2019 BIGDATA SHOWCASE S8.

GCP를 이용한 빅데이터 파이프라인 설계.

클라우드 컨설팅팀 권수용
sykwon@rockplace.co.kr
2019.05.23

rockPLACE

Copyright ⓒ 2019 Rockplace Inc. Strategic Planning Team All right Reserved

rockPLACE

목차

1. 시작하기에 앞서...
2. 파이프라인?
3. 파이프라인 설계를 위한 Step 1
4. 파이프라인 설계를 위한 Step 2
5. 파이프라인 설계를 위한 Step 3
6. 파이프라인 설계를 위한 Step 4
7. 파이프라인 설계를 위한 Step 5
8. 파이프라인 설계 시 놓치기 쉬운 필수 고려사항
9. 마치는 글

Copyright ⓒ All right Reserved by 2019 Rockplace Inc.

시작하기에 앞서...

OpenSource PLACE, ROCKPLACE

rockPLACE

GCP는 타 클라우드 사업자의 서비스와 가장 큰 차이점으로 서비스의 독립성이 강하다는 것 입니다..

- GCP는 각각의 서비스의 독립성이 높으며, 이로 인한 다양한 조합 및 결합이 가능합니다.
- 단 어떻게 적용하느냐에 따라 적용 방법이 수 없이 나오고 그 효과가 다 다르므로 충분한 사전 지식이 필요합니다.
- 그렇기 때문에 파이프라인 설계 시 더 많은 고민과 노력을 기획 단계에서 갖추고 진행하는 것이 무엇보다 중요합니다.

Copyright ⓒ All right Reserved by 2019 Rockplace Inc.

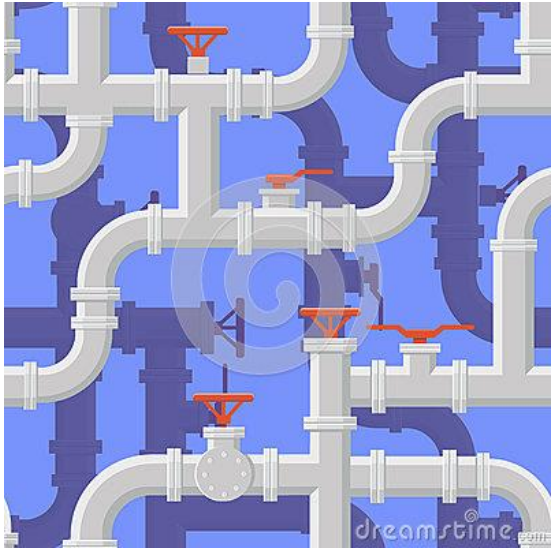
파이프라인?

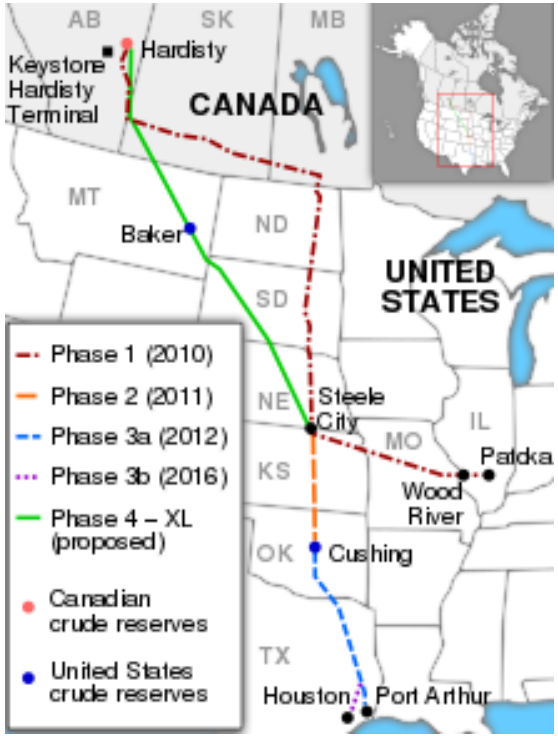
OpenSource PLACE, ROCKPLACE

rockPLACE

1. 사전적 정의

컴퓨터 과학에서 파이프라인(영어: pipeline)은 한 데이터 처리 단계의 출력이 다음 단계의 입력으로 이어지는 형태로 연결된 구조를 가리킨다.(by Wikipedia)





Copyright ⓒ All right Reserved by 2019 Rockplace Inc.

파이프라인?

OpenSource PLACE, ROCKPLACE

rockPLACE

2. 파이프라인의 중요성

파이프라인이 고려되지 않고 진행된 프로젝트는 운영과 유지보수 및 필요에 따른 변경에 절대적으로 취약해질 수 밖에 없음.

전통적인 형태의 빅데이터 구축 사업의 맹점.

1. 년 단위의 예산 책정 및 집행, 구축 프로세스를 따르는 중/대형 SI 형태의 빅데이터 사업으로는 목표한 빅데이터 시스템 구축 및 운영을 위한 명확한 예산 및 기술을 정의하기 힘듦.

