

하둡을 이용한 파일 분산 시스템 보안관리체제 구현

팀명 Time machine

지도교수 유승재 교수님

팀장 김민우
팀원 임채웅
 이준호
 성민용
 황지연

목 차

요 약 문	2
I. 연구 계획	3
1. 연구 목적	3
2. 연구 개요	3
3. 관련 연구	8
(1) 하둡이란?	
(2) PHP란?	
(3) DB(DataBase)란?	
4. 개발 구성	9
II. 하둡을 이용한 파일분산 시스템	10
1. 실행 결과	10
2. 소스코드	12
3. 결 론	26
III. 부 록	27
1. 발표 자료(PPT)	27

요 약 문

1. 연구 배경 및 목적

최근 빅 데이터가 뜨거운 화제로 떠오르고 있는데 2012년부터 시작하여 현재 2014년까지 빅 데이터 시장과 기술이 꾸준히 상승 중이며 2017년까지 계속 상승하고 고도화 될 것으로 예상되고 있습니다. 이와 같은 특성을 가진 하둡을 좀더 실용적으로 활용하고자 PHP를 이용한 웹하드형식의 프로그램을 개발하여 실용성과 보안성을 높였습니다.

2. 연구 내용

사용자가 웹 서버에 로그인하여 다양한 서비스를 활용하고 웹서버와 연동된 하둡 서버에서 업로드된 파일을 분산 저장됩니다. 사용자가 웹서버에 접속하여 파일을 업로드하면 스토리지 서버에 블록 단위로 분산저장 됩니다. 그 파일을 임시 스토리지 서버에 이중화를 하여 스토리지 서버가 다운이 되더라도 파일에 자료유실을 방지 할 수 있습니다.

3. 연구 결과

웹하드 형식의 웹서버를 구축하여 하둡과 연동하여 사용자가 웹서버에 회원가입을 한 후 로그인을 하여 업로드, 다운로드, 삭제를 할 수 있으며 업로드한 서버가 다운되더라도 다른 서버에 이중화가 되어 자료 손실을 방지 할 수 있도록 하였습니다.

I. 연구 계획

1. 연구 목적

지구촌은 인터넷과 모바일 기기로 실시간 끊임없이 소통하고 있다. 소통은 흔적(데이터)을 남깁니다. 통화와 문자메시지, 이메일, 동영상, 블로그, 트위터, 페이스북, 유튜브 등 ‘손가락’의 흔적들이 세계 곳곳의 데이터망에 자동으로 저장되고 있습니다. 바야흐로 ‘빅 데이터(Big Data)’ 시대입니다. 그러나 빅 데이터는 어마어마하게 많은 양의 데이터만을 의미하지 않습니다. 디지털 환경에서 기하급수적으로 늘어나는 빅 데이터에는 또 다른 특징이 있습니다. 정치·경제·사회·문화 할 것 없이 규모가 방대하기도 하지만 데이터 생성 주기가 짧고 데이터 속도는 빨라지고 있습니다. 문자와 영상, 그림, 음악까지 데이터 종류도 다양해지고 있습니다. 전문가들은 ‘빅 데이터’를 “정보화 사회의 원유(Oil)”에 비유하고 있습니다. 기름이 없으면 기계가 작동하지 않듯, 기름이 없으면 부가가치가 높은 각종 제품을 만들어내지 못하듯, 디지털시대에 빅 데이터만큼 중요한 자산은 없다는 것입니다. 어떤 사람들이 어떤 분야에 관심이 있는지, 자주 찾는 정보는 무엇인지, 한번 찾으면 얼마 동안 머무는지, 경쟁사회에서 살아남으려면 개인이건 기업이건 빅 데이터를 부가가치가 높은 소중한 자산으로 만들어야 경쟁사회에서 살아남을 수 있다는 얘기입니다. 방대한 규모의 빅 데이터를 쓸모 있고 가치있게 만드는 사람들, ‘빅 데이터 분석가(Big Data Analyst)’ 또는 ‘디지털 사이언티스트(Digital Scientist)’들이 주목받고 있습니다. 이런 것 들을 하둡을 이용하여 사용 할 수 있습니다. 앞으로의 비전이 발전 가능성이 크고 빅 데이터 시대에 살고 있으며 사용자들이 웹하드 형식이라 쉽게 사용 할 수 있습니다.

2. 연구 개요

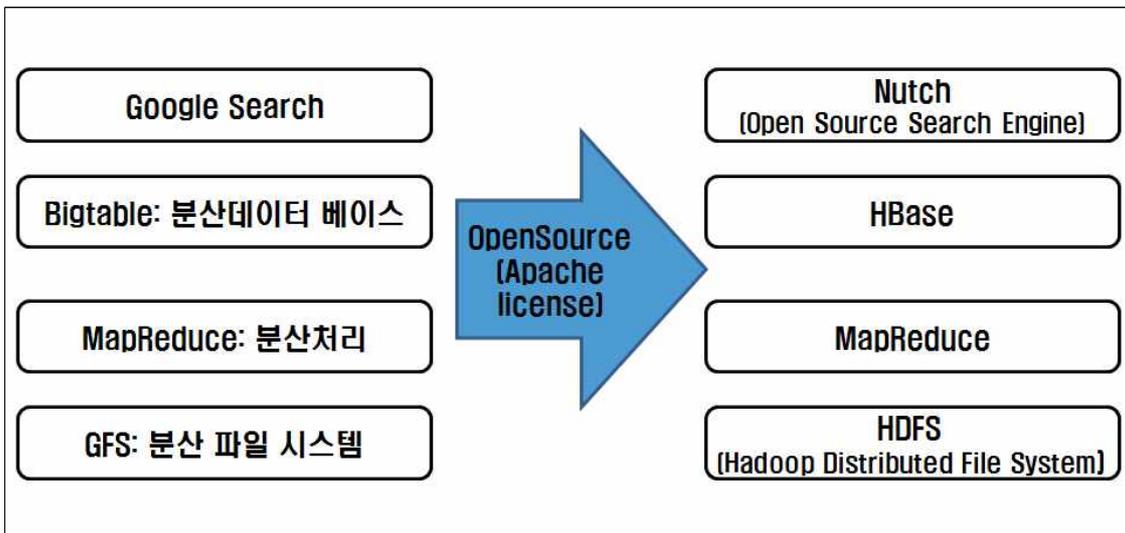
하둡은 비즈니스에 효율적으로 적용할 수 있도록 다양한 서브 프로젝트가 제공됩니다. 이러한 서브 프로젝트들이 상용화되면서, 하둡 에코 시스템(Hadoop ECO system)이 구성됐습니다. 참고로 하둡 에코 시스템은 하둡 생태계라고 표현되기도 합니다. 분산 데이터를 저장하는 HDFS와 분석 데이터를 처리하는 MapReduce가 하둡 코어 프로젝트에 해당하며, 나머지 프로젝트는 모두 하둡의 서브 프로젝트입니다. 이러한 이유가 앞으로 하둡이 미래에 전망이 있다는 것입니다. 웹서버에 사용자가 회원가입을 하면 데이터베이스에 데이터가 저장되어 정보가 저장됩니다. 로그인한 후 업로드, 다운로드, 삭제를 쉽게 사용 할 수 있습니다.

3. 관련 연구

(1) 하둡이란?

