

구글 클라우드
컨설팅
쇼케이스
Season4

빅데이터 & ML 내재화를 이루어보자

#Bigdata #ML #내재화 #Opensource #GCP

2019년 6월 13일(목) 14:00 - 17:00

도리출, 세미나실 3층 [신청하기](#)

(서울시 강남구 테헤란로 138 삼성타워 3층)

빅데이터 & ML 내재화를 이루어보자

클라우드 컨설팅팀 권수용
sykwon@rockplace.co.kr
2019.06.13

시작하기에 앞서...

OpenSource PLACE, ROCKPLACE

GCP는 타 클라우드 사업자의 서비스와 가장 큰 차이점으로 서비스의 독립성이 강합니다.

- 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.
- 그렇기 때문에 파이프라인 설계시 더 많은 고민과 노력을 기획 단계에서 갖추고 진행하는 것이 무엇보다 중요합니다.



금일 거버넌스는 다루지 않지만...

- ✓ **Cloud & The Other** 어떠한 형태로 IT를 운영하시더라도 거버넌스 정의는 필수
- ✓ 거버넌스를 통해 조직, 업무, 보안, 기타 관련된 모든 제반의 기반을 세우세요.
- ✓ 거버넌스는 최대한 단순하게, 하지만 명확하게 정의하셔야 합니다.
- ✓ 거버넌스 없이 IT 시스템을 구축/운영하게 될 경우 표준화 및 효율화를 위해 향후 더 많은 시간과 비용이 투입되게 됩니다.

잊지말자 거버넌스!!!

내재화란?

OpenSource PLACE, ROCKPLACE

[국립국어원]

우리말샘 (6건)

내재·화(內在化)

내재·화 「001」 「명사」 어떤 성질 따위가 사물이나 일정한 범위의 안에 들어 있게 됨. 또는 그렇게 되게 함.

분류: 우리말샘

내재화·되다(內在化되다)

내재화·되다 「001」 「동사」 어떤 성질 따위가 사물이나 일정한 범위의 안에 들어 있게 되다.

분류: 우리말샘

내재화·하다(內在化하다)

내재화·하다 「001」 「동사」 어떤 성질 따위가 사물이나 일정한 범위의 안에 들어 있게 되다. 또는 그렇게 되게 하다.

분류: 우리말샘

[Wikipedia]

Internalization

From Wikipedia, the free encyclopedia

Internalization (or **internalisation**) is the process of making something internal, with more specific meanings in various fields. It is the opposite of externalization.

Contents [hide]
1 Psychology and sociology
1.1 As a symptom
2 Biology
3 Economics and management
4 Finance
5 See also
6 Notes
7 References

Psychology and sociology [[edit](#)]

Main article: Internalisation (sociology)

In psychology, internalization is the outcome of a conscious mind reasoning about a specific subject; the subject is internalized, and the consideration of the subject is internal. Internalization of ideals might take place following religious conversion, or in the process of, more generally, moral conversion.^[1] Internalization is directly associated with learning within an organism (or business) and recalling what has been learned.

In psychology and sociology, internalization involves the integration of attitudes, values, standards and the opinions of others into one's own identity or sense of self. In psychoanalytic theory, internalization is a process involving the formation of the super ego.^[2] Many theorists believe that the internalized values of behavior implemented during early socialization are key factors in predicting a child's future moral character. The self-determination theory^[3] proposes a motivational continuum from the extrinsic to intrinsic motivation and autonomous self-regulation. Some research suggests a child's moral self starts to develop around age three.^[4] These early years of socialization may be the underpinnings of moral development in later childhood. Proponents of this theory suggest that children whose view of self is "good and moral" tend to have a developmental trajectory toward pro-social behavior and few signs of anti-social behavior.

In one child developmental study,^[5] researchers examined two key dimensions of early conscience – internalization of rules of conduct and empathic affects to others – as factors that may predict future social, adaptive and competent behavior. Data was collected from a longitudinal study of children, from two parent families, at age 25, 38, 52, 67 and 80 months. Children's internalization of each parent's rules and empathy toward each parent's simulated distress were observed at 25, 38 and 52 months. Parents and teachers rated their adaptive, competent, pro-social behavior and anti-social behavior at 80 months. The researchers found that first, both the history of the child's early internalization of parental rules and the history of his or her empathy predicted the children's competent and adaptive functioning at 80 months, as rated by parents and teachers. Second, children with stronger histories of internalization of parental rules from 25 to 52 months perceived themselves as more moral at 67 months. Third, the children that showed stronger internalization from 25 to 52 months came to see themselves as more moral and "good". These self-perceptions, in turn, predicted the way parents and teachers would rate their competent and adaptive functioning at 80 months.

