

6.5 탐색 파일럿 실행 3단계-데이터 탐색&처리 4

데이터셋 추가

이번 6장에서 두 개의 데이터셋을 추가로 적재한다고 했다. 바로 “스마트카 마스터 데이터”와 “스마트카 차량용품 구매 이력 데이터”다. 해당 데이터는 외부 데이터로 간주하고 C://예제소스/bigdata2nd-master/CH06 경로에 미리 생성해둔 CarMaster.txt와 CarItemBuyList_202003.txt를 이용한다. 휴 파일 브라우저의 업로드 기능을 이용해 이 두 개의 파일을 적재하고 하이브의 External Table로 정의한다.

6.5 탐색 파일럿 실행 3단계-데이터 탐색&처리 4

데이터셋 추가

01. 휴의 파일 브라우저 기능을 실행한다. 그리고 디렉터리의 위치를 /pilot-pjt/collect로 이동한다.

파일 브라우저

파일 이름 검색 ⚙️ 작업 ✕ 휴지통으로 이동 ⬆️ 업로드 ➕ 새로 만들기

홈 / pilot-pjt / collect 휴지통

<input type="checkbox"/>	이름	크기	사용자	그룹	권한	날짜
<input type="checkbox"/>	↑		flume	supergroup	drwxr-xr-x	March 11, 2020 08:20 PM
<input type="checkbox"/>	.		flume	supergroup	drwxr-xr-x	March 11, 2020 08:20 PM
<input type="checkbox"/>	car-batch-log		flume	supergroup	drwxr-xr-x	March 11, 2020 08:20 PM

표시 45 / 1 항목 페이지 1 / 1 ⏮️ ⏪ ⏩ ⏭

그림 6.53 파일 브라우저를 이용한 데이터셋 추가 1

6.5 탐색 파일럿 실행 3단계-데이터 탐색&처리 4

데이터셋 추가

02. /pilot-pjt/collect 위치에서 “car-master” 디렉토리를 생성한 후 “car-master” 디렉터리로 이동한다.

파일 브라우저

파일 이름 검색 ⚙️ 작업 ✖ 휴지통으로 이동 ⬆ 업로드 1 + 새로 만들기

🏠 홈 / pilot-pjt / collect

2

- 파일
- 디렉토리

<input type="checkbox"/>	이름	크기	사용자	그룹	권한	날짜
<input type="checkbox"/>	↑		flume	supergroup	drwxr-xr-x	March 11, 2020 08:20 PM
<input type="checkbox"/>	.		flume	supergroup	drwxr-xr-x	March 21, 2020 02:29 AM
<input type="checkbox"/>	car-batch-log	3	flume	supergroup	drwxr-xr-x	March 11, 2020 08:20 PM
<input type="checkbox"/>	car-master		admin	supergroup	drwxr-xr-x	March 21, 2020 02:29 AM

그림 6.54 파일 브라우저를 이용한 데이터셋 추가 2

6.5 탐색 파일럿 실행 3단계-데이터 탐색&처리 4

데이터셋 추가

03. 업로드 기능을 이용해 C://예제소스/bigdata2nd-master/CH06/CarMaster.txt 파일을 HDFS의 /pilot-pjt/collect/car-master 경로에 업로드한다.

파일 브라우저

파일 이름 검색 작업 휴지통으로 이동 **1** 업로드 새로 만들기

홈 / pilot-pjt / collect / **car-master** 휴지통

	이름	크기	사용자	그룹	권한	날짜
<input type="checkbox"/>	↑		flume	supergroup	drwxr-xr-x	March 21, 2020 02:29 AM
<input type="checkbox"/>	2 .		admin	supergroup	drwxr-xr-x	March 21, 2020 02:33 AM
<input type="checkbox"/>	CarMaster.txt	126.9 KB	admin	supergroup	-rw-r--r--	March 21, 2020 02:33 AM

그림 6.55 파일 브라우저를 이용한 데이터셋 추가 3

6.5 탐색 파일럿 실행 3단계-데이터 탐색&처리 4

데이터셋 추가

04. 파일 브라우저에서 앞서 업로드한 CarMaster.txt 파일을 열어본다. /pilot-pjt/collect/car-master/CarMaster.txt 파일을 선택한다.

