

아파치 스파크

데이터브릭스 커뮤니티

데이터브릭스 커뮤니티

- 워커 노드 없이 6GB로 제한된 단일 클러스터
- 커뮤니티 에디션 협업이 불가능한 기본 노트북

데이터브릭스 커뮤니티

문서 링크

<https://docs.databricks.com/getting-started/quick-start.html>

데모용

스파크 SQL

문장

```
CREATE TABLE diamonds USING CSV OPTIONS (path "/databricks-  
datasets/Rdatasets/data-001/csv/ggplot2/diamonds.csv", header  
"true")
```

```
SELECT color, avg(price) AS price FROM diamonds GROUP BY color  
ORDER BY COLOR
```

왜 스파크를 사용해야 하는가?

1. 추상화
2. 통합 플랫폼
3. 사용 편의성