

TOROS: Python Framework for Recommender System

kakao

© Kakao Corp.

추천시스템과 파이썬이 만났을 때 생긴 일화에 대해서 공유하고자 합니다

Recommender Systems Everywhere!

The image shows the Kakao homepage. At the top right are '로그인' (Login) and language selection buttons ('지역선택' and 'Language(KO)'). The Kakao logo and '서비스' button are on the left. Below is a grid of service icons:

카카오톡	카카오스토리	같이 가치 with kakao	카카오택시	카카오지하철	카카오드라이버	슬러시	카카오버스
카카오증권	카카오뮤직	카카오내비	카카오플레이스	브런치	플레인	카카오티비	카카오그룹
카카오스타일	카카오홈	카카오페이지	카카오아지트	카카오게임	카카오계정		

A large central image features a hand holding a white smartphone displaying the KakaoTalk app interface with a yellow background and a speech bubble icon.

내안닌국 내뇨 메신저
전 세계 어디에서나 즐기는 1:1 및 그룹채팅

kakao

© Kakao Corp.

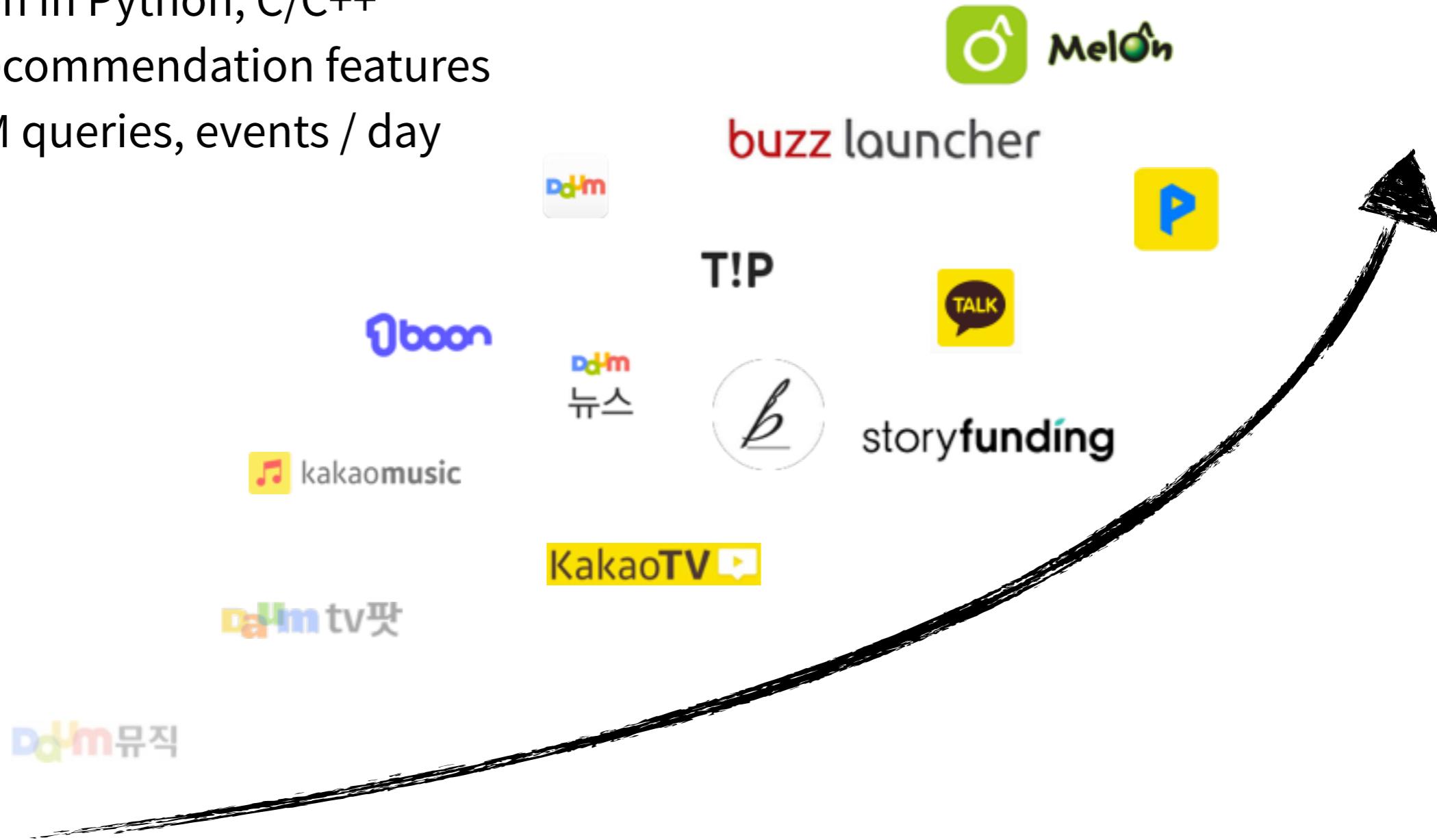
TOROS: A Recommender Systems Framework



TOROS: A Recommender Systems Framework



- Written in Python, C/C++
- +30 recommendation features
- +100M queries, events / day



TOROS: A Recommender Systems Framework



DduM 뉴스 ≡ **kakaoTV**

- 이 시각 추천뉴스
- 산자부의 거짓
- '누진제 개편' 대
- "아무렇게나 주 물의 삭발식"
- 폭염에 아열대
- [단독] 항공사에
- 인생은 한번이
- [단독] 사람잡는
- '시원한 바람'
- 물에 넣으면 지

1boon

- 추천영상
- EDITOR. 자리글 대
- [MBC] '양궁' 동점 승윤
- 양궁 (리우올림픽)
- 사소하지만 쓸 알팁
- [SBS]'만만치 않은 세트
- 양궁 (리우올림픽)
- 보양왕 삼계탕
- 귀지를 통해 태
- 라면으로 보는

DduM

- KT LTE
- ■ ■ ■ ■ TALK
- 세계에서 가장 도시 10
- 나를 위한 추천
- 제 누진제 완화 요금 절감 효과
- 기준 19,570 개편 16,310
- 유명인사 21 관
- news
- 점점 뭐라도 소
- 구
- 추
- 안녕
- 파란달
- 공대생

brunch beta

- 당신을 위한 브런치
- [제주 대정] 과거 현재 그리고 미래의 이야기가 있다-5
- 당진 웨袜마을에서
- 네가 지금 이행이나 할 때
- 바다가 꽃 떨어지는 곳 여수 향일암
- 제주올레 21코스, 그곳엔 행복한 미소를 만 있었다.
- 마음의 고향 강경.

