

# 오픈소스 개발자를 꿈꾸는 후배들을 위한 조언

최현식

# 소개

- 최현식 (Hyunsik Choi)
  - Director of Research, Gruter Corp (2013. 6 ~)
  - PhD (Computer Science & Engineering, 2013. 8)
  - Apache Tajo (incubating) PPMC member & committer (2013.3 ~ )
  - Apache Giraph PMC member and committer (2011. 8 ~ )
  - Mentor of Google Summer of Code (2013)
- Contacts
  - Email: [hyunsik AT apache.org](mailto:hyunsik@apache.org)
  - Linkedin: <http://linkedin.com/in/hyunsikchoi/>
  - Twitter: @hyunsik\_choi

# 목차

## 1. 아파치 타조의 간단 소개

- 1) Tajo Team
- 2) 아키텍처
- 3) Roadmap

## 2. 오픈소스 참여하기

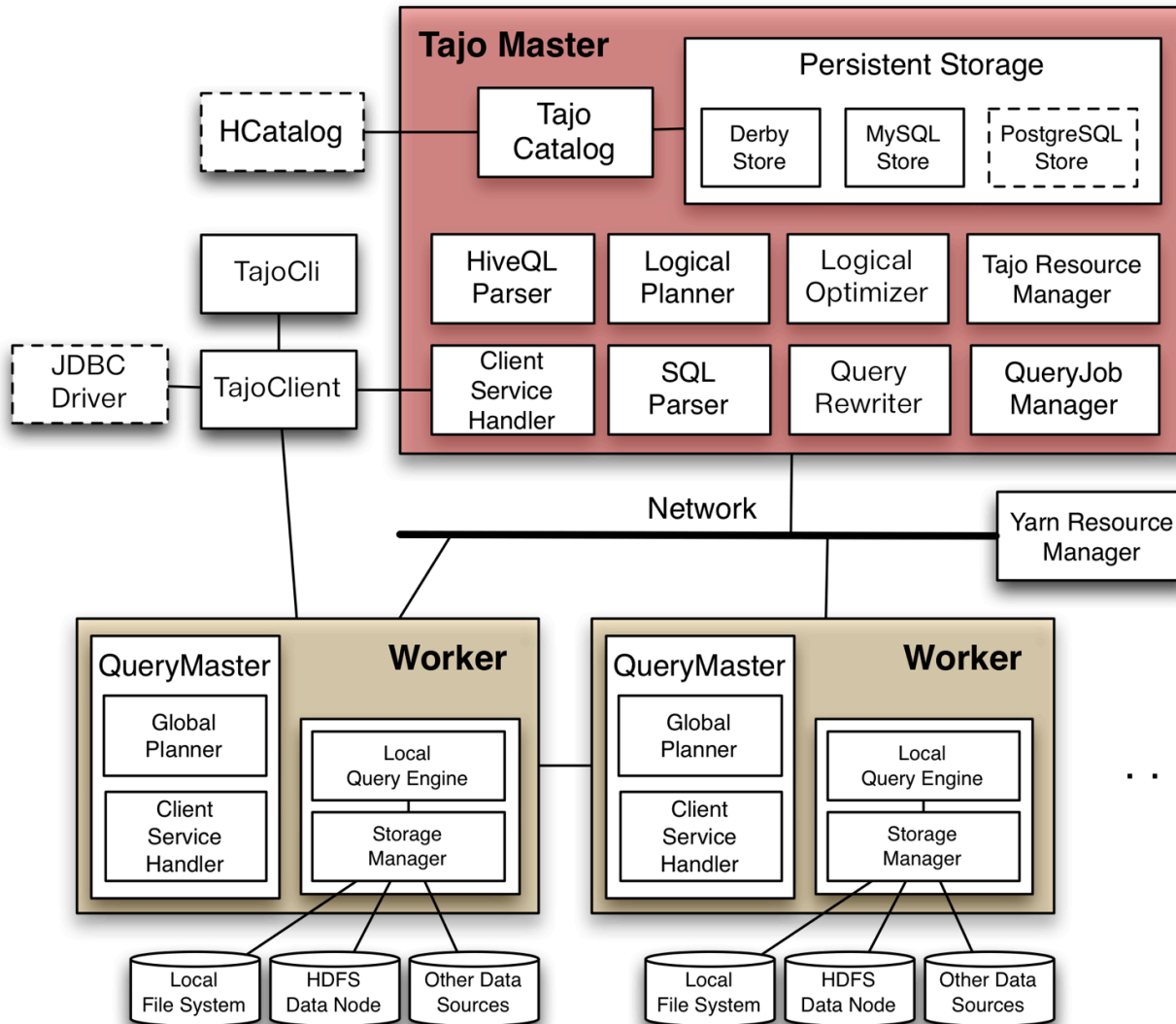
- 1) 좋은 프로젝트 고르는 방법
- 2) 좋은 프로젝트 만들기
- 3) 프로젝트 참여
- 4) 오픈소스 프로젝트 개발자의 중요 덕목

