California Water Drought Analysis

Project Link: http://50.112.164.74:8080/cmpe272/
GitHub Link: https://github.com/kemiporto/cmpe272-group12

Rucha Brahmankar Darrel Pham Claudia Furushima Porto Yasuhiko Sasaki Wei An Liao

CMPE272 / Team 12
Graduate School of Software Engineering
San Jose State University
San Jose, California, USA

***Abstract*— California is facing one of its worst droughts in more than 100 years and one with no end in sight. Conserving water has never been so important as it is becoming one of the most stressed resources. Our solution is divided into two parts. In one part we are focusing on the total water usage per household along with current drought monitoring in California. In the second part We propose a highly-flexible crowd-sourcing system for water data collection and analyses. The system combines cloud computing resource (Amazon EC2) with NoSQL (MongoDB) and big data analysis (Hadoop/Map Reduce) to monitor and extract informative insight from the huge water-related data.**

Index Terms—Water drought, MongoDB, Hadoop/Map Reduce.

I. Introduction

Water plays an indispensable and irreplaceable role in our life. However, some of areas in the world are suffered from the drought because of geographical environments and anthropic factor. Now, California is facing one of the worst droughts in more than 100 years and one with no end in sight. As we know, although more and more companies are doing research for solving drought problem, there is no fundamental solution to it because of the difficulty in the integration of data collection and analytics for scattered and heterogeneous data. Therefore, we propose a system that makes user aware of total water used by him in a day with ways user can conserve water. It also enables crowd-sourcing data easily and displays informative chart diagrams as a result of big data analysis using Hadoop. The system is deployed to one of the servers in cloud (Amazon EC2), naturally having a robust elasticity in case of flash crowd. One of the main features of our system is correlation analysis. Users can apply their own data to our correlation-analysis engine and get the result in real time to help them make decision of policy to solve water drought and water waste problem.

We have done an analysis on California Drought Situation. Our application provides a few solutions to the user, which will help fight the drought. It also provides information on the current situation on the water levels. Users can use our quiz to get insight on how efficient he/she is using his/her water.

In the paper, section II will describe our system's web GUI portal. Section III, IV, V and VI will show what system features are, architecture, technologies, system design and its deployment. Section VII will provide several use-cases to show how users can perform their own analyses. In the end, we describe our future work and conclusions.

II. Water Drought Analytics Portal

Although the volume of data related to "water" is overwhelming in the web space, those data are scattered, making it significantly difficult for users to find useful data and analyze them [1-7]. Our proposing system (Fig. 1) is intended for fostering the convenient place for water analyses where users can perform their own analyses in a one-stop service way.

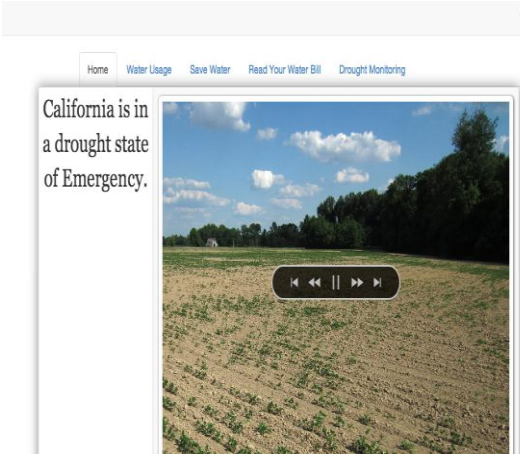


Fig. 1 Water analysis portal

At the same time, the data currently available are not sufficient for certain types of analyses. For example, even if we want to analyze household-based water consumption, such kind of data belongs to only the water companies and is not open to public. This situation makes it difficult to leverage advanced analytics technologies to be utilized in the field of household-based big data analyses. The water usage calculator helps the user to calculate such value just by answering several simple questions.

The crowd-sourcing data site featuring our system enables the collection of such household-based data and makes it easy for users (especially researchers and students) to perform various types of big data analyses.

III. System Behavior

The home page of our web system introduces the user with various parts of the application with information on each section at the bottom. There are four tabs apart from Home. Water Usage, Save Water, Drought Monitoring and Save Water Save California.

