

파이프라인 기법을 통한 워크플로우 시스템의 처리율 향상

김대순, 윤찬현
한국과학기술원 전기 및 전자공학과
e-mail:{sundae21, chyoun}@kaist.ac.kr

Improvement on Throughput of Workflow System through Pipelining Scheme

Daesun Kim, Chan-Hyun Youn
Dept. of Electrical Engineering, KAIST

요 약

워크플로우 시스템은 시스템 구성과 관리, 운용에 있어서 편의성을 갖기 때문에 점차 적용 범위를 늘려가고 있는 추세이다. 하지만 워크플로우는 시스템의 효율성을 고려하지 않기 때문에, 분산 컴퓨팅 시스템과 같은 경우 워크플로우 도입에 따른 시스템 효율성 저하가 발생할 수 있다. 이러한 문제를 해결하고자 본 논문에서는 워크플로우 시스템에 파이프라인 기법을 도입하여 시스템의 전체 사용률을 높이고, 처리율을 향상시키는 방법을 제안한다.

1. 서론

워크플로우 시스템은 시스템 구성과 관리, 운용에 있어서 상당한 편의성을 갖고 있다. IT 기술에 관한 전문 지식이 있는 사용자가 아니더라도, 워크플로우를 조작하여 손쉽게 시스템을 구성하고 원하는 동작을 얻어낼 수 있다. 이러한 장점 때문에, 워크플로우 시스템은 비즈니스 영역에서부터 시작하여 과학 분야 등으로 점차 적용 범위를 늘려가고 있다. 로봇이나 자동차와 같은 시스템에도 워크플로우를 이용한 소프트웨어를 적용하려는 움직임도 일어나고 있다[1].

하지만 워크플로우는 작업의 프로세스를 정의할 뿐, 워크플로우에 따라 구현될 시스템의 효율성은 고려하지 않았다. 워크플로우 도입으로 인해 발생하게 될 시스템의 효율성 저하는 불가피한 것으로 인지되고 있으며, 이는 워크플로우 도입을 가로막는 요인 중 하나이다.

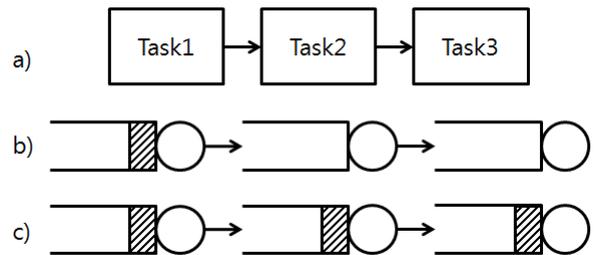
이러한 문제를 해결하고자 본 논문에서는 워크플로우 설계 단계에서 파이프라인 스테이지(Pipeline stage)를 설정하고, 워크플로우 시스템에 파이프라인 기법을 적용하여 시스템의 처리율을 향상시키는 방법을 제안한다.

2. 파이프라인 기법 도입 워크플로우 시스템

2.1. 적용 대상 시스템

본 논문에서 파이프라인 기법의 적용 대상이 되는 시스템은, 각각의 태스크(Task)를 서로 다른 노드(Node)에서 담당하여 처리하는 형태의 워크플로우 시스템이다. 파이프라인 기법을 적용하려면 각 태스크가 독립적으로 수행될 수 있어야 하기 때문에 이러한 조건이 따르게 된다.

위의 조건을 만족하는 노드 형태의 예를 들면, 분산 컴퓨팅 시스템의 개별 머신[2]이나, 웹서비스, 쓰레드, 또는



(그림 1) a) 워크플로우 예시

b) 파이프라인 기법 도입 전 워크플로우 시스템

c) 파이프라인 기법 도입 워크플로우 시스템

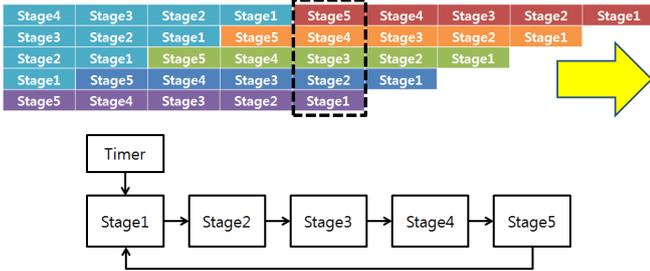
프로세스와 같은 형태가 될 수 있다.

2.2. 파이프라인 적용 목적

(그림 1-a)와 같은 워크플로우를 따라 시스템을 구성할 때, 파이프라인 기법을 적용하지 않은 기존의 워크플로우 시스템은 (그림 1-b)와 같이 Task1을 처리하는 동안 Task2, Task3를 처리하는 노드는 유휴상태로 대기하게 된다. 파이프라인 기법을 적용하는 목적은, 이와 같이 워크플로우 시스템이 하나의 태스크를 처리하고 있는 동안 태스크 처리에 관여하지 않는 나머지 노드들이 유휴상태로 돌입하는 것을 막고, 이어지는 다른 작업들을 처리하게 함으로써 (그림 1-c)의 시스템과 같이 시스템 전체의 사용률을 높여 처리율을 향상시키는 것이다.

