University of Oklahoma

Big Data

Olivia Long, Evan Langenberg, Karen Li, Marshall Mann, Quannah Lindsey

MIS 3353 – Database Management

Swetha Siripurapu

October 23, 2020



*“Don’t compromise on excellence.”*

# **Executive Summary**

The purpose of this project is to provide the Elysian Fly Company with a new relational database. This new database will allow Elysian Fly Company to improve their efficiency, customer segmentation, and ultimately increase profitability. The database built by our company should allow Elysian Fly Company to do their overall business better and therefore give customers a better experience. The scope of this project includes an Entity Relationship Diagram along with assumptions made to form the ERD.

Customer segmentation has been a key issue that needed to be addressed. This project allows for easier customer segmentation by giving customers, orders, and trips attributes that will be put into the database. The ERD created has set the groundwork for building the Elysian Fly Company database. With all of the data collected, Elysian Fly Company will be able to output reports on all aspects of their company, including but not limited to sales, customers, trips, and vendors.

Big Data has collectively logged 16 hours of work into the creation of the ERD. This time logged includes working together as a team as well as individual hours. During this time the team met with the client, made assumptions about the case, completed the ERD design, and prepared a write up. In total phase one of this project is expected to cost Elysian Fly Company $400.

**Contents**

[**Executive Summary**](#_heading=h.1fob9te) **1**

[**Get to Know the Team: Big Data**](#_heading=h.tm6m9ipnsbx5) **4**

[**Conceptual Design**](#_heading=h.2et92p0) **5**

[The Client Meeting](#_heading=h.tyjcwt) 5

[Q&A During the Meeting & Information We Learned](#_heading=h.3dy6vkm) 5

[**Significant Assumptions**](#_heading=h.z2b07xe87ds0) **7**

[What is an ERD? Why is it necessary?](#_heading=h.2s8eyo1) 7

[Business Cycles Used](#_heading=h.17dp8vu) 8

[**ERD Created**](#_heading=h.v5b7qcy36g8y) **9**

[Changes made to generic ERDs](#_heading=h.79jl69rfi2nd) 10

[**Logical Design**](#_heading=h.26in1rg) **13**

[Normalization](#_heading=h.5ksxhqp56tsa) 13

[**Normalized Relations (Primary Keys are underlined; Foreign Keys italicized)**](#_heading=h.35nkun2) **14**

[Differences between ERD and Normalized Relations](#_heading=h.7ns1ktppvei1) 17

[Referential Integrity](#_heading=h.3esb0hw2rrrp) 17

[**Physical Design and Implementation**](#_heading=h.1ksv4uv) **18**

[Data Dictionary](#_heading=h.9nf7jhuizhwt) 18

[Denormalization](#_heading=h.44sinio) 18

[Implemented Physical Design](#_heading=h.2jxsxqh) 19

[Challenges Faced/Addressed During Implementation](#_heading=h.hggtkator1y) 20

[Strengths and Weaknesses Encountered During Implementation](#_heading=h.4i7ojhp) 20

[**Specific SQL Statements Requested**](#_heading=h.hfzjxvrmkcc5) **21**

[Three Additional Queries](#_heading=h.2bn6wsx) 27

[**User Documentation**](#_heading=h.qsh70q) **29**

[**What We Learned Throughout This Process**](#_heading=h.4vpnidh287rl) **31**

[**Appendix**](#_heading=h.1pxezwc) **35**

[Team Contract](#_heading=h.49x2ik5) 35

[Data Dictionary Model](#_heading=h.2p2csry) 36

[Project Management](#_heading=h.147n2zr) 38

# **Get to Know the Team: Big Data**

|  |  |
| --- | --- |
|  | **Name:** Evan Langenberg  **Major:** Accounting  **Minor:** MIS  **Year in School:** Senior  **Internship Experience:** None yet **Background:** Evan is from Houston, Texas, and has extensive experience with SQL, SAP, Mendix, PowerBI, and VBA. |
|  | **Name:** Quannah Lindsey  **Major:** BBA-MIS, MS-MIT  **Year in School:** Senior  **Internship Experience:** None  **Background:** SQL, PowerBI, Python, VBA, R |
|  | **Name:** Karen Li  **Major:** MIS  **Year in School:** Junior  **Internship Experience:** None  **Background:** Karen is originally from Tulsa, OK. She moved to Norman, OK to pursue a Bachelor’s Degree in MIS at the University of Oklahoma. Throughout her college years, Karen has had experience working with SQL, VBA, Excel, and C#. |
|  | **Name:** Olivia Long  **Major:** MIS & Energy Management  **Year in School:** Junior  **Internship Experience:** None  **Background:** From Dallas Area, knowledge in SQL, C#, VBA |
|  | **Name:** Marshall Mann  **Major:** MIS & Energy Management  **Year in School:** Senior  **Internship Experience:** O&G experience  **Background:** from OKC area, SQL, Python, C# |

# **Conceptual Design**

The purpose of conceptual design is to visually show the entities, relationships, and attributes of the content of the database that needs to be created. Moreover, through conceptual design, we aim to abstract the structure of our database by analyzing the information we have been given. During this process we will work to construct our rough drafts of an ERD through hashing out assumptions derived from the case. By participating in conceptual design we aim to help Elysian Fly Company become more profitable and efficient by creating a holistic and thorough ERD to meet their needs.

## **The Client Meeting**

This section contains statistics about the client meeting including when we met with the client, where we met with the client, who interviewed the client and the identity of the client interviewed.

* Meeting Time: 10/6/20, 3:00-3:30 pm
* Location: Zoom
* Interviewers: Olivia Long, Quannah Lindsey, Karen Li, Evan Langenberg and Marshall Mann
* Interviewee: Sam Ferreira

## **Q&A During the Meeting & Information We Learned**

The following are the questions we asked during our client meeting, notated as numbers, followed by the answers, indicated by letters.