# Tajo란?

- Tajo
  - 하둡 기반의 대용량 데이터웨어 하우스 시스템
  - 2010년 부터 리서치 프로토타입으로 개발 시작
  - 아파치 인큐베이션 프로젝트 (올해 3월 ASF 인큐베이션 채택)
- Features
  - SQL 표준 호환
  - 질의 전체를 분산 처리
  - HDFS가 기본 스토리지
  - 관계형 모델 (Nested model로 확장 계획 중)
  - ETL 뿐만 아니라 low-latency 질의(100 ms ~ hours)
  - UDF 지원
- News
  - 0.2-incubating: 11월 20일 릴리즈
  - 0.8-incubating: 12월 릴리즈 계획

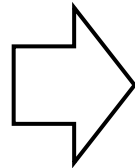
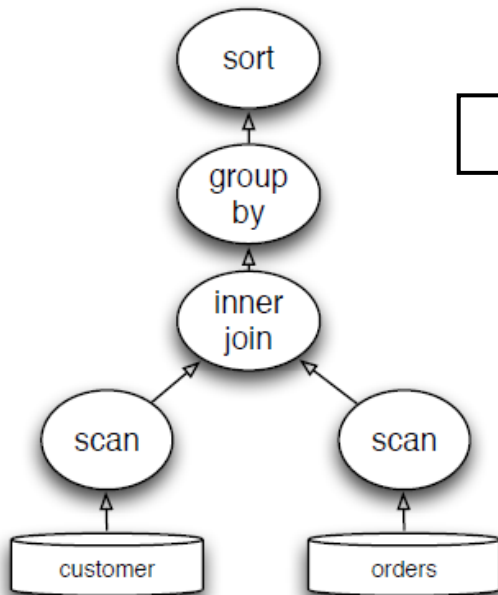