2. 사업 진행중 추가로 발생하는 내/외부적 요구사항에 따른 유연한 대응을 하기가 거의 불가능함.

3. 기술 내재화, 시스템 유지보수, 시스템 확장의 어려움을 내포하고 있음.

Copyright ⓒ All right Reserved by 2019 Rockplace Inc.

파이프라인?

OpenSource PLACE, ROCKPLACE

rockPLACE

3. 빅데이터 파이프라인 설계시 사전 고려사항 (권장사항)

파이프라인 설계시 파이프라인에 적용할 기술적인 요소보다 기반이 되는 부분에 대한 점검 및 사전 준비가 더욱 중요합니다.

- ✓ **최초 도입단계에서는 이용할 기술을 예상하되 확정하지는 말 것을 권장합니다.**
  - ➢ 미리 정해놓을 것이 많을 수록 구현은 쉬워지지만 활용은 점점 어려워 집니다.
- ✓ **IT 조직에서 담당 및 처리 가능한 범위를 확인하고 진행하는 것을 권장합니다.**
  - ➢ 내재화를 이루기 위함입니다.
- ✓ **언제나 운영환경이 변경될 수 있음을 가정하고 진행하시는 것을 권장합니다.**
  - ➢ 비즈니스 및 사회적 상황에 따라 언제든지 운영 환경이 변경될 수 있기 때문입니다.
- ✓ **종속 기술을 최소화 하는 것에 집중하며, 반드시 필요한 부분에는 종속을 과감히 수용하여야 합니다.**
  - ➢ 선택과 집중을 통한 프로젝트 성공
- ✓ **언제든지 파이프라인 및 구성요소는 변경될 수 있음을 이해하고 진행하여야 합니다.**
  - ➢ 비즈니스 및 활용에 집중
- ✓ **수행 기간을 최대한 길게 설정하시고 각 Step 단계를 꼼꼼히 진행하시길 권장합니다.**
  - ➢ 구현보다 거버넌스, 표준화, 분석/설계에 집중할 수록 실제 비즈니스 및 IT 운영에 도움이 됩니다.

Copyright ⓒ All right Reserved by 2019 Rockplace Inc.

파이프라인 설계를 위한 Step 1 (1/3)

OpenSource PLACE, ROCKPLACE

rockPLACE

1. 목표 설정

- ✓ 기술/방법/도구/비용 등을 고려하지 말 것.
- ✓ 무조건 기획, 설계단을 길게 볼 것.(최소 3개월)
- ✓ 조직 또는 이용자로써의 목적을 위한 목표를 추정
- ✓ 구현이 가능할지 불가능할지는 현 단계에서 따지지 말 것
- ✓ 1개의 목표는 절대 안됨. 최대한 많이, 여러 이해 관계자와 함께 고민할 것
- ✓ 기존 유사 시스템을 보유하고 있을 경우 목표설정보다는 As-Is 파악이 선행필수

Copyright ⓒ All right Reserved by 2019 Rockplace Inc.

파이프라인 설계를 위한 Step 1 (2/3)

OpenSource PLACE, ROCKPLACE

rockPLACE

2. As-Is 파악

- ✓ **기존 유사 시스템 보유 여부 확인.**
- ✓ **원천 데이터 생성 부분에 대한 현황 파악.**
- ✓ **내부 IT 조직의 구조 및 구성원의 Skill Set 확인.**
- ✓ **가용 가능 예산 및 기술 로드맵 (최소 3년치) 확인**
- ✓ **기존 유사 시스템을 보유하고 있을 경우 목표설정보다는 As-Is 파악이 선행필수**

Copyright ⓒ All right Reserved by 2019 Rockplace Inc.

파이프라인 설계를 위한 Step 1 (3/3)

OpenSource PLACE, ROCKPLACE

rockPLACE

3. 내재화 / 외주 운영방안 설계

- ✓ **빅데이터 및 ML 시스템의 경우 내재화가 최우선 고려사항이 되어야 함.**
- ✓ **내재화 / 외주에 대한 범위 파악**
- ✓ **내재화 / 외주에 대한 기본 역할 및 목표 설정 (*주 : 상세가 아님)**
- ✓ **내재화 / 외주에 대한 기술 습득 및 전수방안 설정**

Copyright ⓒ All right Reserved by 2019 Rockplace Inc.

파이프라인 설계를 위한 Step 2 (1/3)

OpenSource PLACE, ROCKPLACE

rockPLACE

1. 빅데이터 구축 방법론 정의

구분	Traditional	예:) Zero Based
예산	Total 산정최종 목표를 이루기 위한 HW/SW/Dev 비용	Case By Case 단위 목표별 HW/SW/Dev 비용 산정
수행절차	사업 구성단계 All Fix 사업 진행 중 변경 불가능	Agile Process 사업 진행 중 유연한 변경 가능
이용기술	사업 구성단계 All Fix 사업 진행 중 변경 불가능	각각의 목표별 기술 정의 사업 진행 중 유연한 변경 가능 ** Platform 구성기술 사전 Fix 필요
유지보수	외주(수행) 업체 중심	내재화 중심
확장	유연성이 부족함	확장 최적화
최우선 필요요건	대규모의 사전 사업예산	기업의 IT 정책 및 문화, 역할 변경에 따른 포용성

Copyright ⓒ All right Reserved by 2019 Rockplace Inc.