하둡(Hadoop)은 여러 컴퓨터로 구성된 클러스터를 이용하여 큰 사이즈의 데이터(빅 데이터)를 처리하기 위한 분산처리 프레임워크로 여러대의 서버들로 scale out 할 수 있도록 디자인 되어진 Apache Open Source Project 중 하나입니다. 현재 빅 데이터 분석에 가장 선호되어지고 사용되고 있는 솔루션이라고도 할 수 있습니다. 실제로 야후와 페이스북 등에 사용되고 있습니다. 하둡의 주요 구성요소로는 하둡 분산 파일 시스템인 HDFS(Hadoop Distributed File System), NoSQL 데이터베이스인 HBase, 대용량 데이터 분석 및 처리 프레임워크인 MapReduce가 있으며, 기타로 Pig, Hive, Zookeeper, Sqoop 등의 Hadoop을 기반으로 하는 다양한 오픈소스 분산 처리 프로젝트가 존재합니다. HDFS와 HBase는 각각 구글 시스템 소프트웨어의 구글 파일 시스템 GFS(Google File System)과 구글 데이터 베이스인 빅테이블(BigTable)의 Clone Project입니다. 하둡은 기본적으로 x86 기반의 서버들로 스토리지(HDFS)를 구성하고 HDFS에 저장된 데이터를 분산하여 분석 처리할 수 있는 Java기반의 MapReduce 프레임워크를 제공합니다.

[그림.1]은 JAVA기반의 MapReduce 프레임워크의 구조를 보여주고 있습니다.



[그림1] MapReduce 프레임워크

그렇다면 빅 데이터란 무엇인가?

빅 데이터 또한 클라우드 컴퓨팅과 같이 다양한 정의가 있습니다. 대부분 비슷비슷하지만 다양한 관점에서의 정의가 있어 하나의 문장으로 정의하기가 어려운 점이 있습니다. 이에 대표적인 빅 데이터의 정의를 몇 가지 소개해 드리도록 하겠습니다.

세계에서 가장 큰 인터넷 사전인 Wikipedia에서는 빅 데이터를 "기존 데이터베이스 관리도구의 데이터 수집, 저장, 관리, 분석의 역량을 넘어서는 대량의 정형 또는 비정형 데이터의 집합이며, 이러한 데이터로부터 가치를 추출하고 결과를 분석하는 기술"을 빅 데이터 (or 빅 데이터 기술) 라고 의미하였습니다. 여기서 기존 데이터베이스 관리도구는 RDBMS 를 이야기합니다.

2011년 Mckinsey에서는 데이터베이스 관점에서 "일반적인 데이터베이스 S/W(RDMS)가 저장, 관리, 분석할 수 있는 범위를 초과하는 규모의 데이터"라고 빅 데이터를 정의하였으며, IDC에서는 업무 수행관점에서 "다양한 종류의 대규모 데이터로부터 저렴한 비용으로 가치를 추출하고 초고속 수집, 저장, 발굴, 분석을 지원하도록 고안된 차세대 기술 및 아키텍처"라고 빅 데이터를 정의하였습니다.

위에서 정의한 빅 데이터는 빅 데이터를 바라보는 차원에서의 정의가 있을뿐 정량적인 차원에서의 정의는 없습니다. 실제로도 그렇습니다. 굳이 빅 데이터를 절대적인 크기로 정의한다면.. 주로 컨설팅 업체들에서는 (대표적으로 엑센츄어) 약 20TByte 이상의 데이터를 빅 데이터로 분류한다고 합니다.

빅 데이터 정의에 대해서 정리해 본다면 "서버 한대로 처리할 수 없는 규모의 데이터", "기존의 소프트웨어로는 처리할 수 없는 규모의 데이터", 뒤에서 차츰 설명 드리겠지만 "Scale-up 보다는 Scale-out을 지향", "3V(Volume, Velocity, Variety) 중 2V를 만족하는 데이터" 정도가 될 것 같습니다.

여기서 이야기하는 "Scale-up 보다는 Scale-out을 지향"의 의미는 하둡에서 지향하는 것인데요. X86 급의 저사양 서버에서 데이터를 처리하고 데이터를 처리하기 위한 사양이 부족하다면 장비의 사양을 고사양으로 높이는 것(Scale-up)이 아니라 Memory 또는 저장장치, Server의 대수를 늘리는 것(Scale-out)입니다. 이유는 고사양의 장비를 추가 또는 교체하는 것보다는 x86급의 저사양 서버 여러 대를 두어 처리하는 것이 금액적인 부분이나 성능적인 부분에서 더욱 효율적이기 때문입니다.

3V는 빅 데이터를 특징을 이야기 하는 것인데요. 흔히 3V중 2V를 만족할 경우 빅 데이터라고 이야기 하고 있습니다.

(2) PHP 란?

PHP (PHP: Hypertext Preprocessor)는 프로그래밍 언어의 일종이다. 원래는 동적 웹 페이지를 만들기 위해 설계되었으며 이를 구현하기 위해 PHP로 작성된 코드를 HTML 소스 문서 안에 넣으면 PHP 처리 기능이 있는 웹 서버에서 해당 코드를 인식하여 작성자가 원하는 웹 페이지를 생성합니다. 근래에는 PHP 코드와 HTML을 별도 파일로 분리하여 작성하는 경우가 일반적이며, PHP 또한 웹서버가 아닌 php-fpm(PHP FastCGI Process Manager)을 통해 실행하는 경우가 늘어나고 있습니다.

또한 PHP는 명령 줄 인터페이스 방식의 자체 인터프리터를 제공하여 이를 통해 범용 프로그래밍 언어로도 사용할 수 있으며 그래픽 애플리케이션을 제작할 수도 있습니다.

많은 서버 측 오픈 소스 소프트웨어는 PHP로 구현되었다. PHP를 바탕으로 하는 프로그램 중 대표적인 예로 블로깅 도구 워드프레스, SIR의 그누보드4와 위키백과를 구동시키는 미디어위키를 들 수 있습니다. 대한민국의 PHP로 만든 게시판 중 유명한 것에는 그누보드4, 제로보드XE 등이 있습니다.

PHP는 마이크로소프트의 ASP.NET, 비주얼 베이직, 매크로미디어의 콜드퓨전, 썬 마이크로시스템즈의 자바나 오픈 소스 커뮤니티의 파이썬, Perl, 루비에 대한 대안으로 생각될 수 있습니다.

텍스트, 특히 HTML의 처리에 강점을 가지고 있다. URL의 파싱이나 폼 처리, 정규 표현식 등이 그 한 예입니다. 또한 다양한 데이터베이스를 지원하므로 데이터베이스와 사용자간의 다리 역할도 잘 수행합니다.

PHP는 사용 허가서의 규정을 따라 릴리즈 된 자유 소프트웨어입니다. 그러나 PHP 사용 허가서는 PHP라는 단어의 사용에 제한을 두는 규정을 가지고 있기 때문에 GNU 일반 공중 사용 허가서와 호환되지 않습니다.

(3) DB(database) 란?

1) 데이터베이스의 개념 및 정의

기존의 파일시스템에서는 중복된 정보를 효율적으로 처리하는 것이 관점 연구 분야였습니다. 이 중복을 피하여 정보를 일원화(一元化)하여 처리를 효율적으로 수행하기 위해서 서로 관련성을 가지며 중복이 없는 데이터의 집합을 유지하는 것을 데이터베이스(database)라고 합니다. 그리고 데이터베이스를 관리하는 시스템을 데이터베이스 관리시

시스템(DBMS: Database Management System)이라 합니다. 데이터베이스의 대표적인 구조는 트리구조(tree structure)인 구조적 데이터베이스 외에 관계형 데이터베이스(relational database)와 객체지향 데이터베이스 구조로 대별합니다.

2) 데이터베이스의 구성

데이터베이스는 특정 조직의 업무를 수행하는 데 필요한 상호 관련된 데이터들의 모임으로, 최소의 중복으로 통합(integrated), 저장(stored)된 운영(operation) 데이터로 구성됩니다.

