**Analyzing Vendor Payments for a fiscal year using HiveQL**

Submitted by:

Ejaaz Ahamed, Shaik (305087398)

Kiran Kumar Reddy, Thikkam Reddy (304470977)

Mounika , Vallabhaneni (304746525)

**Abstract**

The vendor payment dataset contains payments made by state agencies and institutions of higher education that are processed through the Office of Management and Enterprise Services. These payments reflect disbursements from a state fund for the purchase of goods received, services performed, reimbursements, and payments to other governments. Travel reimbursements include the Zip Code of the State Capital and not the individual’s address. For security and privacy reasons, payments related to “confidential matters” are included in the dataset; however, the payee name is replaced with “Protected Information” and the address is changed to the State Capitol address. Also, some state agencies are authorized to make payments outside of the Office of Management and Enterprise Services from certain authorized funds. In these cases, the agency reports only the total amount of the expenditures to the Office of Management and Enterprise Services but does not report the detailed vendor payment information. Therefore, these payment details are not included in the dataset.

**Introduction**

Vendor payments reflect disbursements from a state fund for the purchase of goods received, services performed, payments to other governments so as to analyse the total sum of payment spend by various organizations with respect to cities, states and also the average expenditure required for next fiscal year. Analysis of data will be based on the funds allocated to respective departments and their business units.

**Background/Existing Work**

Keeping a track of the payments made by each vendor plays a major role in estimating the average expenses for the next fiscal year. Earlier there are ways where people use to track the payments in search of papers or receipts. But now a new approach was proposed using HiveQL which tracks the yearly vendor payments as per city, state, and department and business unit standards.

**Our Work**

Primarily Cloudberry for Azure Blob Storage is downloaded to our local and has been installed. Later it has been configured to our storage details of Azure. A Hadoop HDInsight cluster has been created and deployed over the Azure cloud and the same was displayed in the Cloudberry local. The dataset is loaded to the /*Root/ (Cluster container name)/HdiSamples* path. A table is created using the Hive query and the data is loaded into the table with the help of TEZ engine. Once this flow is done an ODBC driver is configured in our PC with the cluster credentials and URL. On the other hand Excel 2013 is opened in the local and the queries are executed using the given path, *Data/From Other Sources/Microsoft Query/ (ODBC Configured connection)*. Once the queries are processed they are elevated to a graphical representation using the properties of Excel. Analysis of data in various aspects has been done using various queries.

**Tools Used**

1. HiveQL
2. ODBC driver.
3. Excel 2013
4. Hadoop HDInsight cluster
5. Worker nodes for cluster: 4
6. Head Nodes for cluster: 2
7. Cloudberry Azure Blog Storage

**GitHub details**

The project is uploaded in the GitHub for further references. Details include the queries and documentation.

* <https://github.com/ejyshaik/Vendor-Payments>

**Dataset details**

Our dataset is about the vendor payments where each state was disbursed with some funds which the state uses for their purpose. Looking in depth at the dataset, it consists of various states, cities, departments for which the fund is disbursed, fund details and descriptions, payment amount, business unit and their agency names. The dataset is of about 1,048,576 records and of 428.5 MB in size. The following is the dataset URL for reference.

* <http://catalog.data.gov/dataset/state-of-oklahoma-vendor-payments-fiscal-year-2015>

**Graphical representation of queries**

The below graphical representation shows the total sum of the payment made by different business units for one financial year.

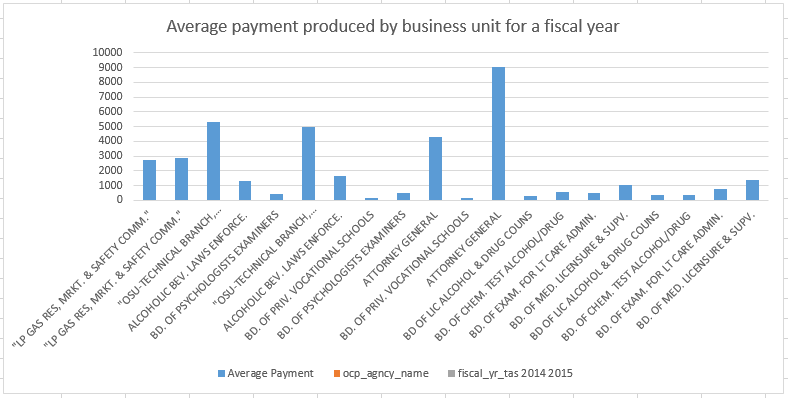


Figure 1. Average payment produced by business unit for a fiscal year.

The following figure shows the pictorial representation of the total amount of payment produced by various departments.