1. *Water Usage*: Water usage page helps user find out the amount of water being used per day. This is done via a Quiz where in we have questions to ask user. After a lot of study we figured out that these are the most common usage area in daily life. The questions are categorized in six parts. If user does not want to answer any specific part he has an option to do that. This quiz allows user for the flexibility he wants to have. We have kept it in mind that not every user wants to spend time in answering all the questions.

The average amount of water for almost all the activities are baseline figures which are common all over the US. This quiz is not restricted to only the California State.

The unit for the total water used is in Gallons. Once the total usage is calculated, a pie chart is created for each section, where user can see the pictorial representation of the water use per categories. This is done using highcharts.

2. *Save Water*: Getting the numbers on water usage still doesn't solve the water drought crisis. User should know the area where the water usage is more compared to others. If the use the water is not pressing, he should take actions to prevent it. We are proving a database where user has access to all the water saving tips. It is very convenient for a user to have a repository or quick access to these tips as per the category. Suppose, he needs to save water outdoor, a click to Outdoor tips will make it even more convenient and familiar to the water saving tips.

3. *View your Water Bill*: User can view his bill and do several analyses by himself. Most of the users don't even know if they can view their water bills. We just pay the utility bills online without knowing the distribution. But the water suppliers do provide the overall usage of water and amount user pays every month in the bill. User needs to have a convenient access to view his bill and do the analysis on his own.

4. *Drought Intensity Analysis*: We downloaded the precipitation, ground water level data from water.ca.gov. The data was processed using hadoop. We did the analysis of a few important parameters that can help monitor the drought situation.

- Precipitation and ground water level per year : As the precipitation increases, the ground water level also increases. This is plotted for every year. As the precipitation decreased in 2013 and 2014,the level of ground water is much less this year.

- Precipitation by month: The precipitation level was also monitored for each year on monthly basis. It was observed that during the Spring time the level of water is higher than any other part of the year.

- Overall California Drought intensity per day : We have six different level of intensities defined. Red being the worst depicts, that part of the year is in the worst drought situation. This intensity is plotted for every day. We observed that for current year, the intensity is red throughout the year. Last year being in Orange which is worst than other levels but better than red level.

For the water calculator, we developed it in javascript. Bootstrap templates were used. Below baselines have been taken into consideration.

- Efficient shower head water usage: 2 Gallons per minute.
- Efficient shower head water usage: 4 Gallons per minute.
- Half tub Bath water usage: 20 gallons
- Full tub Bath water usage: 40 gallons
- Efficient Washing Machine water usage : 40 gallons per load
- Non efficient Washing Machine water usage : 60 gallons per load
- Efficient Dishwasher water usage : 10 gallons per load
- Non Efficient Dishwasher water usage : 30 gallons per load
- Toilet Flush : 3 gallons per minute
- Lawn Sprinkler : 2.5 Gallon per minute
- Watering Plants: 1 Gallon per minute
- Other usage: cooking, drinking : 7 gallons per person

There is not much we can do by ourselves regarding the drought situation. We cannot control temperature or rain levels. The only thing each and every one of us can really do is to use water more consciousness (Fig. 2, 3). The *Save Water* section contains hints for the user to change his daily routines or house

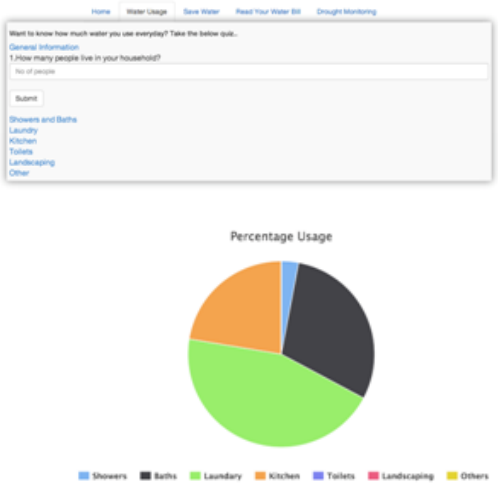


Fig. 2 Items in "Water Usage" menu

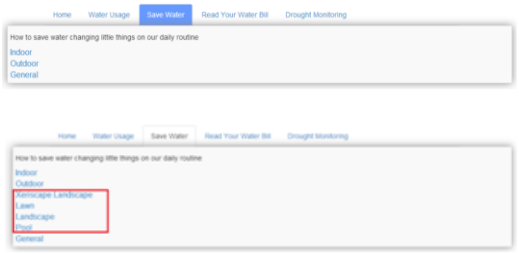


Fig. 3 Items in "Save Water" menu

equipments in order to reduce water waste. The hints are separated in indoor, outdoor and general sections.