① 통합 데이터(integrated data)

중복을 배제하나, 경우에 따라 불가피하게 중복을 허용하는 데이터입니다. 이러한 의도적 중복은 항상 파악하여 관리할 수 있으며, 이러한 것을 최소의 중복 또는 통제된 중복이라고 합니다.

② 저장 데이터(stored data)

컴퓨터의 저장매체에 저장하여 관리하는 데이터를 의미합니다.

③ 운영 데이터(operation data)

단순한 데이터의 집합이 아니라 그 조직의 기능을 수행하는 데 없어서는 안 될 필수 데이터들을 의미합니다.

④ 공용 데이터(shared data)

어느 하나의 응용프로그램이나 응용시스템을 위한 데이터가 아니라, 그 조직의 여러 사용자와 여러 응용시스템들이 서로 다른 목적으로 데이터를 공동으로 이용할 수 있게 합니다.

3. 데이터베이스의 특징

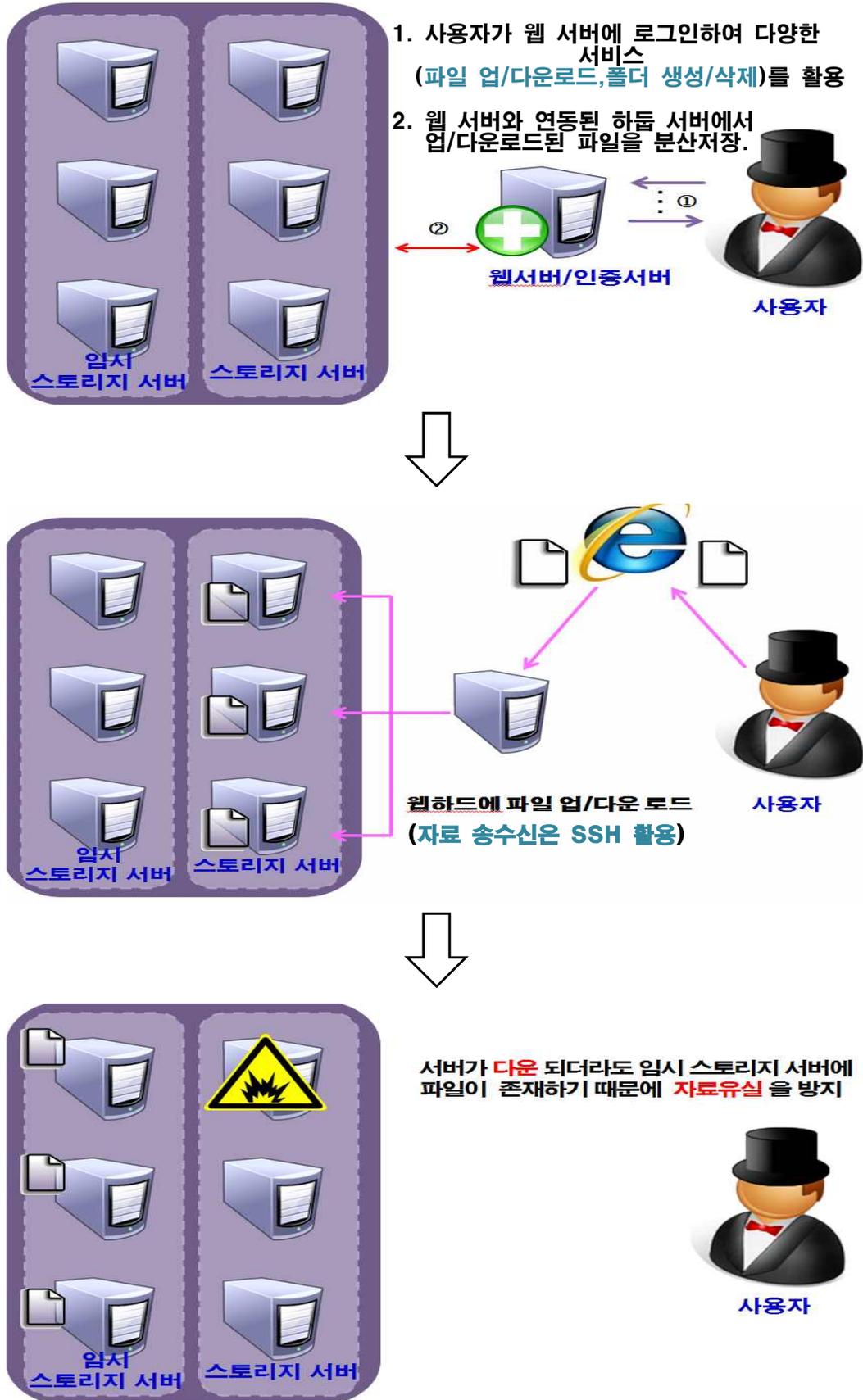
- ① 실시간 접근성(real time accessibility) : 사용자의 질의에 대하여 즉시 처리하여 응답하는 특징을 갖습니다.
- ② 계속적인 진화(continuous evolution) : 삽입, 삭제, 갱신을 통하여 항상 최근의 정확한 데이터를 동적으로 유지하는 특징이 있습니다.
- ③ 동시 공유(concurrent sharing) : 여러 사용자가 동시에 원하는 데이터를 공유할 수 있는 특징을 갖습니다.
- ④ 내용에 의한 참조(content reference) : 데이터베이스에 있는 데이터를 참조할 때 튜플(tuple)의 주소나 위치가 아닌 사용자가 요구하는 데이터

내용에 따라 참조하는 특징을 갖습니다.

- ⑤ 데이터 논리적 독립성(independence) : 응용프로그램과 데이터베이스를 독립시킴으로써 데이터의 논리적 구조를 변경시키더라도 응용프로그램은 변경되지 않는 특징을 갖습니다.

4. 개발 구성

시스템 개요



II. 하둡을 이용한 파일분산 시스템

1. 실행 결과

■ 로그인 페이지



Time Machien 로그인

아이디:

비밀번호:

로그인

회원아이디가 없을 경우: [무료 회원가입](#)

Copyright © 2014 by Time Machine All rights reserved.

■ 회원가입



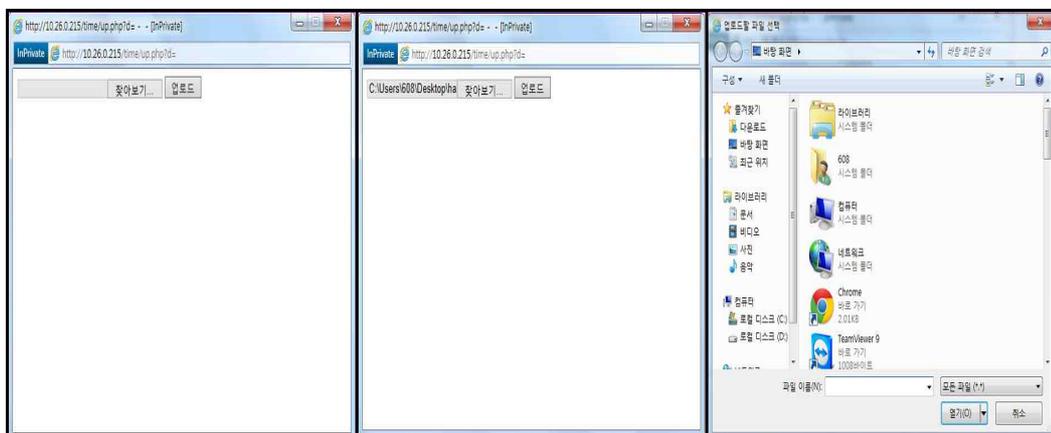
회원가입

아이디 *	<input type="text" value="junho"/>	중복체크 (영문이나 알파벳 최대 16자리로 입력하세요!)
이름 *	<input type="text" value="이준호"/>	
이메일 *	<input type="text" value="wnsgh1074@naver.com"/>	
비밀번호 *	<input type="password" value="****"/>	(최대 12자리 까지 가능합니다.)
비밀번호 확인 *	<input type="password" value="****"/>	(최대 12자리 까지 가능합니다.)

