aliz

Tomorrow's technology delivered today

Dagrofa

Goals

Dagrofa is Denmark's largest food wholesaler with 20 % market share. Dagrofa has 14000 employees and it is behind two successful department store chains, more than 500 grocery stores and 450 self-employed merchants.

To operate successfully on a large scale Dagrofa needs to make informed, data-driven decisions and for that constant market analysis. The company puts great value on big data analytics that allows them to adjust their strategies based on actionable insights. In the previous phases of the project a data warehouse was built for Dagrofa. The solution already supported analysis of historical and current state data and it allowed a wide variety of stream and batch jobs. But the existing system was also a basis for improvements to be even more aligned with Dagrofa's goals.

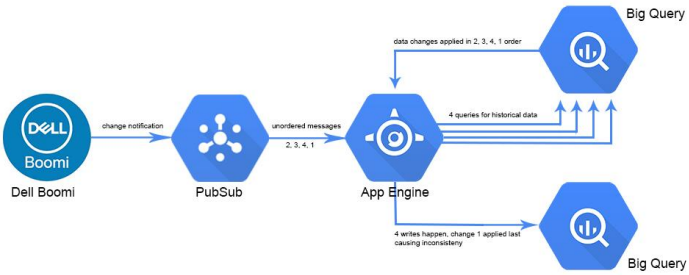
Challenge: Out of order message processing	Solution: efficient, consistent message processing with time windows
In the previous solution a **Dell Boomi data preparation** service pushed a large amount of messages to servlets. The data was processed, deserialized and validated on **Google App Engine.** The target Big Query table was queried to reevaluate history, then came postprocessing and finally the results were **written to BQ.** The solution had multiple weaknesses. • Large spikes: GAE handled badly when thousands of messages arrived within minutes • Out of order data: some changes came in quick succession • Concurrency: messages were lost when more arrived at the same time • High cost: every single message started BQ queries Dagrofa wanted to get consistent and correct data despite parallelism, while lowering BQ costs by skipping the unnecessary queries.	In response to the problems of the customer the Aliz team made a plan to replace the existing system with a new, better one within 6 months by the following steps. • Collect issues from the client caused by the insufficient previous solution, specify requirements and estimate cost reduction • Replace GAE processing with a **Dataflow** pipeline • Batch messages before processing, reorder them within a **5-minute sliding window** • Cut expenses: reduce query cost by querying once per window, and by reducing the duplications in the resulting data • Optimize post processing and BQ writes • Start to use new solution for 25 different type of entities

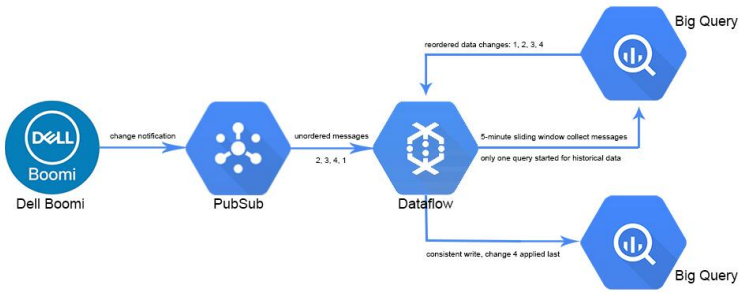
aliz

Tomorrow's technology delivered today

Dagrofa

Solution Architecture





Business value

With the newly implemented solution Dagrofa's data became more reliable. Right now if the relative data latency is under 6 hours, then unordered executions cause no inconsistencies. Dataflow also allows to avoid problems caused by splitting messages between multiple paths or parallel executions. The results remain consistent even when a message and a user modify the same data at the same time - previously this resulted in one of the changes being lost.

The increased correctness of the most important data types benefited the client's decision making by more accurate comparisons, trend analysis and as a stable, valid source for consumers of the data, it serves as the basis for automated analytical solutions.

One of the most valuable data types for Dagrofa describes their items in various stores and how they are handled. The store item type was in the highlight of attention during the entire development process. The item data often arrives in large spikes containing ten thousands of messages, while the changes most frequently affect only the item range column. In this case it was very important to apply the messages in the correct order and to keep the item history consistent because complex post processing follows the inserts that prioritizes the store items.

Another major factor was the significant cost reduction. By lowering the number of queries and duplicate data while also optimizing the processing the solution cut costs by 35%. Beside the obvious savings the maintenance of the system became easier. The unified process will make debugging faster, and it is also easier to introduce further changes and improvements to the system.