일한 방법은 경영자들의 “통계분석, 원인분석, 대책”이 담겨있는 보고서를 요구하는 것.
- 점점 강도가 세지면, 현업담당자들은 할 수 없이 “통계/데이터마이닝”을 배울 수 밖에 없다
- 데이터 분석능력은 기업의 자산(core competency)이다.
 - ✓ 업무를 모르면 데이터 분석을 할 수 없다.
 - ✓ 업무 내용은 끊임없이 변한다.
 - ✓ 분석 방향은 수시로 변할 수 있다.
 - ✓ 따라서 데이터 분석은 외부의 도움은 받을 수 있으나 내부 자산화를 해야 한다. (“원인분석-예측 모델”)
- 스스로 데이터 분석 능력을 키우기 위해서는 “현업담당자들을 위한 분석 워크샷”을 하는 것이 좋다

현업 담당자들을 위한 분석 워크샷



신규TF



- “현업담당자들을 위한 분석 워크샷” 이 필요하다
 - ✓ 팀별로 분석가를 차출
 - ✓ 자신의 일을 하면서 오후시간 할애
 - ✓ 팀의 일 중에서 하나를 골라서 데이터 분석
 - ✓ 분석전문가를 초빙해서 통계/분석/예측 방법에 대한 교육을 받은 후에
 - ✓ 담당 조교를 배정해서 분석하는 방법을 코칭하도록 함
 - ✓ 1주일에 한번, 전문가의 코멘트
 - ✓ 2~3개월 후에, 최종보고
- 이것이 잘 수행 되다면 빅데이터에 대한 테마와 전체 마스터 플랜도 가능하다. (컨설팅업체 필요)

IT 담당자들의 생각



- 대부분의 기업 및 기관의 IT부서가 빅데이터 TF를 리드 한다.
 - ✓ IT부서가 제일 많이 안다고 생각한다.
 - ✓ 실제로는 빅데이터에 대해서 아는 것이 없다
 - ✓ Hadoop도 잘 모른다. 그리고 배우기를 꺼려 한다.
- IT벤더를 불러서 빅데이터에 대한 설명과 사례를 듣는다
 - ✓ 이해는 가나, 실제 현업 담당자들에게 타당성을 설득하기 힘들다.
 - ✓ 현업부서에서 sponsor를 해야 프로젝트가 시작되는 구조이다
- 벤더들에게 POC(Proof of Concept)를 해달라고 한다
 - ✓ 실제 데이터로 하기에는 어려움(데이터 보안)
 - ✓ 데이터를 generation
 - ✓ 결과?

IT 담당자들의 생각

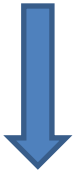


- 벤더 선정을 위한 BMT(BenchMark Test)
 - ✓ 무한 경쟁, 엄청난 비용
 - ✓ 결론은 테스트 분야별로 얼마나 속도가 나오는가?
 - ✓ 이것이 빅데이터 성공에 얼마나 중요할까?
- 빅데이터 테마선정, 분석능력 강화, 활용 방안 등 중요한 분야는 IT담당자들은 잘 모른다
 - ✓ 이것이 빅데이터 성공의 중요한 요소인데... (POC, BMT 등으로 시간, 노력을 소비)
 - ✓ 현업부서는 이 부분에 대한 명확한 리더십을 원하는데, 답변이 시원하지 않다
- 일은 하고 있되 총체적 난국 상황...