파일 브라우저

뒤로 홈

파일 편집

새로 고침

바이너리로 보기

다운로드

최종 수정 날짜
2020.03.21 오전 2
시 33분 +09:00

사용자
admin

그룹
supergroup

크기
126.95 KB

모드
100644

/ pilot-pjt / collect / car-master / CarMaster.txt

A0001	여 32 미혼 서울 프리랜서 1000 2009 F
A0002	남 53 미혼 충남 주부 2500 2015 A
A0003	여 62 기혼 대전 회사원 2500 2012 B
A0004	남 31 미혼 광주 공무원 2000 2010 D
A0005	남 67 미혼 대구 공무원 1700 2002 C
A0006	여 30 미혼 인천 전문직 2000 2016 D
A0007	남 61 미혼 전남 개인사업 1700 2003 E
A0008	여 20 미혼 충북 개인사업 1500 2013 G
A0009	여 60 미혼 경남 프리랜서 3500 2015 D
A0010	여 69 미혼 제주 개인사업 1200 2003 A
A0011	남 29 기혼 충남 주부 1000 2008 G
A0012	여 53 미혼 세종 학생 2500 2006 E
A0013	여 43 미혼 인천 전문직 1500 2016 E
A0014	남 63 미혼 대구 공무원 1500 2006 C
A0015	남 45 미혼 충남 프리랜서 1700 2012 F
A0016	여 17 미혼 강원 회사원 2000 2001 G
A0017	남 48 미혼 충북 프리랜서 1000 2010 B
A0018	여 70 기혼 경북 개인사업 2000 2004 H
A0019	남 32 기혼 인천 주부 1700 2004 C

그림 6.56 파일 브라우저를 이용한 데이터셋 추가 4

- “차량번호” | “성별” | “나이” | “결혼여부” | “지역” | “직업” | “차량용량” | “차량연식” | “차량모델”

6.5 탐색 파일럿 실행 3단계-데이터 탐색&처리 4

데이터셋 추가

05. 두 번째 추가 데이터셋인 CarltemBuyList_202003.txt 파일도 같은 방식으로 업로드한다. 휴의 파일 브라우저를 이용해 HDFS 상의 /pilot-pjt/collect 경로에 “buy-list” 디렉터리를 생성하고, C://예제소스/bigdata2nd-master/CH06/CarltemBuyList_202003.txt 파일을 /pilot-pjt/collect/buy-list 디렉터리로 업로드한다. 업로드 결과가 그림 6.57처럼 표시될 것이다.

