**Flood Insurance Claims Database Design & Implementation**

Haider Abbas, Humdan Ahmed, Mitchell Kalbach,

Teaven Tedros, Zack Richards, Zach Wild

(Team 2)

# 

# 

# 

# 

# 

# 

# Table of Contents

[**Table of Contents**](#_uhgmoe2tyhrm) **1**

[**Introduction**](#_gan6s8heogei) **2**

[**Database Description**](#_nvfuz1vnfe2a) **3**

[Logical Design (include a PNG of your ERD in this section)](#_7e3zyci4jtkt) 3

[Physical Database](#_gpxhmxy43nxs) 4

[Sample Data](#_c8f3irx0dl1r) 4

[Views / Queries](#_vn0bvbrvh9al) 5

[**Changes from original design**](#_eyh0xt72m4fr) **5**

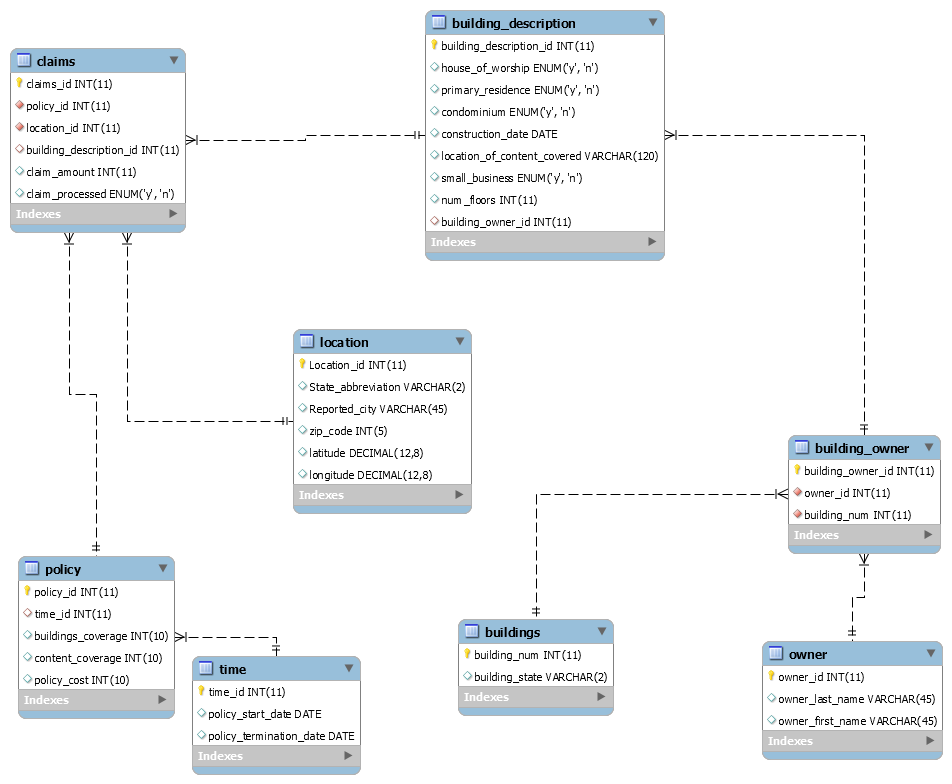
[**Lessons Learned**](#_iagxd6qzpe58) **5**

[**Potential Future Work**](#_j9lh8tvx98ov) **6**

# Introduction

In areas with unpredictable weather, information regarding severe flooding and the insurance policies that relate to those floods are important. Having appropriate insurance policies is a major concern for people living in high flood risk areas. Because of this, we decided to create a database which offers insight into how those floods impact the insurance policies and costs. In addition to flood data, building information is also included in our dataset, which can be particularly helpful when analyzing what types of buildings are damaged and how much they are damaged. Flood damage information concerning buildings is necessary when determining insurance costs. The target audience for who would primarily be using our database would be insurance companies as well as the government. The insurance companies would use this database to determine what areas of the country should have higher premiums for flood insurance. The government can use the database to see what areas of the country are higher risk for flooding and approximately how much money from FEMA should be allocated in the case of a flood. The average homeowner can also use the database to determine if they are at risk of flooding and whether they should purchase insurance. Our plan with this project was to mitigate some of these concerns and to offer information to insurance companies and government organizations that could improve the current situation and allow them to make more effective choices when it comes to making business decisions in this area.

# Database Description

Logical Design

Since our database is meant to keep track of flood insurance claims, the central entity in our database is the claim. Each claim has an associated policy, building, and location. The policy has its own self-contained details, the buildings have a number of dimensions they can vary along (as well as owners), and the location is meant to represent where the policy is located geographically.