When you click on Indoor or Outdoor it shows you the house room in which those suggestions apply, so the user can easily find and target which modifications he wants to look for.

On the picture above (Fig. 3) we highlighted the part that is expanded when the user clicks on the Outdoor part.

On the drought monitoring we can find data about California drought situation, like precipitation levels for the past few years and drought intensity. There is also some analysis and suggestions so the user can better understand and be more aware of current situation.

View Your Water Bill: To make user aware of how to read his water bill (Fig. 4). The slide show points out the main section in your water bill you should be looking at. You should compare previous water usage to your current water usage and how much the water company if charging you for the level of usage. If you use more than the allowed baseline, your water company can charge you a lot more per unit than if you were to use less.

The *Drought Monitoring* section describes previous rain fall and how your usage is comparable to past rain fall (Fig. 5, 6). If you are not conserving during low periods of rainfall, you should start thinking about changing your habits so you can contribute.

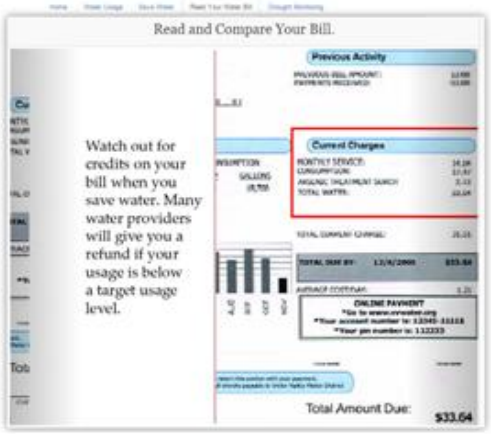


Fig. 4 Water bill

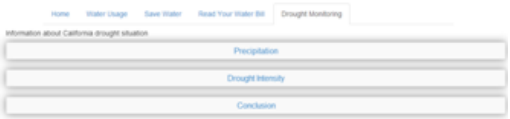


Fig. 5 Items in "Drought Monitoring" menu

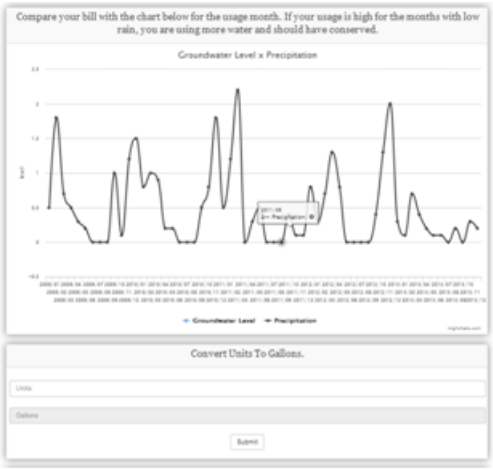


Fig. 6 Groundwater level versus precipitation

Fig.7 Water provider map

You can locate a water company near you and find out more about the current baseline and how much is being charge

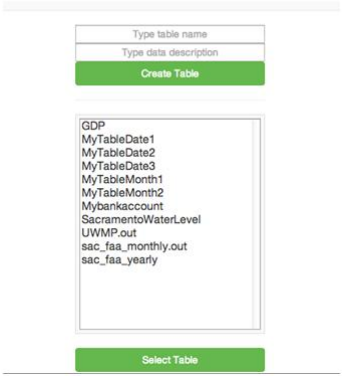
Fig. 8 Homepage of previous system

for each baseline (Fig. 7). They can also give you tips on what you can do to reduce water usage.