파이프라인 설계를 위한 Step 2 (2/3)

OpenSource PLACE, ROCKPLACE

rockPLACE

2. 빅데이터 구축 방법론 특장점 명세화 및 공유 – 이후 모든 프로세스에 기반이 됨

- ✓ **On-Premises, Public Cloud, Private Cloud, Hybrid Cloud, Virtualizaion 구분 없는 적용 가능**
  - ➢ **시스템 형태의 제약을 받지 않음.**
- ✓ **기업 및 조직이 보유한 인력, 예산에 따라 유동적인 절차 및 범위의 설계 가능.**
  - ➢ **상황별 최적화 방안 수립 및 때론 작게 따론 크게. 기업 및 조직의 요구 반영 가능.**
- ✓ **각 사업을 Cycle화 할 수 있으며, Cycle을 수행할 수록 기업 및 조직의 내재화 강화**
  - ➢ **시스템의 기술 및 운영, 관리 내재화**
- ✓ **각 빅데이터 사업은 초기에는 상호 독립적이나, 이를 연계 및 활용할 수 있는 소규모 사업 또는 프로젝트를 통해 빅데이터 시스템의 무한 확장 가능.**
  - ➢ **확장성**

Copyright ⓒ All right Reserved by 2019 Rockplace Inc.

파이프라인 설계를 위한 Step 2 (3/3)

OpenSource PLACE, ROCKPLACE

rockPLACE

3. Public Cloud 도입에 대한 특장점 분석 : 예) On GCP

- ✓ **개발, 테스트, PoC 시스템의 운영계 전환이 손쉬움.**
  - ➢ **전환 간편화, 비용 최적화**

- ✓ **모든 서비스, 플랫폼을 직접 개발, 구현할 필요가 없음.**
  - ➢ **대부분의 Big Data Workflow 서비스를 GCP에서 제공**
  - ➢ **DataProc(Hadoop), DataFlow(Pre/Post Processing),BigQuery(DW), DataStudio(BI)**
  - ➢ **개발기간 및 플랫폼/솔루션 구축 부담 최소화, 운영 및 관리 자동화**

- ✓ **구글의 글로벌 네트웍을 이용한 전세계 수준의 빅데이터 시스템 구축 가능**
  - ➢ **지역에 국한되지 않는 빅데이터 시스템 구축**

- ✓ **각 빅데이터 사업은 초기에는 상호 독립적이나, 이를 연계 및 활용할 수 있는 소규모 사업 또는 프로젝트를 통해 빅데이터 시스템의 무한 확장 가능.**
  - ➢ **Google Cloud의 대부분의 기술은 Open Source 기반 기술을 사용함.**
  - ➢ **OpenSource 기술을 통해 보다 신속하고 편리한 확장이 가능하며, 필요시 타 시스템 이전 가능**

Copyright ⓒ All right Reserved by 2019 Rockplace Inc.

파이프라인 설계를 위한 Step 3 (1/4)

OpenSource PLACE, ROCKPLACE

rockPLACE

1. Cloud Provider 선정

In Memory	Relational		NoSQL		Object	Block	Warehouse
Cloud Memorystore	Cloud SQL	Cloud Spanner	Cloud Datastore	Cloud Bigtable	Cloud Storage	Persistent Disk (GCE)	Big Query
Good for: Web/mobile apps, g aming	**Good for:** Web framew orks	**Good for:** RDBMS+scale, HA, HTAP	**Good for:** Hierarchical,m obile, web	**Good for:** Heavy read + write, events,	**Good for:** Binary or object d ata (BLOB)	**Good for:** Local VM file storage	**Good for:** Data Warehou se
Such as: Game state, user ses sions	**Such as:** CMS, eCom merce	**Such as:** User metadata, Ad/Fin/MarTech	**Such as:** User profiles, Game State	**Such as:** AdTech, Fina ncial, IoT	**Such as:** Media, analytics, a rchive/backup	**Such as:** Application da ta/binaries	**Such as:** Analytics, Da shboards