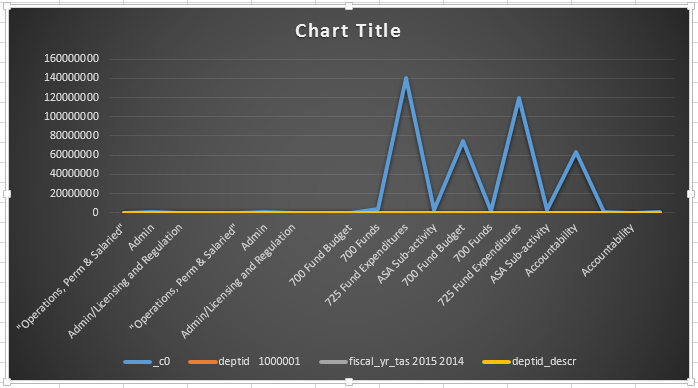
****

Figure 2. Total amount of payment produced by various departments.

The amount of vendor payments done for a fiscal year 2015 for cities Los Angeles, San Francisco and San Diego is also analysed. Below is the pictorial representation.

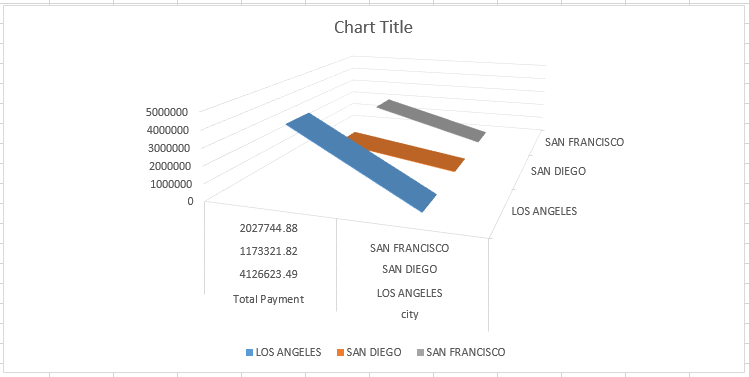
****

Figure 3. Payment produced by cities Los Angeles, San Francisco and San Diego.

Total payment analysis for state wise is also analyzed. Below is the graphical representation.

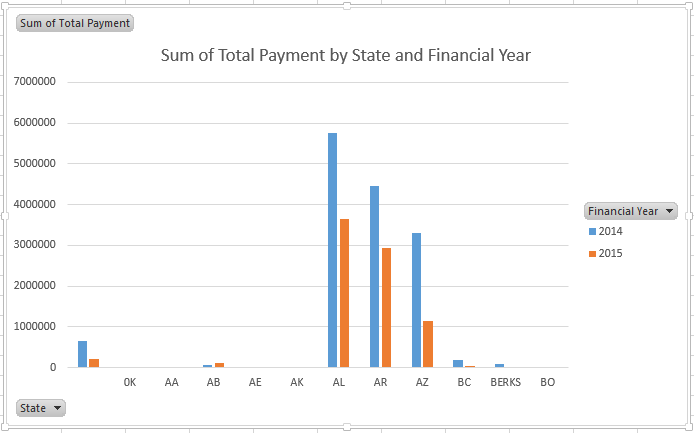


Figure 4a. Payment produced by each state on FY basis.

The same was elevated on the US map.

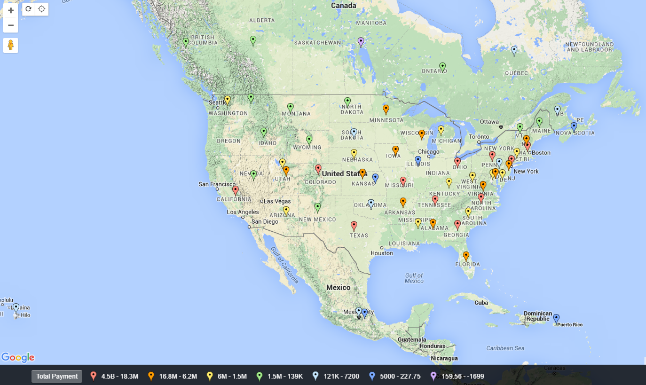


Figure 4b. Payment produced by each state on FY basis elevated on the US map.

An analysis is also done to find out the top five business units for a fiscal year 2015.

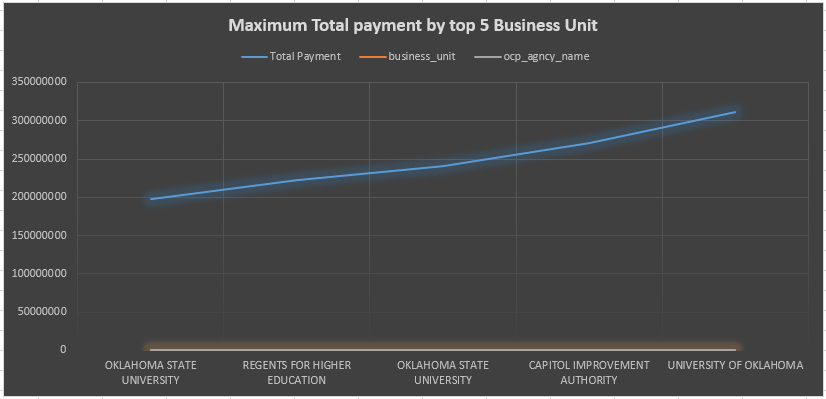


Figure 5. Top 5 business unit names grouped by expenditures spent.