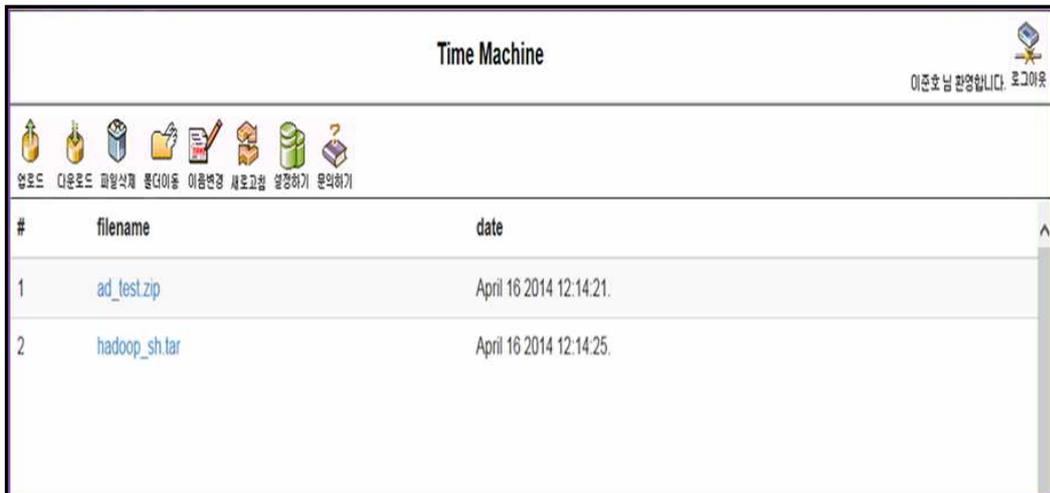
등록 취소

Copyright © 2014 by Time Machine All rights reserved.

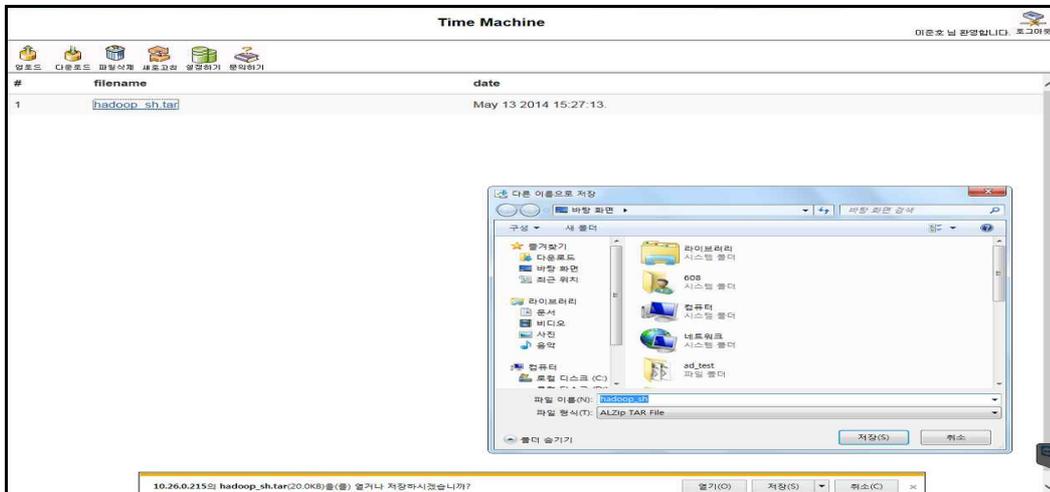
■ 업로드



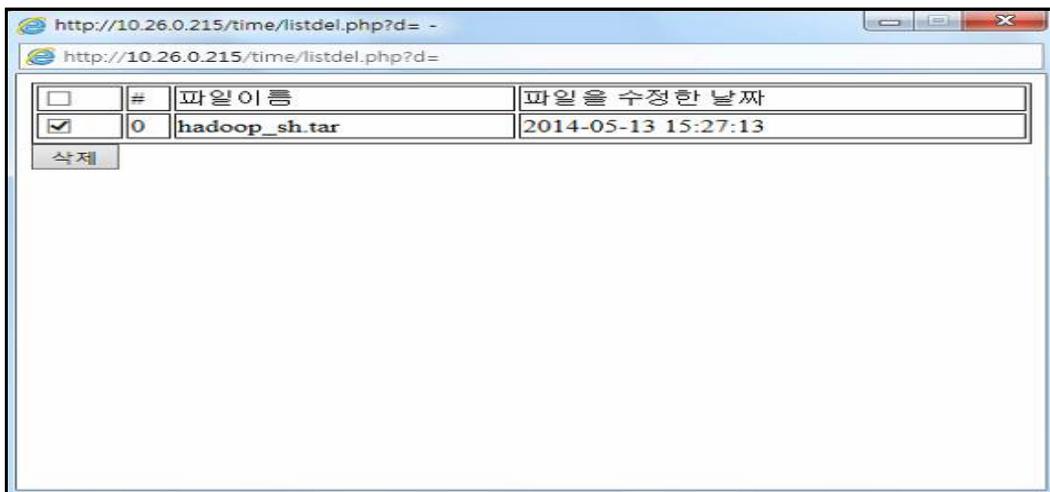
■ 파일리스트



■ 다운로드



■ 파일 삭제



2. 소스코드

■ 회원가입

```
<?
  if($_REQUEST["exec"]=="join"){
    require_once("database.inc");
    $conn = get_connection("time_machine");

    // 회원 중복 가입 체크
    $sql = "select count(*) cnt from member where
member_id='".$_REQUEST["member_id"]."";
    $rs = mysql_query($sql);
    if(mysql_errno()==0 && mysql_num_rows($rs)>0){
      $row = mysql_fetch_array($rs);
      if($row["cnt"] > 0){
?>
        <script language="javascript">
          alert("회원아이디 '<?= $_REQUEST["member_id"] ?>'는 벌써 사용중인
아이디입니다.");
          history.back();
        </script>
<?
      }
    }

    mysql_free_result($rs);

    $sql = "insert into member values ";
    $sql = "('" . $_REQUEST["member_id"] . "','" . $_REQUEST["name"] . "','" . $_REQUEST["passw
d"] . "','" . $_REQUEST["email"] . "', now(), null)";

    mysql_query($sql);

    if(mysql_errno() == 0){
      require_once("config.inc");

      // 하둡 계정 생성
      exec("mkdir hadoop/'".$_REQUEST["member_id"]."';sudo useradd
".$_REQUEST["member_id"]."; /usr/local/hadoop/bin/hadoop fs -mkdir
/user/'".$_REQUEST["member_id"]."; sudo /usr/local/hadoop/bin/hadoop
fs -chown '".$_REQUEST["member_id"].':'".$_REQUEST["member_id"]."
/user/'".$_REQUEST["member_id"].'");
```

```

        session_start();

        $_SESSION["isLogin"] = true;
        $_SESSION["member_id"] = $_REQUEST["member_id"];
        $_SESSION["name"] = $_REQUEST["name"];
    ?>
    <script language="javascript">
        alert("회원가입이 정상적으로 처리되었습니다.");
        location.replace("./");
    </script>
<?
    }else{
        echo("[DB Error]<br>".mysql_error());
    }
    mysql_close($conn);
    exit;
}
?>
<html>
<head>
<title>사용자 등록</title>
<style type="text/css">
    table, th, td, input { font-size:9pt; }
</style>
<script language="javascript">
    function initPage(){
        joinForm.member_id.focus();
    }

    function checkForm(form){
        if(form.member_id.value.length==0){
            alert("아이디를 입력하세요!");
            form.member_id.focus();
            return false;
        }

        if(form.passwd.value.length==0){
            alert("비밀번호를 입력하세요!");
            form.passwd.focus();
            return false;
        }

        if(form.passwd2.value.length==0){
            alert("비밀번호 확인을 입력하세요!");