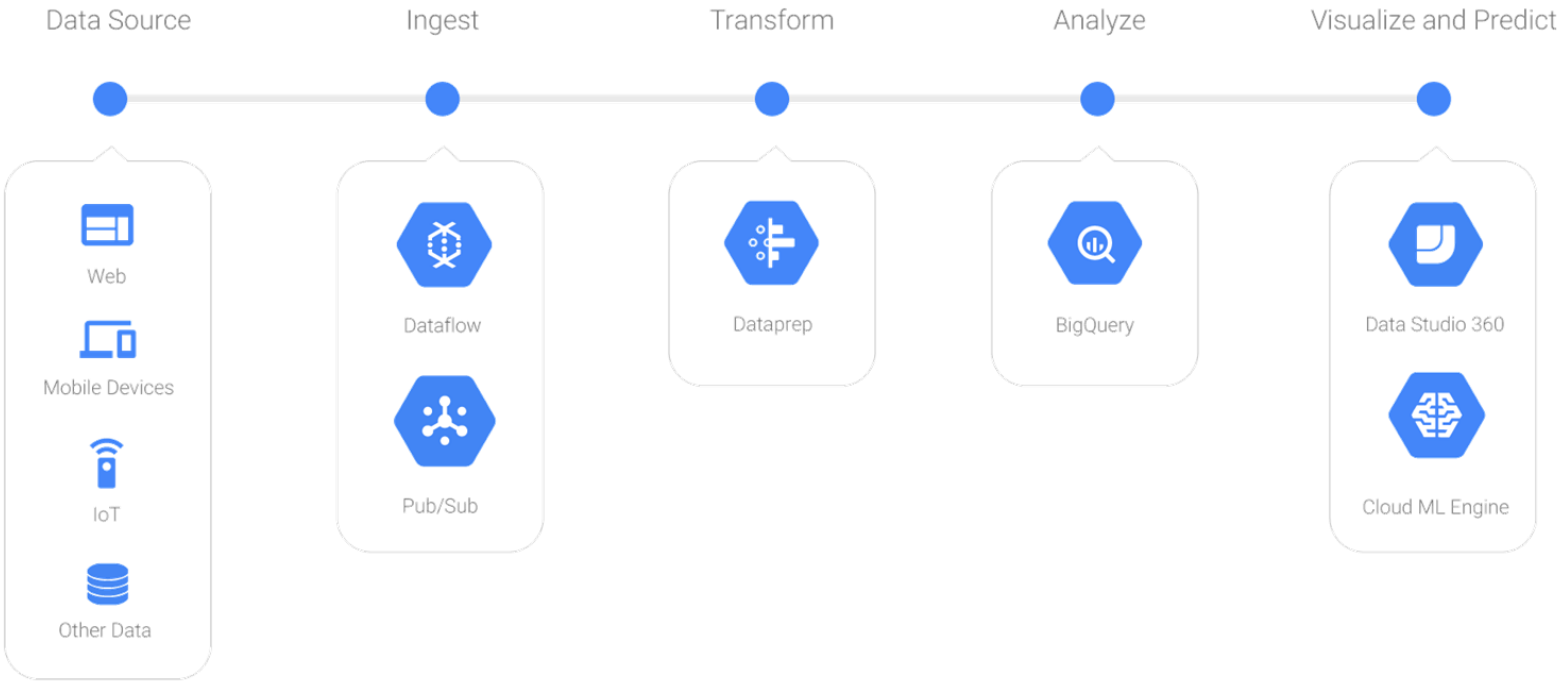
Copyright ⓒ All right Reserved by 2019 Rockplace Inc.

파이프라인 설계를 위한 Step 3 (2/4)

OpenSource PLACE, ROCKPLACE

rockPLACE

2. Cloud Provider의 주요 빅데이터 파이프라인 구성 및 장/단점 파악
 - 기술 제약요소, 종속요소, 실시간 및 배치 프로세스 적용에 따른 변동사항, 성능 임계치 등의 확인이 필수.
 - 놓치기 쉬운 Cost 구간에 대한 리스크 요소 사전 도출이 중요함



Copyright ⓒ All right Reserved by 2019 Rockplace Inc.