1. On the guide services section, would you like us to model the different lakes and rivers that the guides will take clients to?

a. Need to know the location where the trips go.

b. GPS location calculation is fine

2. Would it be safe to say a customer can only be one of the 3 types?

a. More than likely yes, locals who are regulars.

b. Distant customer who moves to the local area can be reclassified

3. What is your criteria for most preferred vendors? Number of sales?

a. Most preferred vendors would be the number of sales and quality

b. Once they are in the system they are most likely your vendors until you have to redo their contract

c. We can have multiple vendors.

d. A lot of deals and sells a lot of products.

e. If you are a vendor that does a lot of sales.

f. You may reorder the preferred vendor before the less preferred vendor because you know the rate of sales.

g. Reorder points are very important.

4. The effect of flyers on sales

a. They want email addresses to facilitate marketing of catalogs

b. If it is a repeat customer you can track their purchases. You don’t have to ask the customer for information. If we don’t have customer info they will input it, if they do have information they are a repeat customer.

5. What is the criteria for giving out discounts to customers?

a. Discounts are mostly seasonal

b. Sometimes there are seasonal discounts. We come up with discounts when customers go five times, holiday specials, Seasonal visitors are given a discount.

6. Does the firm want to further investigate the pattern between type of customers and type of guides?

a. A customer is allowed to request a guide

b. When a customer calls in

c. Really nice if a customer can request for a guide

d. If a customer is seasonal, more than likely they are busy during the season and might not get the client they want.

7. Are employees also customers as well

a. Yes

b. We tell the employees not to take off during peak season

c. Employees get a discount

d. Employees are assigned to multiple roles and are not worried about employees being a customer.

8. Do you record multiple phone numbers and addresses?

a. One time customer: local address where you are staying

b. Multiple addresses and phone numbers

9. On the fly bundles, can you do any combo on the bundles or are they limited to one type of fly?

a. Determine how many flies the customer wants to buy

b. Normal bundles are plastic fly

c. You are allowed to customize the combo and say I will take a premade package or customize hooks from a particular vendor and put together the rest of the items in the kit.

d. DIY bundle where a customer is allowed to put together whatever he wants.

10. What does systematically control purchasing processes mean?

a. On purchasing,

b. Purchasing is done by authorized individuals

c. Needs to make sure there is a person who is doing this

d. Separation of duties

e. Who is qualified to make a purchase.

f. Do not share questions with any of the other teams.

11. What quantitative factors are involved in measuring salespeople and vendor manager success?

a. Sales measures success for both categories

12. What kind of information would you like us to collect about your customers?

a. Everything, CC info, etc.

13. DIY bundles is one of the production cycles and you will tell them if your ERD belongs to one of the cycles.

# **Significant Assumptions**

In this section, we will discuss the significant assumptions that helped shape key decisions in designing this ERD. A significant assumption could be quantified as an assumption derived from information in our case study that bears significant weight with regard to particular design choices. The following assumptions will be assigned respective numbers one through five.

1. The **OfferingBundle** is an associative entity that provides a copacetic way to track product bundles such as our lure kit by aggregating several products together and returning a unique product ID to be used in offerings.
2. **CustomerType** is a reference table that was created in order to account for the assumption that customer types are relatively static, as described by the firm. Therefore, when sorting and segmenting our customers, we can sort by the type of customer they are in one singular location.
3. **Fly** is not limited to just flies themselves, the entity Fly can also track other products/parts of kits such as feathers, hooks, flies that can eventually be subcomponents of a Bundle. Fly is set up this way because a Bundle often requires more than just flies.
4. An **Employee** can be either a regular employee, or a guide who leads Customers on Trips. This is denoted by the EmpGuide binary attribute on the Employee table.
5. **SalesChannel** is the reference table that connects to our SalesOrder table and delineates sales based on one of the three sales channels: in-person, phone, or online.
6. We do not track the payment info or **PmtInfo** for our invoices regarding materials purchased. To reflect the payment information of materials we purchased, we simply refer to the cost of those materials which is tracked for each fly as **FlyCost**.

## **What is an ERD? Why is it necessary?**

An ERD can be defined as a visual representation of a relational database that shows how each piece fits together. More specifically, it shows how the nouns and verbs of business relate to one another. For example, imagine that you are a manager and want to create a diagram of the team you manage. What is the relationship between the team and the employees that make it up?

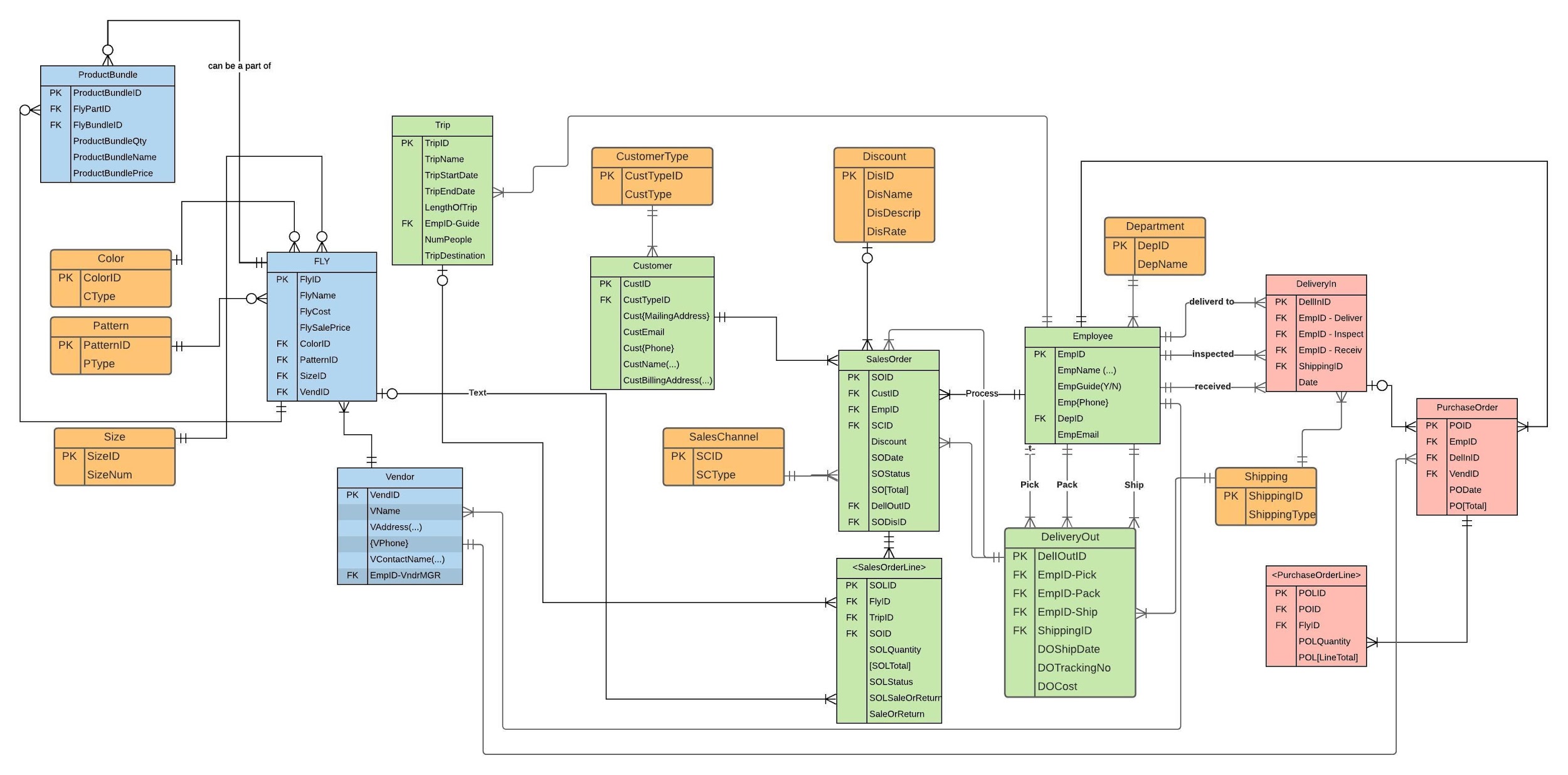
One team has many employees, but depending on your company’s structure, an employee could also be on many teams, or maybe they are only assigned to one at a time. These important distinctions are important components of an ERD, and show why ERDs are so important to showing relationships between data.