먼저 빅데이터 TF 리더십을 기획부서에 넘겨라



주도권
이양



신규TF



- 빅데이터는 IT 프로젝트가 아니다
 - ✓ 기업의 경영층, 현업부서, IT부서등 전체 부서가 연관된 프로젝트이다
 - ✓ 조직 역학상 IT부서가 이를 조정할 수 없다
 - ✓ 따라서, 기획본부 또는 사장 직할로 팀을 만들어야 한다.
- 빅데이터 TF팀의 역할
 - ✓ Fact based 경영 방안, 빅데이터 테마 선정, 기업내 분석능력 강화, 빅데이터 활용 방안 및 변화관리등을 주도해야 한다
 - ✓ “현업분석 능력강화를 위한 워크샷” 주도
 - ✓ 이를 위해서는 당연히 마스터 플랜이 필요하다
- IT팀의 역할
 - ✓ 데이터 거버넌스, 표준화, 빅데이터 인프라설계 및 튜닝, 확장방안
 - ✓ 빅데이터 기술 획득 및 데이터 처리, 요약, 통합 방안 수립

IT담당자들은 빅데이터 IT 기술을 익히는데 주력하라



- Hadoop 및 연관기술을 습득
 - ✓ Hadoop, MR, Hive, Pig, Zookeeper, Flume, etc...
 - ✓ NoSQL계열 DB: Mongo DB, Hbase, Cassandra DB
 - ✓ 90년대 초반 RDB와 SQL을 배워야 살았듯이 지금은 Hadoop을 배워야 한다.
 - ✓ SQL on Hadoop 을 중점적으로. (현재 추세는 Hadoop을 Full SQL을 사용할 수 있도록 가는 방향이다.)
 - ✓ RDB 처럼 Hadoop Admin을 할 수 있어야 한다
- 데이터 통합(data integration) 방법을 배워야 한다
 - ✓ R 과 Hadoop을 연결하는 방법
 - ✓ 기존 RDB와 Hadoop을 연결하는 방법
 - ✓ Hadoop과 ODBC/JDBC를 사용하여 기존 BI too
과 연결하는 방법

신규TF: 빅데이터는 one-time project가 아니다



- One-time project
 - ✓ 한번 도입하고 끝나는 프로젝트
 - ✓ 예산 세우고 구축하고, 그 다음 해부터 유지보수하는 프로젝트
- 빅데이터는 계속 하는 프로젝트이다.
 - ✓ 기업의 데이터는 계속 생기고, 적재하고, 가공해야 한다.
 - ✓ 빅데이터의 새로운 테마를 개발하고, 계속 분석하고, 예측해야 한다
 - ✓ 새로운 데이터가 계속 입력이 되고, 적재가 된다
 - ✓ 인프라도 서서히 늘려가야 한다
 - ✓ 분석 결과를 활용하고, 그 활용 현장에서 습득한 노하우가 분석에 다시 적용 되어야 한다.
 - ✓ 빅데이터 이것의 반복이다.
- 따라서 빅데이터는 변화관리가 생명이다.
- 주체는 신규TF에서

신규TF: 빅데이터의 수준은 전문가의 수준이다



- 빅데이터 프로젝트

- ✓ 패키지가 아니다. (참고: ERP프로젝트는 패키지내부의 프로세스를 이해하고 이를 따라가면 성공)
- ✓ 정해진 왕도가 없다.
- ✓ 전문가의 경험과 수준에 의해서 결정될 수 밖에 없다.
- ✓ 전체적인 로드맵을 가지고 각 팀이 소통하는 것이 중요하다.
- ✓ 빅데이터 추진 조직의 효율성, 프로젝트 추진 방법론, 데이터 거버넌스, 변화 관리 등이 중요해진다
- ✓ 사람을 키우고, 외부 전문가들, 외부 파트너들의 협조가 핵심 성공요인이다.