# Tajo 아키텍처

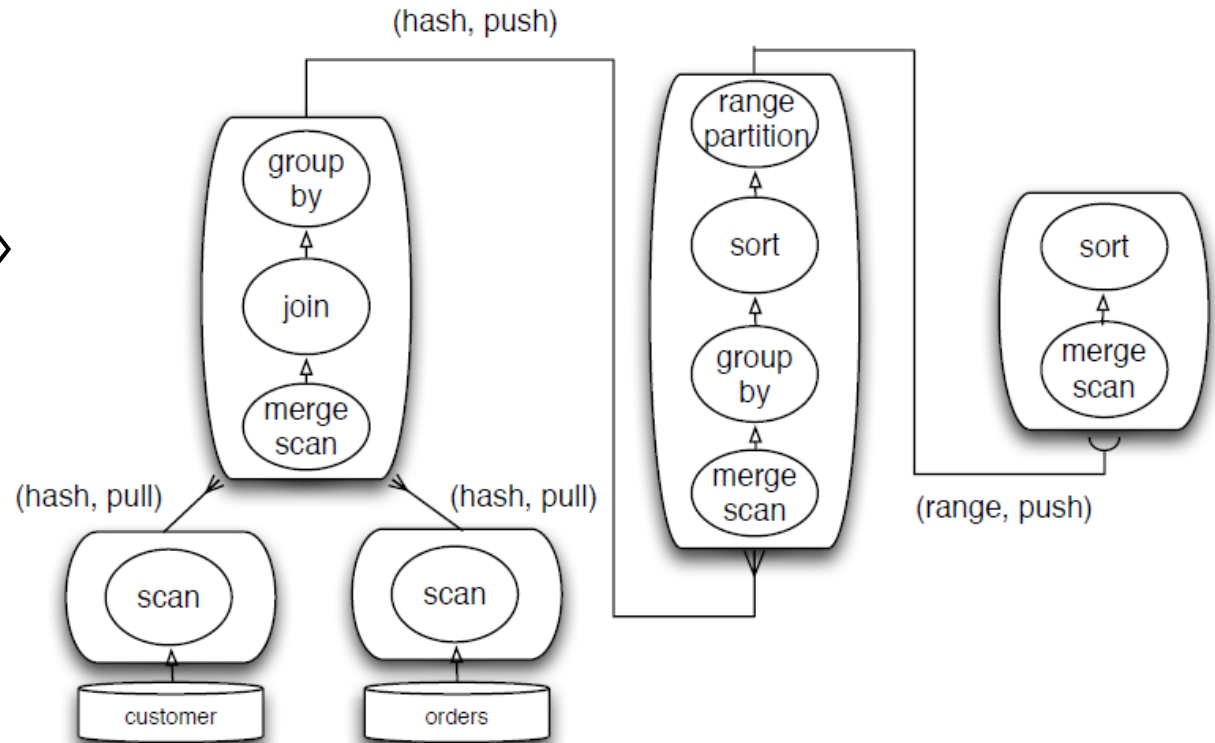


# Tajo 분산 처리 모델에 기반한 분산 실행 계획의 예제

논리 실행 계획  
(logical plan)



분산 실행 계획  
(distributed execution plan)



select col1, sum(col2) as total, avg(col3) as average from r1, r2  
where r1.col1 = r2.col2 group by col1 order by average;

# Roadmap

- 2013:11.20: 0.2 공식 릴리즈
- 2013.12 : Apache Tajo 0.8 Release
  - SQL 지원 향상, 안정성 향상
  - Outer join 지원
  - Hadoop 2.2.0 GA 포팅
  - Table Partitioning: Hash, Range, List, Column (Hive style)
  - Hcatalog 액세스 지원
- 2014 Q1: Apache Tajo 1.0-alpha release
  - 강력하고 다양한 rewrite rule 추가
  - 윈도우 함수 지원 (window function)
  - JIT query compilation & vectorized 엔진 프로토타입