## **Business Cycles Used**

Our ERD incorporated all three business cycles: the revenue cycle, expenditure cycle, and lastly, production cycle. The firm, Elysian, utilizes a modified version of the revenue cycle. The reason that we knew to include this cycle is because Elysian sells both products and services for profit, and thus requires a unique model of their revenue cycle. Elysian also has an expenditure cycle, as the firm does have to procure products and raw materials from vendors to manufacture and sell. Moreover, as previously mentioned, Elysian manufactures goods from raw materials, thus they require a modeling of their production flow in the ERD.

# **ERD Created**

*Our ERD:*



## Changes made to generic ERDs

|  |  |  |
| --- | --- | --- |
| **Change #** | **Original ERD** | **Updated ERD** |
| **Differences b/t Customer Entity:**   * Customer is now attached to the reference table CustomerType |  |  |
| * SalesOrder now has a reference table attached to it named Sales Channel |  |  |
| * Employee table is connected to the Reference table Department also DeliveryOut, and DeliveryIn. |  |  |
| * Product bundle is an associative entity that bundles products together. |  |  |
| * Product now has the color, pattern, and size reference tables attached |  |  |
| * DeliveryOut is now connected to Shipping. |  |  |
| * DeliveryIn is now attached to Shipping. |  |  |
| * Fly and Trip are now not part of a supertype formerly called Product, and are their own entities. |  |  |
| * SalesOrder has a reference table attached to it named Discount |  |  |

# **Logical Design**

Logical design is the phase where normalization occurs. This phase of logical design serves as a blueprint to guide developers. Through normalization, the logical design phase helps clear up any redundancies and makes the database more efficient for whenever reports are run. In this project, logical design is used by normalizing the relationships created for Elysian Fly Company. For this project form 3NF is used to normalize the data, this means that all columns are atomic, there are no partial functional dependencies, and there are no transitive dependencies. By using 3NF the relations should have no data duplication problems and they should be atomic.

## **Normalization**

Normalization is the process used to ensure that a database outputs reliable reports by attaining data integrity. It also helps ensure that when the database is outputting reports, that it is efficient in that process. Normalization will ensure the atomicity of columns and eliminate data duplication issues.

# **Normalized Relations (Primary Keys are underlined; Foreign Keys *italicized*)**

TColor(ColorID, CType)

TPattern(PatternID, PType)

TSize(SizeID, SizeNum)

TSalesChannel(SCID SCType)

TCustomerType(CustTypeID, CustType)

TDepartment(DepartmentID, DepName)

TShipping(SID, ShippingType)

TEmployee(EmpID, EmpFirstName, EmpLastName, EmpGuide(Y/N), EmpHomePhone, EmpCellPhone, EmpEmail, *EmpDepID* )

Foreign Key EmpDepID references TDepartment

Not Null

On Delete No Action

TVendor(VendorID, VName, VAddress, VPhone, VContactName, *EmpID-VndrMGR)*

Foreign Key references TEmployee

Not Null

On Delete No Action

TFly(FlyID, FlyName, FlySalePriceFlyCost, *ColorID, PatternID, SizeID, VendorID*)

Foreign Key ColorID references TColor

Not Null

On Delete No Action

Foreign Key PatternID references TPattern

Not Null

On Delete No Action

Foreign Key SizeID references TSize

Not Null

On Delete No Action

Foreign Key VendorID references TVendor

Not Null

On Delete No Action

TProductBundle(ProductBundleID, *FlyBundleID, FlyPartID*, ProductBundleQty,

ProductBundleName, ProductBundlePrice)

Foreign Key FlyID-Bundle references TFly

Not Null

On Delete No Action

Foreign Key FlyID-Part references TFly

Not Null

On Delete No Action

TTrip(TripID, TripName, TripStartDate, TripEndDate, *EmpID-Guide*, NumPeople)

Foreign Key EmpID-Guide references TEmployee

Not Null

On Delete No Action

TCustomer(CustID, *CustTypeID*, CustMailSteet, CustMailCity, CustMailState, CustMailZip CustEmail, CustCellPhone, CustHomePhone, CustFirstName, CustLastName, CustBillStreet, CustBillCity, CustBillState, CustBillZip)