kakao

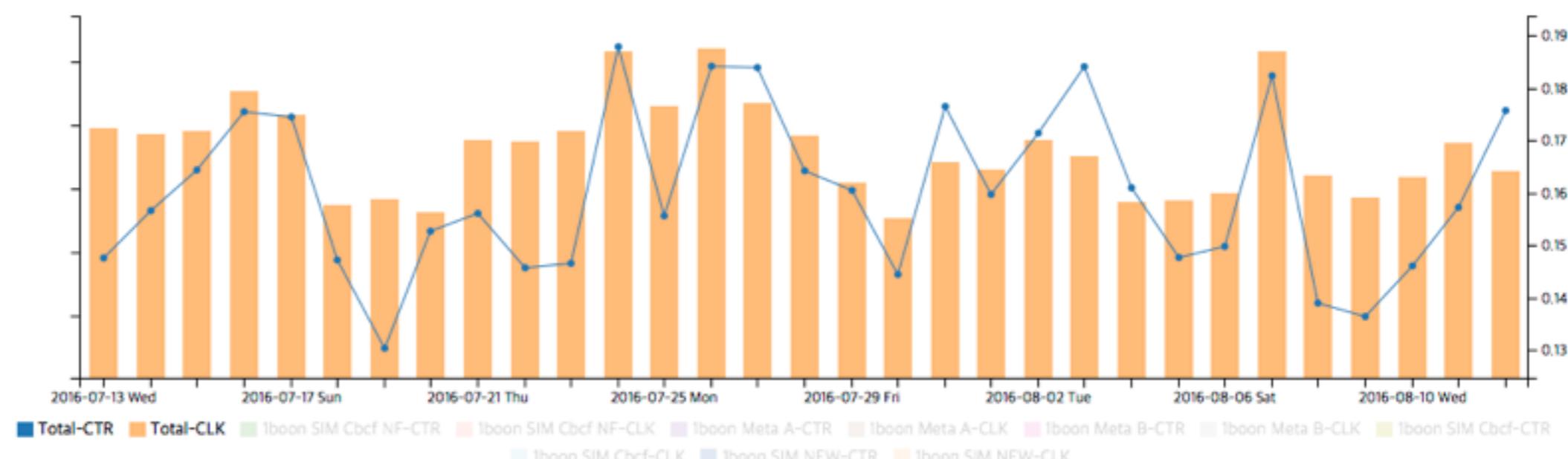
© Kakao Corp.

[Dashboard](#)[Statistics](#)[Toros](#)[1boon](#)[Brunch](#)[DaumApp](#)[DaumBlog+](#)[DaumEncyclopedia](#)[DaumMovie](#)[tvPot](#)[KakaoTV](#)[Melon](#)[KakaoMusic](#)[KakaoPage](#)[MediaDaum](#)[MediaDaum-PRMC](#)[Daum-CarMagazine](#)[TIP](#)[KakaoTalk](#)[Youtube](#)

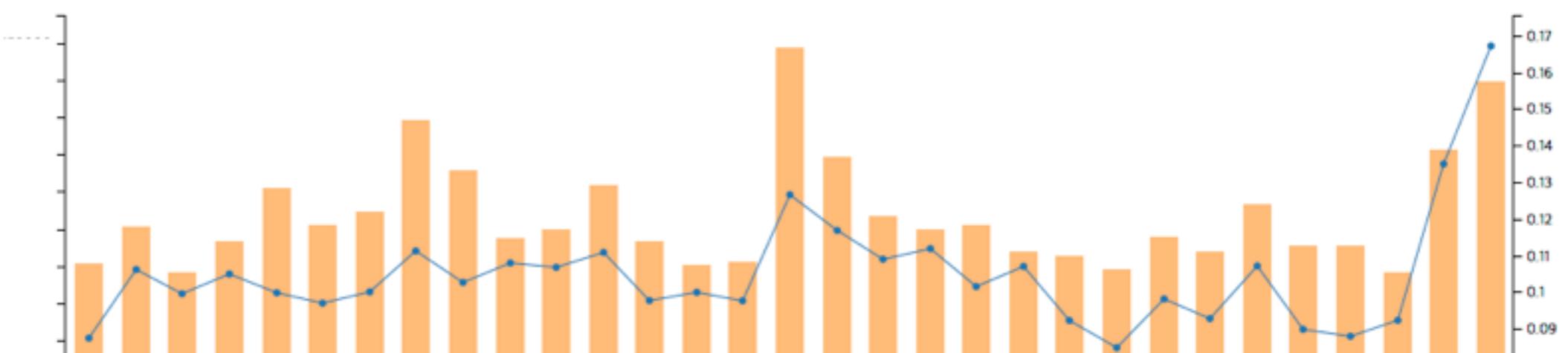
Overall

[1](#) [Month](#) [Daily](#) [2016/08/12\(금요일\)](#) [Submit](#)[3 Month Weekly](#) [3 Month Daily](#) [1 Week Hourly](#) [3 Day Hourly](#)

1boon 추천글 Click / Ctr



Brunch 연관글 Click / Ctr



[Dashboard](#)[Statistics](#)[Toros](#)[1boon](#)[Brunch](#)[DaumApp](#)[DaumBlog+](#)[DaumEncyclopedia](#)[DaumMovie](#)[tvPot](#)[KakaoTV](#)[Melon](#)[KakaoMusic](#)[KakaoPage](#)[MediaDaum](#)[MediaDaum-PRMC](#)[Daum-CarMagazine](#)[TIP](#)[KakaoTalk](#)[Youtube](#)

