Introduction and Scope

AmerisourceBergen is a pharmaceutical distribution company based in Pennsylvania. In 2018, it ranked 12th on the Fortune 500 list, with annual revenue of $153 Billion. While handling the distribution for as much as 20% of the US pharmaceutical market creates many opportunities, it also generates a host of challenges, such as:

1) Drug manufacturing relies on complex chemical interactions that are heavily affected by the environment in which they are produced;
2) The opiod epidemic has brought increased scrutiny and legal action against distributors, and AmerisourceBergen in particular has recently come under heavy criticism.

The scope of this paper will be based on the assumption that AmerisourceBergen is interested in developing a data platform that can address both of these challenges. From a purely business standpoint, the system should be able to process orders at the clinic level, so that demand can be accurately predicted and supply chain efficiency can be increased. The company would also like this demand-level data to be passed to key manufacturers, so that the variable quality processes inherent in drug manufacturing can be implemented with expected demand in mind.

On a moral or public stewardship level, AmerisourceBergen would like a system that can aid in the prediction and flagging of over-prescribing activity at clinics around the country. The system should monitor prescription activity for odd behavior and unusual changes, much in the same way that the financial industry monitors transactions for likely fraud.

Finally, the company would like to ingest or scrape data from police departments and treatment organizations to both build a model linking prescription rates and abuse, as well as monitor the effectiveness of the company's anti-abuse initiatives.

Database Design

For our storage database, we will deploy MongoDB, an open source leader in NoSQL, document-oriented storage. MongoDB is scalable and well-suited to the high frequency data logging we will be handling as we process large volumes of disparate data from clinic-level systems that most likely will not conform to an "ideal" structure.

Additionally, as we build out our capabilities, it is likely that our schemas will be fluid as we expand scope. Document-oriented NoSQL storage is ideal for this, and MongoDB is proven to handle it well. As we will be analyzing these data for the above stated reasons, we will also rely on frequent aggregations, and views are better suited to this purpose than relational, SQL-based storage.

Batch vs Stream Processing Considerations

Clinic-level order processing will be handled via stream processing, as the sales, ordering, and inventory systems from which the data will be gathered generate transactions in near real-time.