내재화를 하지 않을 경우 어떤 일이 벌어질까요?

OpenSource PLACE, ROCKPLACE

- ✓ 빅데이터 및 머신러닝 프로젝트의 성공률이 30% 이하로 내려갑니다
- ✓ 시장의 트렌드 및 조직의 비즈니스 상황에 대처가 불가능합니다.
- ✓ 조직이 빅데이터와 머신러닝을 컨트롤 할 수 없습니다.
- ✓ 구축을 하더라도 시간이 가면 갈수록 돈 잡아먹는 하마로 변질됩니다.
- ✓ 한번의 실패로 이후 모든 로드맵이 변경되며, 기존의 것들이 무의미해 집니다.



**여러분의 조직은 안 겪으실 것 같으시죠?
"99% 확률로 경험하시게 될 것 입니다."**

내재화를 안 하면 왜 이런 일이 발생하게 될까요?

- ✓ 빅데이터 및 머신러닝의 출발은 모호함에서 시작하기 때문입니다.
- ✓ 일반 IT 프로젝트와 달리, $1+1 = 2$ 라는 딱 떨어지는 경우는 존재하지 않습니다.
- ✓ 빅데이터와 머신러닝은 생물과 같아서 수많은 변수를 가지고 있기 때문입니다.
- ✓ 프로젝트 진행 중, 프로젝트 종료 후 운영중이라도 수많은 환경적, 비즈니스적 영향을 받으며 계속 주시하고 변경하고 검증해야만 가치가 있기 때문입니다.
- ✓ 여러 변수에 대해 대응하는 시간이 지체될수록 활용 성공율은 0%에 수렴하게 됩니다. 따라서 이것저것 따질 겨를이 없습니다.
- ✓ 내부에서 검증 및 활용성을 입증하지 못한다면 거짓된 정보를 바탕으로 잘못된 의사결정을 위한 도구로 전락할 수 밖에 없습니다.

**시장에는 수많은 약장수들이 존재합니다
물론 그 중에 저도 포함됩니다.**

반드시 내재화를 해야 하나요?

OpenSource PLACE, ROCKPLACE

rockPLACE

네. 이것은 MUST 입니다.

내재화를 하기 위해서 기본적으로 뭘 갖춰야 하나요? **rock**PLACE

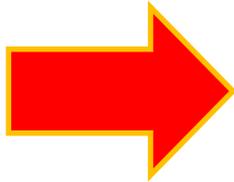
OpenSource PLACE, ROCKPLACE

✓ 기획능력

✓ 설계능력

✓ 개발능력

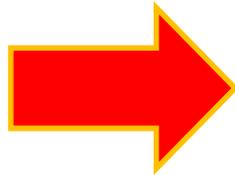
✓ 검증능력



운영능력

우리는 하고 있는데 무엇이 잘 못 되었다는 거죠?

- ✓ 기획능력
- ✓ 설계능력
- ✓ 개발능력
- ✓ 검증능력



소싱... 소싱... 소싱...

견적에 기반한 처리

Copy & Paste

다시 말씀드립니다.
빅데이터와 머신러닝은 $1+1=2$ 아닙니다.

기본적인 것들을 얼마나 갖추면 되나요?

- ✓ 기획능력 : 목표 정의, 활용방안 정의
- ✓ 설계능력 : 파이프라인 설계
- ✓ 개발능력 : REST API, JAVA, Python
- ✓ 검증능력 : 커뮤니케이션

다 내재화 할 수는 없잖아요?

rockPLACE

OpenSource PLACE, ROCKPLACE

물론입니다 !!!

단, 기본은 반드시 지켜주세요 !!!

**1년, 2년, 3년...
반드시 내 것으로**

STEP 1 내재화를 하기위해 가장 먼저 뭐부터 할까요?