Foreign Key CustTypeID references TCustomerType

Not Null

On Delete No Action

TDiscount(DisID, DisName, DisDescrip, DisRate)

TSalesOrder(SOID, *SOCustID*, *SOEmpID*, *SOSCID*, *SODelOutID*, *SODisID*, SODate, SOStatus)

Foreign Key SOCustID references TCustomer

Not Null

On Delete No Action

Foreign Key SOEmpID references TEmployee

Not Null

On Delete No Action

Foreign Key SOSCID references TSalesChannel

Not Null

On Delete No Action

Foreign Key SODelOutID references TDeliveryOut

Not Null

On Delete No Action

Foreign Key SODisID references TDiscount

Not Null

On Delete No Action

TSalesOrderLine(SOLID, *SOFlyID*, *SOID*, SOLQuantity, SOLStatus, SOLSaleOrReturn)

Foreign Key SOLID references TSalesOrder

Not Null

On Delete No Action

TDeliveryIn(DelInID, *DelEmpID-Deliver*, *DelEmpID-Inspect*, *DelEmpID-Receive*, *DelInShippingID*, Date)

Foreign Key DIEmpID-Deliver references TEmployee

Not Null

On Delete No Action

Foreign Key DIEmpID-Inspect references TEmployee

Not Null

On Delete No Action

Foreign Key DIEmpID-Receive references TEmployee

Not Null

On Delete (No Action)

Foreign Key DIShippingID references TShipping

Not Null

On Delete No Action

TPurchaseOrder(POID, *POEmpID*, *PODelInID*, *POVendID*, PODate)

Foreign Key PODelInID references TDeliveryIn

Null Allowed

On Delete Set Null

Foreign Key POEmpID references TEmployee

Not Null

On Delete No Action

Foreign Key POVendID references TVendor

Not Null

On Delete No Action

TPurchaseOrderLine(POLID, *POPOID,* POLQuantity)

Foreign Key POPOID references TPurchaseOrder

Not Null

On Delete No Action

TDeliveryOut(DelOutID, *DelEmpID-Pick*, *DelEmpID-Pack*, *DelEmpID-Ship*, *DelOutShippingID*, Date)

Foreign Key DOEmpID-Pick references TEmployee

Not Null

On Delete No Action

Foreign Key DoEmpID-Pack references TEmployee

Not Null

On Delete No Action

Foreign Key DoEmpID-Ship references TEmployee

Not Null

On Delete No Action

Foreign Key DOShippingID references TShipping

Not Null

On Delete No Action

## **Differences between ERD and Normalized Relations**

The major difference between an entity relationship diagram and normalized relations is that when normalizing, attributes are assigned to specific fields and relationships are further broken down. For example, a multi value attribute is acceptable to be included in an ERD, but when normalization occurs that multi value attribute has to be taken into parts of the subjects that it represents. Another major difference is that ERDs are a diagram showing the relationships between entities, while normalized relations dig into those relationships and help ensure that there are no redundancies. It is beneficial to have normalized relations to ensure that there will be no redundancy in the data and ensure the atomicity of the data. Ensuring these two factors helps confirm that when reports are outputted, they will be accurate and efficient.

## **Referential Integrity**

Referential Integrity refers to the association between entities. This means that if there is a relationship among entities, the foreign key shown has to match up with a valid primary key, or else it must be null. This ensures that a reference from a row in one table to a row in another table is valid. The formula for Referential Integrity is:

Foreign key [name] references [table name mandatory side]

[Not Null/Null Allowed]

On Delete (Restrict/Cascade/Set Null)

On Update Cascade

By adding this into the normalization process, the principal end of the constraint, the dependent end of the constraint, and the referencing property on the dependent end is specified.

# **Physical Design and Implementation**

Physical design is a step in the database design process where the design of the database is integrated into a RDBMS. An RDBMS is a relational database management system. The purpose of the physical design aspect helps ensure that the database outputs reports efficiently. Physical design is platform specific which means that the design will vary depending on which server it is built for. This is important for Big Data to use to ensure that Elysian Fly Company has a database that outputs reports correctly and in a timely fashion.

## **Data Dictionary**

A data dictionary is an output resulting from the physical design process. The data dictionary reflects the field name, data type, size (if applicable), whether it is null or not null, the primary key range, references, and sample data. This helps the database developers create sample data, determine how the database should be implemented, and determine how the data fields should be used. Big Data used a data dictionary to help create sample data tables. These tables were then transferred to be used to create the required queries. An example of the data dictionary Big Data used is shown below.



In this example the table TFly is shown. The attributes of Fly are listed along with each of their data types. For FlyName and FlyCost, the sizes are shown because they are a varchar and double. The “F200-299” represents the primary key range. Color, Pattern, Design, and Vend ID show where they are referenced from since they are foreign keys. The far-right column shows examples of how the data will be displayed. All of these components are helpful to give developers a visual of the data and make the overall database creation process easier.

## **Denormalization**

Denormalization is the process of reducing the normal form of tables within the database. The process of denormalization allows data duplication, which is the reverse of what normalization does. Denormalization allows for the database to process queries more efficiently. Because denormalization allows for data duplication and condensing fields, problems may arise when using this process. There are guidelines as to whether to implement denormalization or not. These guidelines are if the entries for a denormalized field are the same, and if a reference table is referenced by only one other table and that its attributes are not usually included in a WHERE clause.

## **Implemented Physical Design**

## 

## 

## **Challenges Faced/Addressed During Implementation**

Challenges that we faced included missing attributes within our data once we started writing queries, deciding whether or not to implement denormalization, and making sure everything was cohesive. We overcame the missing attributes by adding attributes into our database to ensure our queries could run effectively. When determining whether to normalize or denormalize we also evaluated our data to see if we needed to have it in 2NF or 3NF to make our database more accessible.

## **Strengths and Weaknesses Encountered During Implementation**

Strengths that Big Data possessed throughout implementation included creating the data dictionary, working well together, and testing our queries efficiently. Some weaknesses of Big Data while building the ERDs and writing queries when implementing this database included creating the original queries, recalling information from previous modules, and working with a new database system that was unfamiliar to us.