신규TF: 사내 빅데이터 전문가 양성



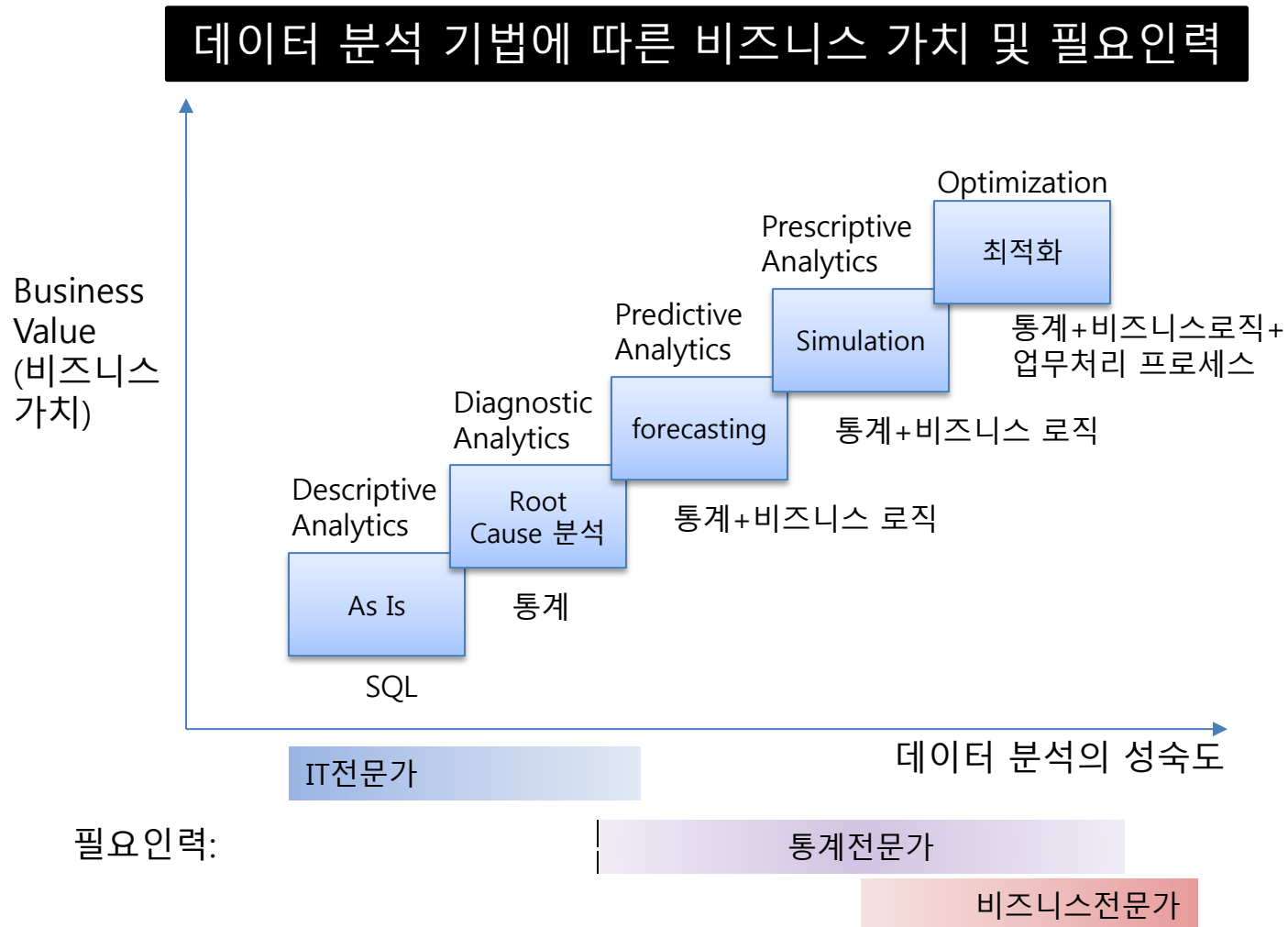
- 빅데이터 전문가는 하나가 아니다
 - ✓ 빅데이터 IT전문가
 - ✓ 빅데이터 분석 전문가
 - ✓ 소셜미디어 분석 전문가
 - ✓ 빅데이터 도메인 전문가(산업 또는 업무)
 - ✓ 빅데이터 전략전문가
 - ✓ 시각화 전문가
- 내부 빅데이터 전문가와 outsourcing
 - ✓ 전문가 양성을 위한 내부/외부교육
 - ✓ 인재 확보
 - ✓ 외부 전문가 풀과 제휴
 - ✓ 외부 파트너기업과 제휴

의사결정 !!