“당신의 AS - IS 확인”

구분	TASK	설명
현황분석 Part 1	IT 및 관련 조직 정보 분석	- 주요 조직구성 파악
	IT 조직 구성원 역량 파악	- IT 구성원에 대한 기술 역량 분석을 통한 기술 스펙 확인
	IT 조직 및 구성원 간 역할 분석	- IT 조직 및 구성원의 주요 업무 및 역할 확인
빅데이터 및 머신러닝 Needs 분석	아이템 수집	- 다양한 아이템 수집을 통한 향후 발전방안 모색
	우선순위 분석	- 아이템 분석을 통한 요구사항 분석
	주요 필요 기술 분석	- 아이템 구현을 위한 주요 필수기술 도출
현황분석 Part 2	아이템 실현화 가능성 분석	- 아이템 실현 가능여부에 대한 1차 검토
	내재화 가능 부분 분석	- 실현 가능한 아이템에 대한 내재화 가능 여부 및 가능 부분 도출
	카테고리 분석	- 아이템 별 카테고리 편성을 통한 주요 비즈니스 매핑

현업부서 + IT 조직



**Workshop, Brainstorming
Seminar, Community**

기술 로드맵 및 조직구성안 수립

구분	TASK	설명
기술 로드맵 설계	빅데이터 도입을 위한 기술 로드맵 설정	- 빅데이터 도입을 위한 기술 로드맵 설계 ... 현재 당장 내재화 불가능한 부분에 대한 도입/습득에 대한 로드맵
	머신러닝 도입을 위한 기술 로드맵 설정	- 머신러닝 도입을 위한 기술 로드맵 설계 ... 현재 당장 내재화 불가능한 부분에 대한 도입/습득에 대한 로드맵
	오픈소스 도입을 위한 기술 로드맵 설정	- 오픈소스 도입을 위한 기술 로드맵 설계 ... 현재 당장 내재화 불가능한 부분에 대한 도입/습득에 대한 로드맵
조직 구성안 설계	조직 구성안 제언	- 내재화를 위한 최적 조직 구성안 설계 및 제언
	조직 역할 제언	- 내재화를 위한 최적 조직 별 역할 설계 및 제언
	조직 주요 구성원 별 역할 제언	- 내재화를 위한 최적 조직 구성원별 역할 설계 및 제언
DevOps 기본 방안 설계 (Option)	DevOps 정책 기획 - Part 1	
	DevOps 정책 설계 - Part 1	
	DevOps 방안 기획 - Part 1	

업무 체계 수립

구분	TASK	설명
필수 업무 프로세스 분석	As-Is 업무 프로세스 분석	- 기존 조직 IT 업무 프로세스 분석
	As-Is 업무 프로세스 Risk 분석	- 기존 조직 IT 업무 프로세스에 대한 주요 Risk 사안 분석
	필수 수정 필요 업무 프로세스 분석	- 신기술 도입 및 운영을 위해 필수적으로 변경이 필요한 업무 프로세스 분석
필수 업무 프로세스 수행 방안 설계	조직 구성안 제언	- 조직 구성 개선안 제언
	조직 역할 제언	- 조직 역할 개선안 제언
	조직 주요 구성원 별 역할 제언	- 조직 구성원 별 개선 역할 제언
	To-Be 업무 프로세스 제언	- To-Be 업무 프로세스 제언
DevOps 수행 방안 설계 (Option)	DevOps 정책 기획 - Part 2	
	DevOps 정책 설계 - Part 2	
	DevOps 방안 기획 - Part 2	

시스템 도입/구축방안 수립

구분	TASK	설명
도입 Item 분석/선정	내재화 가능 범위 분석	- 아이템 도입 및 운영 시 내재화 가능 범위 분석
	비용 타당성 분석	- 아이템 도입에 따른 비용 타당성 분석
	구축 이후 활용 및 발전방향 분석	- 이후 발전방향 분석
시스템 플랫폼(환경) 분석/선정	도입가능 플랫폼 검토	- 도입이 가능한 시스템 형태 분석(On-Prem, Cloud, 기타 등등)
	도입가능 플랫폼 타당성 분석	- 도입이 가능한 시스템에 대한 도입 타당성 분석
	도입 플랫폼 발전방향 분석	- 도입 플랫폼에 대한 이후 발전방향 분석/설계
시스템 적용 기술 분석/선정	적용 필요 솔루션/서비스 검토	- 도입에 필요한 적용 솔루션/서비스 분석
	적용 필요 솔루션/서비스 타당성 분석	- 도입에 필요한 적용 솔루션/서비스에 대한 비용 타당성 분석
	구축 이후 활용 및 발전방향 분석	- 도입 후 활용 및 발전방향 분석/설계
To-Be Base Architecture 수립	개념 인프라 구성도 설계	- 개념 인프라 구성도 설계
	개념 솔루션/서비스 구성도 설계	- 개념 솔루션/서비스 구성도 설계
수행방안 수립	R&R 정의	- 조직/구성원 간 필수 R&R 정의
	수행방안 수립	- 수행방안 수립