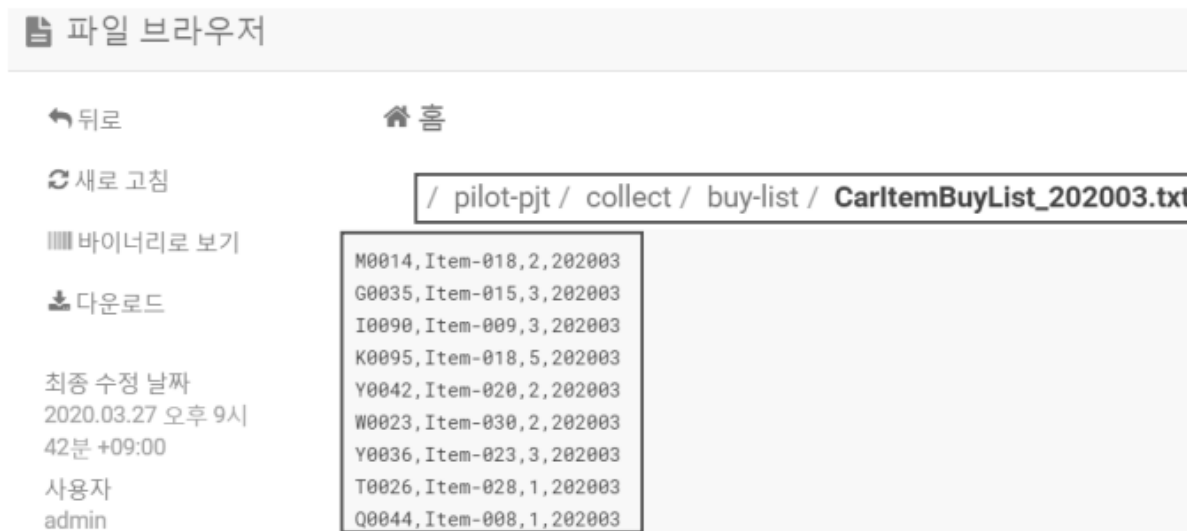


그림 6.57 파일 브라우저를 이용한 데이터셋 추가 5

스마트카의 상품 구매 이력 데이터셋의 스키마 구조는 “,”(콤마)로 분리되고 총 4개의 항목이 있다.

- “차량번호”, “구매 상품코드”, “만족도(1-5)”, “구매월”

6.5 탐색 파일럿 실행 3단계-데이터 탐색&처리 4

데이터셋 추가

06. 추가한 두 개의 데이터셋인 “스마트카 마스터”와 “스마트카 상품 구매 이력”을 하이브의 External 테이블로 정의한다. 휴 상단 메뉴의 [Query Editor] → [Hive]를 선택해 하이브의 쿼리 편집기로 이동한 후 스마트카 마스터 테이블인 SmartCar_Master를 생성한다.