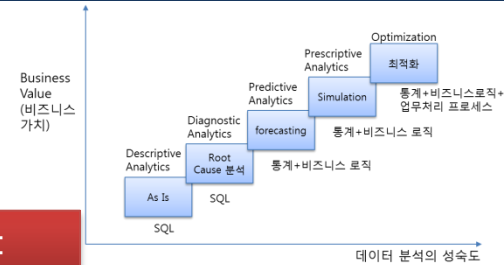
사람들은 의사결정을 할 때 (예: 물건을 살 때),
무의식적으로 어떤 과정을 거치는가?

기업은 의사결정을 할 때 (예: 매출 목표 달성을 위한 영업전략 수립),
무의식적으로 어떤 과정을 거치는가?

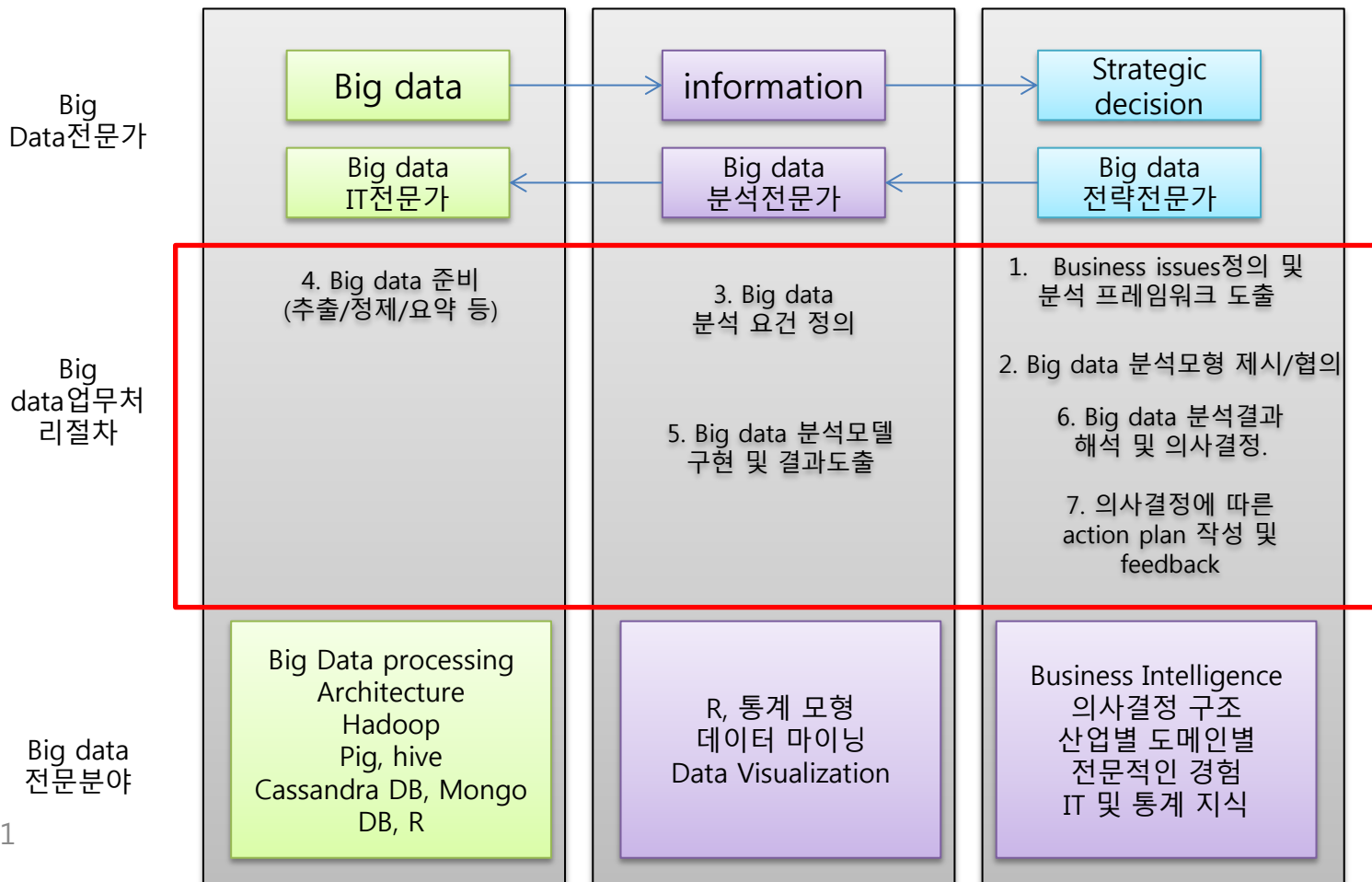