Actually we enhanced our system in the last stage of this project period. As a reference, our previous system *Save Water Save California* home page is provided in Figs. 8 to 11. The home page of our site is shown in Fig. 8. As you see in the top of the page, "adding data" and "viewing data" are the primary functions of the system from the user perspective. Any contributions from the users are collected through the button "Add Data" and any useful contributions (feedbacks) to the users are embedded in the button "View Crowd Source Data".

HomePage: Homepage has two options for user to enter the data

1.Add data: Click on this button to create a table of users water usages. This link navigates user to enter his table name and



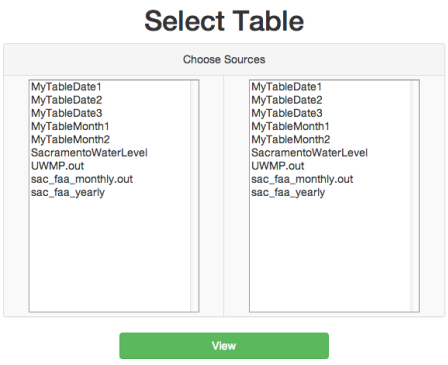


Fig. 9 Creating table and choosing sources

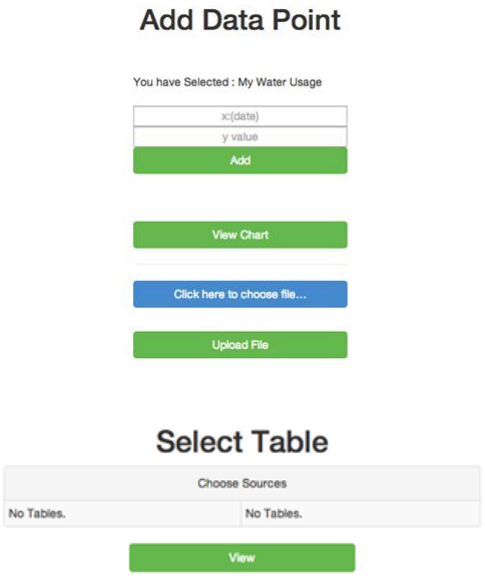


Fig. 10 Adding data and selecting table

keep adding data for himself. The x-axis can be date, month or year. The created table is stored in the database and use can access it anytime in the future. Mongo DB is used to store the data into the database.

2.View Crowd Source data: Click this if user table is already in the database or wants to compare two tables already in the database. User is navigated to SelectTable page.

AddNewData Page: New data is added here. If user selects an option to view chart he can see the difference of usage per month.

SelectTable Page (Fig. 11): On this page, the user can select the table already existing in the database. In this way user doesn't loose the track of his old data. Suppose there are two users, and they want to compare their own data together, they can do that.

View Chart: Google API is used to view charts.

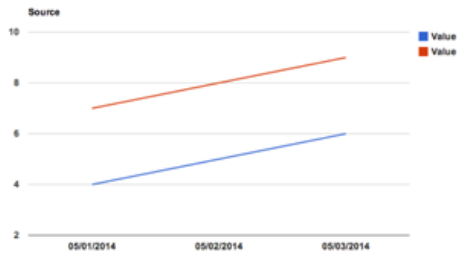


Fig. 11 Graph representation

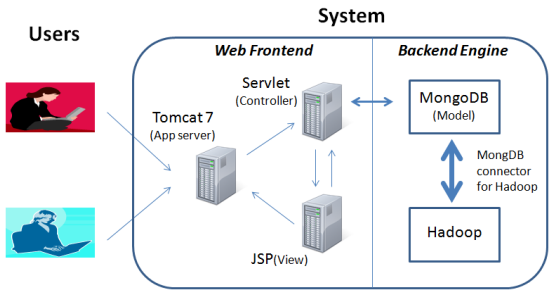


Fig. 12 System architecture

Simultaneously two charts can be viewed together (Fig. 11).

IV. System Overview - Architecture and Technologies

To achieve the purpose described in the previous section, we developed a highly flexible system. The crowd-sourcing data collection system architecture is depicted in Fig. 12 where a typical web system (frontend) is combined with a data-intensive analytics engine (backend). The web system consists of application server (Tomcat7), JavaEE servlet, and JSP (Java server pages). On the other hand, the backend engine consists of MongoDB and Hadoop, and MongoDB connector for Hadoop is used between the two. The technologies we used in the development are listed in Table I.