```

```

    form.passwd2.focus();
    return false;
}

if(form.passwd.value != form.passwd2.value){
    alert("비밀번호화 확인 비밀번호가 일치하지 않습니다.");
    form.passwd2.value="";
    form.passwd2.focus();
    return false;
}
}

function dupCheck(){
    var nWidth = 350;
    var nHeight = 120;
    var nTop = (screen.height - nHeight) / 2;
    var nLeft = (screen.width - nWidth) / 2;

    var strURL = "dupcheck.php";
    if(joinForm.member_id.value.length>0)
        strURL += "?exec=check&member_id=" + joinForm.member_id.value;
    window.open(strURL, "dupCheck",
"height="+nHeight+",width="+nWidth+",top="+nTop+",left="+nLeft);
}
</script>
</head>
<body topmargin="0" bottommargin="0" leftmargin="0" rightmargin="0"
onLoad="initPage();">
<form name="joinForm" action="join.php" method="post"
onSubmit="return(checkForm(this));">
<input type="hidden" name="exec" value="join">
<table border="0" width="750" cellpadding="5" cellspacing="0" align="center">
<tr><td height="10"></td></tr>
<tr>
<td align=left colspan="2">
<table border="0" cellspacing="0" width="100%"
cellpadding="0">
<td>
<table border="0" width="100%" cellpadding="0" cellspacing="0">
<tr>
<td bgcolor="#C3D9FF" style="padding:5 5 5 5;"&nbsp;</td>
</tr>
</table>
</td>
</tr>
</table>

```

```

        </td>
    </tr>
</table>
        </td>
    </tr>
<tr>
    <td>
        <fieldset style="padding:10 10 10 10;">
            <legend></legend>
            <table border="0" width="100%" cellpadding="5" cellspacing="0">
                <tr><td height="30"></td></tr>
                <tr>
                    <td align="center" valign="middle">
                        <table border="0" width="600" cellpadding="3" cellspacing="1"
bgcolor="#CCCCCC">
                            <colgroup>
                                <col nowrap align="center" width="150" bgcolor="#EEEEEE"
style="color:gray;">
                                    <col nowrap bgcolor="#FFFFFF">
                                </colgroup>
                                <tr>
                                    <td>아이디 *</td>
                                    <td>
                                        <input type="text" name="member_id" size="12"
maxlength="16">
                                        <input type="button" value="중복체크"
onClick="dupCheck();">
                                        (영문이나 알파벳 최대 16자리로 입력하세요!)
                                    </td>
                                </tr>
                                <tr>
                                    <td>이름 *</td>
                                    <td>
                                        <input type="text" name="name" size="20" maxlength="50">
                                    </td>
                                </tr>
                                <tr>
                                    <td>이메일 *</td>
                                    <td>
                                        <input type="text" name="email" size="60" maxlength="50">
                                        <i>
                                    </td>
                                </tr>
                                <tr>
                                    <td>비밀번호 *</td>

```

```

        <td>
            <input type="password" name="passwd" size="12"
maxlength="12">
                (최대 12자리 까지 가능합니다.)
            </td>
        </tr>
        <tr>
            <td>비밀번호 확인 *</td>
            <td>
                <input type="password" name="passwd2" size="12"
maxlength="12">
                    (최대 12자리 까지 가능합니다.)
                </td>
        </tr>
        <tr>
            <td colspan="2" bgcolor="#FFFFFF" align="left"
valign="bottom">
                </td>
        </tr>
    </table>
</td>
</tr>
<tr>
    <td align="center" height="40">
        <input type="submit" value=" 등 록 " >
        <input type="button" value=" 취 소 "
onClick="location.replace('./');">
    </td>
</tr>
</form>
</table>
</fieldset>
</td>
</tr>
<tr><td height="10"></td></tr>
<tr>
    <td align="center">
        <table border="0" width="750" cellpadding="0" cellspacing="0">
            <tr>
                <td bgcolor="#C3D9FF" style="padding:5 5 5 5;">&nbsp;</td>
            </tr>
        </table>
    </td>
</tr>
</tr>

```

```

<tr>
  <td align="center">
    Copyright © 2014 by Time Machine All rights reserved.
  </td>
</tr>
</table>
</body>
</html>

```

■ 로그인

```

<?
require_once("config.inc");
require_once("database.inc");

session_start();

if($_SESSION["isLogin"]){
  echo("<script>location.replace(\".\");</script>");
  exit;
}

$configFile = getcwd().getFileSeparator()."time_machine.php";

if($_REQUEST["exec"]=="login"){
  $sql = "select passwd, name from member where
member_id='".$_REQUEST["member_id"].'";

  $conn = get_connection("time_machine");
  $rs = mysql_query($sql);

  if(mysql_errno() == 0 && mysql_num_rows($rs)>0){
    $row = mysql_fetch_array($rs);
    if($_REQUEST["passwd"]==$row["passwd"]){
      $userDir = getIni($configFile,
"disk.root").getFileSeparator().$_REQUEST["member_id"];

      $_SESSION["isLogin"] = true;
      $_SESSION["userDir"] = $userDir;

      if(!file_exists($userDir)) mkdir($userDir);
      $nextUrl = ".";
      $_SESSION["member_id"] = $_REQUEST["member_id"];
      $_SESSION["name"] = $row["name"];

```

```

        echo("<script>location.replace(\"./\");</script>");

    }else{
        echo("<script>alert(\"비밀번호가 다릅니다.\");history.back();</script>");
    }
}else{
    echo("<script>alert(\"".$_REQUEST["member_id"]."님은 사용자 정보에
없습니다.\");history.back();</script>");
}
mysql_free_result($rs);
mysql_close($conn);
}
?>
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN">
<html>
<head>
    <title>Time Machine</title>
    <style type="text/css">
        table, th, td, input, a { font-size:9pt;}
    </style>
    <script language="javascript">
function initPage(){
    loginForm.member_id.focus();
}

function checkForm(form){
    if(form.member_id.value==""){
        alert("회원아이디를 입력하세요!");
        form.member_id.focus();
        return false;
    }

    if(form.passwd.value==""){
        alert("비밀번호를 입력하세요!");
        form.passwd.focus();
        return false;
    }
    return true;
}
    </script>
</head>
<body bgcolor="#ffffff" onLoad="initPage();">
<div align="center">
<table cellspacing="0" width="750" cellpadding="0" border="0">