결과 Share

Query: ⚡ 행운을 빌어요 - Beginner's Luck 페퍼톤스 3780554



▶ Ready, Get Set, Go! - Colorful Express 페퍼톤스 1055513

Youtube 검색



▶ 캠퍼스 커플 (With 옥상달빛) - 하이파이브 (HIGH-FIVE) 페퍼톤스 4814828

Youtube 검색



▶ 100퍼센트 - Flavor 보드카 레인 1961266

Youtube 검색



▶ 취미는 사랑 (Album Ver.) - 가을방학 가을방학 2960593

Youtube 검색



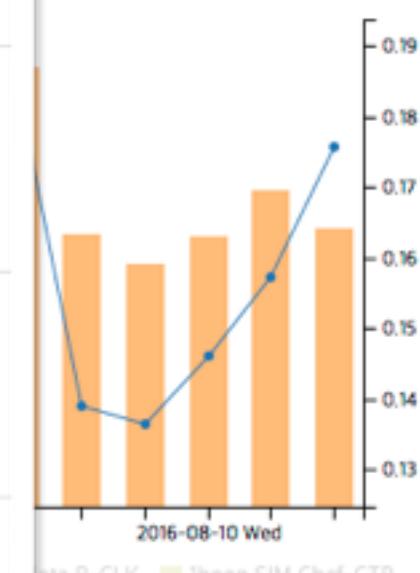
▶ 팝콘 - Life 데이브레이크 (DAYBREAK) 2669855

Youtube 검색

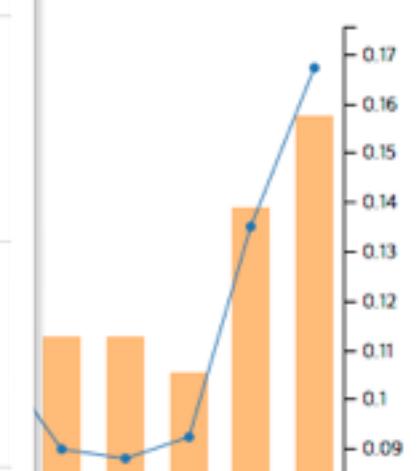


▶ 골든 글러브 - Just Pop 마이 앤트 메리 533075

Youtube 검색



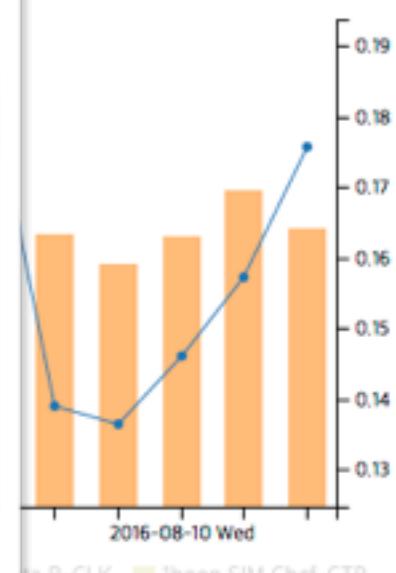
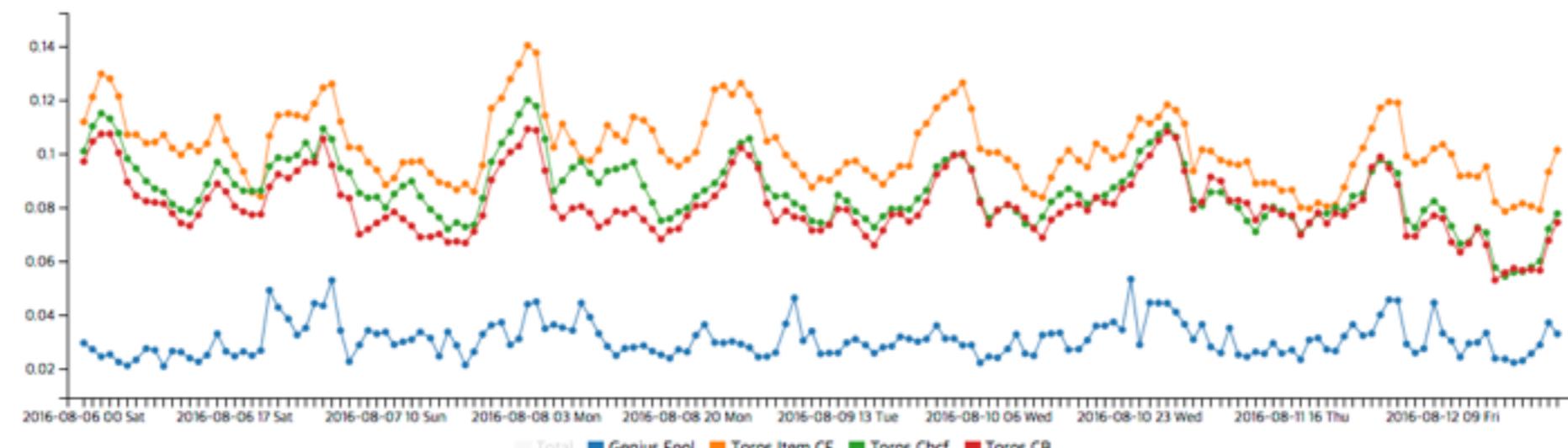
eta B-CLK 1boon SIM Cbcf-CTR



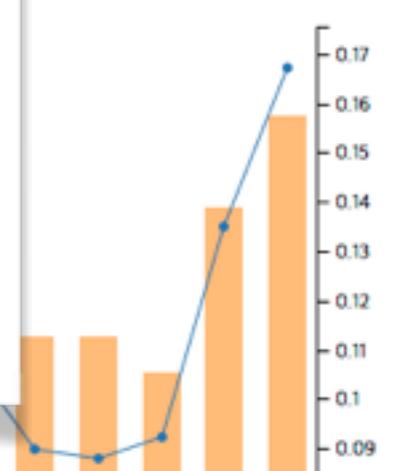
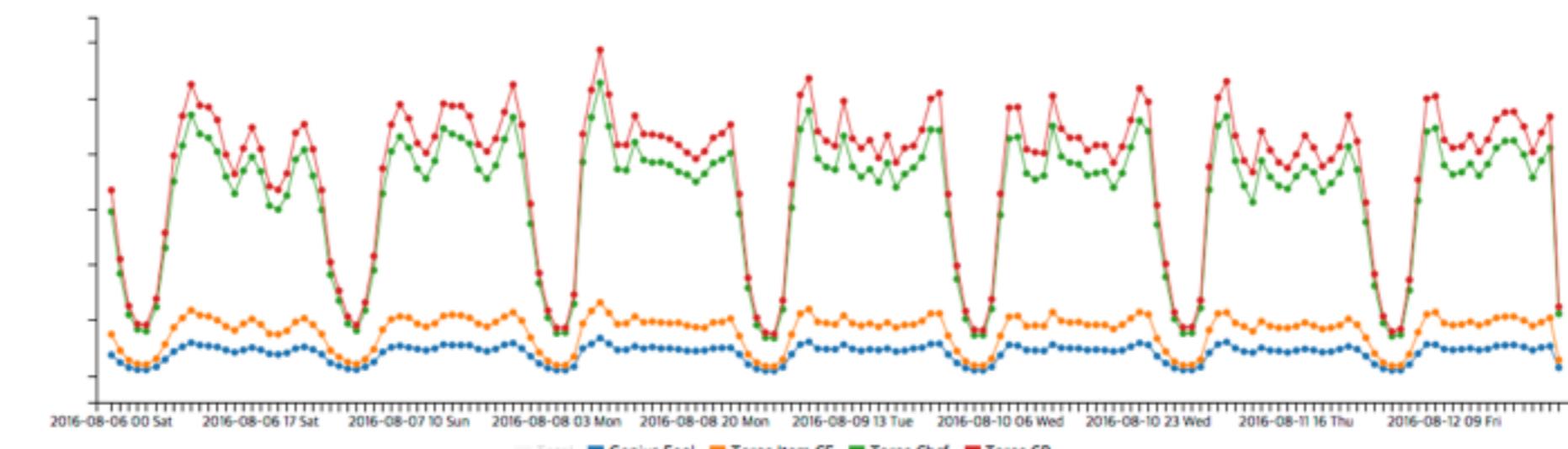
[Dashboard](#)[Statistics](#)[Toros](#)[1boon](#)[Brunch](#)[DaumApp](#)[DaumBlog+](#)[DaumEncyclopedia](#)[DaumMovie](#)[tvPot](#)[KakaoTV](#)[Melon](#)[KakaoMusic](#)[KakaoPage](#)[MediaDaum](#)[MediaDaum-PRMC](#)[Daum-CarMagazine](#)[TIP](#)[KakaoTalk](#)[Youtube](#)