# **Specific SQL Statements Requested**

In this section we have displayed the specific programs we were asked to execute by Elysian Fly Company. These queries were written and tested in Microsoft SQL Server Management Studio by the team. Included below is the question asked, the SQL code, and a glimpse of the output is displayed to give a visual on what the report ran.

|  |  |  |  |
| --- | --- | --- | --- |
| **Query #** | **Question** | **SQL** | **Partial Output** |
|  | What is the total sales(in dollars) by customer state per year? | SELECT SUM((FlySalesPrice\*SOLQuantity)+(T.Price\*SOLQuantity)) as "Total Sales", Year(SODate) as "Year", CustBillState  FROM T\_Customer C JOIN T\_SalesOrder SO ON C.CustID=SO.SOCustID JOIN T\_SalesOrderLine SOL ON SO.SOID=SOL.SOID JOIN T\_Fly F ON SOL.FlyID=F.FlyID  JOIN T\_Trip T ON SOL.TripID = T.TripID  GROUP BY Year(SODate), CustBillState |  |
| 2. | What is the total sales(in dollars) by vendor per year? We must be able to calculate profit. | SELECT SUM(SOLQuantity\*FlySalesPrice) as "Total Sales", (FlySalesPrice-FlyCost) as Profit, VName, Year(SODate) as "Year"  FROM T\_SalesOrder SO JOIN T\_SalesOrderLine SOL ON SO.SOID=SOL.SOID Join T\_Fly F ON SOL.FlyID=F.FlyID JOIN T\_Vendor V ON F.VendID=V.VendID  GROUP BY VName, FlySalesPrice, FlyCost, Year(SODate) |  |
| 3. | What are the ten highest selling patterns, sizes, and pattern-size-color combinations in a given year? | SELECT Top 10 Count(FlyBundleID) as "Combinations", PatternID, SizeID, ColorID  FROM T\_SalesOrder SO Join T\_SalesOrderLine SOL ON SO.SOID=SOL.SOID  JOIN T\_Fly F ON SOL.FlyID=F.FlyID JOIN T\_ProductBundle PB ON F.FlyID=PB.FlyPartID  WHERE Year(SODate)=2019  GROUP BY PatternID, SizeID ColorID |  |
| 4. | What is the number of times each product(fly) was sold? We want to see also those flies that have never been sold so that we can discontinue them. | SELECT F.FlyID, Count(F.FlyID)  FROM T\_Fly F, T\_SalesOrderLine SOL  WHERE F.FlyID=SOL.FlyID  AND (SELECT F.FlyID  FROM T\_Fly  AS Fly, T\_SalesOrderLine AS SOL  WHERE SOL.FlyID = Fly.FlyID  AND SOL.FlyID = NULL) |  |
| 5. | Total sales (in dollars) for each channel per month | SELECT (SOL.SOLQuantity\*Fly.FlySalesPrice) as Sales, SO.SOSCID as [Sales Channel], Month(SO.SODate) as Month  FROM T\_SalesOrderLine as SOL, T\_Fly as Fly, T\_SalesOrder as SO  WHERE Fly.FlyID = SOL.FlyID  AND SOL.SOID = SO.SOID  GROUP BY (SOL.SOLQuantity\*Fly.FlySalesPrice), SO.SOSCID, Month(SO.SOdate); |  |
| 6. | 10% of the products that have the highest margin. | SELECT TOP 10 PERCENT WITH TIES Fly.FlyID AS ProductID, (Fly.FlySalesPrice-Fly.FlyCost) as Margin  FROM T\_Fly as Fly  ORDER BY [Margin] desc; |  |
| 7. | What are the ten most popular (units sold) DIY flying materials? | SELECT TOP 10 Fly.FlyID as [DIY Material], COUNT(SOL.FlyID) as [Times Sold]  FROM T\_Fly as Fly, T\_ProductBundle as PB, T\_SalesOrderLine as SOL  WHERE Fly.FlyID = PB.FlyPartID  AND PB.FlyPartID = SOL.FlyID  GROUP BY Fly.FlyID; |  |
| 8 | The number of distinct products managed by each vendor manager. | SELECT E.EmpID, E.EmpFirstName, E.EmpLastName, COUNT(F.FlyID) Products\_Managed\_By\_Manager  FROM T\_Employee E JOIN T\_Vendor V ON E.EmpID=V[EmpID-VndrMGR] JOIN T\_Fly F ON V.VendID=F.VendID  GROUP BY E.EmpID, E.EmpFirstName, E.EmpLastName |  |
| 9 | The upcoming, scheduled guided trips (i.e., the guided trips that have already been sold) for each guide, including the guide’s name, the trip destination, the customer name, and the number in the customer’s party. | SELECT SO.SOLID, T.TripID, E.EmpGuide, E.EmpFirstName, E.EmpLastName, T.NumPeople, T.TripDestination  FROM T\_Employee E JOIN T\_Trip T ON E.EmpID=T.[EmpID-Guide] JOIN T\_SalesOrderLine SO ON T.TripID=SO.TripID  WHERE EmpGuide=’YES’ |  |
| 10 | Number of trips and the number of customers taken on fishing trips by each guide in the past 6 months. | SELECT COUNT(SOL.TripID) Num\_Trips, COUNT(C.CustID) Num\_Customers, T.[EmpID-Guide], T.TripStartDate  FROM T\_Customer C JOIN T\_SalesOrder SO ON C.CustID=SO.SOCustID JOIN T\_SalesOrderLine SOL ON SO.SOID=SOL.SOID JOIN T\_Trip T ON SOL.TripID=T.TripID  WHERE TripStartDate between ‘2020-01-01’ and ‘2020-06-01’  GROUP BY T.[EmpID-Guide], T.TripStartDate |  |
| 11 | Names and email addresses of all customers who made purchases in a given month. We need to be able to enter the month. | CREATE Procedure Query11  @month int  As  SELECT C.CustFirstName, C.CustLastName, C.CustEmail, SO.SODate  FROM dbo.T\_Customer AS C INNER JOIN dbo.T\_SalesOrder AS SO ON C.CustID = SO.SOCustID  WHERE (MONTH(SO.SODate) = @month);  Go  Declare @month int  Set @month = '12'  exec Query11 @month |  |
| 12. | Number of times used and dollars spent on each shipping vendor and shipping type by vendor. | Select S.Shipping\_Type, S.Shipping\_Rate, COUNT(DO.[DelIEmpID-Ship]) as TimesUsed, S.Shipping\_Rate \* COUNT(DO.[DelIEmpID-Ship]) as DollarSpent  From T\_SalesOrder SO Join T\_DeliveryOut DO ON SO.SODellOutID = DO.DelOutID  Join T\_Shipping S ON DO.DelOutShippingID = S.SID  Group By S.Shipping\_Type, S.Shipping\_Rate |  |
| 13. | Invoice lines for a given sales invoice number and given customer name  *(Refer to our last assumption with reference to how we track payment info/costs)*  . | SELECT C.CustID, C.CustFirstName, C.CustLastName, SO.SOID, SOL.FlyID, SOL.SOLQuantity, FlyCost  FROM T\_Customer C Join T\_SalesOrder SO ON C.CustID = SO.SOCustID Join T\_SalesOrderLine SOL ON SO.SOID = SOL.SOID Join T\_Fly F ON SOL.FlyID = F.FlyID  WHERE C.CustFirstName = 'Iris'  AND C.CustLastName = 'Brown'  AND SO.SOID = ‘806’  AND C.CustID = ‘630’ |  |
| 14. | A number of times a discount was applied to a sales order. List all the information about the discount, the total amount saved by customers that used the discount. | SELECT D.DisID, D.DisName, D.DisDescrip, (SUM(F.FlySalesPrice \* SOL.SOLQuantity)\* D.DisRate) as AmtSaved  FROM T\_Discount D Join T\_SalesOrder SO ON D.DisID = SO.SODisID  JOIN T\_SalesOrderLine SOL ON SO.SOID = SOL.SOID  JOIN T\_Fly F ON SOL.FlyID = F.FlyID  GROUP BY D.DisID, D.DisName, D.DisDescrip, D.DisRate |  |