```

```

        <tr>
            <td align=left colspan="2">
                <table border="0" cellspacing="0" width="100%"
cellpadding="0">
                    <tr>
                        <td>
                            <table border="0" width="100%" cellpadding="0" cellspacing="0">
                                <tr>
                                    <td bgcolor="#C3D9FF" style="padding:5 5 5:5">&nbsp;</td>
                                </tr>
                            </table>
                        </td>
                    </tr>
                </table>
            </td>
        </tr>
    </table>
<table width=100% align=center cellpadding=3 cellspacing=2 border=0>
    <tr>
        <td align=center>
            <table cellpadding="0" cellspacing="0" border="0" width="750">
                <tr>
                    <td rowspan="3" width="20" nowrap><spacer
type="horizontal" width="20"></td>
                    <td width="450" valign="top">
                        <table cellpadding="0" cellspacing="0" border="0" width="450"
style="line-height:200%:">
                            <tr>
                                <td><img src=images/Time-Machine5.png></td>
                            </tr>
                        </table>
                    </td>
                </tr>
                <td rowspan="3" width="20" nowrap><spacer
type="horizontal" width="20"></td>
                <td rowspan="3" width="240" valign="top">
                    <form name=loginForm action="login.php" autocomplete=off
method=post>
                        <input type=hidden name="exec" value="login">
                        <div>
                            <table width="100%" cellpadding="0" cellspacing="0" border="0">
                                <tr>
                                    <td><img src=images/crn_r_top_l.gif width="10" height="10"
alt="" border="0"></td>
                                    <td width="100%" bgcolor="#eeeeee">

```



```

                <tr><td
                    colspan="2"><br><hr
                        size="1"
noshade><br></td></tr>
                <tr>
                    <td colspan="2">
                        <b>회원아이디가 없을 경우:</b>
                    </td>
                </tr>
                <tr>
                    <td colspan="2" align="right">
                        <a href="join.php">무료 회원가입</a><br><br><br>
                    </td>
                </tr>
            </table>
        </div>
    </td>
    <td
        style="width:10px"
        background=images/brdr_r.gif
nowrap></td>
    </tr>
</table>
<table width="100%" cellpadding="0" cellspacing="0" border="0">
    <tr>
        <td><img src=images/crn_r_bot_l.gif width="10" height="10"
border="0"></td>
        <td
            width="100%"
            bgcolor="#eeeeee"
background=images/brdr_b.gif width="100%">
            <spacer type="block" width="100%" height="10"></td>
        <td><img src=images/crn_r_bot_r.gif width="10" height="10"
alt="" border="0"></td>
    </tr>
</table>
</form>
</div>
    </td>
</tr>
<tr>
    <td>
        <table cellpadding="0" cellspacing="0" height="1"
width="450" border="0">
            <tr><td
                height="5"
                nowrap><spacer
type="block" height="5"></td></tr>
            <tr><td
                nowrap
                height="1"
                width="450"
                bgcolor="#cccccc"><spacer type="block" width="450" height="1"></td></tr>
            <tr><td
                height="5"
                nowrap><spacer
type="block" height="5"></td></tr>
        </table>
    </td>
</tr>

```

```

        </td>
      </tr>
    </table>
  </td>
</tr>
<tr>
</tr>
<tr align="right">
  <td height="80">&nbsp;</td>
</tr>
<tr>
  <td align="center">
    <table border="0" width="750" cellpadding="0" cellspacing="0">
      <tr>
        <td bgcolor="#C3D9FF" style="padding:5 5 5 5;">&nbsp;</td>
      </tr>
    </table>
  </td>
</tr>
<tr>
  <td align="center">
    Copyright © 2014 by Time Machine All rights reserved.
  </td>
</tr>
</table>
</body>
</html>

```

■ 파일 리스트

```

<html>
<head>
<!-- Latest compiled and minified CSS -->
<link
                                                                    rel="stylesheet"
href="//netdna.bootstrapcdn.com/bootstrap/3.1.1/css/bootstrap.min.css">

<!-- Optional theme -->
<link
                                                                    rel="stylesheet"
href="//netdna.bootstrapcdn.com/bootstrap/3.1.1/css/bootstrap-theme.min.
css">
</head>
<body>
<!-- Latest compiled and minified JavaScript -->
<
      s      c      r      i      p      t
src="//netdna.bootstrapcdn.com/bootstrap/3.1.1/js/bootstrap.min.js"></scri

```

```

pt>
<?php
    session_start();

    if(!$_SESSION["isLogin"]){
        echo("<script>alert(\"로그인 하세
요!\");top.location.replace(\".\");</script>");
        exit;
    }
//위의 link와 script 태그는 bootstrap을 적용하기 위해 사용
echo '<table class="table table-striped">'; // /
테이블 시작
echo '<thead>';
//테이블 머리(맨 윗줄) 시작

echo '<tr>';
echo '<th>';
echo '#';
echo '</th>';
echo '<th>filename</th>';
echo '<th>date</th>';
echo '</tr>';
echo '</thead>';
//테이블 머리 끝

echo '<tbody>';
//테이블 몸통 시작

$dh = @opendir("./hadoop/$_SESSION[member_id]");
//디렉토리 열기

$i = 0;
while(($file = readdir($dh)) != false) { // /
    디렉토리에서 파일을 가져옴(파일을 다 가져오면 false반환됨)
        if($file == '.' || $file == '..') continue;
        $file_path =
"./hadoop/$_SESSION[member_id].DIRECTORY_SEPARATOR.$file;
        if(is_file($file_path)) {
            $i++;
            echo '<tr>';
            echo '<td>';
            echo "$i";
            //몇번째인지 출력

            echo '</td>';
            echo '<td>';
            echo "<a href='$file_path'>$file</a>"; // /
파일 다운로드 링크 생성
            echo '</td>';
            echo '<td>';

```

```

        echo date("F d Y H:i:s.", filemtime($file_path)); //파일 최종
수정날자 출력
        echo '</td>';
        echo '</tr>';
    }
}
closedir($dh);
//디렉토리 닫음
echo '</tbody>';
//테이블 몸통 마침
echo '</table>';
//테이블 마침
//echo "$ii";
?>
</body>
</html>

```

■ 업로드

```

<?
session_start();

if(!$_SESSION["isLogin"]){
}
$save_dir = "hadoop/$_SESSION[member_id]";
if(is_uploaded_file($_FILES["file"]["tmp_name"]))
{
    $dest = $save_dir.$_FILES["file"]["name"];

    if(!move_uploaded_file($_FILES["file"]["tmp_name"], $dest))
    {
        die("파일을 지정한 디렉터리에 저장하는데 실패했습니다.");
    }
}
exec("sudo /usr/local/hadoop/bin/hadoop fs -put
/var/www/html/time/'$dest' /user/'$_SESSION[member_id]':sudo
/usr/local/hadoop/bin/hadoop fs -chown '$_SESSION[member_id]'
/user/'$_SESSION[member_id]/*");

echo "업로드 완료!";
?>

```

■ 삭제

```
<?
$chkvalue = $_POST['checkbox'];
$file_path = $_GET['file_path'];

for ($i = 0; $i < count($chkvalue); $i++) {

    unlink($file_path . str_replace("|", " ", $chkvalue[$i]));

}
exec("/usr/local/hadoop/bin/hadoop fs -ls $file_path");
echo "삭제 완료</br>";

?>
```

3. 결 론

웹하드 서비스를 통해 인터넷 공간에 파일을 저장하거나 문서 편집이 가능하며 타인과 자료를 공유할 수 있습니다.

과거처럼 휴대용 디스켓이나 외장 하드 등을 가지고 다닐 필요가 없고, 웹상의 저장공간에 파일들을 저장해 놓았다가 언제 어디서든 다운받아 사용할 수 있습니다. 분실 혹은 도난의 위험이 적어 최근 들어 많은 사용자들이 증가하고 있는 추세입니다.

아파치 하둡이란 대량의 자료를 처리할 수 있는 큰 컴퓨터 클러스터에서 동작하는 분산 응용 프로그램을 지원하는 프리웨어 자바 소프트웨어 프레임워크입니다. 원래 너치의 분산 처리를 지원하기 위해 개발된 것으로, 아파치 루씬의 하부 프로젝트입니다.

분산처리 시스템인 구글 파일 시스템을 대체할 수 있는 하둡 분산 파일 시스템과 맵리듀스를 구현한 것입니다.

