프로젝트 중간보고서

네임노드 최적화를 통한 마스터 메모리 절감

텍스트, 엠블럼, 로고, 원이(가) 표시된 사진

자동 생성된 설명]

**팀: 6조 이석진, 문무현**

**제출 일시: 2024 - 05 - 18**

목차

1. 팀원 소개

2. 작품 소개

2.1 핵심 문제

2.2 해결 방법

2.3 SW 설계

2.3.1 기능 요구사항

2.3.2 비기능 요구사항

2.3.3 SW 설계도

2.3.4 최종 산출물 구성

3. 예상되는 문제

3.1 Risk Analysis

3.2 Risk Reduction Plan

4. 테스트 방법

5. 참고 자료

**네임노드 최적화를 통한 마스터 메모리 절감**

**팀원 소개**

201911265 이석진 201910162 문무현

**작품 소개**

람다 아키텍처에서 카프카로 제공되는 작은 데이터들을 배치 레이어에서 모아서 처리를 하기때문에 크게 유용하지 않을 수 있다. 하지만 실시간 레이어에서 처리되는 실시간 데이터는 사이즈가 작다.  이 데이터와 통합하여 최종 사용자에게 제공하는 방식에서는 작은 파일로 처리된다. 이 부분에서 병합을 통해서 네임노드의 메모리를 절감할 수 있다.

핵심 문제 : 마스터 노드의 메모리 부족으로 인한 파일 저장 한계

참고)클라우데라에서는 파일 2M개당 1G의 램을 요구합니다.

해결 방법 : 작은 파일을 하나의 블록으로 통합함으로서 블록의 개수를 줄인다. 줄어든 블록의 개수만큼 마스터노드의 메모리가 절약된다. 해당 블록에 대한 메타데이터는 슬레이브 노드가 가지고 있도록한다.

SW 설계

기능 요구사항 : 블록매칭을 통한 블록 수 절감, 해당 블록의 메타데이터를 전달한다.

비기능 요구사항 : 얼마만큼 많은 소형파일들을 저장하며 슬레이브 메모리를 어떻게 절약할 수 있을 것인가? 파일 1개당 4바이트를 활용하여서 위치를 지정 12비트는 블록 단편화에 20비트는 블록을 지정하는데에 사용한다. 64MB 단위로 블록시 64MB/(2^12) = 16KB이다. 총 저장가능한 용량은 64MB\*2^20 = 64TB 저장이 가능하다. 블록사이즈에 따라서 달라지며 사용자의 기호에 맞게 단편화 사이즈도 조절이 가능하다.

SW 설계도 : 기본적으로 하둡의 HDFS 구성과 동일합니다. 아래는 메타데이터를 브로드캐스팅하는 과정을 콜라보레이션 다이어그램으로 표현하였습니다.

텍스트, 스크린샷, 도표, 폰트이(가) 표시된 사진

자동 생성된 설명

최종 산출물 구성

인프라 레이어 : hypervisor VM에 Ubuntu를 활용하여서 진행예정입니다.

플랫폼 레이어 : 자바와 하둡을 활용할 예정입니다.

어플리케이션 레이어 : 개발 모듈이름은 BC4RMM(block compression for reduction master’s memory)이 자바와 하둡만을 활용합니다.

**예상되는 문제**

**Risk Analysis**

메타데이터 전달 : 처음에 어떻게 전송하고 어떻게 이 메타데이터를 업데이트 할 건인가?

전달된 파일을 처리하는 방법 : 블럭 내에서 원하는 파일위치를 어떻게 찾을 것인가?

블록 매칭 알고리즘 : 블록 사이즈와 파일 사이즈에 따라 어떻게 최적화 할 수 있는가?

**Risk Reduction Plan**

슬레이브로 메타데이터 전달 : 소켓 **통신을 통해서 처리를 할 정입니다**

전달된 파일을 처리하는 방법 : 4바이트의 정규화된 정보를 처리하여 최적화할 예정입니다.

블록 매칭 알고리즘 : 그리디 알고리즘을 적용시킬 계획입니다.

**테스트 방법**

알고리즘은 워드카운터로 활용할 예정이며 테스트 데이터는 1MB 데이터 1M개와 블록 사이즈의 파일 2K개로 구성할 예정입니다. 측정할 대상은 마스트노드와 슬레이브 노드의 메모리의 점유량과 처리하는데 총 소요되는시간을 측정합니다.

**스케줄**

스케줄을 9-10주차에 파일형식을 고안하며 11-12주차에 파일 브로드캐스팅을 구현하며 12-13주차에 블록매칭 알고리즘을 완성합니다. 14,15주차에는 구현과 테스트 튜닝을 진행할 예정입니다.

**참고자료**

참고자료로 클라우데라 하둡 세팅 가이드라인 를 활용하였습니다.

<https://docs.cloudera.com/HDPDocuments/HDP2/HDP-2.5.0/bk_command-line-installation/content/configuring-namenode-heap-size.html>

(2021년 자료)