Data Warehouse Modernization use case

Customer: Large Government Agency

- **Solution Vision:** Construct a scalable and efficient data center that will allow gathering and analysis of nationwide water-related information for the general purpose of providing a sustainable water resource management and flood prevention system.
- **Mission:** Provide an environment to extend the capabilities and scalability of existing DWH solution including upgrading the infrastructure and services provided by the existing DWH and adding tiers to handle big data, management of content delivery and more.
- **Initial System State:** The system state at the beginning if the projects was a traditional Datawarehouse system based on a PostgreSQL RDBMS that uses ETL capabilities to load daily data.

  Reports that were provided looked and a very narrow geographical reach (due to amount of data) and in order to allow the analysis of cross-region data, a scalable infrastructure was required that to be capable of processing large amounts of structured data.

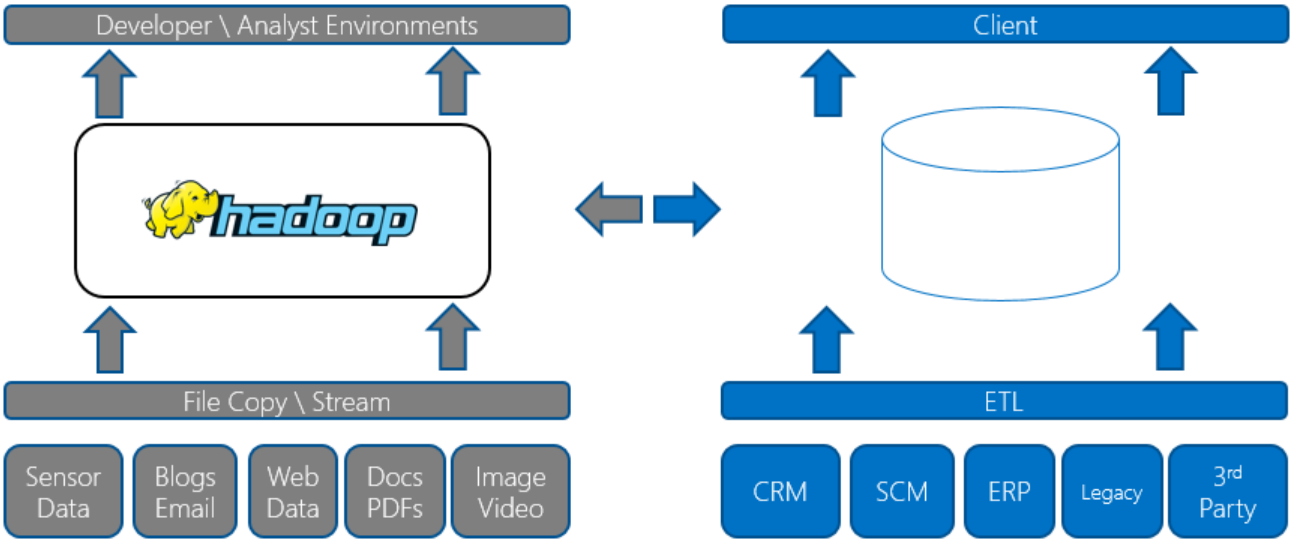
Use Case

In order to achieve the solution vision of allowing the enrichment of the existing DW with capabilities that will allow the ability to analyze huge amounts of data with the ability provide forecasts, recommendations and reports, we have constructed a data platform that relies on Lambda architecture and implements three main types of patterns:

- **Pre-Processing** - using big data capabilities as a “landing zone” before determining what data we moved to the data warehouse.
- **Offloading** - moving infrequently accessed data from data warehouses into enterprise-grade Hadoop.
- **Exploration** - using big data capabilities to explore and discover new high value data from massive amounts of raw data and free up the data warehouse for more structured, deep analytics.

The following drawing illustrates in High Level the approach that we have used when constructing the system:

© 2018 Matrix BI. All rights reserved.



By implementing the above, we have managed to construct a multi-layer data platform that is capable of providing traditional Datawarehouse capabilities aside Big Data and advanced Analytics capabilities and in addition, the ability to provide resilience of data (High Availability).

We have also achieved:

- Near real time analysis capability (1 Min) using Hbase & Storm.
- Self Service Analytics capabilities with Integrated Dashboards through the use of Datawarehouse and OLAP cubes
- Map Reduce as an ETL framework to join non-structured data and Analyze large amounts of geographical data with data from the Datawarehouse.

© 2018 Matrix BI. All rights reserved.