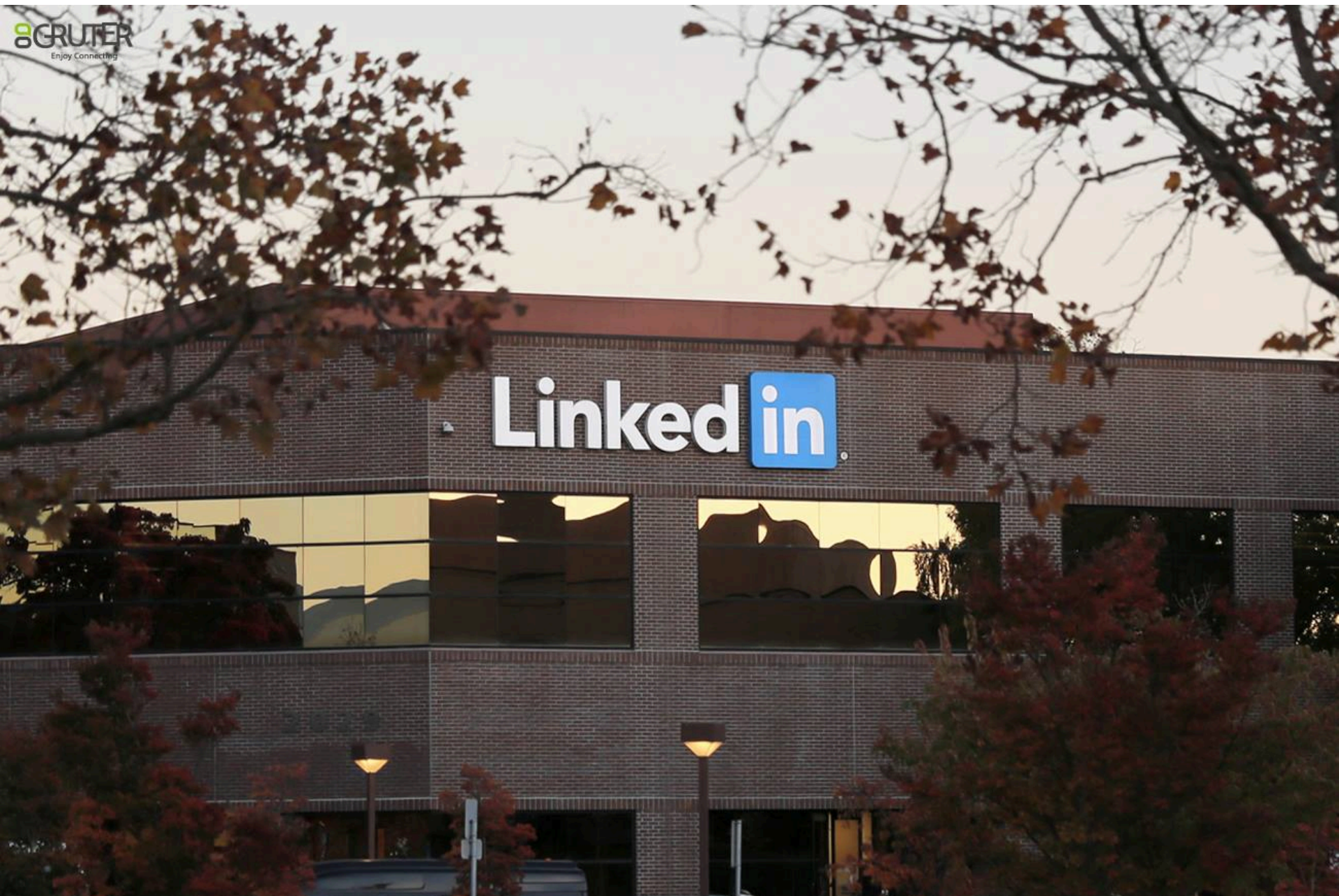
# Tajo Team

- 초기 코드는 고려대학교 데이터베이스 연구실 공헌
- 2013년 3월 아파치 인큐베이션
- 다양한 소속의 멤버들이 팀 구성
  - Estay, Gruter, LinkedIn, Kakao, Korea Univ., Platfora, SK telecom
- 9명의 PPMC & Committer
- 10여명의 Contributors



