



Introducing Brooklin

Lee Dongjin | dongjin@apache.org



배경: 자주 받는 질문들

- Kafka Cluster 간에 Mirroring을 해야 하는데 뭐가 좋나요?
 - Kafka → Kafka
- Kinesis 나 Pub/Sub으로 구현된 서비스를 Kafka로 옮겨야 하는데 어떻게 해야 하나요?
 - Kinesis → Kafka
 - Kafka → Kinesis
- Storage의 변경 사항을 받아보고 싶은데 + Storage 간에 동기화를 해야 하는데…
 - MySQL → Kafka
 - Kafka → BigQuery

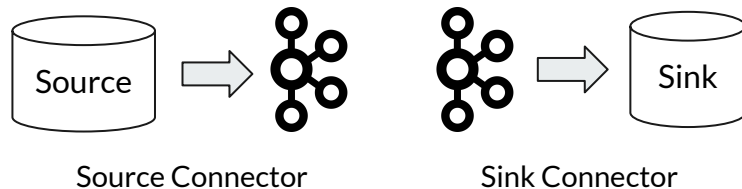


현재까지의 해결책 (1): Mirroring

이름	개발	유/무료	비고
MirrorMaker	Kafka Community	무료	<ul style="list-style-type: none">• Kafka Built-in
MirrorMaker 2	Kafka Community (Cloudera)	무료 (?)	<ul style="list-style-type: none">• Kafka Built-in• WIP (2.4.0)
Mirus	Salesforce	무료	-
uReplicator	Uber	무료	<ul style="list-style-type: none">• Chaperone 기반• 범용성...?
Replicator	Confluent	유료	-

현재까지의 해결책 (2): Migration & etc.

- Kafka Connect
 - “외부 Storage / Messaging System의 변경 사항을 Kafka Topic으로 보내 주거나, 그 반대의 작업을 해 주는 System”
 - “적당한 Connector만 있으면 뭐든지 끌어오거나 부을 수 있습니다.”
 - 웬만한 건 다 plugin 형태로 나와 있음.
 - 필요하면 다른 거 참고해서 만들면 됨.



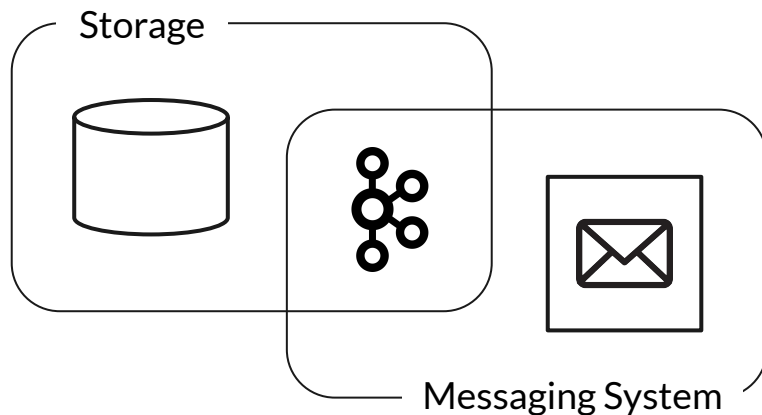


문제: 한두개라면 모르겠는데...