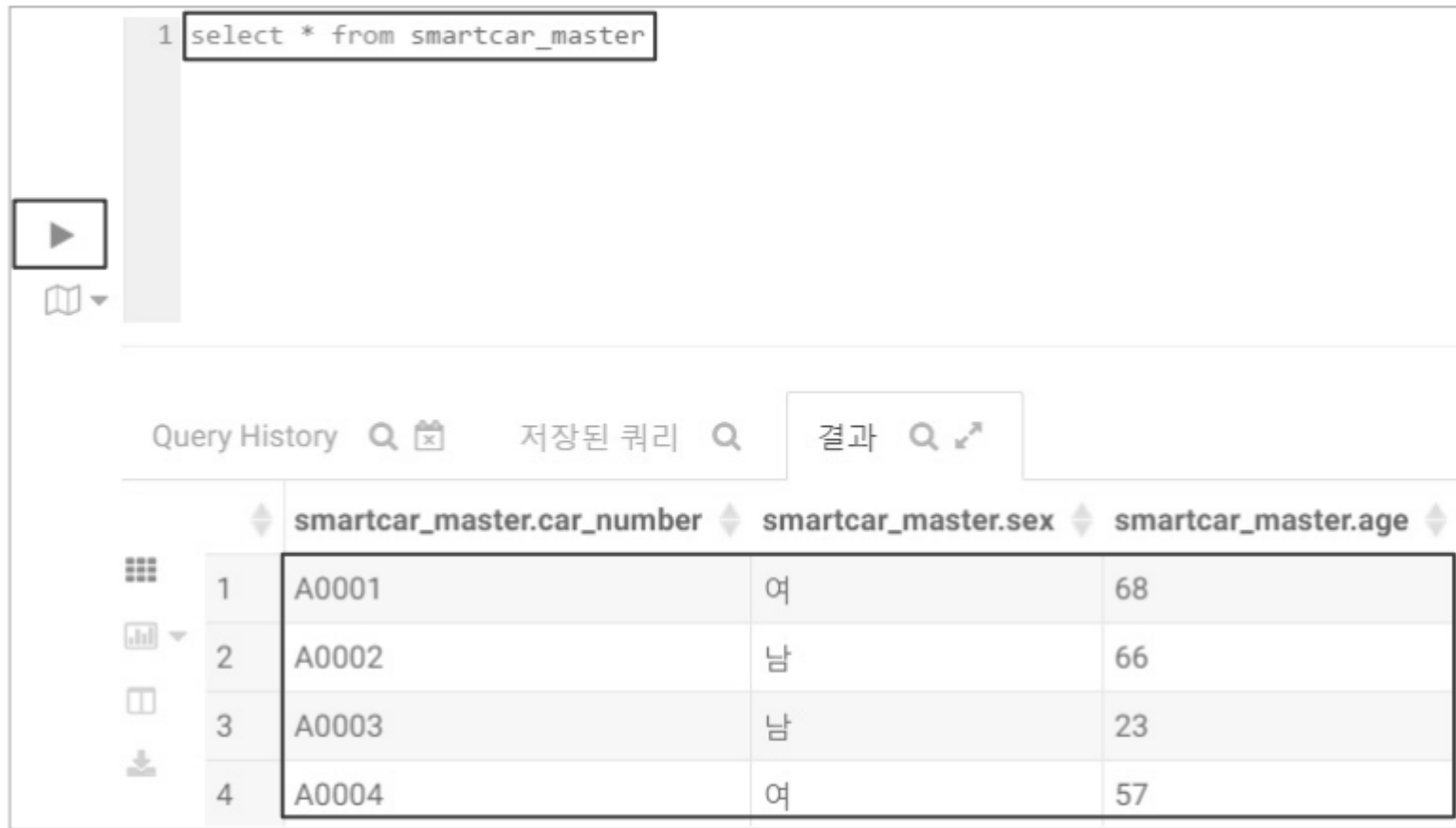
```
1 CREATE EXTERNAL TABLE SmartCar_Master (  
2   car_number string,  
3   sex string,  
4   age string,  
5   marriage string,  
6   region string,  
7   job string,  
8   car_capacity string,  
9   car_year string,  
10  car_model string  
11 )  
12 row format delimited  
13 fields terminated by '|'   
14 stored as textfile  
15 location '/pilot-pjt/collect/car-master'
```