TABLE I. Technologies

Name	Technology type
JavaEE	Development platform
Eclipse	IDE
MongoDB	Database (NoSQL)
Hadoop	Scalable distributed computing
git	Distributed version control system
GitHub	Project hosting using git
Web Designing	HTML , Bootstrap, JSP and Servlets

V. System Design

A. Use-case diagram (Fig. 13)

Users can create tables that carries compatible data with one another to compare them and get insight. Data can be of different categories, but if they have any kind of correlation, the users can compare the two tables to gain insight.

B. Sequence diagram (Fig. 14)

Creating a new table will save the table to Mongodb and the new table is pushed to the webpage to display this new table. The user selects the table to go to a different page to add x and y data points. The user can view a chart of all of the data points currently existing in the table.

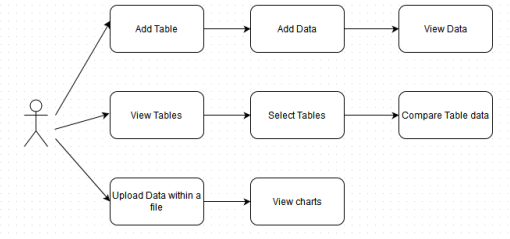


Fig. 13 Use-case diagram

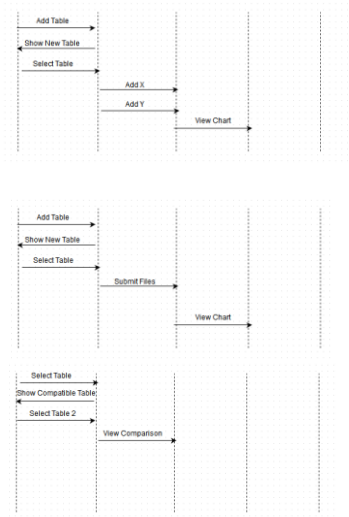


Fig. 14 Sequence diagram

The user can also upload a file containing comma delimited values. The user can go through the same process as if he/she is adding new data points, but instead upload a file to us.

The user can also select two tables to compare. When selecting the first table, a list of compatible tables who show on a new column. The user can select an item on the new column to view a comparison chart. This comparison chart should provide some insight on how these two data correlate.

VI. Deployment

The AWS EC2 was used for the deployment target. Apache Tomcat 7, MongoDB, Hadoop, and the connector between the last two were manually setup in the server in advance. Then, generated war file of our system was deployed to the server.

VII. Analysis Examples

On the drought monitoring: we can find data about California drought situation, like precipitation levels for the past few years and drought intensity. There is also some analysis and suggestions so the user can better understand and be more aware of current situation. We analyzed the data about precipitation (Fig. 15):

We can see that in the last year we had much less rain than in the previous years. California has every year a season of no

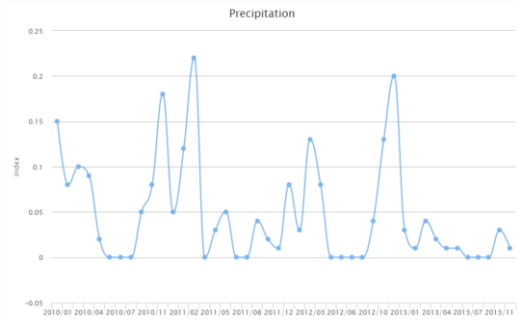


Fig. 15 Precipitation analysis

rains, but there is usually a few months during the year which precipitation gets above 0.05. Those months are responsible to increase the groundwater level enough to help us maintain the drought on a controlled way.

The problem we are facing now is mainly caused by the fact that the precipitation doesn't achieve 0.05 since January 2013.

On the analysis of California drought intensity:

The chart in Fig. 16 classifies the intensity of the drought in California in six levels: from *no drought* (nothing) to *exceptional drought*. The graph shows the percentage of California in each intensity level. We can see that we've always had periods of drought, but they would be mostly between the *moderate drought* and *severe drought* intensity. But for the year 2014 we can see that an average of around 20% of California is in an exceptional drought state and around 50% on an extreme drought state.