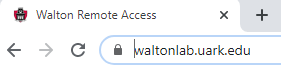
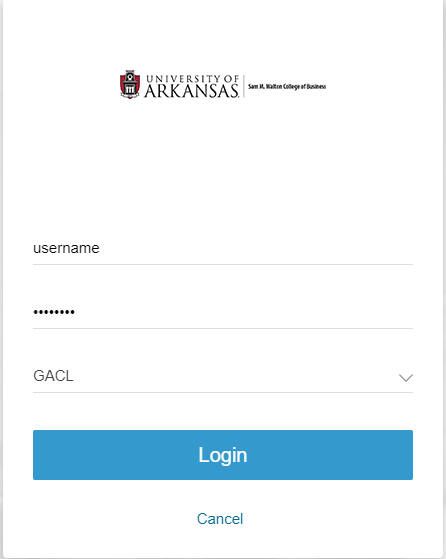
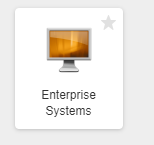
## **Three Additional Queries**

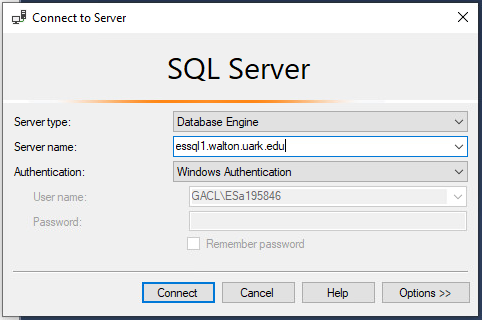
In the section below Big Data has created three additional queries that we felt would be useful for Elysian Fly Company. These queries were made by evaluating the needs of Elysian Fly Company and seeing what best fit them. These three unique queries should help Elysian Fly Company evaluate how they can be more effective and efficient. Below, the question, why it’s important, the SQL code, the output, and a recap of the findings are displayed.

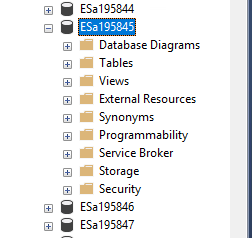
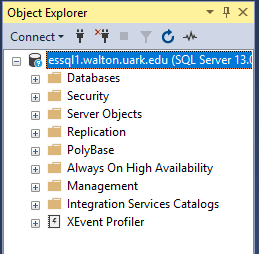
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Query # | Question | Why is this important | SQL | Partial Output | Recap of Findings |
| 1 | Which employees are most involved with processing SalesOrder?  This query captures information from when the employee first began work at Elysian. | We need to know this so we can see which employees are the most active within the organization when it comes  to getting SalesOrders through our system. | SELECT SO.SOEmpID AS [Employee ID], Emp.EmpFirstName AS [First Name], COUNT(SO.SOID) AS [Sales Orders Processed]  FROM T\_Employee Emp, T\_SalesOrder as SO  WHERE Emp.EmpID = SO.SOEmpID  GROUP BY SO.SOEmpID, Emp.EmpFirstName  ORDER BY [Sales Orders Processed] desc; |  | We found that most of the employees seem to process a fairly equivalent amount of Sales Orders. |
| 2 | Who is a local customer that we can incentivize to go on more guided trips? | This exquisite code exists to inform local patrons of upcoming guided trips as well as frequent shopper discounts. | Select CustID, CustMailStreet, CustMailCity, CustMailState, CustMailZip, CustEmail, CustFirstName, CustLastName  From T\_Customer  Where CustTypeID=102; |  | There were a significant number of local clients. |
| 3 | What is the average profit made from the sale of Flies each year by customer state and which states do we need to target more? | This output is important because it can show Elysian what states are bringing in the most profit, and which ones they need to reach more. | Select AVG(FlySalesPrice-FlyCost) as "Average Profit", Year(SODate) as "Year", CustBillState  From T\_Customer C JOIN T\_SalesOrder SO ON C.CustID=SO.SOCustID JOIN T\_SalesOrderLine SOL ON SO.SOID=SOL.SOID JOIN T\_Fly F ON SOL.FlyID=F.FlyID  Group By Year(SODate), CustBillState  Order By "Average Profit" asc |  | The top 5 lowest profiting states were shown, which can be alleviated by active marketing. |

# **User Documentation**

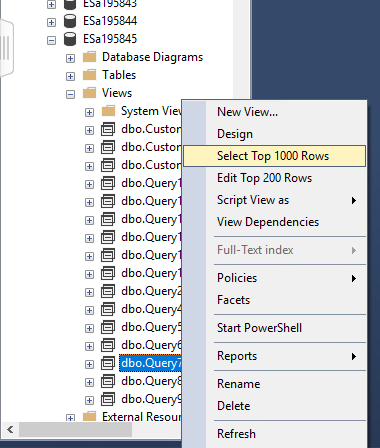
In this section, we will detail to our end-user how to access their new database, and secondly, the step-by-step instructions to view relevant queries/views to find information the client has requested. The instructions will encompass the entire process, from loading up the Virtual Machine all the way through to viewing the results of queries created within the database.