하둡 분산 파일 시스템(HDFS, Hadoop distributed file system)은 하둡 프레임워크를 위해 자바 언어로 작성된 분산 확장 파일 시스템입니다.

HDFS은 여러 기계에 대용량 파일들을 나눠서 저장을 합니다. 데이터들을 여러 서버에 중복해서 저장을 함으로써 데이터 안정성을 확보 하였습니다.

웹하드를 이용한 파일 업/다운로드 서비스와 하둡의 파일분산 기능,서버 이중화의 기능을 접목시켜 데이터들을 여러 서버에 중복해서 저장을 함으로써 데이터 안정성을 확보 하였습니다.

Ⅲ. 부 록

1. 발표 자료(PPT)



지도교수 : 유승재 교수님
7조 타임머신 (Time Machine)

목 차

1. 조원 소개 및 역할
2. 주제 선정
3. 연구진행상황 및 운영 절차
4. 작품 진도 점검

조원 소개 및 역할

김민우

총괄 책임 / 웹서버 구축

임채웅

하둡 구축 / 웹서버 연동

성민용

하둡 구축 / 자료조사

이준호

웹서버 구축 / 자료조사

황지연

하둡 구축 / 자료조사

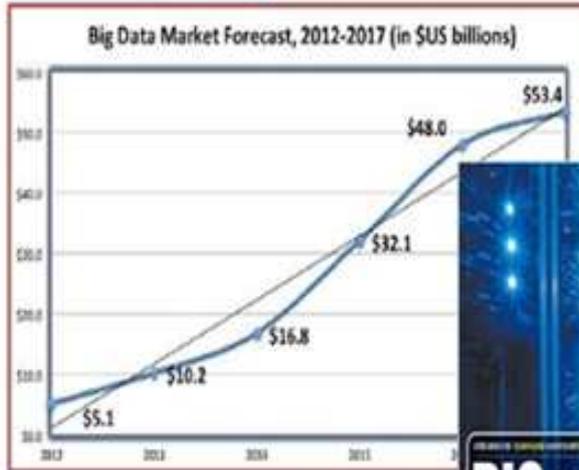
주제 선정

주제

- ◆ **주제 : 하둡을 이용한 파일 분산 시스템 보안관리체제 구현**
- ◆ **하둡(Hadoop) 특성**
 - 분산파일 시스템을 구현할 수 있는 최신 기술
 - 가용성과 안정성을 높인 클라우드 기반 스토리지 시스템
 - 상대적으로 저렴한 비용의 서버 여러 대로 시스템 구성 가능

주제 선정

주제 선정 이유



2012~2017까지 빅데이터
기술과 시장의 전망

SNS 사용으로 인해 증가한
방대한 양의 정보들



연구 결과 및 운영 절차

파일분산시스템

◆기본체계 : 아파치 하둡

- 대량의 자료를 처리할 수 있는 컴퓨터 클러스터에서 동작하는 분산응용프로그램을 지원하는 자유 자바 소프트웨어 프레임워크

◆시스템 구축

- 하둡 스토리지 서버와 암시스토리지서버로 이중화
- 웹 하드 서비스 페이지 제작, 서버와 연동

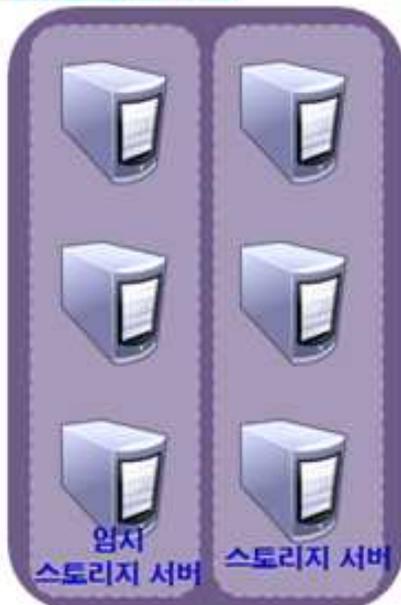
연구 결과 및 운영 절차

개발 환경

- ◆ 운영체제 : Linux (CentOS-5.5)
- ◆ 개발언어 : 웹 스크립트 , PHP
- ◆ 프로그램 : Hadoop-1.2.1

연구 결과 및 운영 절차

사용자 인증

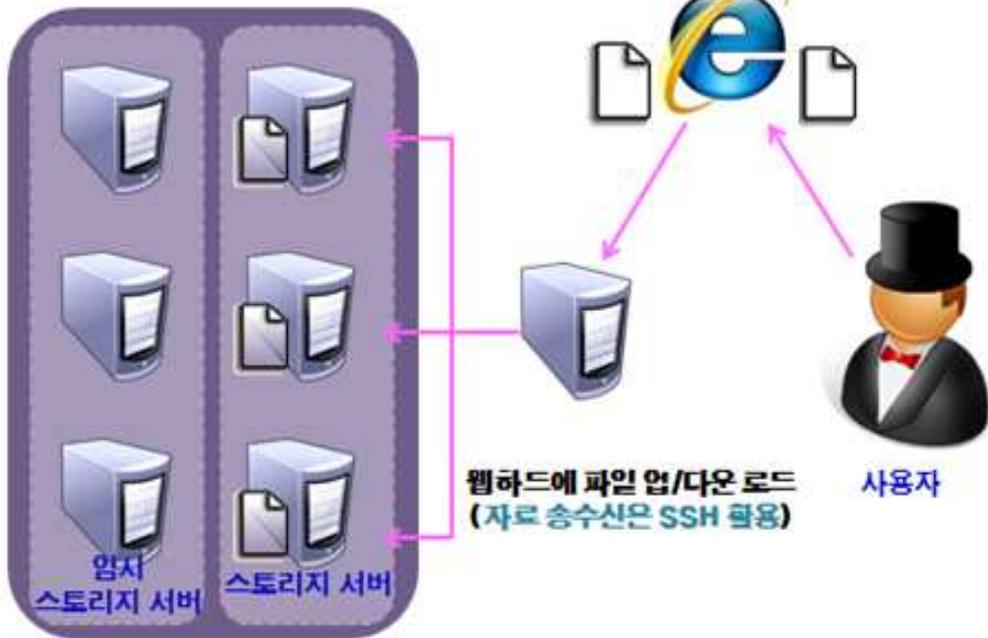


1. 사용자가 웹 서버에 로그인하여 다양한 서비스 (파일 업로드, 다운로드, 파일삭제)를 활용
2. 웹 서버와 연동된 하둡 서버에서 업/다운로드된 파일을 분산저장.



연구 결과 및 운영 절차

Hadoop을 이용한 파일 분산



연구 결과 및 운영 절차

하둡의 이중화



연구 결과 및 운영 절차

로그인



The login page features a large image of a green camera with a circular arrow on its lens. To the right is a login form titled "Time Machine 로그인". It contains two input fields for "아이디:" and "비밀번호:", followed by a "로그인" button. Below the form, there is a link for "회원아이디가 없을 경우: 무료 회원가입" with a hand cursor pointing to it. The footer contains the text "Copyright © 2014 by Time Machine All rights reserved."

Time Machine 로그인

아이디:

비밀번호:

로그인

회원아이디가 없을 경우: [무료 회원가입](#)

Copyright © 2014 by Time Machine All rights reserved.

연구 결과 및 운영 절차

회원가입



The registration page has a tab labeled "회원가입". It contains a form with the following fields: "아이디" (ID) with the value "junho" and a "중복체크" (check for duplicates) button; "이름" (Name) with the value "이준호"; "이메일" (Email) with the value "wngph1074@naver.com"; "비밀번호" (Password) with four asterisks; and "비밀번호 확인" (Confirm Password) with four asterisks. Below the form are "등록" (register) and "취소" (cancel) buttons. The footer contains the text "Copyright © 2014 by Time Machine All rights reserved."