그림 6.58 하이브 External 테이블 생성 - 스마트카 마스터 정보

6.5 탐색 파일럿 실행 3단계-데이터 탐색&처리 4

데이터셋 추가

07. “스마트카 마스터” 테이블의 데이터를 하이브 쿼리를 이용해 확인한다.



The screenshot shows a Hive query interface. At the top, a SQL query is entered in a text box: `1 select * from smartcar_master`. Below the query box is a play button icon. Underneath, there are tabs for 'Query History', '저장된 쿼리' (Saved Queries), and '결과' (Results). The '결과' tab is selected, displaying the query results in a table format. The table has three columns: 'smartcar_master.car_number', 'smartcar_master.sex', and 'smartcar_master.age'. There are four rows of data.


	smartcar_master.car_number	smartcar_master.sex	smartcar_master.age
1	A0001	여	68
2	A0002	남	66
3	A0003	남	23
4	A0004	여	57

그림 6.59 하이브 External 테이블 조회 – 스마트카 마스터 정보

6.5 탐색 파일럿 실행 3단계-데이터 탐색&처리 4

데이터셋 추가

08. “스마트카 차량용품 구매 이력”에 대한 하이브 테이블인 SmartCar_Item_BuyList를 생성한다.



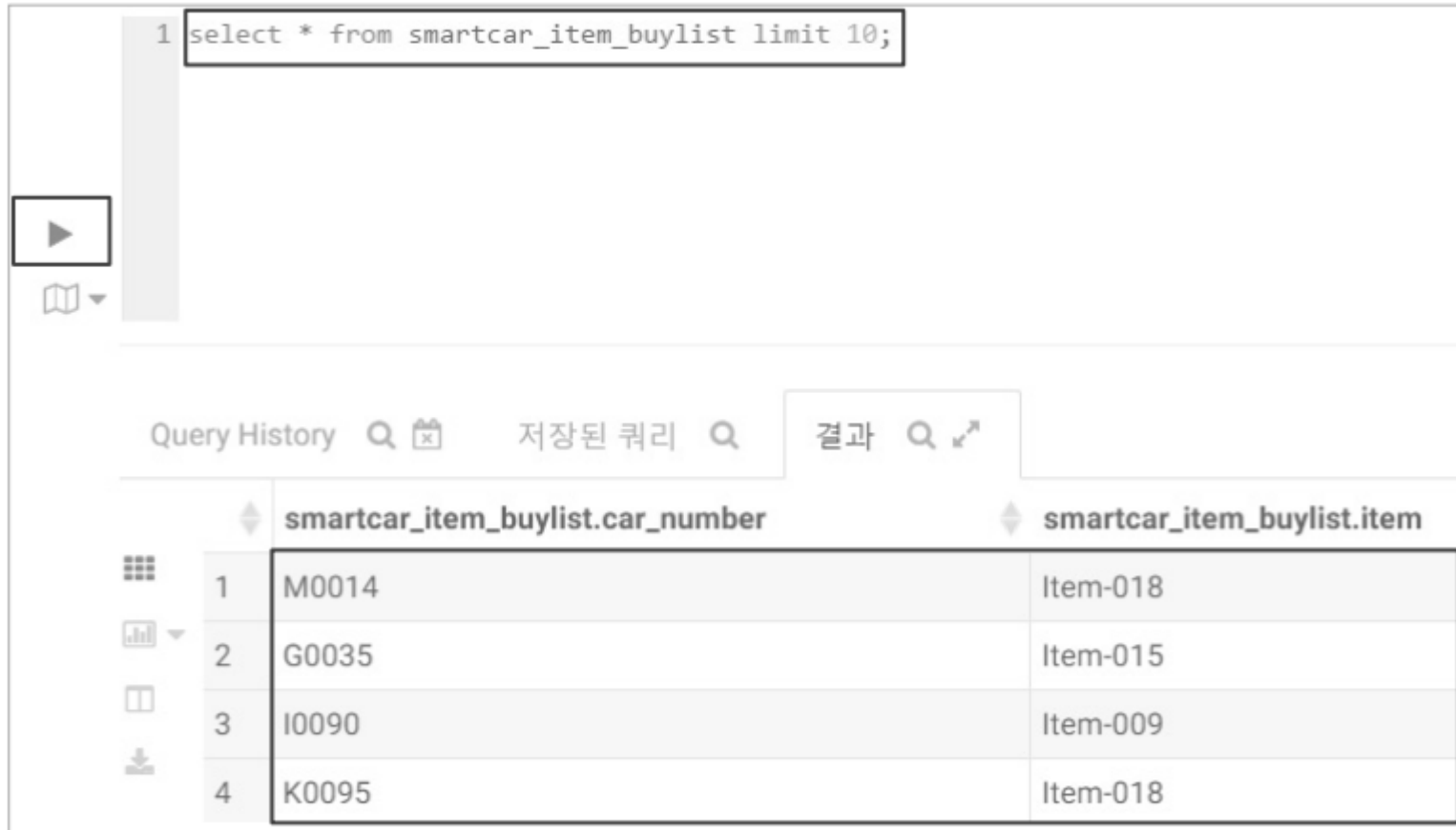
```
1 CREATE EXTERNAL TABLE SmartCar_Item_BuyList (  
2   car_number string,  
3   item string,  
4   score string,  
5   month string  
6 )  
7 row format delimited  
8 fields terminated by ','  
9 stored as textfile  
10 location '/pilot-pjt/collect/buy-list'
```

그림 6.60 하이브 External 테이블 생성 - 스마트카 상품 구매 이력 정보

6.5 탐색 파일럿 실행 3단계-데이터 탐색&처리 4

데이터셋 추가

09. “스마트카 상품 구매 이력” 테이블의 데이터를 하이브 쿼리를 이용해 확인한다.



The screenshot shows a Hive query interface. At the top, a SQL query is entered in a text box: `1 select * from smartcar_item_buylist limit 10;`. Below the query box is a play button icon. Underneath, there are tabs for 'Query History', '저장된 쿼리' (Saved Queries), and '결과' (Results). The '결과' tab is selected, displaying a table with two columns: 'smartcar_item_buylist.car_number' and 'smartcar_item_buylist.item'. The table contains four rows of data.

	smartcar_item_buylist.car_number	smartcar_item_buylist.item
1	M0014	Item-018
2	G0035	Item-015
3	I0090	Item-009
4	K0095	Item-018

그림 6.61 하이브 External 테이블 조회 – 스마트카 상품 구매 이력 정보

6.5 탐색 파일럿 실행 3단계-데이터 탐색&처리 4

 데이터셋 추가

실습