The same analysis was carried out to find out the top five departments on the amount of payments made.

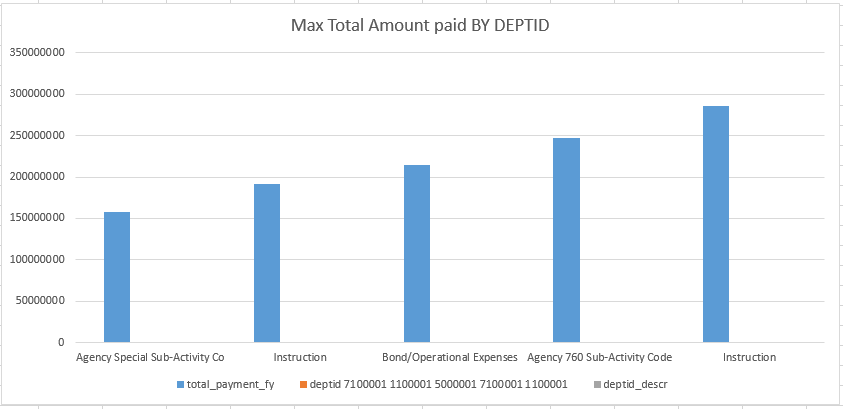


Figure 6. Top 5 departments on the basis of amount of expenditure spent.

An analysis has been performed to find out the payee occurences for the particular business units which in turn means that the number of times the expendiure spent for a particuar business unit by a payee.

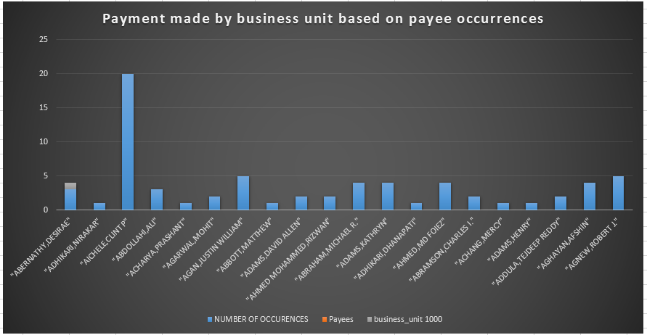


Figure 7. Payee occurrences for the particular business units.

Analytic for the data set is also carried out by estimating the average amount of payment for the upcoming financial year. Pictorial representation follows,

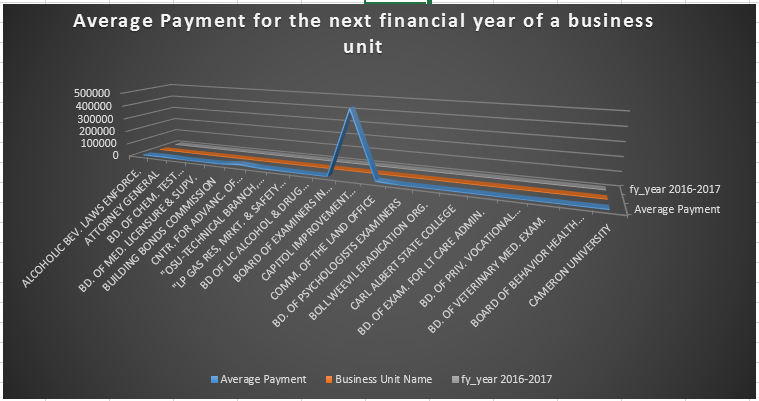


Figure 8. Average amount of payment for the next financial year.

**Conclusion**

From above statistics a clear insight on the average amount of payments for various departments and their numerous business units has been articulated. An average payment amount for a fiscal year of various business units has been calculated to cut down the unnecessary costs and Individual monthly and fiscal year payment details of funds allocated, business units and departments of their respective States and Cities has been cited. Average amount of payment for the next year is also calculated which will serve as an analytics purpose in fuure.

**References**

* <http://catalog.data.gov/dataset/state-of-oklahoma-vendor-payments-fiscal-year-2015>
* <http://www.cloudberrylab.com/free-microsoft-azure-explorer.aspx>
* <http://portal.azure.com>
* <https://www.microsoft.com/en-us/download/details.aspx?id=36434>
* <https://cwiki.apache.org/confluence/display/Hive/LanguageManual>
* <http://hortonworks.com/blog/evaluating-hive-with-tez-as-a-fast-query-engine/>