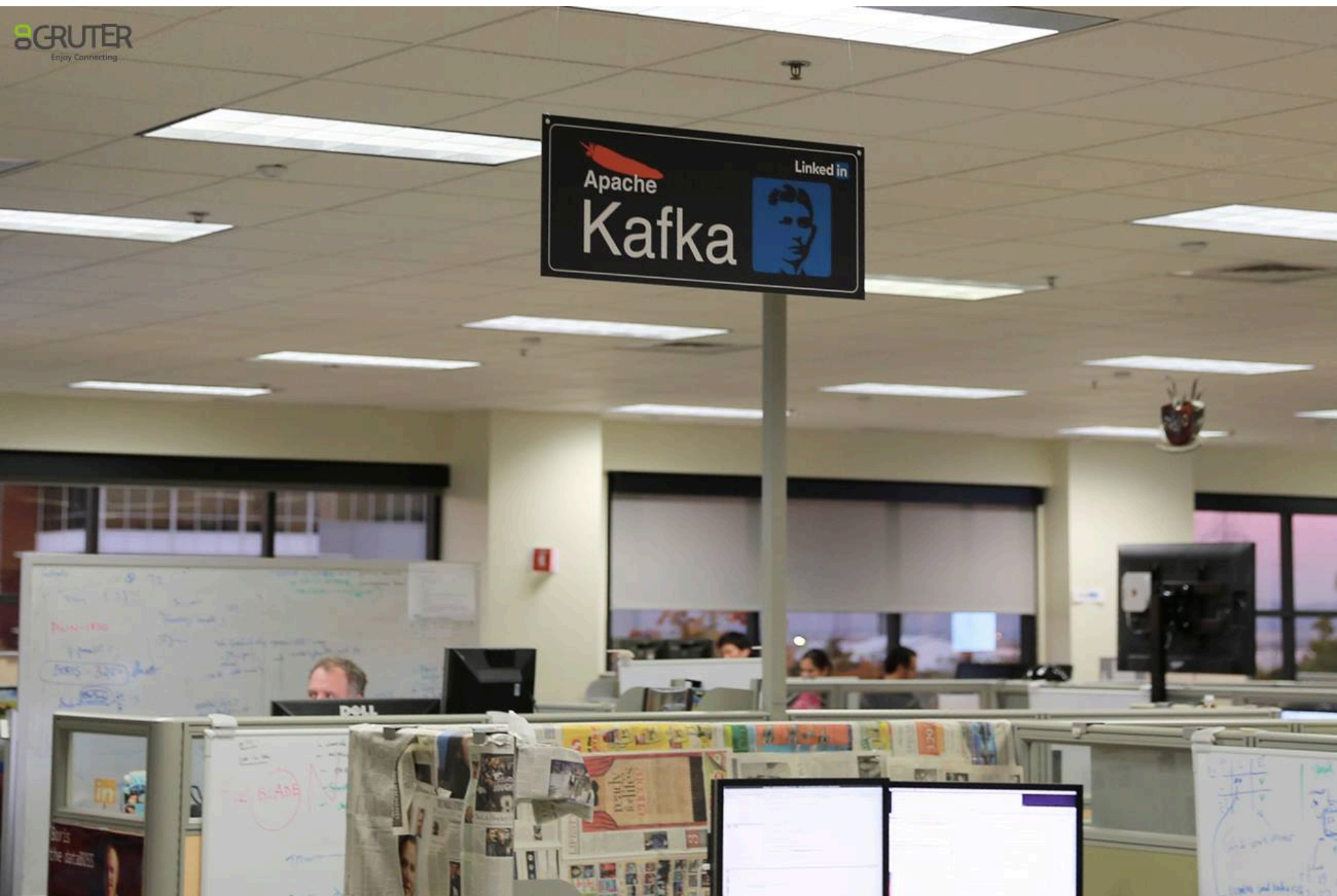
# 최근 활동

- 11.20 - Tajo 0.2.0-incubating Released
- 11.04 - Presentation at Bay Area Hadoop User Group  
- LinkedIn Special Event
- 10.15 - Presentation at Devview 2013
- 05.27 - Two projects accepted to Google Summer of Code 2013
- 04.09 - Demonstration at IEEE ICDE 2013
- 03.07 - Tajo Project enters incubation



LinkedIn

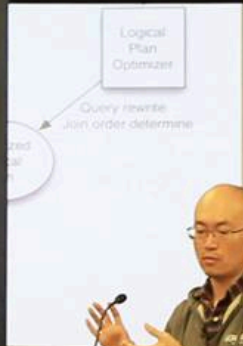




**Hadoop Use Group:**  
**Apache, Tajo and Samza**  
**AT**  
**Linked in**







# Linked in



# 참여 하기

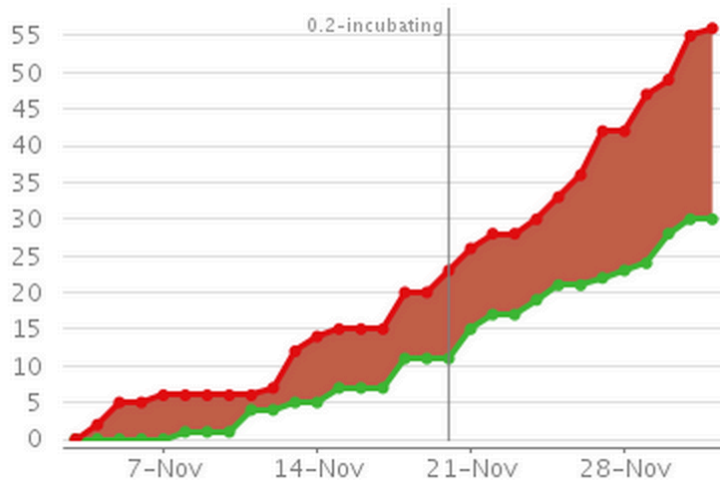
- Tajo Home
  - <http://tajo.incubator.apache.org>
- Getting Started
  - <http://tajo.incubator.apache.org/tajo-0.2.0-doc.html#GettingStarted>
- Checkout the development branch
  - <http://tajo.incubator.apache.org/source-code.html>
- Jira – Issue Tracker
  - <https://issues.apache.org/jira/browse/TAJO>
- Join the mailing list
  - [tajo-dev-subscribe@incubator.apache.org](mailto:tajo-dev-subscribe@incubator.apache.org)

# 좋은 프로젝트 찾기

1. 우선 (1) 좋아하는 것, (2) 잘하는 것을 명확하게 인식
  - 적절한 프로젝트 후보를 선택
2. 프로젝트 주제 평가
  - 프로젝트 목표 (goal)이 명확한가?
  - 프로젝트 철학, 원칙이 적절한가?
3. 참여 가능한가?
  - 커뮤니티 주도인가?
  - 기업 주도인가?
4. 커뮤니티가 얼마나 활발한가?
  - 메일링 리스트 트래픽
  - 이슈 트래커