회원가입

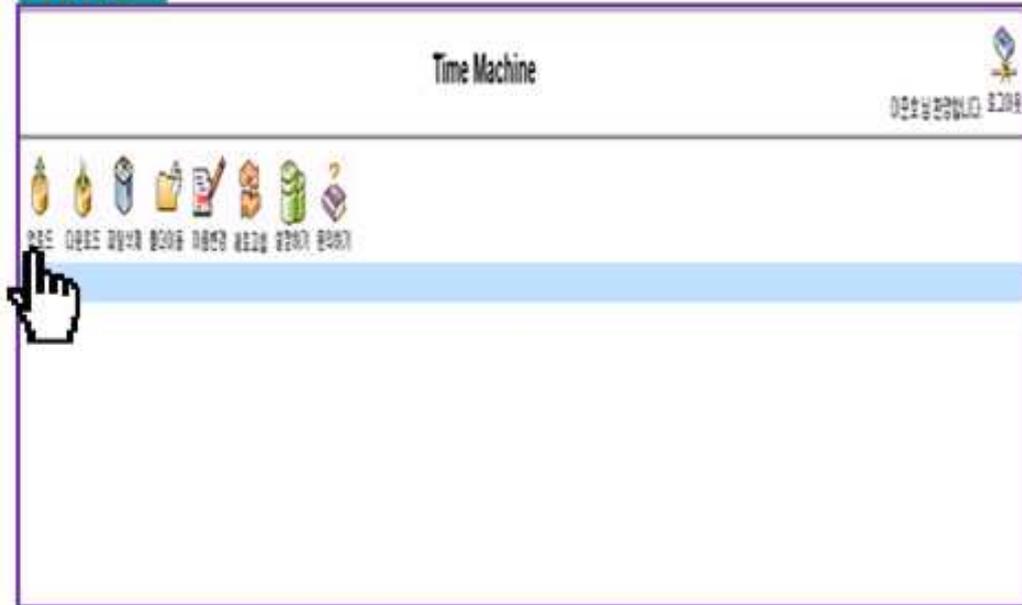
아이디	junho	중복체크 (영문이나 알파벳 최대 16자리로 입력하세요!)
이름	이준호	
이메일	wngph1074@naver.com	
비밀번호	****	(최대 12자리 까지 가능합니다.)
비밀번호 확인	****	(최대 12자리 까지 가능합니다.)

등록 취소

Copyright © 2014 by Time Machine All rights reserved.

연구 결과 및 운영 절차

서비스



연구 결과 및 운영 절차

업로드



연구 결과 및 운영 절차

업로드 결과 출력

Time Machine		
		이민호님 관리자 ID: 8309
		
#	filename	date
1	이민호.txt	May 20 2014 15:31:10
2	홍보보.txt	May 20 2014 15:32:05
3	4학년.txt	May 20 2014 15:31:45

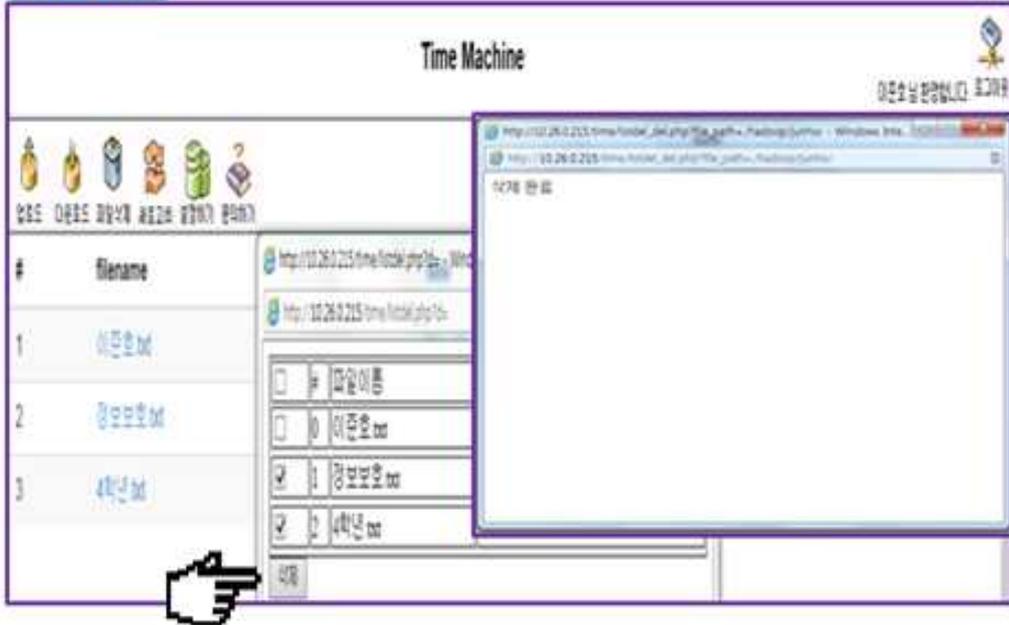
연구 결과 및 운영 절차

파일 삭제

Time Machine		
		이민호님 관리자 ID: 8309
		
#	filename	date
1	이민호.txt	May 20 2014 15:31:10
2	홍보보.txt	May 20 2014 15:32:05
3	4학년.txt	May 20 2014 15:31:45

연구 결과 및 운영 절차

파일 삭제



연구 결과 및 운영 절차

삭제 결과 출력



연구 결과 및 운영 절차

하둡 상태창

NameNode 'localhost.localdomain:9000'

Started: Fri Mar 14 06:02:37 KST 2014
 Version: 1.2.1, r190192
 Compiled: Mon Jul 22 15:29:04 PDT 2013 by matt
 Upgrades: There are no upgrades in progress.

[Browse the filesystem](#)
[NameNode Logs](#)

Cluster Summary

6 files and directories, 1 blocks + 7 total, Heap Size is 48.0G

Configured Capacity	14.2G
DFS Used	50 KB
Non DFS Used	4.52 GB
DFS Remaining	13.75 GB
DFS Used%	0%
DFS Remaining%	95.22%
Live Nodes	1
Dead Nodes	0
Decommissioned Nodes	0
Number of Under-Replicated Blocks	0

NameNode Storage:

Storage Directory	Type	State
/home/hadoop/hdfs/tmp/dfs/name	NAME_NODE_EDITS	Active

This is [Apache Hadoop](#) release 1.2.1

Contents of directory [/user/junho](#)

Go to:

[Go to parent directory](#)

Name	Type	Size	Replication	Block Size	Modification Time	Permission	Owner	Group
하둡호.1.txt	File	0 KB	1	64 MB	2014-05-20 15:58	rw-r--r--	junho	junho

[Go back to DFS home](#)

Local logs

Log directory

This is [Apache Hadoop](#) release 1.2.1

하둡 디렉토리
파일 리스트

하둡 네임노드 상태창

결론 및 의견

- ◆ 하둡을 이용한 파일분산시스템 보안관리체제 구현함으로써
 - 대용량 (빅데이터) 자료를 적은 비용으로 관리 및 활용이 가능한 분산시스템을 구축
 - 특히 로그인 및 SSH를 활용한 보안체제를 운영하면서 스토리지 이중화로 메인 서버가 다운되더라도 자료 유실을 예방
- ◆ 하둡을 이용한 분산시스템을 구축을 통해 스토리지 이중화 및 보안 관리에 대한 기술을 습득하고 빅 데이터 자료관리의 중요성을 체험



Q & A 감사합니다