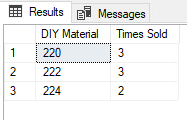
First, you need to open a browser and go to the website “ **https://waltonlab.uark.edu** “. When the website opens, select the “**Access Virtual Desktop With Web Browser**” option, this is located on the right side of the screen. Then click **“accept”**, this will prompt you to enter your credentials. You need to make sure that the domain is **“GACL”**.  Click **“Login.”** Once you are logged in, click on the “**Enterprise Systems**” icon. This will take you to a virtual desktop. From here, go to the **Windows Start** button in the lower left corner of the screen and click on it. Scroll to find the **search tool** and type in “**SSMS**”, the program titled **“Microsoft SQL Server Management Studio 18”** will pop up, click open.

Once that opens, a pop up will appear prompting you to enter a server name, enter server name “**essql1.walton.uark.edu**” and click connect.

Once connected, you are on the correct server. On the left side of the screen, there is a pop up with several different folders, expand the **Databases** folder. Next, scroll down on the list of databases and find **ESa195845**, once found click the plus button to expand. 

From here you can look at the implemented physical design under the **Database Diagrams** folder, each table under the **Tables** folder, and all queries under the **Views** folder. In order to access a View, right-click on the View corresponding to the query you wish to execute. Click the “**Select Top 1000 Rows”** option to display the results in a column/row display form.





# **What We Learned Throughout This Process**

Throughout this process Big Data has successfully overcome several obstacles together. As a team, Big Data learned how to create solutions together, manage our time, and ensure that all of the clients needs were met. While working on this project, Big Data became engrossed in ensuring that every detail was perfect so that we could output our best work, however that sometimes developed into overthinking tasks that otherwise would have been simple.

As a team we also learned to take a step back and make sure that we were solving problems in the most efficient way while utilizing each team member’s strengths and working as a team. Written below are a few lessons learned throughout this experience from each team member, as we would like to reflect on our accomplishments and what we learned from them.

|  |  |
| --- | --- |
| **Member Name:** | **What you learned:** |
| Olivia Long | While working on this extensive project, I have learned several things about database creation and working with a team. At first I was not completely solid on ERD work, but as we got into creating our own my confidence increased. A large part of this can be attributed to the help from my team, we all worked together and helped explain different things to each other, increasing all of our knowledge together. I think for me that was the biggest take away of the project, when working as a team we built on each member's strengths and helped each other learn, making all of us much more knowledgeable by the end of the project. A personal accomplishment I was proud of was being able to execute my queries after going through much trial and error. I was also proud that I was able to create my own custom query that worked and would be helpful to the client. Another personal accomplishment I had was being able to explain what we were doing in words. I tend to struggle with explaining things, but as I wrote about denormalization, referential integrity, and physical design, I became even more confident in my knowledge. Overall this project has significantly increased my knowledge in databases and has taught me how to work effectively with a team while being successful. |
| Marshall Mann | This project really felt like a capstone to this course. It was a practical application of what we learned in class. Reflecting on our journey at the end of it, I think about how impossible it would seem that we did all of this. If I had understood the magnitude of this project at the beginning I wouldn’t have believed it. I liked how we got to apply everything we learned in class. I learned how to create an ERD and that sometimes I have a tendency to go too far into detail. I learned that large projects can be broken up into small chunks. This project would have seemed overwhelming without periodic milestones. The milestones really helped to make it manageable. I learned that communication and setting micro deadlines is important when working as a team. Meeting briefly and assigning parts helps keep the project moving along. I enjoyed using Microsoft SQL Server and actually implementing our database. Being able to run queries was very satisfying. I found generating all of the data for the tables to be interesting because sometimes I could use the generated data website but for things like lures I researched actual popular lures to add to our product table. Generating the data was challenging but rewarding because you really had to use your head. |
| Evan Langenberg | Finally coming to the end of this project, I am incredibly proud with how we worked as a team to complete all the tasks that were placed before us. Even at individual levels of competency that varied depending on the unit we were in, we all managed to leverage our strengths and be incredibly vulnerable to improve and move forward. We encountered lots of roadblocks in Milestone I, overthinking and overanalyzing our requirements and making our workload way too intense. However, after refining our processes and having informal group roles being established, our team evolved into a productivity powerhouse. We all are incredibly thankful that we were paired together, as none of us have ever had a group in our careers as students that has been this successful and easy-going.  For individual achievements, I believe I grew exponentially as not only a leader as I became more confident in my understanding and mastery of the course material, but as a future member of the workforce. This project was as close a simulation to real-life, remote-working scenarios that are more common now post-COVID than any other era in history. Moreover, I improved at my ability to break down large goals into attainable, measurable steps, and I also improved my ability to delegate. This project had many moving parts, with instructions and data being referenced from multiple documents from multiple sources, and one of the keys to our success was properly sifting through all the information in front of us to derive the tasks at hand. Finally, this project was a massive boost to my interpersonal skills. From interacting with the professor regularly to receive guidance and direction, and working with my colleagues in Big Data, my ability to liaise and collaborate with others to reach a common goal improved dramatically during the course of this project. |
| Karen Li | I have learned a lot while working on this project this semester. To me, this class was more than a class where you would only sit down and take notes. This class and project has provided me experience to simulate what real life would look like if I was working under a company. While working on the project, I have definitely improved my communication and team working skills. Without a team, this project wouldn’t have been as successful as it is right now and it would have been more stressful if I was working on it alone. Thus, I am very grateful for our team and each of the members. In milestone 1, we faced many challenges when we were building our ERD. There were instances where we were over-complicating things when it could’ve been easily fixed by adding an attribute or creating a reference table. A good lesson I took away from that is that it is okay to take a step back in our thought process and ask for help. Another skill I have improved on is time management. This project would have been impossible to complete if it was done in one sitting. I have learned that setting weekly deadlines and distributing equal tasks among our team members has helped our project move forward and feel a lot less intimidating. The project has also helped me understand the course material better while applying what we learned in the classroom. One last skill I have gained while working on this project is using the SQL Server Management Studio. Prior to this class, I had zero experience with this software. I liked how easy it was to create and assemble your database and implement it. |
| Quannah Lindsey | This process has been an eye opening experience and I have learned so much. During milestone 1 I learned that it is important to keep the ERD as easy and streamlined as possible and it will make things easier in the long run. When the ERD was constructed it was neat to see how all of the tables fit together and were able to be linked with their primary and foreign keys as well as the reference tables. During milestone 2 normalization I struggled because the concept of normalization can seem difficult and it really is until you have actually worked through the process. I enjoyed getting to see everything broken down and written out to see how everything was again connected. I learned during the normalization process that it is important to stay organized so that it is easier to write everything down. Milestone 3 has been the most enjoyable for me because you actually get to see all of the data come together and get to see how everything is working together and linked physically. I learned how to troubleshoot our queries until I was able to get the results that the company was asking for and although it was stressful when they were not working it was exciting when I figured it out and was able to move onto the next query. The process overall has taught me life skills in the field that I wouldn't have gotten from other classes. This process has taught me patience and a different way of thinking about things and how data tables work together to create a whole process. This team has been instrumental in my success in this process and I can honestly say I wouldn’t have it any other way. As a team we worked very well together and we all had a helpful attitude when someone was struggling or needed help. We struggled through some of the processes, but we struggled as a team and my only negative is that I cannot keep this team throughout the rest of my college career. |

