

[4장. 구조적 API 개요]

* Spark.

DataFrame, Dataset 처리 과정
정의하는 분산 프로그램
모델.

- 라임형 / 비라임형 API 개념 / 차이점
- 핵심 용어
- 구조적 API 해석 / 클래스의 실행 방식

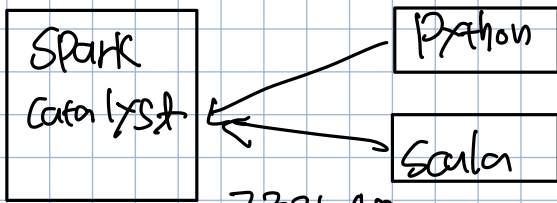
- DataSet
 - DataFrame
 - SQL 테이블 / 뷰
- 배치 ↔ 스트리밍
쉽게
변환 가능

4.1

DataFrame, DataSet 모두 권한을 가지는 분산 데이터
→ 데이터 생성을 위한 지면 연산 실행 계획.

4.2 구조적 데이터 타입 개요.

Spark 실행 계획 수립 / 처리에 사용하는 자체 데이터 타입정보를 가지고 있는 카탈리스트



구조적 API → Spark data type 연산.

Engine.
사용

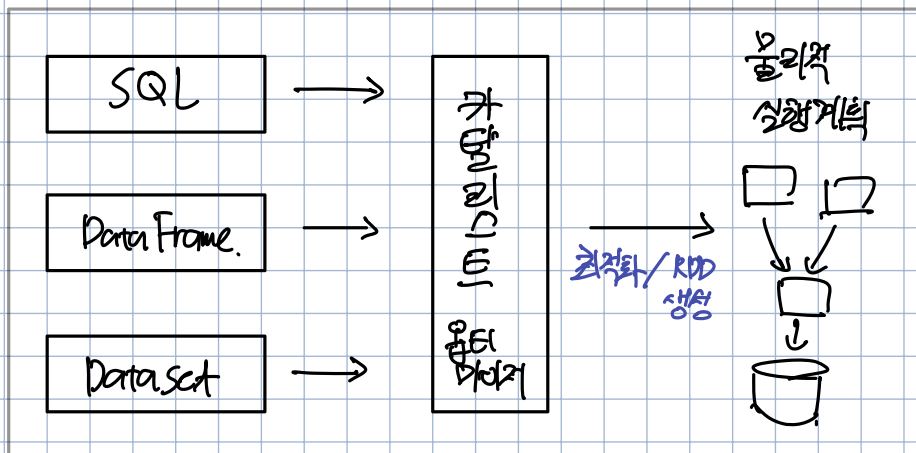
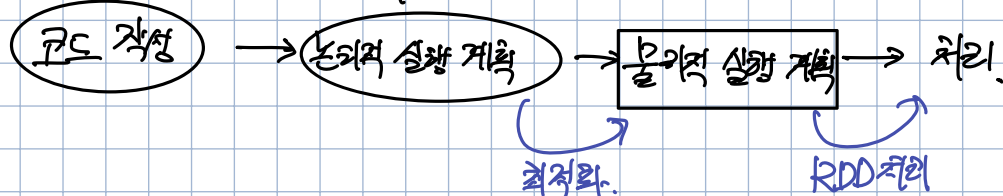
4.3.1 DataFrame vs Dataset

	스키아 불치여부 확인	지원 언어
DataFrame	권리형	Python, R → 권리형...
Dataset	권리형 타입	Java, Scala, JVM

✓ Row 타입 구성, 연산에 최적화된 인메모리 format

< 4.4. 관련 API 실행 과정 >

- Spark 코드 → 클러스터 실행 처리.



4.4.1 논리적 실행 계획, 물리적 실행 계획