시스템 운영방안 수립

구분	TASK	설명
시스템 운영방안 설계	시스템 운영에 대한 R&R 분석/설계	- 운영과 관련한 R&R 정의
	시스템 관리에 대한 R&R 분석/설계	- 관리와 관련한 R&R 정의
시스템 개선방안 설계	시스템 개선에 대한 R&R 분석/설계	- 시스템 개선과 관련한 R&R 정의
	시스템 개선에 대한 적용방안 분석/설계	- 개선 결과물 적용방안 정의
	시스템 개선에 대한 관리방안 분석/설계	- 개선 결과물에 대한 관리방안 정의

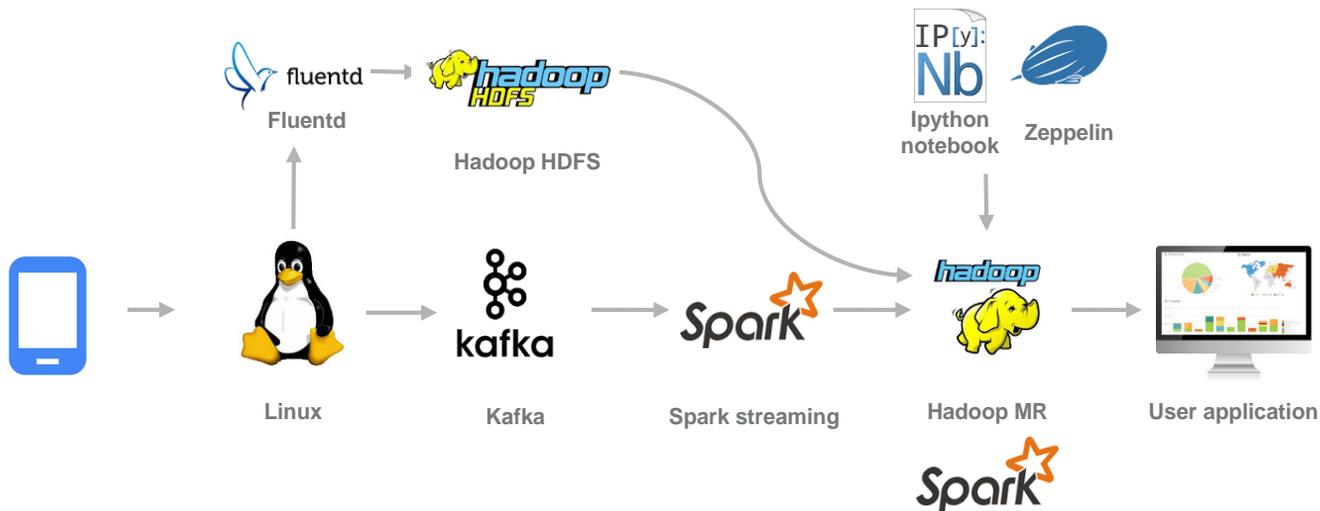
- ✓ 기법을 정해놓고 하는건 쌍팔년도 방식
- ✓ 우선 목표와 활용방법에 따른 표준화 우선
- ✓ 통계학과 ML 알고리즘은 나~~중에
- ✓ 통계학과 ML 알고리즘은 외주이용? OK !!
- ✓ 통계학과 ML 알고리즘보다는 기본에 충실

기업이라면 무중속에 대한 고려를 반드시 하자

OpenSource PLACE, ROCKPLACE

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.