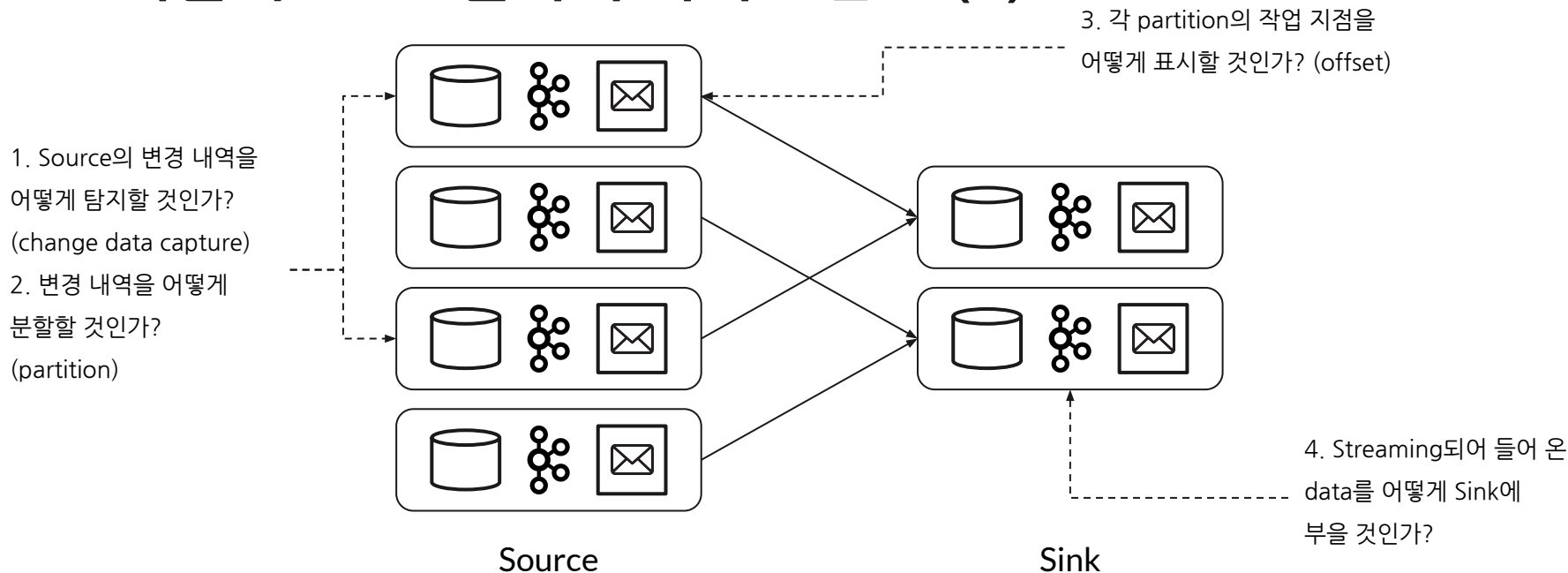
- "... 연결되어야 하는 Kafka Topic / Storage / Messaging System이 N개가 되면?"
- MirrorMaker
 - Point to Point
- Kafka Connect
 - 별도 Connect Cluster + Kafka Cluster
- 자원 활용...? 유지 관리...?
- 한 줄 요약: 문제가 산으로 갑니다.

해결책: 조금 멀리서 바라보면... (1)

- ... 본질적으로는 똑같은 문제
 - Source → Sink 로 Data Streaming을 해 주는 문제
 - Source, Sink: Storage 혹은 Messaging System



해결책: 조금 멀리서 바라보면... (2)





Linkedin Brooklin

- 다양한 Storage / Messaging System 사이의 실시간 data streaming을 위한 범용 분산 Framework
 - 안정성(reliable), 확장성(extensible), 규모 가변성(scalable)
 - Connector, Transport Provider, DataStream
- Linkedin에서 다양한 Storage / Messaging System들을 연결하기 위해 개발 (2016 ~)
 - <https://engineering.linkedin.com/blog/2017/10/streaming-data-pipelines-with-brooklin>
- 오픈 소스로 공개 (2019.6)
 - <https://engineering.linkedin.com/blog/2019/brooklin-open-source>



Connector & TransportProvider

- Connector
 - Storage / Messaging System의 변경 사항을 Data Stream으로 변환.
 - cf: KafkaMirrorMakerConnector: Kafka Mirroring 전용
- TransportProvider
 - Data Stream의 내용을 Storage / Messaging System에 씀.

System	Connector	TransportProvider
Directory	DirectoryConnector	DirectoryTransportProvider
File	FileConnector	-
Kafka	KafkaConnector	KafkaTransportProvider

DataStream

- "Source에서 Sink로의 Data Streaming 작업"
- 동적으로 정의할 수 있음.
 - Connector, TransportProvider 사용

이 둘은 새로운 시스템이 도입되지 않는 한 계속
그대로 사용할 수 있고...

DataStream = { identifier, Connector, TransportProvider, 기타 설정 }

... 따라서 이 둘만 잡아 주면 된다!



... 직접 보여드리겠습니다.

- Directory → Directory



결론

- Brooklin = Kafka Connect의 더 유연하고 일반적인 버전 + Mirroring도 할 수 있음.
 - cf: MySQL → BigQuery
- 아직 초기
 - 2019년 6월 첫 공개, 9월 현재 1.0.0.
 - File, Kafka 정도만 제공되는 중
 - MySQL 등: 현재 작업중 ...
 - Kafka Mirroring 정도라면...?



질문?

- Brooklin Project
 - code: <https://github.com/linkedin/Brooklin/>
 - wiki: <https://github.com/linkedin/Brooklin/wiki>
 - chat: <https://gitter.im/linkedin/brooklin>
- Brooklin Docker
 - <https://github.com/dongjinleekr/brooklin-docker>
 - <https://hub.docker.com/r/dongjinleekr/brooklin>