

1. 문자열 'Big data processing'를 변수에 저장하고 그것을 한 문자씩 출력하는 프로그램을 작성하시오.

Answer:

```
str = 'Big data processing'

for s in str:
    print(s, end=',')
```

executed in 11ms, finished 16:00:05 2021-03-22

B,i,g, ,d,a,t,a, ,p,r,o,c,e,s,s,i,n,g,

2. 1 부터 50 까지 자연수의 합을 구하고 결과를 출력하는 프로그램을 작성하시오.

Answer:

```
result = 0

for i in range(1, 50 + 1):
    result += i

print(result)
```

executed in 8ms, finished 16:00:46 2021-03-22

1275

3. 피보나치 수열을 구하는 함수를 정의하고, 그 함수를 이용해 수열의 40 번째 값을 구하시오. 피보나치 수열은 첫째 항과 둘째 항이 1 이고 세번째 항부터는 바로 앞의 두 항의 합인 수열이다(1, 1, 2, 3, 5, 8, 13, ...).

Answer:

```
fibonacci = []

for x in range(0, 40):
    if x < 2:
        fibonacci.append(1)
    else:
        fibonacci.append(fibonacci[x - 2] + fibonacci[x - 1])

print(fibonacci[40 - 1])
```

executed in 13ms, finished 16:08:57 2021-03-22

102334155

4. 하둡과 하둡 에코시스템에 대해 중학생 동생에게 설명한다고 가정하고, 자신의 언어로 설명해 보시오.

Answer:

하둡은 2006 년 야후의 더그 커팅이 '넛치'라는 검색기능을 개발하는 과정에서 대용량의 체계적으로 구조화 되어있지 않은 데이터(비정형 데이터)를 기존의 관계형 데이터 베이스 기술로는 처리가 힘들다는 것을 깨닫고, 새로운 기술을 찾는 중 구글에서 발표한 GFS 와 MapReduce 관련 논문을 참고하여 개발하였습니다. 이후 아파치 재단의 오픈 소스로 공개 되었습니다.

하둡은 하나의 성능 좋은 컴퓨터를 이용하여 데이터를 처리하는 대신, 적당한 성능의 범용 컴퓨터 여러 대를 집산화(클러스터화)하고, 큰 크기의 데이터를 집단에서 병렬로 동시에 처리하여 처리 속도를 높이는 것을 목적으로 하는 분산처리를 위한 오픈소스 프레임워크라고 할 수 있습니다.

하둡의 주요 모듈은 총 5 개로 구성됩니다. (모듈이란 구성 단위라고 생각하시면 됩니다.)

- Hadoop Common  
하둡의 다른 모듈을 지원하기 위한 공통 컴포넌트 모듈
- Hadoop HDFS  
분산저장을 처리하기 위한 모듈  
여러 개의 서버를 하나의 서버처럼 묶어서 데이터를 저장
- Hadoop YARN  
병렬처리를 위한 클러스터 자원관리 및 스케줄링 담당
- Hadoop Mapreduce  
분산되어 저장된 데이터를 병렬 처리할 수 있게 해주는 분산 처리 모듈
- Hadoop Ozone  
하둡을 위한 오브젝트 저장소

하둡의 장단점

- 장점  
오픈소스로 라이선스에 대한 비용 부담이 적음  
시스템을 중단하지 않고, 장비의 추가가 용이(Scale Out)  
일부 장비에 장애가 발생하더라도 전체 시스템 사용성에 영향이 적음(Fault tolerance)  
저렴한 구축 비용과 비용대비 빠른 데이터 처리  
오프라인 배치 프로세싱에 최적화
- 단점  
HDFS 에 저장된 데이터를 변경 불가  
실시간 데이터 분석 같이 신속하게 처리해야 하는 작업에는 부적합  
너무 많은 버전과 부실한 서포트  
설정의 어려움