1	Week	Hourly	2016/08/12(금요일)	<input type="button" value="Submit"/>
3 Month Weekly	3 Month Daily	1 Week Hourly	3 Day Hourly	

CTR



Impression



그런데 왜 파이썬이어야 했죠?

~~솔직히 처음에는 시스템이 이렇게 커질지 몰랐어요~~

Music, ItemCF, MovieLens,
Rating, Paper,...



kakao

© Kakao Corp.

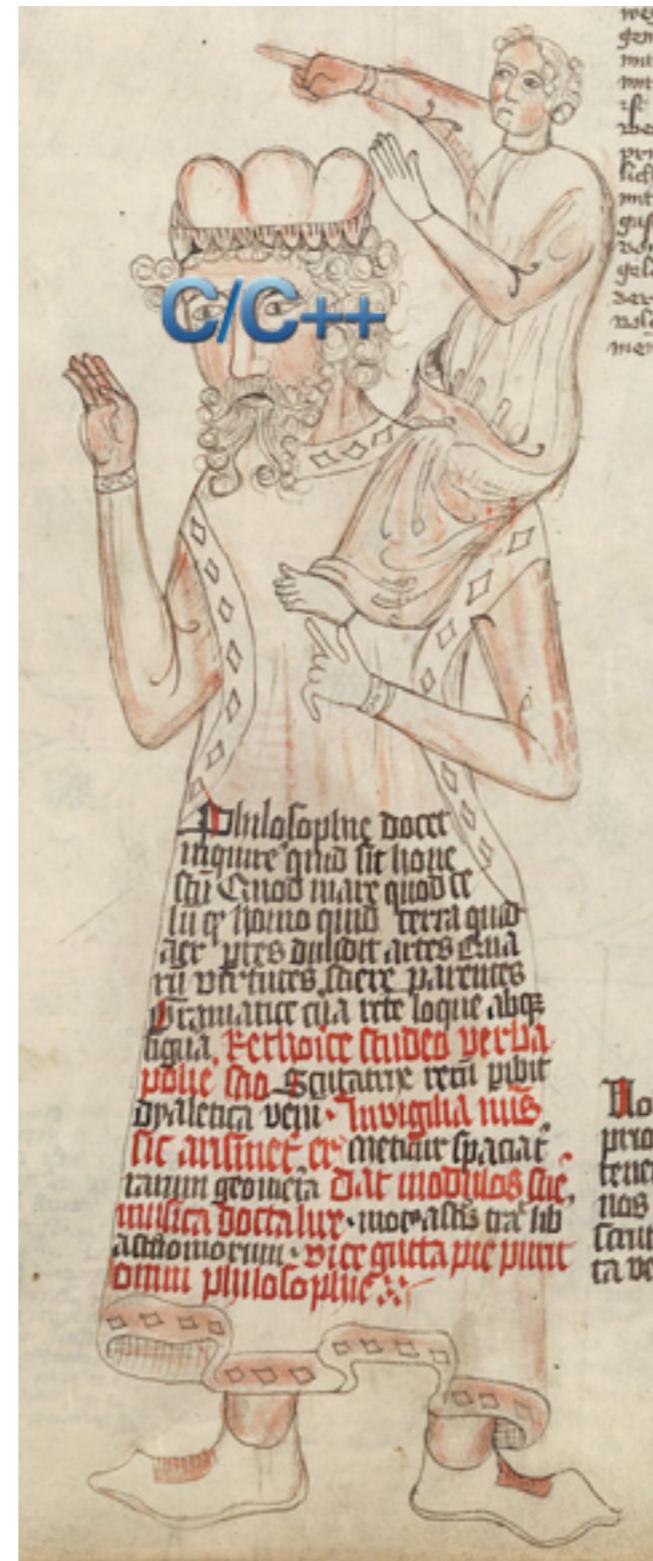
* <https://www.sitepoint.com/full-stack-developer/>

* <http://wallpapers.ws/10852-sea-water-stone.html>

그래도 우리가 계속 파이썬으로 간 이유

1. 빠른 개발속도와 비교적 저렴한 유지 비용
2. 폭넓은 라이브러리: front-end to machine learning
3. 활발한 커뮤니티
4. and..

standing on the shoulders of ...