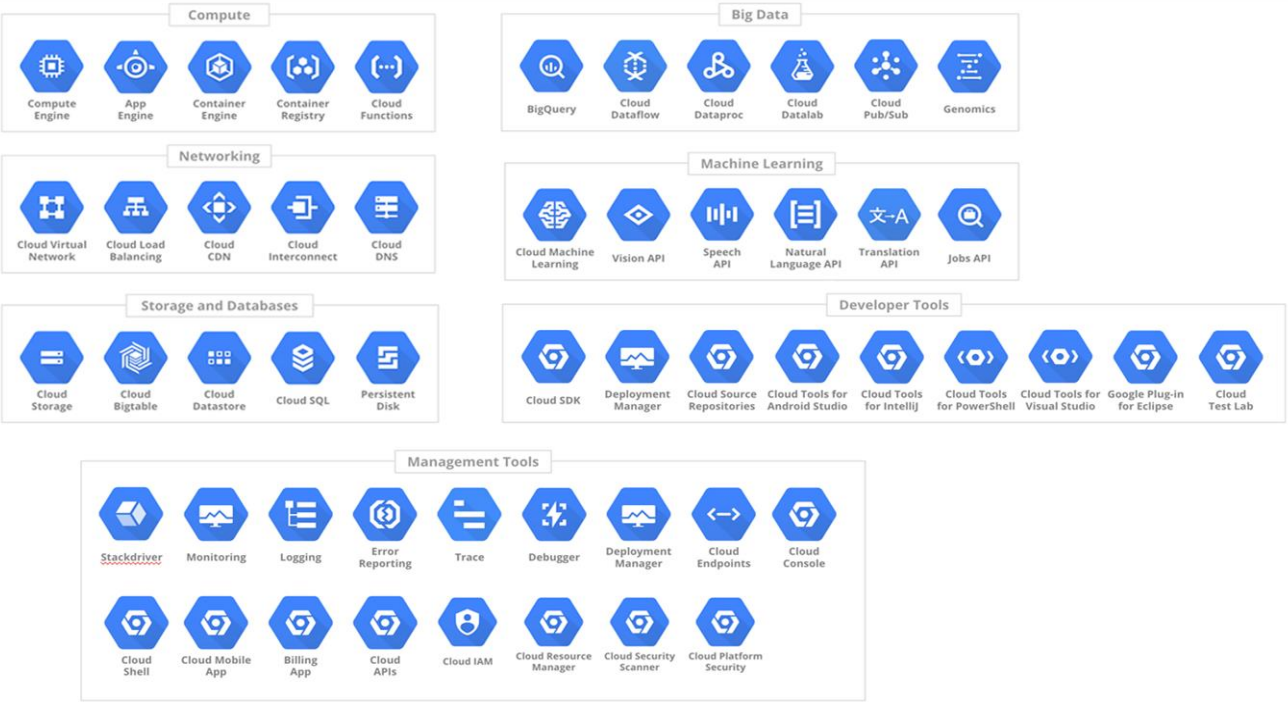
파이프라인 설계를 위한 Step 3 (2/4)

rockPLACE

3. GCP 서비스의 특징 이해

- GCP는 각각의 서비스의 독립성이 높으며, 이로인한 다양한 조합 및 결합이 가능합니다.
- 단 어떻게 적용하느냐에 따라 적용 방법이 수 없이 나오고 그 효과가 다 다르므로 충분한 사전 지식이 필요합니다.

GCP는 독립적입니다



Copyright ⓒ All right Reserved by 2019 Rockplace Inc.

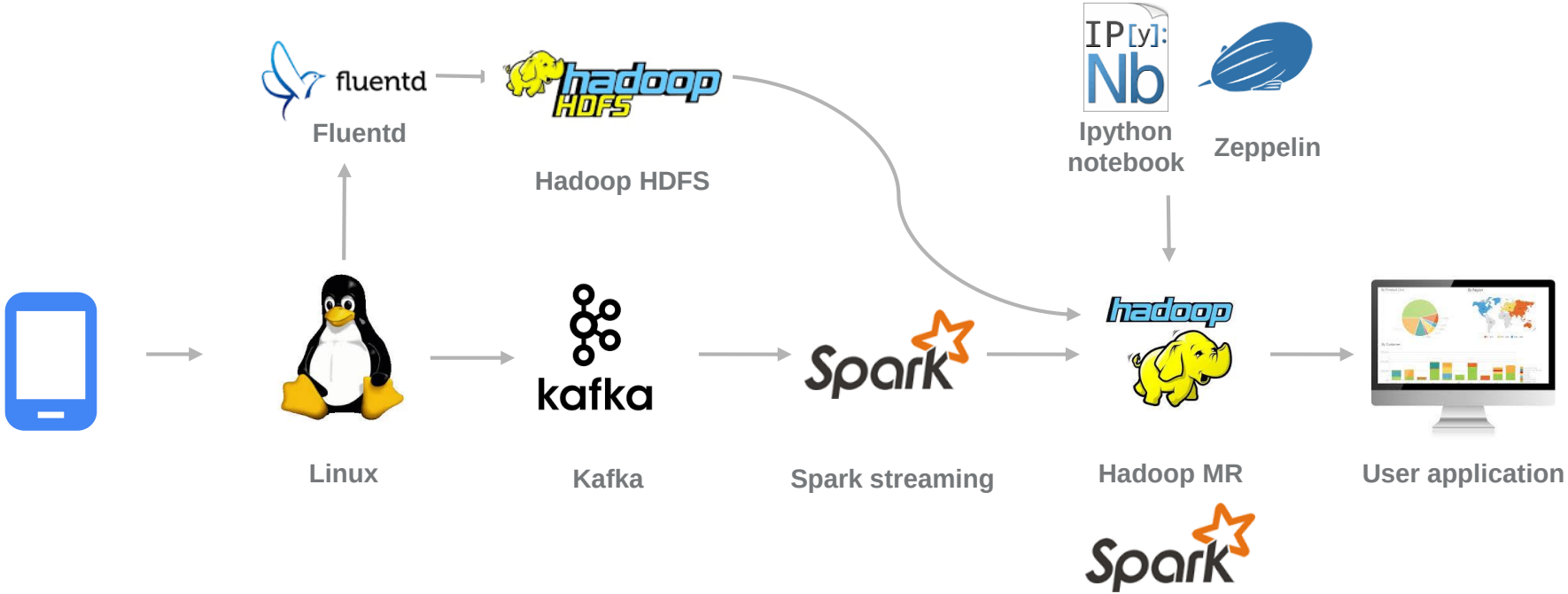
파이프라인 설계를 위한 Step 3 (4/4)