First off the buildings as mentioned vary in a number of ways that are important in assessing a claim. Some attributes are the number of floors in a building, when it was built, or whether it houses a business. These are encapsulated in the *building\_description* table which is meant to store the descriptions of the buildings. This table is linked to the *building\_owner*table. This table is meant as a linking table for associating owners with their buildings and enabling a one to many relationship between owners, buildings, and building descriptions. Each owner has a unique entry in the *owner* table. The owner is represented by a first name, last name, and owner id. Each unique building has an entry in the *buildings*table.

Another part of the claim is the location. This is represented with a couple of attributes; latitude, longitude, state, city, and zip code. This is a relatively simple piece of information but has enough information to be kept in an entity independent of the claim.

Lastly the policy attached to a claim is made of two entities; the policy and a time entity. The time entity is meant to keep track of when the policy starts and expires. The policy, which is linked to claims, has information associated with it such as the cost of the policy, the coverage on the building, and the buildings contents’ coverage.

## Physical Database

Right now and during the development process each of us kept a copy of the database on their personal computer. We would share sql files which contained the table definitions, sample data, and views. This worked well for the scope of the project. A better solution would be to have a shared database on a server where we could collaborate. We tried this but had too much difficulty getting it working.

## Sample Data

Our sample data was based off a dataset we retrieved from FEMA’s national flood insurance program’s dataset. The problem with this data is that the anonymization removed any information that would identify who owned which properties. This meant that we couldn’t associate a policyholder with multiple policies, buildings, or claims. Since this has a lot more utility, and is what would actually exist inside of an insurance company, we created our own dataset based off of what information was contained in the FEMA datasets. In an ideal world this is what we would be able to find for all of the actual insurance claims. This limited the amount of data that we could generate for our database but increased the practicality and utility of it.

## Views / Queries

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| View Name | Req A | Req B | Req C | Req D | Req E |
| Claims\_by\_ location | X | X | X |  |  |
| policy\_data\_after\_july09 | X | X |  |  |  |
| Cost\_per\_building | X |  |  | X |  |
| Primary\_residence |  | X |  |  |  |
| bruce\_wayne\_claims | X | X | X | X | X |

# Changes from original design

In terms of our scope and resources, we definitely had to scale down on both. There was too much data to make sense of and for that reason we had to clean out much of the data in our csv file. Because of this, we were not able to get answers to the questions that we wanted to apply to this particular topic and dataset. This is where we think that we would make the cutoff for the deliverable we are working on and place any other desirable functionality in the following sprint if we were developing a program or application.

# Lessons Learned

We learned that planning and time management with this project were imperative. Even though some parts of the project seemed easy enough to complete at a later point, we made efforts to make progress in very small increments consistently every week. We think that the same results could not have been achieved without this system. In order to do this we had to plan how each deliverable would be completed and who would be doing what. This required reassessing while working on deliverables and figuring out which parts had to be completed in order to complete the others. We maintained constant communication so that everyone was on the same page and knew what they had to do for every assignment.

We also learned that we had to be ready to use any resources at any moment because issues will arise where the desired result is not achieved. We made sure that the rubrics were available, notes and lectures were open, and that we all had access to the same resources while we worked.

As we were cleaning the data before importing into workbench, we were deciding on whether to filter the data in excel based on columns that we could remove blanks from and still have enough rows. Another way we thought of filtering was by first deleting the columns we were not going to use and then filter from there. The reason for this is because we felt that the results could be biased in some manner based on the filters we were applying to the data. The resulting data we were left with seemed to have a very low occupancy count and this was the reason we considered filtering in different ways.

One particular thing we learned is that Agile concepts work very well. In the future we wouldn’t choose to have four other projects going on while we are working a sprint or on another project.

# Potential Future Work

Like we mentioned before, we learned a lot throughout the duration of this project. If we were to expand upon this project in the future, we would use what we learned to create an even better research project. We would use the same planning and time management techniques implemented in this project in our future project. We would also centralize our resources for easier use as we found that this increased the speed at which we were able to work at. Also, as mentioned before, we would incorporate Agile concepts as we found that they work very well. If we had an opportunity to do this again, we would improve in certain aspects. The first thing would be getting a better data set because we had some issues with the dataset we used. So, to that, we had to change up the database. Another thing we can do is work together with FEMA to get data specialized to our group and have them use our database to do more work with it. This will be good because we will have a clean dataset and that will allow us to create something that will help them with their company. For the future if we have a better dataset we will be able to create more valuable information for insurance companies.