Aryan Kumar
6.1800 DP Prep Assignment

Introduction
The country of Fictlandia wants to develop a new census system, Census 2.0. In the past, the three levels of government have conducted their census independently, meaning citizens may have received up to three separate census requests in a year (p. 3). The national government recognizes this may decrease census participation, so the new system will collect census information only once and make it available to all levels of government. The government hopes the new system will bring their census system into the modern era, while also providing additional capabilities.

This paper will summarize key elements of the Census 2.0 specification and highlight questions pertaining to its design. We will begin with a discussion of the structure and responsibility of the new census system, followed by an analysis of its use cases and design priorities, and finally conclude by posing additional questions about the system.

System Structure
The country of Fictlandia has three hierarchical layers of government – the municipal, state and national governments. Though census data will be available across all levels of government, it will be collected at the municipal level (p. 3). The municipality is responsible for collecting all the data needed by the various levels of government, and propagating that data throughout the system.

When the census opens, residents will have the option to submit their forms via mail or online to their municipality. Note that the residents do not submit their forms at a uniform time of day; there are periods of greater load on the online form submission system (p. 7-8). The municipality must then process these forms and input the data into the system, using the limited paper scanners and computing resources provided to them by the national government (p. 9-10).

The national government will run its own cloud service, with several data centers throughout the country. The new census system will need a scheme to store data locally for the municipalities, but also to distribute the data across the national government's data centers for their access (p. 8). Our census system will also not need to be concerned with the storage or handling of data within the state government's systems. Instead, the state government will act as a user of our census system, whom we will need to provision with data from municipal machines and/or the cloud service (p. 10).

The census system outlined above will have several key responsibilities. First, it must store census data across the local machines and cloud service for timely querying by the various levels

of government. Importantly, data on this system will need to be kept forever, so the system must be able to handle a growing dataset (p. 8). Secondly, the system will need a method to transmit data efficiently, securely, and completely across the various levels of government. To do so, it can employ encryption but this will come at the expense of three times larger data, so these concerns will need to be balanced (p. 11)

There are a few other reliability concerns our system will need to contend with as well. As mentioned previously, the system will experience greater load during certain periods. People are more likely to fill out the census forms during certain times of the day, and the local government may query the census system more often during election periods (p. 11). The census system should ideally be able to scale to meet the greater load. In addition, local machines can crash occasionally (p. 9), so the system should ideally be resilient to failures of individual components or be able to limit their effect.

Use Cases and Priorities
The census system of Fictlandia is intended to serve the people, and thus its impact on the people must be first and foremost assessed. The Census 2.0 system will be used for the following: the national government will use state population counts to assign legislatures; school boards will use it to ensure children are enrolled in schools; and all levels of the government will use it to project service demands (p. 11-12). It is thus vitally important to ensure the accuracy of the census data for these use cases, even over efficiency or reliability. An inability to access data for a short period of time will not be as much of an issue compared to inaccurate voter rolls, education enrollment, or mis-projected healthcare or education services stemming from erroneous data.

Our secondary priority is to ensure census data is secure. No organization or entity is supposed to be able to access any personally identifiable information for 70 years (p. 5). If any data breaches occur during data transmission or in our databases, citizens' right to privacy will be compromised and there may be other harmful repercussions to them. In addition, this could lower census participation rates in the future due to concerns over system security. Thus, it is crucial our system be as secure as possible, utilizing as much encryption in key aspects of its implementation.

Finally, the census system must be the most responsive to the municipal level. The municipalities will be directly providing citizens services and as such, may query the census system more often and in a time constrained manner (p. 8). This is particularly true for election periods. It is therefore very important we prioritize efficiency and access at a municipal level, over other levels of government when such trade-offs must be made.

Additional Questions

- Do we assume the probability a local machine fails is a uniform distribution? Or, are these resources more susceptible to failure during periods of higher load?
- Is there flexibility in scheduling the sysadmin's or administrative assistant's work schedule? If so, how much flexibility (i.e. can they work weekends or during the night)?
- How much will the census data be queried by the various levels of government, in quantitative terms? Will there be times when particular levels of government will query data more frequently?
- Can the national government's data centers experience failure in computers or networks, similar to how the municipality's resources can?