OpenSource PLACE, ROCKPLACE

rockPLACE

4. Open Source 적용 여부 및 적용 가능 부분 파악

- 내재화 가능한 범위, 기술, 업무 프로세스를 기반으로 적용 기술 및 범위를 검토해야 합니다.
- 각 단위 요소에 대한 기술 제약사항 및 커스터마이징, 성능에 대한 한계성을 사전 확인하여야 하며, 사전 확인이 어려운 부분은 단위 PoC를 통해 요소를 검증하도록 합니다.



Copyright ⓒ All right Reserved by 2019 Rockplace Inc.

파이프라인 설계를 위한 Step 4

OpenSource PLACE, ROCKPLACE

rockPLACE

Big Data Work-Flow 설계 – Step 1 ~ 3 Base

** 본 과정에서 실제로 구현/적용 가능한 목표가 도출됩니다.

** 본 과정에서 Step 1~3에 대한 사항들이 명확해 집니다.

Phase 1.

1) 목표설정

2) 빅데이터 수집 방안 마련

Phase 2.

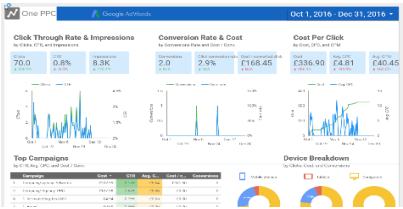
1) 빅데이터 저장관리 방안 마련

2) 빅데이터 처리 방안 마련

3) 빅데이터 분석 방안 마련

Phase 3.

1) 빅데이터 시각화 방안 마련



Copyright ⓒ All right Reserved by 2019 Rockplace Inc.

파이프라인 설계를 위한 Step 5

OpenSource PLACE, ROCKPLACE

rockPLACE

PoC, PoC, PoC … & Go!!!

**** 빅데이터, ML은 성공이 보장된 것이 아닙니다 ****

무조건 된다는 생각은 버리세요
각각의 주요 부분에 대한 PoC는 필수

**** 가슴은 뜨겁게!! 머리는 차갑게!! ****

갖추고자 하는 시스템에 발생하는
리스크 요소를 사전 식별하고
리스크 요소의 해결을 위한 PoC 수행

성공적인 PoC를 통해 리스크가 삭제 또는 최소화
되었다면 이제는 본사업 Go! Go! Go!

**** But & Atention!!!! ****

본 사업은 최대한 작게 설정, 단위 목표별 수행

Copyright ⓒ All right Reserved by 2019 Rockplace Inc.

파이프라인 설계 시 놓치기 쉬운 필수 고려사항(1/4)

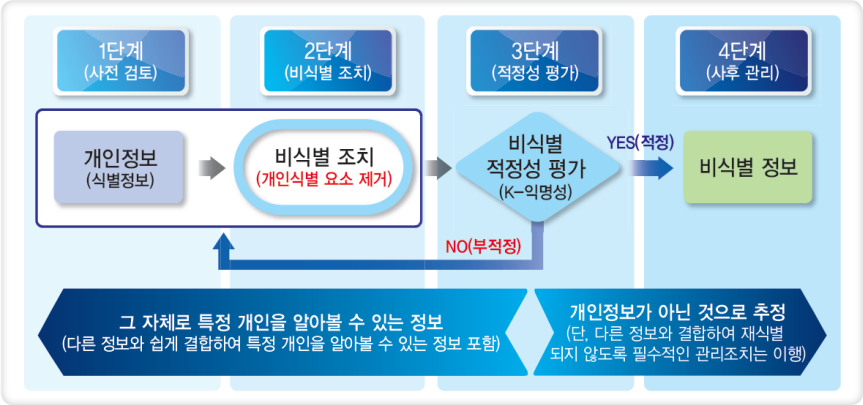
OpenSource PLACE, ROCKPLACE

rockPLACE

1. 개인정보, 식별정보, 보안정책 등 민감사항에 대한 사전 확인 필요

개인정보에 대한 비식별화 조치

● 비식별 조치 및 사후관리 절차 ●



[출처]: 한국교육학술정보원 "개인정보 비식별 조치 가이드라인"

빅데이터 시스템 구축시 연관 조치 필요성 사전 검토 필요

1. Zero Based 빅데이터 구축 방법론은 만능인 것 처럼 보이지만 결코 만능이 아닙니다.

2. 빅데이터 시스템 구축시 발생할 수 있는 개인정보, 식별정보, 기업 내/외부 보안정책 등 연관 조치가 필요할 경우 이를 Zero Based 빅데이터 구축 방법론에 포함시킬 수도, 별도로 진행할 수도 있습니다.

