Cloud Query

Authors: Joshua Eisenberg, Maria Patterson

Abstract
This paper details the structure and creation of the Cloud Query Database and the addition of Cloud Query Metadata to the OSDC. Both of these additions will be based off of the nature of the Cloud Classification Tree.

Key Phrases: graph theory, metadata, classification

Introduction
Currently there is no single command that would allow the users of the OSDC to generate a listing of all data sources that share common properties. There is no way to obtain a list of all the weather data files from a specific location over the course of a certain time period. Similarly there is no way to analyze the statistics of the qualitative nature of the data cloud. The connections between data sets of similar sources cannot be quickly mapped.

In order to map the OSDC's data, it must be classified and ordered into a graph; Specifically this ensemble of information will form a tree. This structure will be known as the Cloud Query Classification Tree (CQT). Cloud Query (CQ) will be a tool where users can obtain information about the elements and paths of the tree by entering simple commands into an OSDC Sullivan Terminal. Finally, information regarding the subgraphs of the CQT will be compiled into a metadata database, Cloud Query Metadata (CQM), which will be added to the OSDC's collection of data sets.

The goals of CQ to establish an efficient means of searching for data sets on the cloud. CQM on the other hand can be analyzed by researchers interested in the relationships between of the CQT's many subgraphs and the connections between the data sets.

Cloud Query Classification Tree - CQCT
The first step in implementing Cloud Query is to build the CQCT. The basic graphical design of the CQCT is shown in figure 1. The data stored in the tree comprises of elements and categories. An element is a series of words that can be described using category. A category is a series of words that describes at least two elements. Functionally, a category describes one feature of a data set. The root category "Source of Data" has two possible elements, "Earth" and "Space." This distinction was made since this idea can be used to separate the locations where the data was actually recorded. Finally, all categories can be elements, except for the root category "Source of Data." Elements that represent specific data files have meta data that are standard for all elements that belong to the same category. The automation of the construction of this table will be slightly altered for each data set, since every data set has different metadata and structure.

Cloud Query - CQ
Figures 2 and 3 demonstrate two simple queries that could be implemented in Cloud Query. Cloud query is the command library that users will use to find data files that share certain properties. Given a text query to the terminal, users can either receive a list of all nodes and files that satisfy the input prompt and a subgraph that represents how the elements and categories and graphically connected.

Cloud Query Metadata - CQM
Data can be obtained by recording the structure of the subgraphs that are returned by Cloud Query. This will be classified in a new data set on the OSDC called the Cloud Query Metadata. These subgraphs can be analyzed for their dependencies, structure and how their paths relate to one another. Researchers can learn about the ways in which petabytes of data are connected. This can be applied to the fields of graph and network theory, and even help optimize the implementation of Cloud Query.

Conclusion
This paper proposed the addition of Cloud Query, which retrieves information based on the data in the OSDC, which is stored in graphical form in the Cloud Query Classification Table. The paths and subgraphs returned by Cloud Query will be stored in the Cloud Query Metatable, which can be analyzed by mathematicians and computer theorist who can benefit from

This project is truly interdisciplinary since Cloud Query can establish a search ability of the entire contents of the cloud. Scientists from any background can find data not necessarily from their field, but with related variables. Additionally, the metadata generated by the classification tree's subgraphs will form it's own data set for storage on the cloud. These proposals will make the cloud better documented and it will help us understand further what kinds of data researchers are actually looking for.