



sw maestro

Honeybee

프로젝트 계획서- version 1.0

- 데이터 시각화
- 후보 사이트 방문객 시각화
- 구매방법 및 구매금액 시각화
- 기타 필요한 정보를 데이터로 부터 시각화

Contents

[Contents](#)

[1. Preface](#)

[1.1. 데이터 시각화](#)

[1.2. 데이터 분석](#)

[1.3. 해결 방안](#)

[2. Section](#)

[2.1. Overview](#)

[2.2. Best Practices](#)

[2.3. 진행사항](#)

[2.4. 요청사항](#)

[3. 문의사항 및 전달사항](#)

[3.1. 데이터베이스 덤프 파일](#)

[3.2. 라이선스 문제](#)

[3.3. 고객정보](#)

1. Preface

1.1. 데이터 시각화

빅데이터 기술로 단순 로그 정보를 분석 함으로써 사용자가 상황 판단을 할 수 있는 시각 정보를 제공 한다.

1.2. 데이터 분석

로그, 데이터베이스의 정보를 기준으로 정보 표현에 필요한 데이터들을 빅데이터 기술로 분석 한다.

1.3. 해결 방안

많은 량의 로그 데이터를 분석 하기 위해 하둡을 기반의 빅데이터 기술 적용



2. Section

2.1. Overview

누보에 저장된 로그 데이터를 분석하여 판매자에게 혹은 누보 관리자가 필요한 정보를 손쉽게 볼 수 있는 시각화 환경을 마련 것을 목표로 한다. 따라서 로그 데이터를 분석하기 위해 빅데이터 기술의 하둡을 사용 하여 데이터를 분석하고 그에 따른 정보 시각화까지 이 프로젝트의 범위가 된다.

2.2. Best Practices

로그 데이터를 분석하고 데이터 시각화 하기 위한 최적의 방법을 수렴하였다.

2.2.1 KendoUI

그래프 시각화 하는 부분은 프로젝트 범위 밖이므로 유료정책 서비스인 KendoUI를 사용 하였으나 그래프 표현에 있어서 잘 만들어진 프레임웍 이기 때문에 앞으로 유지보수 하는데 있어서 더욱 더 많은 익점이 발생 할 수 있다고 판단하여 선택 하게 됐다. 당사가 사용시 KendoUI의 라이선스 구매를 해야 된다.

2.2.2 Hadoop

데이터 분석 서비스에 아키텍처 수립 과정에서 가장 단순 하면서 앞으로 유지보수성을 생각하여 Hadoop이라는 기본 시스템만 사용하게 됐다. Hadoop 사용법은 처음에 설치 과정이 힘드나 앞으로 분석 서비스 사용시 Pig, HIVE같은 손쉬운 방법으로 유지 보수 할 수 있는 방법이 존재 하므로 Hadoop 아키텍처를 선택 하였다.

2.2.3 정보 저장소 MySQL

Hadoop를 통해서 데이터분석이 완료 되면 분석된 데이터를 MySQL에 저장하게 된다. 이렇게 선택한 이유는 Hadoop은 분석 을 시도 시 최소 20초 라는 시간이 걸리기 때문에 앞으로 Hadoop에서 분석 완료된 정보들은 MySQL에 저장하여 어플리케이션에서 사용 할 수 있게 구성 하게 된다.

2.3. 진행사항

현재 화면 UI 프로토 타입이 개발되었고 필요한 기본 데이터 구성까지 진행. 현재 진행

사항은 데이터 수집, 데이터 분석을 하여 UI와 통합하는 과정이 있다.

2.3.1 KendoUI

데이터 시각화 하기 위해서 KendoUI로 데이터를 그래프 형태로 표현을 하게 되었고 Bootstrap로 전체 페이지 구조를 구현 했다.

2.3.2 Hadoop

소프트웨어 마에스트로 2층에 있는 서버 4대에 하둡을 시스템을 구축 하였습니다. 또한 하둡 분석 서비스를 만들기 위한 설계를 만들고 있는 상태이며 테스트 데이터를 수집 하고 있다.

2.3.3 정보 저장소 MySQL

아직 구축을 하지 않는 상태이며 최종적으로 MySQL보다 더욱 좋은 솔루션이 있는지 혹은 Hadoop데이터 분석이 완료 된 데이터를 직접 테스트 할 것이다.

2.4. 요청사항

데이터 분석에서 가장 필요한 실 데이터가 필요 하게 된다. 그렇기 때문에 누보에 불가피하게 데이터 요청을 할 수 밖에 없는 상황 입니다. 또한 구매정보와 유저 접속 정보 데이터를 추출 하기 위하여 누보에 로그인, 로그아웃 서비스를 수정 하게 됩니다.

2.4.1 데이터 요청

데이터 분석 하기 위해 실제 운영되는 데이터가 불가피 하게 필요 합니다. 따라서 프로젝트 진행을 위해 데이터를 요청합니다.

2.4.2 서비스 수정

유저 정보와 같이 데이터 분석을 할려면 로그인, 로그아웃 되는 시점에 누보 사이트에 추가적인 데이터를 입력을 받아야 합니다. 입력 받는 데이터는 session_id와 user_id가 되며 이 데이터는 어느 시점에 어느 장소에서 노부에 접속을 했고 어떤 서비스를 사용 했는지 분석 하기 위한 구조 입니다.

2.4.3 일정 계획

앞으로 프로젝트 완성 11일 정도 남았습니다. 이 기간안에 서비스 완성 해야 되므로 데이터 제공 일정 및 서비스 수정 일정을 잡아야 한다고 생각 됩니다. 그리고 최대한 빠르게 이번주 안에 데이터 수집이 완성이 될 것을 요청 합니다.



2.4.4 데이터베이스 덤프 파일 csv로 수정 요청

현재 데이터베이스 덤프 파일은 수많은 SQL문이 존재 합니다. 원래는 현재 시스템에 맞는 데이터에서 분석을 해야 되지만 프로젝트 시간 관계상 분석 시간이 부족하다고 판단 되어 csv 파일로 필요한 테이블 내용만 덤프 할 수 있도록 요청 합니다.

3. 문의사항 및 전달사항

3.1. 데이터베이스 덤프 파일

데이터베이스 덤프 파일을 전체 백업된 내용을 전달 혹은 변경 사항이 이루어진 부분만 선택하여 하둡시스템에 적용 문제

3.2. 라이선스 문제

SW maestro 주관하여 제작 하였기 때문에 분석 서비스는 SW maestro 개발팀 주관한 오픈소스 프로젝트 입니다. 따라서 당사는 오픈소스 라이선스를 사용하는 권한을 가지게 됩니다. 수정 및 배포는 오픈소스 라이선스를 따릅니다.

3.3. 고객정보

데이터베이스 덤프 파일 분석 결과 고객 개인 정보가 많았습니다. 이러한 정보들을 사용 여부에 대한 문제