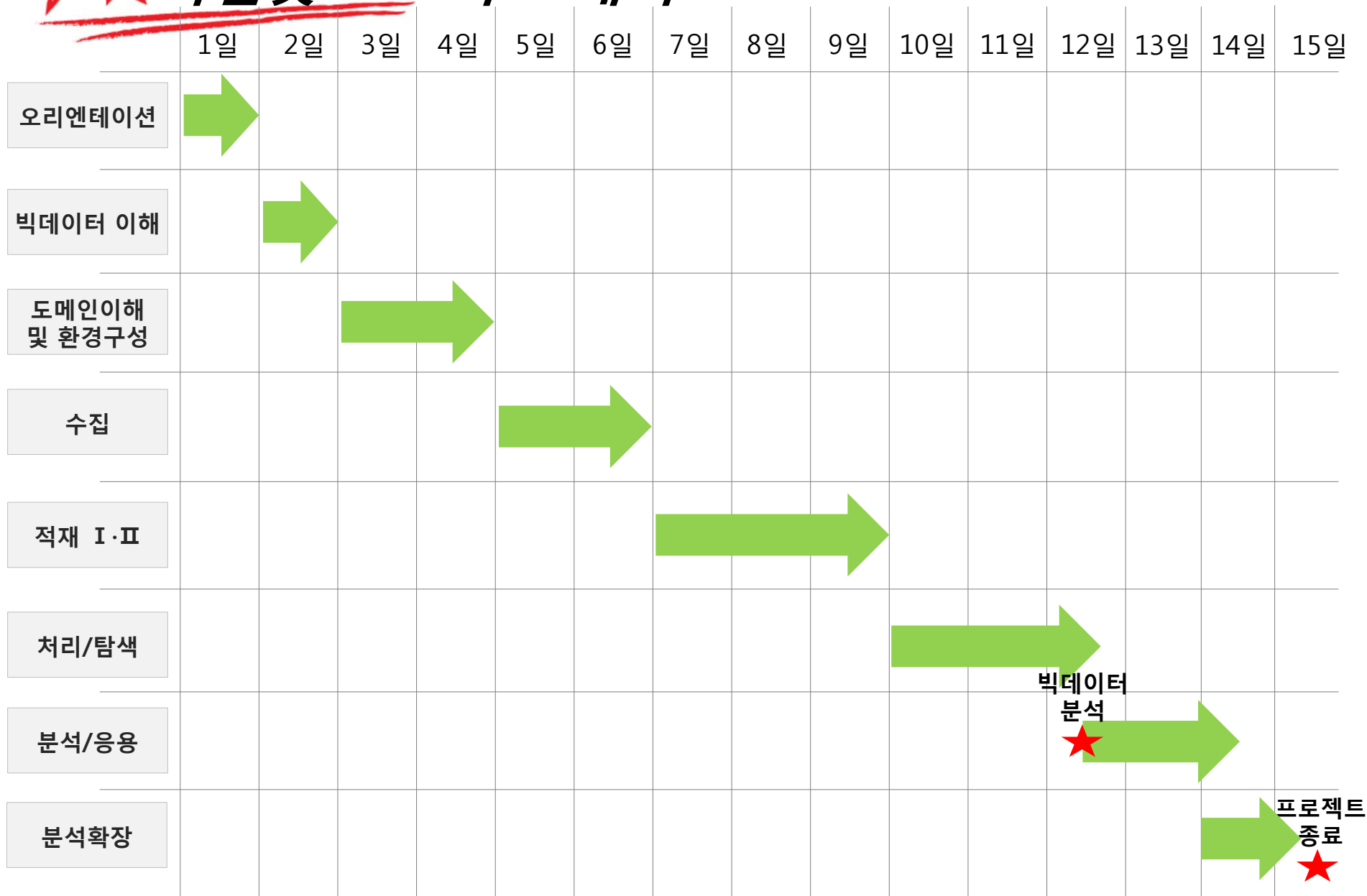


# 파일럿 프로젝트 단계

1. 빅데이터 이해하기
2. 빅데이터 파일럿 프로젝트 (이해 및 환경구성)
3. 빅데이터 수집
4. 빅데이터 적재 I (대용량 로그 파일 적재)
5. 빅데이터 적재 II (실시간 로그/분석 적재)
6. 빅데이터 탐색
7. 빅데이터 분석
8. 분석환경 확장

# 파일럿 프로젝트 계획



빅데이터  
분석

프로젝트  
종료

# 7. 빅데이터 분석

## 빅데이터 분석 소개

### 1. 빅데이터 분석 개요

빅데이터 분석에 대한 기본 정의와 활용 범위를 설명한다.



### 2. 빅데이터 분석에 활용하는 기술

빅데이터 분석에서 사용할 5가지 기술(임팔라, 제플린, 머하웃, 스파크ML, 스쿱)을 소개하고 각 기술별 주요 기능과 아키텍처, 활용 방안을 알아본다.



### 3. 분석 파일럿 실행 1단계 - 분석 아키텍처

스마트카의 빅데이터 분석과 관련한 요구사항을 구체화하고, 분석 요구사항을 해결하기 위한 파일럿 아키텍처를 제시한다.



### 4. 분석 파일럿 실행 2단계 - 분석 환경 구성

스마트카의 빅데이터 분석 아키텍처를 실제로 설치 및 환경을 구성한다. 임팔라, 스쿱, 제플린, 머하웃 순으로 설치하게 된다.



## 7. 빅데이터 분석

### 5. 분석 파일럿 실행 3단계 - 임팔라를 이용한 데이터 실시간 분석

임팔라로 스마트카 데이터셋을 인메모리 기반으로 실시간 조회 및 분석한다. 또한 하이브에서 사용했던 쿼리를 임팔라에서 실행 및 비교한다.



### 6. 분석 파일럿 실행 4단계 - 제플린을 이용한 실시간 분석

제플린의 웹 유저 인터페이스를 이용해 스마트카 운행 지역 분석을 위한 스파크 SQL을 작성 및 실행한다.



### 7. 분석 파일럿 실행 5단계 - 머하웃과 스파크ML을 이용한 데이터 마이닝

스마트카 데이터셋을 이용해 3가지 머신러닝을 실행한다. 추천(Recommendation), 분류(Classification), 군집(Clustering) 분석을 진행한다.



### 8. 분석 파일럿 실행 6단계 - 스콧을 이용한 분석 결과 외부 제공

탐색/분석된 스마트카 데이터셋을 외부 RDBMS로 제공한다.