3. 연관 민감정보 및 정책이 많으면 많을수록, 다양하면 다양할 수록 Zero Based 빅데이터 구축 방법론의 제약사항은 많아지게 됩니다.

4. 민감성이 적은 부분부터 Zero Based 빅데이터 구축 방법론을 이용하여 하나하나 구축한 후 최종적으로 Platform 기반의 내재화된 빅데이터 시스템을 갖추는 것을 권장합니다.

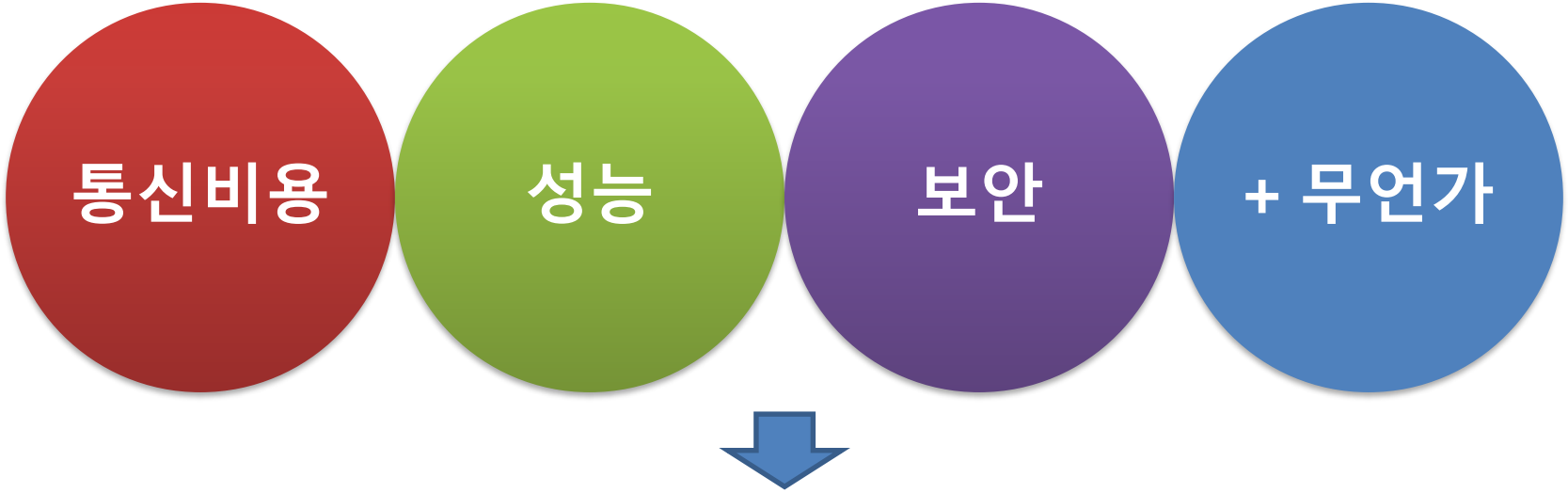
Copyright ⓒ All right Reserved by 2019 Rockplace Inc.

파이프라인 설계 시 놓치기 쉬운 필수 고려사항(2/4)

OpenSource PLACE, ROCKPLACE

rockPLACE

2. 멀티 클라우드



도입 안하니만 못한 상황을 조심하세요

Copyright ⓒ All right Reserved by 2019 Rockplace Inc.

파이프라인 설계 시 놓치기 쉬운 필수 고려사항(3/4)

rockPLACE

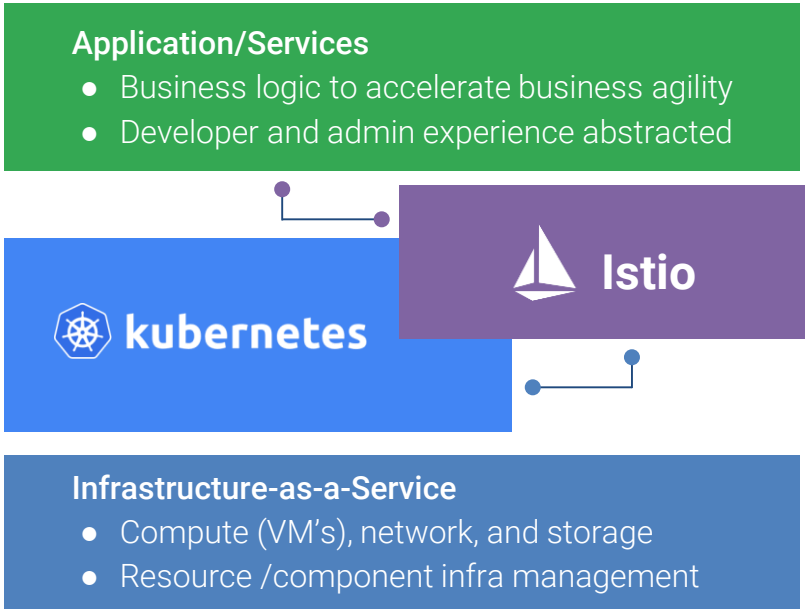
3. Hybrid & Management

It enables developers and operators to define **composable services** that can be managed and consumed as part of a cloud services platf orm or broader SaaS solution