VIII. Ease of Use

- Provide relevant and informative content.
- Easy navigation.
- Promote efficient water usage.
- Create an attractive atmosphere.
- Evoke emotion.
- Provide special awareness with animated pages
- Collapsible accordion to hide an overwhelming amount of data.
- Simple color scheme to focus relevant information.
- Slide show to allow users to concentrate on items of interests.

IX. Challenges

While developing this application, we faced many challenges at each step of *software development life cycle.*

- To segregate the data available on water.ca.gov. Though the data was available to us, we had to convert it into the format which we wanted to use in the application.

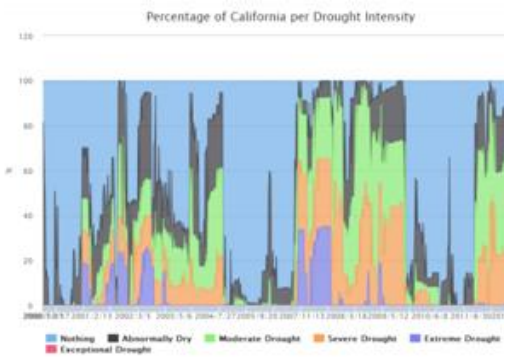


Fig. 16 Drought analysis

- Of all the charts and data available on water.ca.gov it was a challenge for us to pick up which we were able to study properly and put it in the term which would be easier for user to understand. Scientific data is not always easily readable to a layman.
- To limit to the number of use cases and make the application more efficient. This domain is so vast and useful for today that the number of use case we could have implemented is very high. Due to the time constraint we had to restrict ourselves to limited use cases. We could make the system more efficient and usable.

X. Future Work

We have implemented only the basic part of the intended system. Therefore, various features can be attached to the current system to improve user experience, system reliability, and system performance. Several ideas for the future enhancements are listed in Table II.

TABLE II. Future enhancements

Type	Detail
Automation	Automatic data connection between web frontend and backend engine
Integration	Our application can be integrated with any social media page and will give updates of the water usage and how much the user is contributing to fight water drought situation. (Something similar to gaming updates on Facebook)
Analytical Charts	More charts based on the snow level can be added to the application.
Tracking	Keep track of how many users are using the webpage and add an user login feature to improve more on security
Security	Improvise the application in security aspects. Add login features.

XI. Conclusions

We have developed a highly-flexible system to help user be aware of his total water usage and for crowd-sourcing data collection especially useful for water drought analysis. The system accumulates various data straddling multiple areas from basic environmental measures such as precipitation to fine grain data of household water consumption.

Drought monitoring graphs showed that the California water drought is currently a big problem but it can be saved by using it wisely. Although our system is still immature both in the system reliability and the performance as well as the size of data collected, we hope a successful data accumulation will help the analyses in the research field related to water and promote people to reduce the consumption of water under the crisis condition of this drought.

References.

[1] Department of Water Resources, California Cooperative Show Surveys, http://cdec.water.ca.gov/cgi-progs/iodir_ss?s=WSIHIST, 2013

[2] Department of Water Resources, 2010 Urban Water Management Plan Data, http://www.water.ca.gov/urbanwatermanagement2010_Urban_Water_Management_Plan_Data.cfm, 2014

[3] Department of Water Resources, Water Data Library : Continuous Time Series Data, http://www.water.ca.gov/waterdatalibrary/docs/Hydstra/index.cfm?site=12N04E03N003M, 2014

[4] Western Regional Climate Center, SACRAMENTO FAA ARPT, CALIFORNIA - Climate Summary, http://www.wrcc.dri.edu/cgi-bin/cliMAIN.pl?ca7630, 2014

[5] California Nevada River Forecast Center, CNRFC - Weather - Observed Precipitation, http://www.cnrfc.noaa.gov/rainfall_data.php, 2014

[6] Western Regional Climate Center, CalClim: California Climate Data Archive, http://www.calclim.dri.edu/pages/tools.html, 2013

[7] USGS, USGS Water Use Data for California, http://waterdata.usgs.gov/ca/nwis/wu, 2014