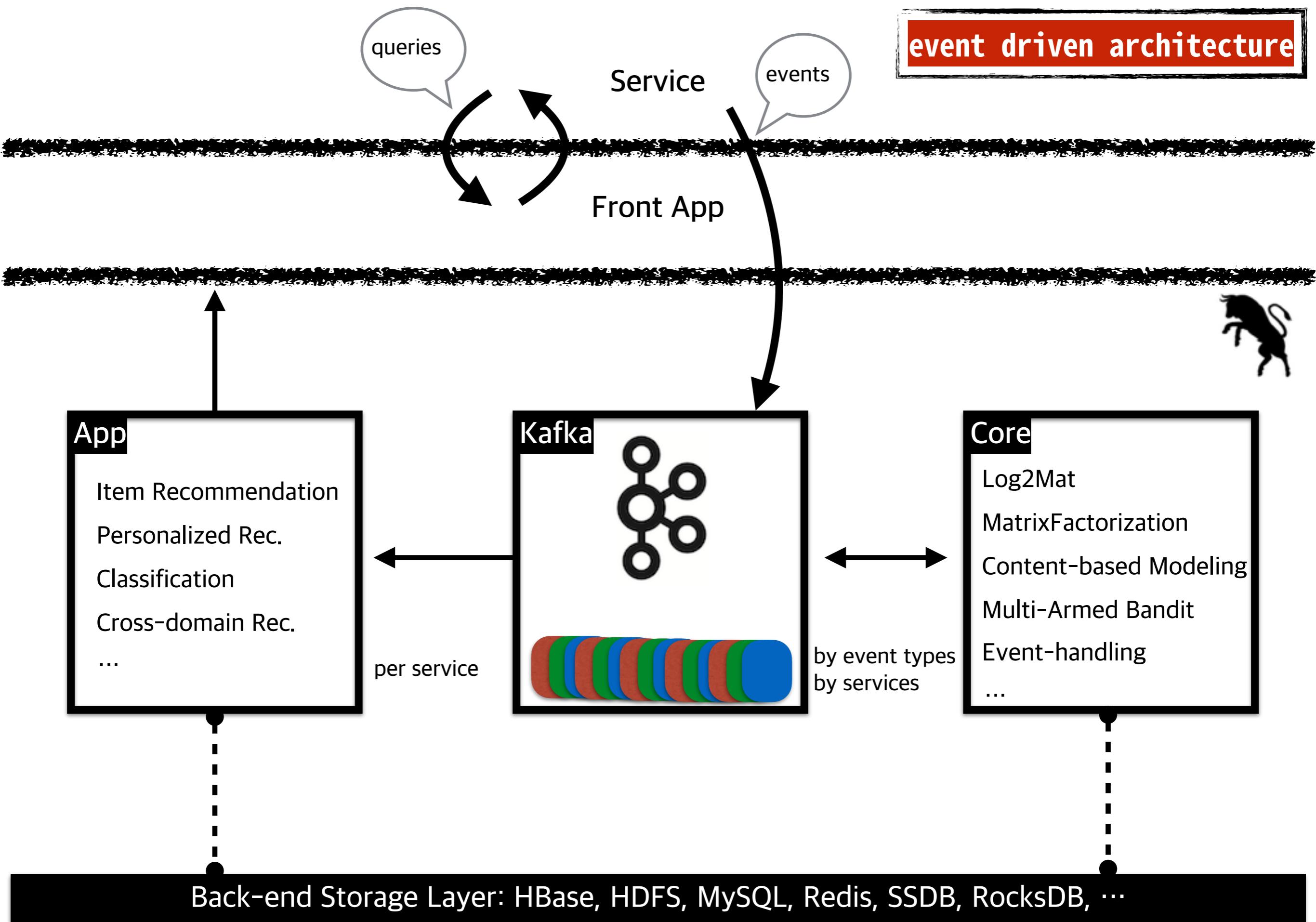




kakao

© Kakao Corp.

event driven architecture

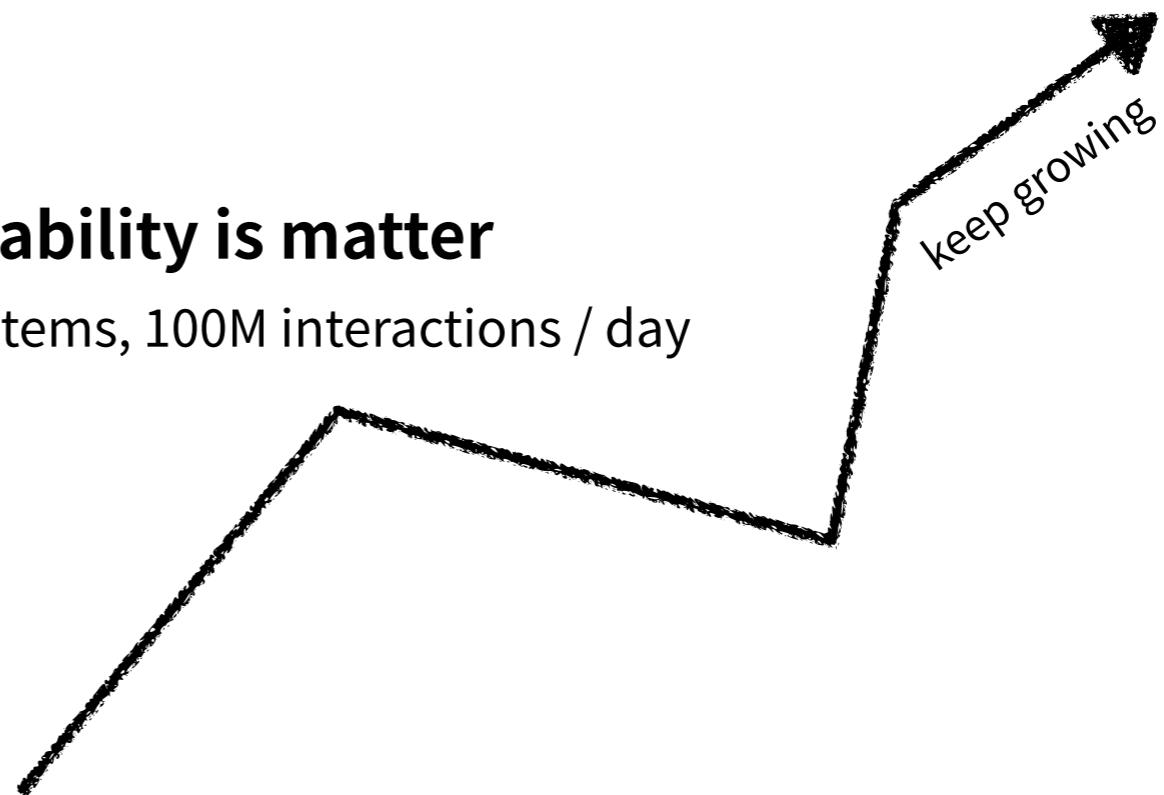


Episode: multiprocessing and Kafka

multiprocessing and Kafka

scalability is matter

+100M users/items, 100M interactions / day



multiprocessing and Kafka

```
>>> import multiprocessing
```



multiprocessing and Kafka

multi-processing은 원래 어려운 문제: load balancing, fault tolerance, scalability, ...
multi-threading, GIL의 기여도 만만치 않음(…)