# 좋은 프로젝트 찾기

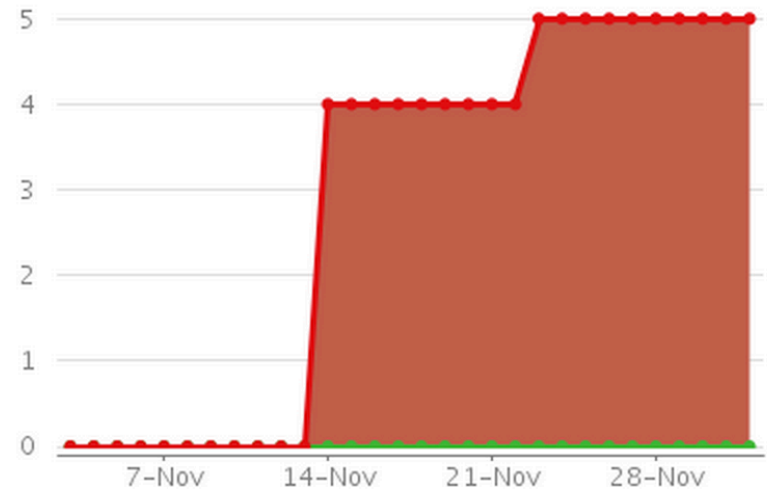
Issues: 30 Day Summary



Issues: **56** created and **30** resolved

Healthy Community

Issues: 30 Day Summary



Issues: **5** created and **0** resolved

Not Healthy Community

V.S.



# 좋은 프로젝트 만들기

1. (1) 좋아하는 것, (2) 잘하는 것을 명확하게 인식
2. 세상이 원하는 것
  - 이왕이면 널리 사용되는 것
  - 커뮤니티 없는 오픈소스는 의미 없다.
3. 프로젝트 목표
  - 내가 만들고자 하는 것이 정확하게 무엇인가?
  - 원칙, 철학을 세우고 항상 고민하자 (성장에 따라 발전하게 된다)
4. 커뮤니티 빌딩
  - 건강한 커뮤니티가 오픈소스 프로젝트의 심장
  - 다른 프로젝트와 연계
  - 블로그, 논문
  - 사용자 호응 (contribution, question)에 빠르고 친절하게

# 프로젝트 참여는?

- 우선 명심할 것
  - 조급해 하지 말라
  - 코어 코드를 고칠 수 있게 되기 까지 오랜 시간이 걸림을 명심하라 (어려운 프로젝트 일 수록 더욱)
  - 작은 것 부터 시작
  - 코드를 읽는 것이 다가 아니다.
    - 시간이 지남에 따라 배경지식 (이론) 습득이 더 중요해짐

# 프로젝트 참여는?

1. 분위기, 진도 파악
  - 메일링 리스트 구독
2. (사용자가 아니었다면) 사용자가 되자
3. 할 수 있는 이슈를 찾아보자
  - 간단한 것 부터 시작해서 프로세스를 익히자
4. 코드 익숙해 지기 (수 주 ~ 1년까지 걸릴 수 있음)
  - 코드 리뷰 보기
5. 의사 결정에 참여
6. 점차 핵심 코드 수정을 시도해보자

**제일 중요한 것은 끈기와 집중력**

# 오픈소스 개발자의 중요한 덕목

- 왜 오픈소스 개발자가 되려고 하는가?
  - 명성 (커미터 감투)?
  - 돈?, 취업?
  - 자기만족?
  - 재미?
- 커미터가 최종 목표여서는 안됨
  - 커미터라는 타이틀은 저절로 따라 오는 것
  - 자신이 현재 어떤 품질의 코드를 생산하는가? 다른 사람들을 어떻게 리딩하느냐가 제일 중요
- 기술적 리딩, 커뮤니티 리딩 자질을 바탕으로 지식을 공유할 줄 아는 사람이 되는 것이 중요

# 오픈소스 개발자의 중요한 덕목

- 열정 + 올바른 가치관
- 커뮤니케이션 능력
  - 다른 사람 배려, 설득, 영어
  - 테크니컬 discussion
  - 자존심을 버려라, 커뮤니티의 최선의 선택이 제일 중요
- 기술적 지식
  - 이론 (도메인 지식, 전공 지식)
  - 프로그래밍

Q & A