# **Appendix**

## **Team Contract**

Team Name: **Big Data**

Team Motto: **“Don’t compromise on excellence.”**

**Team Members**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Name | Email | Phone | Strengths | Availability to Meet |
| Marshall Mann | marshall.w.mann-1@ou.edu | 405-420-6955 | Organization, focus, [I am often away from my phone and email when studying. I will not be able to respond immediately and thus should be contacted for group meetings at least a day in advance.] | Monday 1:40 - 4:30 pm, Tuesday 2:45 - 4:30 pm, Wednesday 1:30-3:45, other times by appointment. |
| Quannah Lindsey | quannah.j.lindsey-1@ou.edu | 580-399-8581 | Detail oriented, time management | Most evenings, weekends (I work during the day when i'm not in class) |
| Evan Langenberg | evan.m.langenberg-1@ou.edu | 281-799-2815 | Time management, being able to empathize and listen, and the desire to learn | Pretty much any day after 5, and the weekends |
| Karen Li | karen.l.li@ou.edu | 918-706-3705 | Organization, Patience, Responsibility, Discipline | Mon 2-6pm, Tues 1-5pm, Wed 2-6pm, Thurs 1-5pm  Fri 2-5pm |
| Olivia Long | olivialong4@ou.edu | 972-567-6163 | Responsibility, work well with others, staying on task | Everyday after 5 except for Thursdays and most weekends |

**Team Expectations for the confidential peer evaluation :**

Clear communication of roles within the team and responsibility for your section, meeting deadlines, doing good quality work, and being on top of communications through Teams or otherwise.

**The behavior for which points will be deducted on the confidential peer evaluation:**

Not finishing your assigned section, bad/absent communication, submitting poor quality work, being disrespectful or being rude to fellow team members.

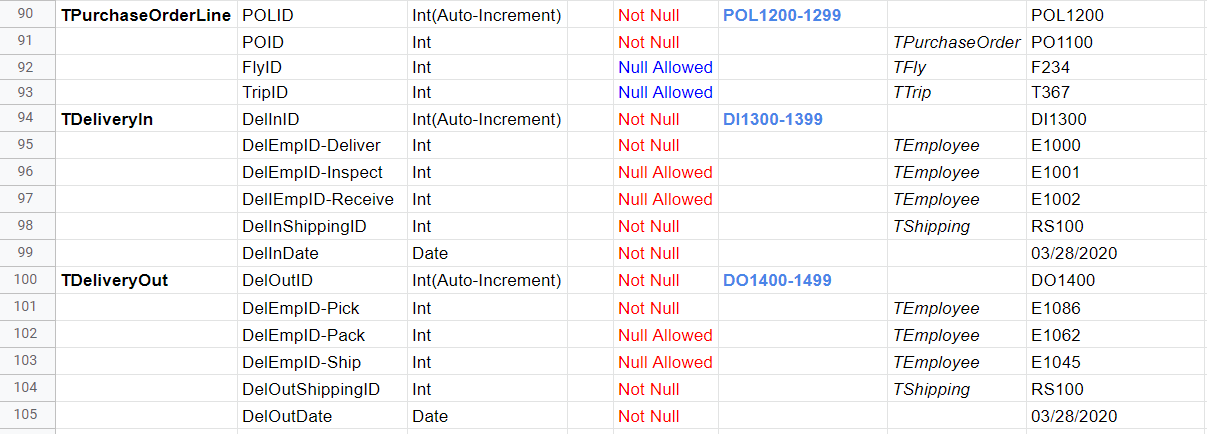
**Presentation Date Preferences (Rank Order Available Dates; make sure you list dates that absolutely don’t work for your team):**

October 23, November 9, November 23

## **Data Dictionary Model**







## **Project Management**

The project management tool helped our team ensure that everyone was doing their part and how much we should charge to Elysian Fly Company for our work. We learned that as we got deeply invested in this project, we were spending large amounts of time working on it, but that we needed to only add minutes spent actually creating and solving solutions. This helped us better manage our time in meetings, and kept our price down so that we would be a more appealing choice to Elysian Fly Company.