multiprocessing and Kafka

Kafka as a Message/Job Queue

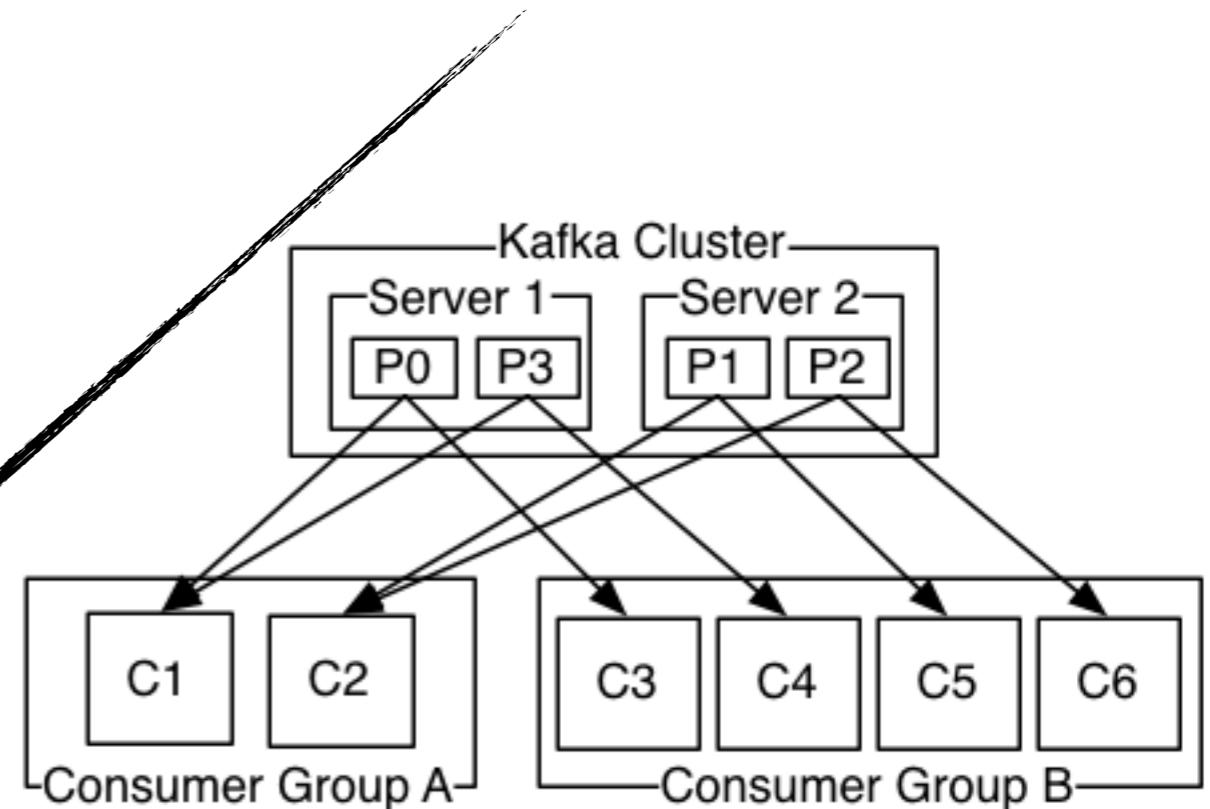
```
>>> import multiprocessing
```

```
>>> pool = multiprocessing.Pool(8)
```

```
>>> pool.map(do_some_work, jobs)
```

```
>>> Q = kafka.get_consumer(group_id='group_a')
```

```
>>> Q.get_message()
```



multiprocessing and Kafka

load balancing(distributed!)
fault tolerance
scalability
simpler/better code/flow

Kafka 토픽 설정에 의존성에 생긴다는 단점이 있지만,
변경이 잣지 않고 얻을 수 있는 장점에 비하면 소소함

Episode: performance

performance



니가 참아 파알못
이자나..

파이썬? 그거 스크립트
언어라서 느리지 않나요?

performance

틀린 얘기는 아니지만…

대부분의 경우는 네트워크, 디스크 I/O 병목, or 바로 당시…

정말로 CPU 연산이 병목인 경우라면…

엄마를 부르자

C/C++

performance

수치연산: Numpy, Scipy, ...

OpenBLAS만 잘 설치해도 10~20배의 성능 향상. 이것만 잘해도 수치 연산으로 마음 아플 일은 없어요.

다른 경우: MatrixFactorization, Kafka, ...

네이티브 C/C++ 라이브러리 쓰자

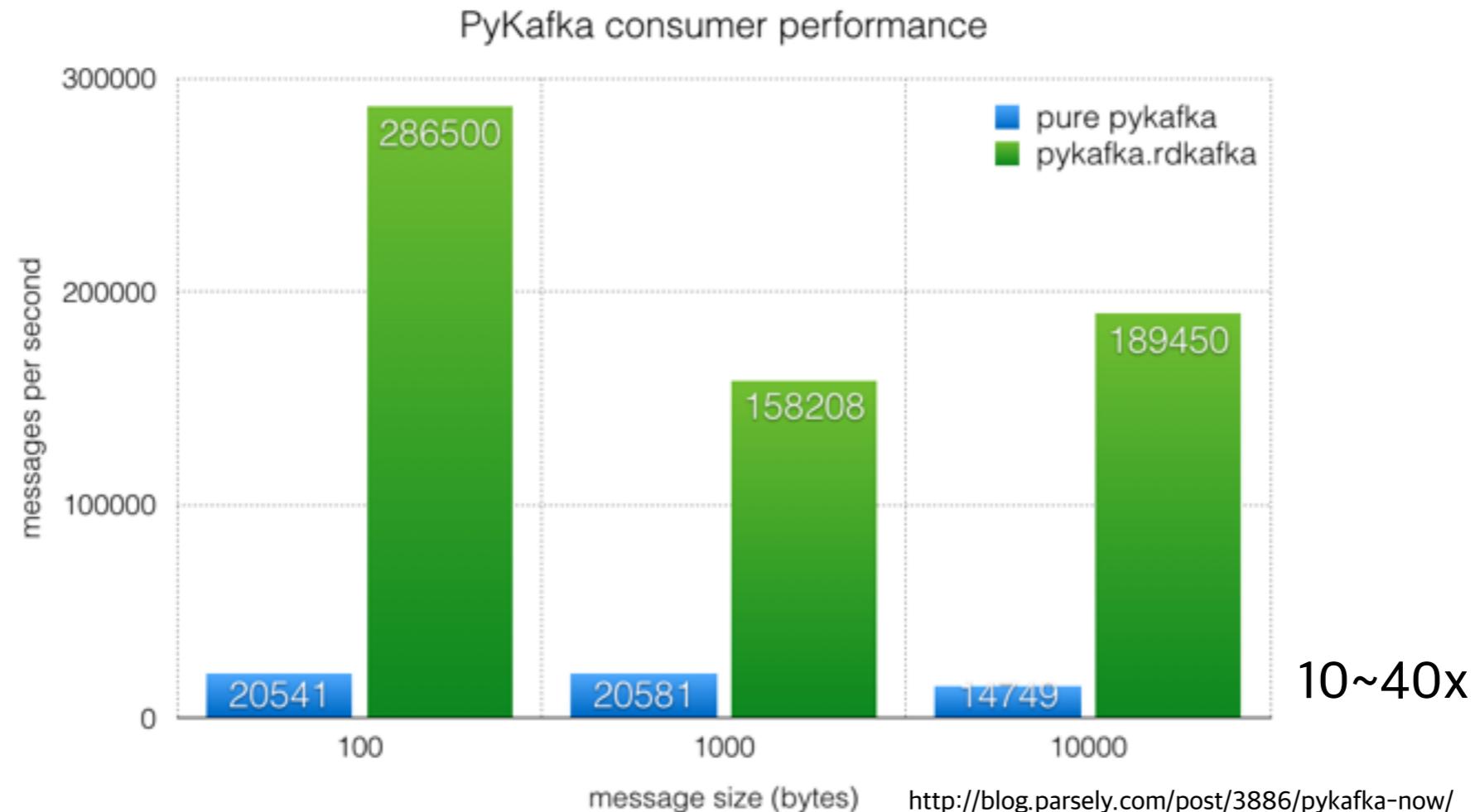


performance

Cython, SWIG, Python Extension...