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



성공적인 빅데이터 & ML을 위한 마음가짐과 행동

OpenSource PLACE, 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!! ****

**갓추고자 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 & Attention!!!! ****

본 사업은 최대한 작게 설정, 단위 목표별 수행

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”

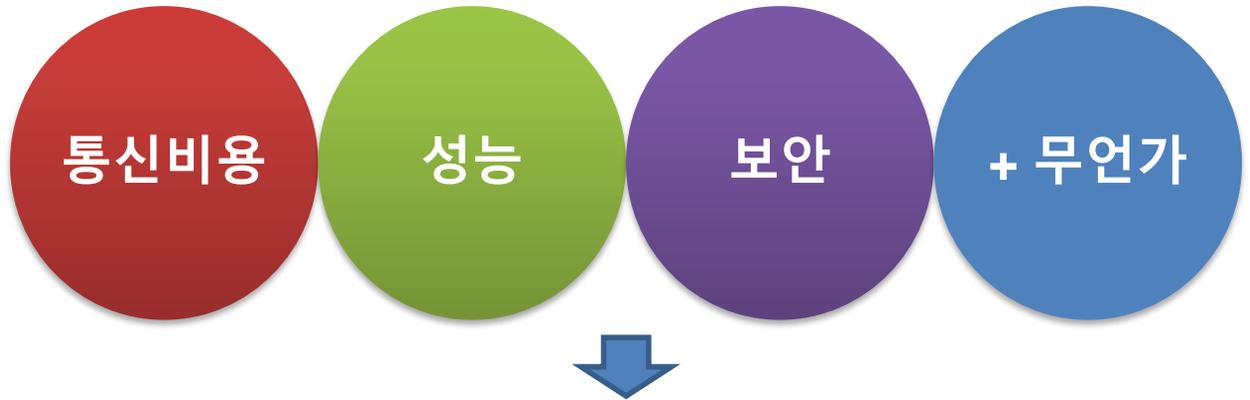
빅데이터 시스템 구축시 연관 조치 필요성 사전 검토 필요

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.

빅데이터 & 머신러닝 중 놓치기 쉬운 필수 고려사항

OpenSource PLACE, ROCKPLACE

2. 멀티 클라우드



도입 안하니만 못한 상황을 조심하세요

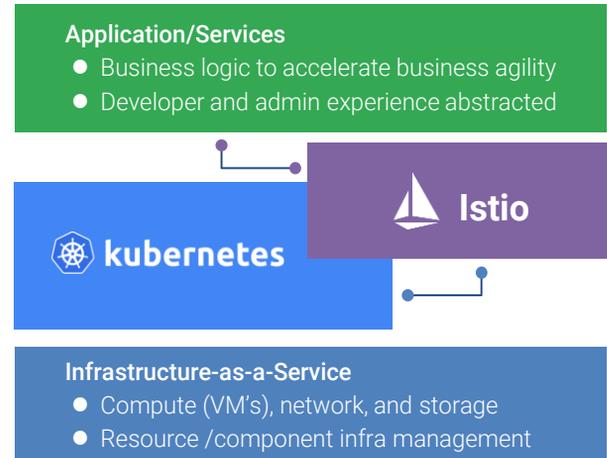
3. Hybrid & Management

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

Istio Service Mesh provides a **transparent** and **language-independent** way to flexibly and easily automate **application network** functions:

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

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



빅데이터 & 머신러닝 중 놓치기 쉬운 필수 고려사항

4. Micro Service Architecture & Dev Ops

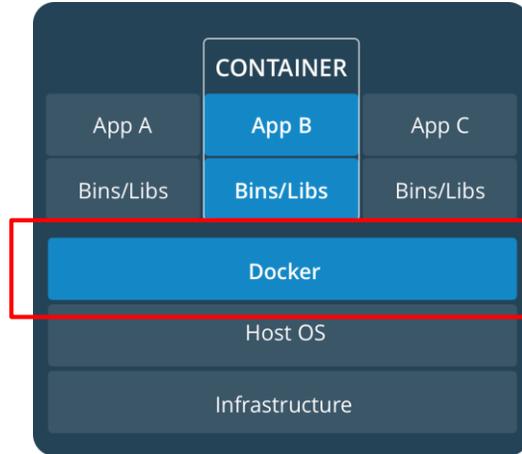
Docker 기술을 이용한 MSA 기법을 이용할 경우 언제 어디서나 확장 가능한 무중단 서비스를 제공할 수 있습니다.

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.



others:
LXC
systemd-nspawn
rkt
...

*Images provided by Docker

다른건 다 잊어도 이것만큼은 기억 부탁 드립니다

목적을 이루기 위한 수단에 얽매이지 마세요.

