Presented by: Ryan Lawson, Christie Sutton, Austin Bowman, and Trey Watson



Photography Data Warehouse Design & NOsql Plan

Image Photography Studios Project Design

East Tennessee State University

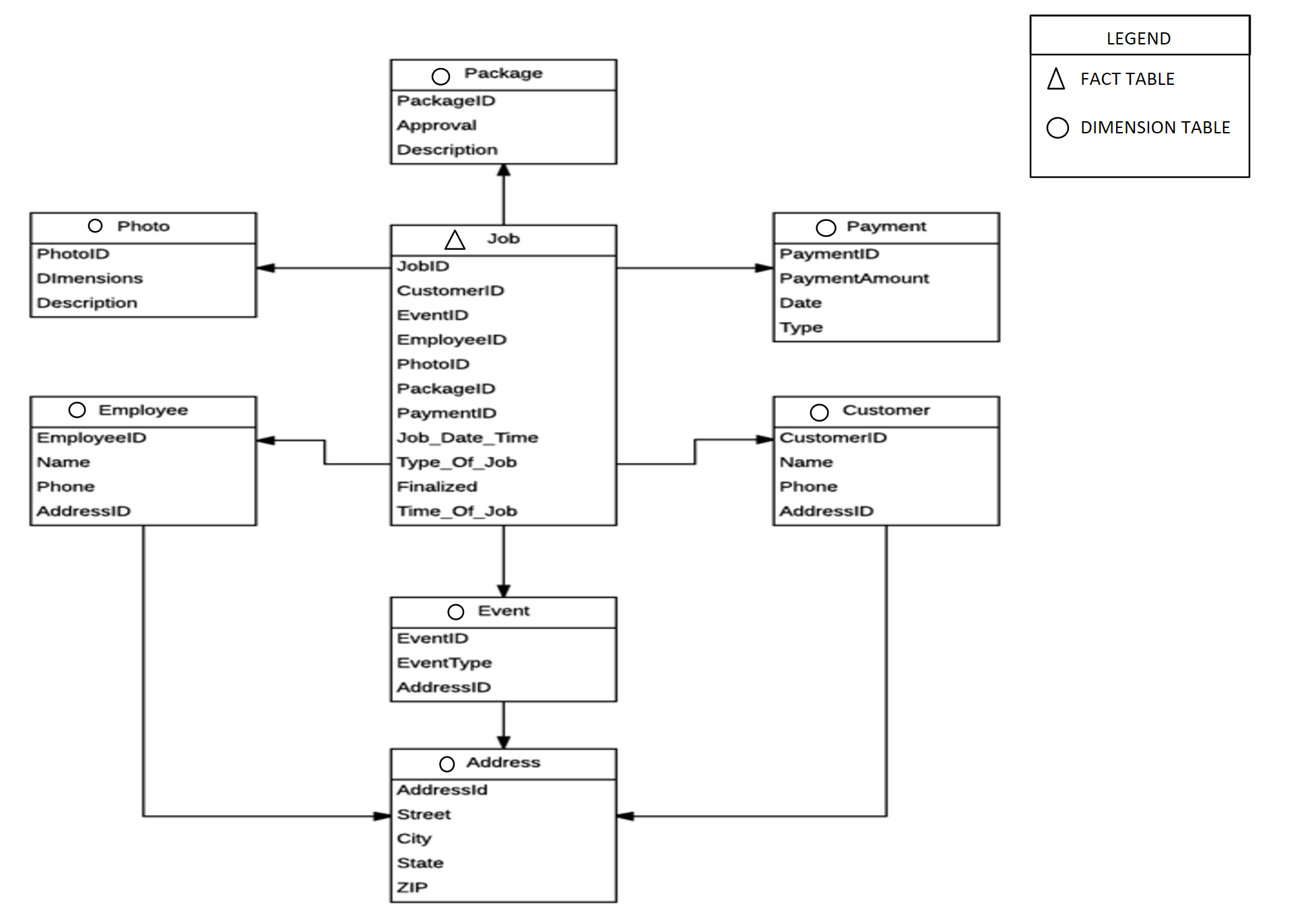
Version 1.4

August 8, 2017

# Photography Data Warehouse Design & NOsql Plan

## Data Warehouse design

### Snowflake schema





### Justifications

The term Data Warehouse was coined by W.H. Inmon who described it as, “A subject-oriented, integrated, non-volatile, time-varying collection of data that is used primarily in organizational decision-making.” It is appropriate to use a data warehouse for Image Photography Studio, because it allows flexibility in reporting, the analysis of business growth, and increases the recognition of underlying data. Image Photography Studios could use the data within the data warehouse to determine the optimal time to expand operations. It can also help to monitor employee performance over time.

### extract-transform-load (ETL) process plan

The Extract-Transform-Load(ETL) process of moving the production database to a data warehouse will consist of a series of insert-select statements. The extraction for the Payment table, the Photo table, and the Package table would be the simplest of tables due to the lack of connections to other tables. The data would not have to be transformed, and the load would be completed with the inserts. The next table would be the address table, and would require multiple points of extractions without duplicate data. The address table extraction would require three tables union together. The transformation would only require the auto-generated ID, which is completed by the row number function; however, no duplicate can be satisfied by the union. Next, Employee would map to the address table requiring a join statement to the new address table with the accurate ID. Now, the photo table would just require the link to jobs, and would not require a difficult ETL process. The Customer table and the Event table would require the same join to Address as with the Employee table. Finally, the fact table Job would be pulling most of the data from the original Job table in the production warehouse.

## Nosql

### To Nosql or to not nosql

NoSQL, specifically MongoDB, would not be a good fit for Image Photography Studio’s. One reason for this is that NoSQL permits any data to be saved anywhere at any time without verification, while an SQL database is more secure because it implements data integrity rules. This would limit the access to the company’s data without verifying who is trying to change or manipulate the data. A second reason being that NoSQL is a newer technology, where an SQL database offers plenty of support, expertise, and tools. NoSQL is more exciting because it is new, but SQL would be better for a company that is just starting out because they offer so much support and tools, when running into any issues that could arise. The final reason would be that NoSQL offers less flexibility when querying data verses an SQL database that allows the use of joins, which provides greater flexibility when manipulating data.

### etl process plan

In the future if Image Photography Studio is moving to the NoSQL, MongoDB, then the plan would be to export all the desired queries to the required format. Then to import them into the desired collections within the NoSQL database. This completes the ETL because MongoDB is already partially set up for a data warehouse.

### Comparisons

The assumption we made with the ETL of both the SQL data warehouse and the NoSQL is that SQL is better to use in the beginning, because converting it to the NoSQL verses starting from scratch is easier. NoSQL is a complex system and requires a better understanding of your documents to query data, while the SQL database is more in depth and simpler by design. Image Photography Studio’s current amount of data being produced is not large enough to need NoSQL therefore it would create unneeded expenses for the company at this time.