난이도는 다소 높지만, 거의 오버 헤드가 없는 수준으로 C/C++ 코드 결합 가능

C/C++ 라이브러리 plug-in 가능 여부도 라이브러리 선택시 중요 체크 포인트!



Episode: caching

caching

“아버님 댁에 캐시 하나 놔드려야겠어요”
caching aka 무안단물



kakao

% 특정 종교 단체와 관계없습니다 %
© Kakao Corp.

caching

흔한 서비스 투입 리소스

8G Ram, 8 VCPU, 50GB SSD

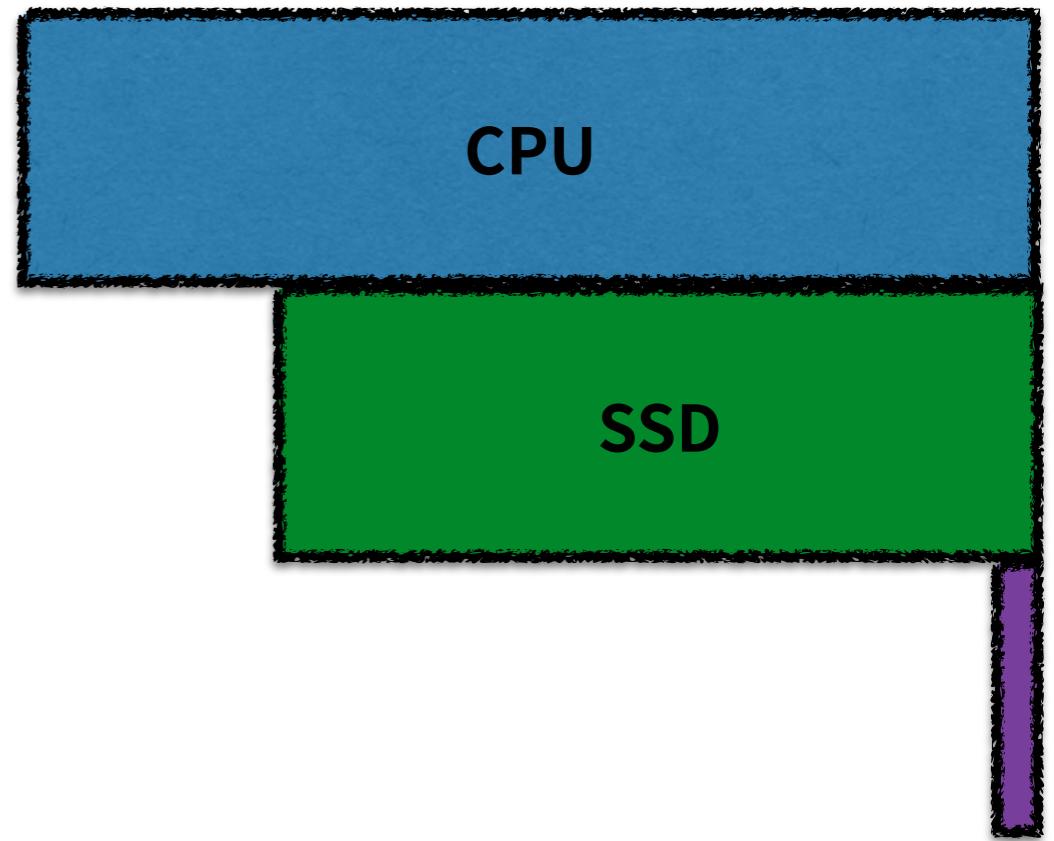
스케일, 모델

+100M items, +2~3 latent factor models, ...

유사 컨텐츠를 찾기 위한 연산 1회

dot products, ensemble, **filtering**, **wrapping**, **serving..** under 100ms

자원의 여유



“RAM? 그런 달달한 것이
남아있진 한가?”

caching: 캐시… 그거 대충하면 되지 않나요

1) 양도 많고…

```
>>> sys.getsizeof(u'') # 52  
>>> sys.getsizeof('') # 37  
>>> sys.getsizeof(0) # 24  
...  
>>> cPickle.loads(open('100M_kv.dict.cPickle').read()) # ???
```



caching: 캐시… 그거 대충하면 되지 않나요

2) 그냥 많고… (Remote Procedure Call)

1회 연산에 100 ~ 3,000 번의 캐시 조회 필요



caching



좋긴 한데, 뭐랄까 “감히 니가?”의 도도한 느낌

- 진입장벽이 낮음, 잘 알려짐
- 비쌈



An alternative to Redis, 100x Redis. ~~역시 대륙!~~

- leveldb



A persistent key-value store for fast storage environments

- leveldb
- SSD
- multi client

caching

RocksDB

자유로운 설정과 상황에 따라 리소스 배분이 용이하고, 다중 프로세스 환경에서 활용도가 높아 굉장히 요긴하게 사용됩니다.

- 메모리 관리
- 빠른 로딩(~ zero)
- 런타임 injection
- 자유도 높은 설정
- ...

Method	Cores	Elapsed(secs)
mput	1	93
sst	1	98
sst	4	47
sst	6	27
sst	12	22

- 32bytes key, 4bytes value
- 6.5M pairs
- not optimized(block cache, ...)



ETC

kakao

© Kakao Corp.

ETC

Python2 or Python3

kakao

© Kakao Corp.

ETC

Python2

kakao

© Kakao Corp.

Python2

최신 좋아합니다. 정말이에요. 믿어주세요.

정말 많은 라이브러리를 쓰고 있고, 개중에는 Python3가 지원하지 않는 것도 있어요.

ETC

Python2

최신



nsuke commented on Nov 4, 2015

The Apache Software Foundation member



I did some research on this.

TType.UTF8 was added years ago but currently not used by any language.

So using this now will result in incompatibility against any other implementation.

정말
어요.

도 있

I rebased [@c4urself](#)'s rebase and fixed both fastbinary (py2) and string/binary (py2 & py3) issues :
[#680](#)



asfgit pushed a commit that closed this pull request on Nov 6, 2015

THRIFT-1857 Python 3 Support ...