분석의 성숙도에 따른 비즈니스 가치



Big Data 업무 프로세스

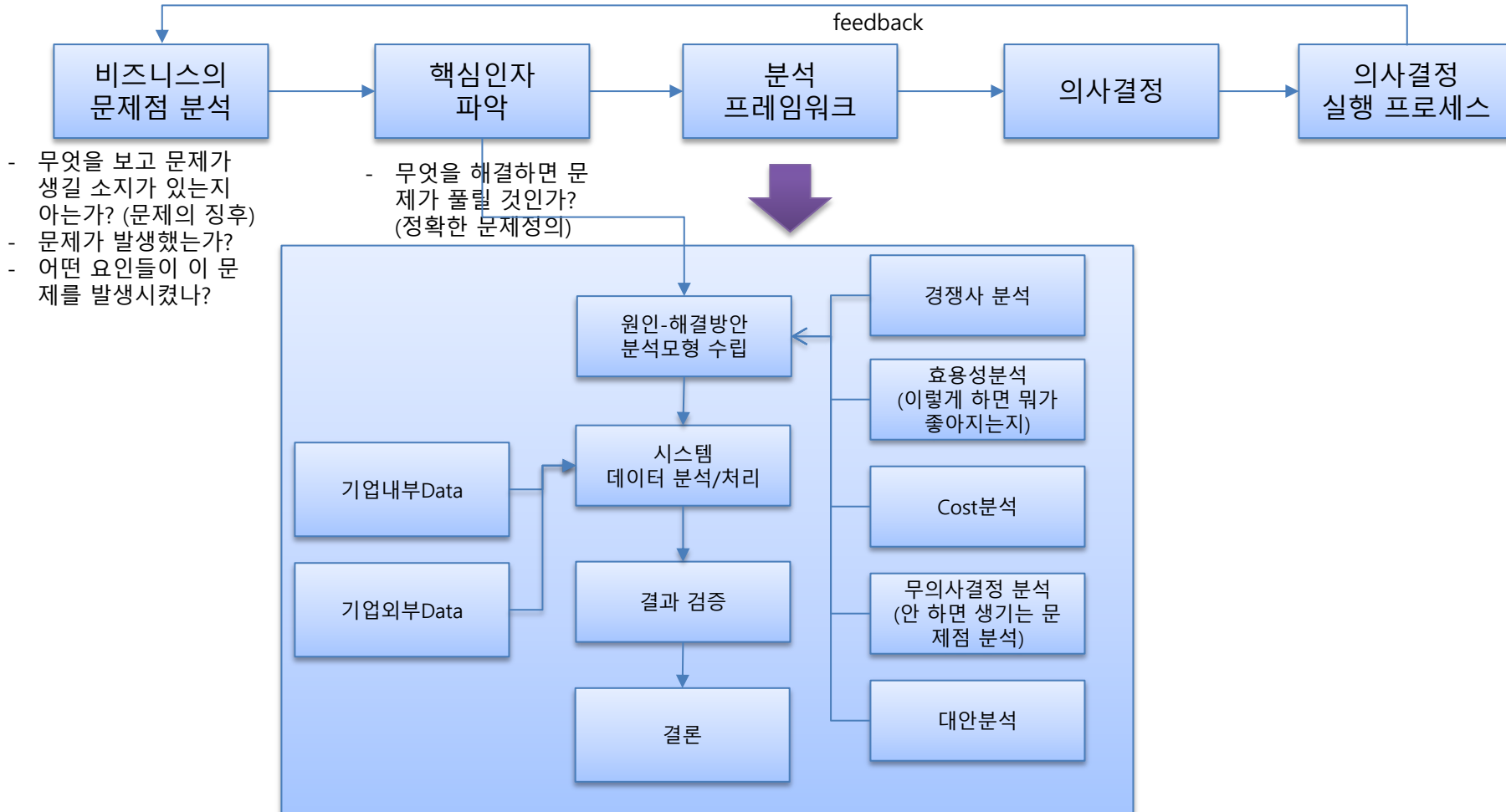


Big data 분석을 위한 전문가의 역할과 분야



"Data Driven Decision Making Process"