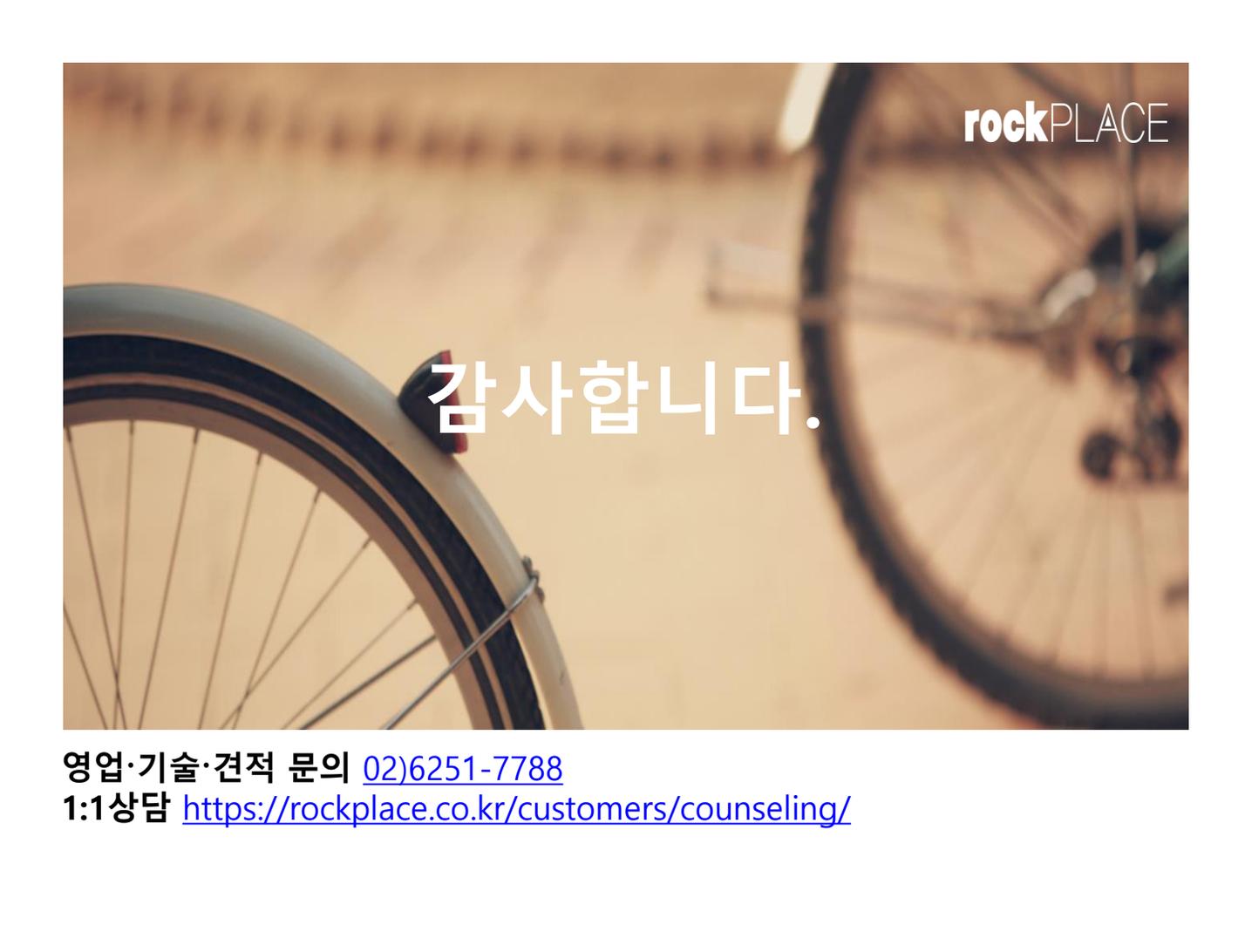
필요한 기술을 직접 구현하려 하지말고 쓸수 있는건 다 가져다 쓰세요.

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

역할에 얽매이지마세요.

모르는 것은 부끄러운 것이 아닙니다. 아는 사람과 협업하세요.

IT로 예술하지 마세요, 돈을 버세요 ^^



rockPLACE

감사합니다.

영업·기술·견적 문의 [02\)6251-7788](tel:0262517788)

1:1상담 <https://rockplace.co.kr/customers/counseling/>