Istio Service Mesh provides a **transparent** and **language-independe nt** way to flexibly and easily automate **application network** function s:

- Control, configure and monitor app-level requests
- Ensure service auth, resilience, routing, observability, fault-i njection
- Automatic policy enforcement at scale

Cloud Services Platform **includes a fully managed implementation of Istio**, integrated with GCP capabilities like Global Load Balancing



Copyright ⓒ All right Reserved by 2019 Rockplace Inc.

파이프라인 설계 시 놓치기 쉬운 필수 고려사항(4/4)

OpenSource PLACE, ROCKPLACE

rockPLACE

4. Micro Service Architecture & Dev Ops

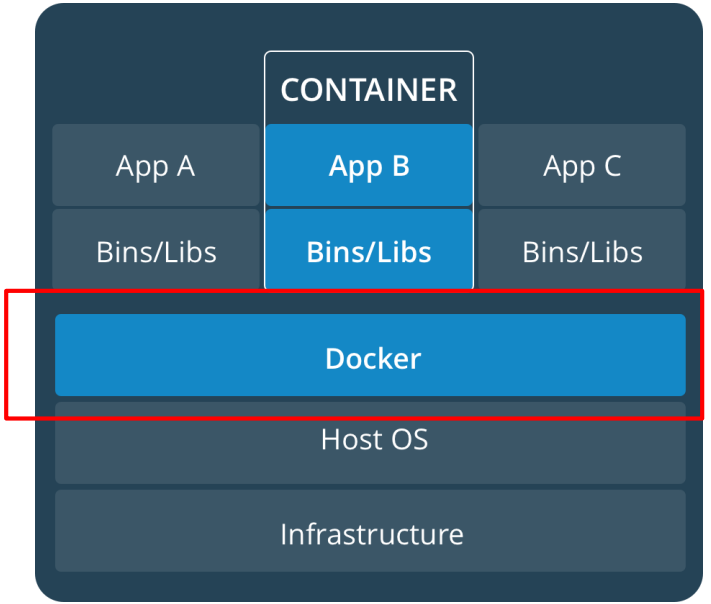
Docker 기술을 이용한 MSA 기법을 이용할경우 언제 어디서나 확장 가능한 무중단 서비스를 제공할 수 있습니다.

신속한 개발, 배포, 확장, 인프라 효율성을 MSA를 통해 경험하세요.

신규 수집 데이터 및 기 수집 데이터의 수집방법, 처리방법에 대한 변경시 독립적인 구성 및 운영, 업데이트가 가능합니다.

Docker와 MSA 결합을 통해 언제나 Hybrid 환경을 통해 원하는 비즈니스를 실현할 수 있습니다

단!!! 관련 기술에 대한 내재화를 하였을 경우에 가능합니다.



docker

others:
LXC
systemd-nspawn
rkt
...

*Images provided by Docker

Copyright ⓒ All right Reserved by 2019 Rockplace Inc.

마치는 글

OpenSource PLACE, ROCKPLACE

rockPLACE

다른건 다 잊어도 이것만큼은 기억 부탁 드립니다 – Part 1.

1. **명확한 목표**를 정하세요.

2. 해당 시스템이 반드시 필요한지 **심각**하게 고민하세요

3. **진행 방안**을 결정하세요

4. 수행 조직과 관리 조직에 **역할을 명확히 설정**하세요

5. **전문가의 말**을 들으시고 주위의 말에 **흔들리지 마세요**

6. **초심**을 잃지 마시고 **단계별**로 **정확**하고 **차분**히 진행하세요

Copyright ⓒ All right Reserved by 2019 Rockplace Inc.

마치는 글

OpenSource PLACE, ROCKPLACE

rockPLACE

다른건 다 잊어도 이것만큼은 기억 부탁 드립니다 – Part 2.

목적을 이루기 위한 수단에 얽매이지 마세요.

필요한 기술을 직접 구현하려 하지말고 쓸수 있는건 다 가져다 쓰세요.

기술는 창의적이지 않습니다. 하지만 빠르고 편리하고 강력하지요.
창의적인 활동에 집중하세요. 여러분이 기술을 이길수는 없습니다.

역할에 얽메이지마세요.

모르는 것은 부끄러운 것이 아닙니다. 아는 사람과 협업하세요.

IT로 예술하지 마세요, 돈을 버세요 ^^

Copyright ⓒ All right Reserved by 2019 Rockplace Inc.

영업기술문의 ☎ 02)6251-7788
고객 1:1 상담 https://rockplace.co.kr/customers/counseling/