분석프레임워크

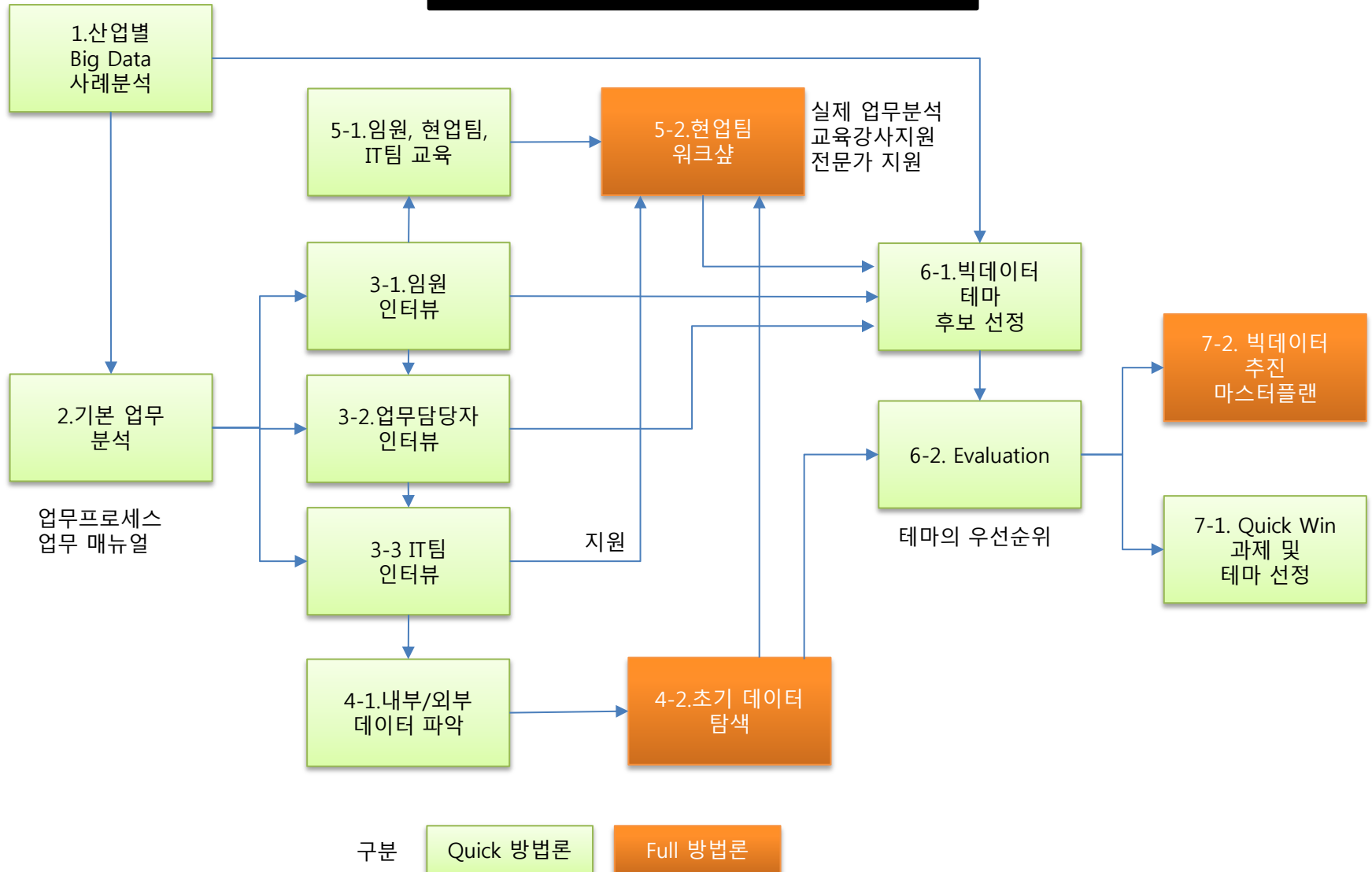


Big Data를 활용한 의사결정 마스터 플랜 수립

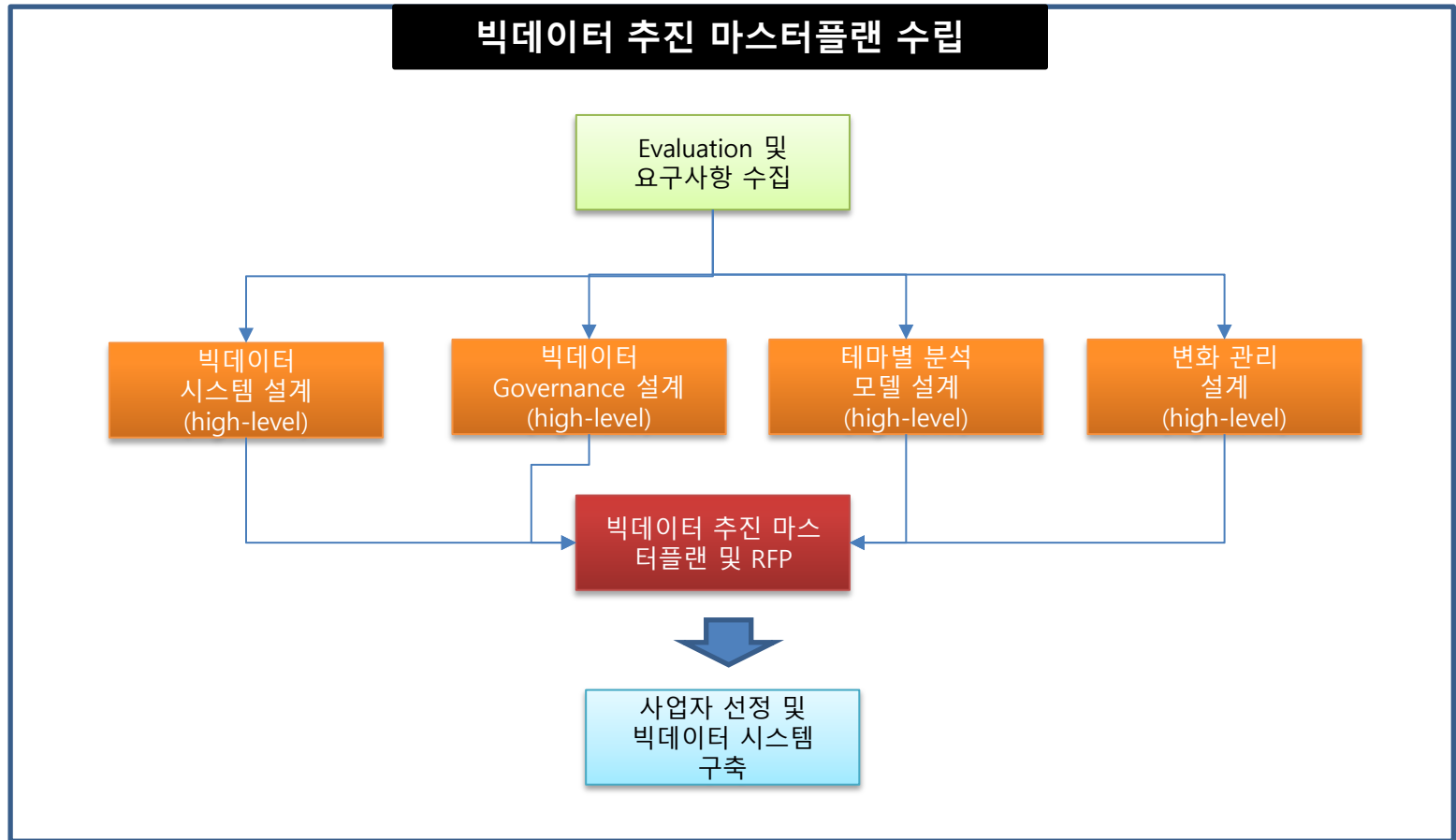


Big Data 테마 도출 방법론

빅데이터 테마 도출 방법론



빅데이터 추진 마스터 플랜 수립



빅데이터 성공을 위하여



- 데이터에 기반을 둔 의사결정 체계
- TF팀은 사장 직속 또는 기획본부에
- 보고서의 빅데이터화 (“통계-원인분석-예측 보고서”)
- 현업부서의 분석능력 강화
- 현업부서의 2~3개월간 워크샵
- IT부서의 빅데이터 기술 습득
- 전문가의 중요성 이해
- 인재 양성/내부외부 교육 프로그램
- 외부전문가/전문기업과 파트너쉽
- 지속적인 활용을 위한 변화관리 주력

해외컨퍼런스 소개 1: 최신 빅데이터 분석 및 IT기술 동향

Strata + HADOOP
CONFERENCE + WORLD