760511f



asfgit closed this in [760511f](#) on Nov 6, 2015



gadLinux pushed a commit to gadLinux/thrift that referenced this pull request on Mar 6

THRIFT-1857 Python 3 Support ...

67e6d14

kakao

© Kakao Corp.

Supervisor: A Process Control System

역사와 전통의…

Supervisor: A Process Control System

Supervisor is a client/server system that allows its users to monitor and control a number of processes on UNIX-like operating systems.

It shares some of the same goals of programs like *launchd*, *daemontools*, and *runit*. Unlike some of these programs, it is not meant to be run as a substitute for `init` as "process id 1". Instead it is meant to be used to control processes related to a project or a customer, and is meant to start like any other program at boot time.

Supervisor

- 중앙집중식 프로세스 관리
- 간결
- 확장성: XML-RPC, 프로세스 상태 체크, 시스템 모니터링, ...
- 검증됨(since 2004)

ETC

```
redirect_stderr = true
process_name=%(program_name)s
numprocs=1

[program:meta.ab.watcher]
command = python MetaWatcher.py ./MetaAB
directory = /daum/kskim/aurochs.app.git/1boon
stdout_logfile = ./Log/meta_ab_watcher.log
stdout_logfile_maxbytes = 50MB
stdout_logfile_backups = 3
redirect_stderr = true
process_name=%(program_name)s
numprocs=1

[program:meta.ab.picker]
command = python MetaPicker.py ./MetaAB
directory = /daum/kskim/aurochs.app.git/1boon
stdout_logfile = ./Log/meta_ab_picker.log
stdout_logfile_maxbytes = 50MB
stdout_logfile_backups = 3
redirect_stderr = true
process_name=%(program_name)s
numprocs=1

[eventlistener:ev]
command = python ev.py
directory = /daum/kskim/aurochs.app.git/
events=PROCESS_STATE,TICK_60

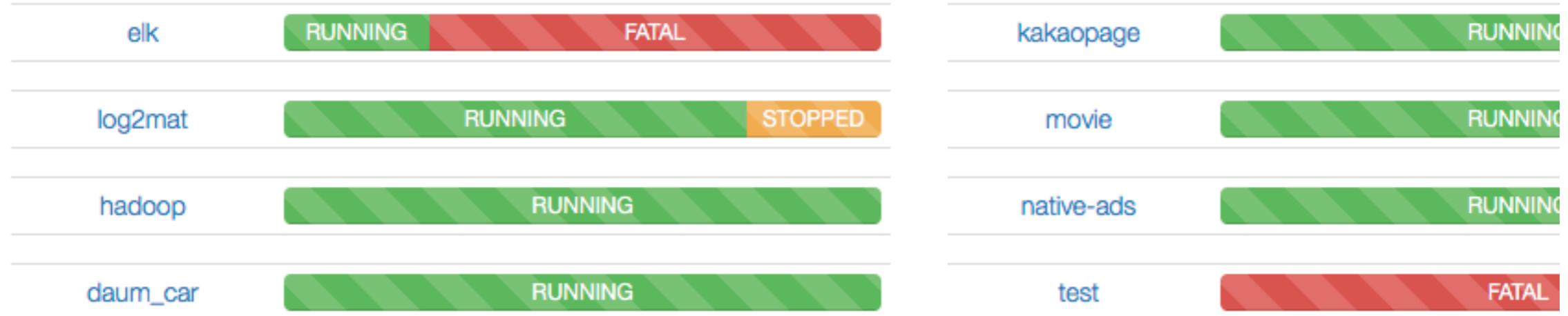
[supervisord]

[supervisorctl]

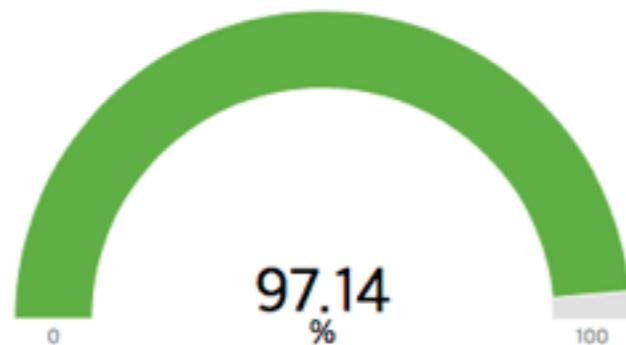
[inet_http_server]
port = *:23231
username = xxx
password = xxx

[rpcinterface:supervisor]
supervisor.rpcinterface_factory = supervisor.rpcinterface:make_main_rpcinterface
```

Overview



[brunch] BRUNCH 태그/유사글/작가/태그



kiev001 [3]		kiev001 [4]	
	RUNNING 10		RUNNING 14
	FATAL 0		FATAL 0
	STOPPED 1		STOPPED 0
bc-rec-tfidf1 [2]		bc-rec-tfidf2 [2]	
	RUNNING 3		RUNNING 3
	FATAL 0		FATAL 0
	STOPPED 0		STOPPED 0

ETC

Supervisor

Processes -

PID	Group	Name	Start	Status	Operation
13595	als	als	2016-08-02 21:35:36	RUNNING	Start Stop Restart View Log
21055	aurochs.app.1b.ad_doc_updater	aurochs.app.1b.ad_doc_updater	2016-07-15 11:05:08	RUNNING	Start Stop Restart View Log
28163	aurochs.app.1b.cf.wv	aurochs.app.1b.cf.wv	2016-08-01 12:54:02	RUNNING	Start Stop Restart View Log
8517	aurochs.app.1b.fb	aurochs.app.1b.fb	2016-07-04 16:11:10	RUNNING	Start Stop Restart View Log
15805	aurochs.app.1b.fe	aurochs.app.1b.fe	2016-07-13 05:32:52	RUNNING	Start Stop Restart View Log
14484	aurochs.app.1b.itest	aurochs.app.1b.itest	2016-07-13 05:25:14	RUNNING	Start Stop Restart View Log
15439	aurochs.app.1b.mp	aurochs.app.1b.mp	2016-07-13 05:30:50	RUNNING	Start Stop Restart View Log
14513	aurochs.app.1b.mp.dump	aurochs.app.1b.mp.dump	2016-07-13 05:25:18	RUNNING	Start Stop Restart View Log
873	aurochs.app.1b.rec.base	aurochs.app.1b.rec.base	2016-08-04 16:06:23	RUNNING	Start Stop Restart View Log

지금까지 우리가 파이썬을 써온 모습



새로운 문제 발견



Accepted

그래서 우리는 파이썬을 계속 쓸까?

ETC

Yes

kakao

© Kakao Corp.

Contacts

김광섭 (lucas.kim) / mrkwangsub@gmail.com

추천기술셀 @ kakao

we are hiring :)