2.3. 파이프라인 스테이지 설정

파이프라인 스테이지를 설정하는 데에 있어서 가장 먼저 고려해야 할 부분은 각 노드의 태스크 처리 시간과 워크플로우의 형태이다.



(그림 2) 5-스테이지 파이프라인 워크플로우 시스템 예시

워크플로우의 형태에서 문제가 되는 부분은 분기(Split)와 병합(Merge)이 일어나는 위치이다. 분기와 병합으로 인해 워크플로우 상에서 파이프라인 스테이지를 설정하기가 어려워질 수 있기 때문이다. 본 논문에서는 워크플로우의 분기와 병합에 의한 문제를 고려하지 않고, (그림 1-a)와 같이 태스크가 순차적으로 수행되는 형태의 워크플로우를 사용할 것이다.

(그림 1-c)의 시스템은 각각의 태스크가 하나의 파이프라인 스테이지로 설정된 상태이다. 이 경우, 모든 태스크의 처리가 끝나야 각각의 결과가 다음 노드로 전달되어 이후의 태스크를 처리할 수 있다. 이러한 상황에서 Task3의 처리 시간이 다른 태스크보다 길다고 하면, 다른 태스크들은 처리가 끝나도 Task3의 처리가 끝날 때까지 다음 노드로 넘어가지 못하고 대기해야 하므로 파이프라인 기법의 적용 효과가 떨어지게 된다. 이처럼 파이프라인 스테이지들의 처리 시간 사이에 편차를 최소한으로 줄여야 노드들이 유휴상태로 대기하는 시간을 최소로 줄일 수 있어 시스템의 전체 사용률을 높일 수 있으므로, 파이프라인 스테이지를 설정할 때 각 노드의 태스크 처리 시간을 고려하여야 한다.

2.4. 시스템 동작

파이프라인 기법을 적용한 워크플로우 시스템은 (그림 2)와 같이, 항상 파이프라인 스테이지와 같은 수의 작업을 동시에 처리한다. 이러한 동작을 위해 처음에는 타이머가 작동하여 최대 스테이지 처리 시간 간격으로 작업을 하나씩 실행시켜서 파이프라인에 작업을 채워나간다. 그래서 파이프라인 스테이지와 같은 수의 작업이 수행되고 나면, 그 이후로는 마지막 스테이지의 작업이 종료되는 시점에 새로운 작업을 실행시킨다.

이와 같이 시스템이 동작할 때, 각 파이프라인 스테이지의 전체 수행 시간에 차이가 생기면 한 노드가 태스크 처리를 마치기 전에 뒤이어 오는 작업의 처리 요청이 먼저 들어올 수 있다. 따라서 각 노드는 작업 처리 요청을 받아들여 버퍼(Buffer)가 반드시 필요하다.

3. 결론

워크플로우 시스템은 시스템 구성, 관리, 운용에 편의성을 제공하여 점차 적용 범위를 넓혀가고 있다. 하지만 적

용 범위의 확장과 함께 워크플로우 시스템의 효율성 문제도 대두되고 있다. 특히 분산 컴퓨팅 환경에서 워크플로우 시스템의 효율성 문제가 두드러지게 나타나며, 본 논문에서는 이를 해결하고자 파이프라인 기법 도입을 제안하였다. 본 논문에서 제안한 방법은, 태스크가 순차적으로 수행되는 형태의 워크플로우를 대상으로, 워크플로우 상에서 각 태스크의 수행 시간을 고려하여 파이프라인 스테이지를 설정하고, 이를 바탕으로 파이프라인을 구성하여 전체 시스템의 사용률을 높이며 결과적으로 시스템의 처리율을 높이는 것이다. 분기와 병합 등으로 인해 복잡한 형태를 가지는 워크플로우에 대해서 파이프라인 스테이지를 설정하는 방안은 향후 연구 주제로 남아있다.

Acknowledgement

본 논문은 2012년도 정부(지식경제부)의 재원으로 한국연구재단-차세대 생명정보 분석을 위한 생명정보학 플랫폼 개발[2012-10040231]과 BK21 사업의 지원을 받아 수행된 연구임.

참고문헌

[1] Wei-Chih Chen, Chi-Sheng Shih, "ERWF: Embedded Real-Time Workflow Engine for User-centric Cyber-Physical Systems," IEEE 17th International Conference on Parallel and Distributed Systems, 2011.
 [2] 김대순, 한영주, 강동기, 윤찬현, "Plug & Play 지원형 차량용 미들웨어," 제 36회 한국정보처리학회 춘계학술발표대회 논문집, 제 18권 제 2호, 2011.
 [3] Timothy M. McPhillips, Shawn Bowers, "An Approach for Pipelining Nested Collections in Scientific Workflows," ACM SIGMOD Record, Vol. 34, No. 3, 2005.