TOOLS AND TECHNIQUES THAT MAKE DATA WORK

2014 STRATA + HADOOP WORLD

대한민국 연수단 일정 (5박 6일)

- 주 최 • 매경미디어그룹 매경닷컴, 한국생산성본부
- 일 정 • 2014년 10월 14일(화) ~ 19일(일) [5박 6일]
- 국 가 • 미 국 (뉴욕)
- 교육대상 • 기업 Big data 분석, 개발, 설계 관련 임직원 • 기업 경영전략, 기획, 경영혁신 관련 임직원
- IT 정책 및 운영관련 중앙정부, 지자체 및 공공기관 담당자

○ 특 장 점

컨퍼런스 후 Wrap-up 세미나

컨퍼런스 기간 동안 매일 저녁 주요 내용에 대한 Wrap-up 세미나 개최



코디네이팅 전문가
장동인 대표

- 현 빅데이터 전문가 협의회 의장
- SAS Korea 부사장
- Oracle Korea 컨설팅 본부 이사
- 삼성, SKKT, 포스코 등 빅데이터 및 CRM 컨설팅수행



연수 결과 보고서 제공

연수 후 컨퍼런스의 핵심내용을 정리한 연수 결과 보고서를 제공

정규 교육 이수 수료증 발급

- 정규교육 이수로 인정하여 한국생산성본부 수료증 발급
- 매경 Big Data Forum 2014 참가특전



○ 교육비용

| 1인 기준 | 2인 이상 기준 | 기타 |
|-------|----------|---|
| 690만원 | 670만원 | * 연수관련 주요 비용 포함 (비자 및 여권발급비, 개인경비는 불포함 VAT 별도) |

해외컨퍼런스 소개 2: 빅데이터 성공사례를 고객이 직접 발표



- The Teradata Partners Conference 2014
- Oct 19-23, Nashville, TN

한국 테라데이타
장동인

Don.chang@teradata.com

donchang@hanmail.net

010-5259-9509

Facebook: <http://www.facebook.com/jang.cloud>