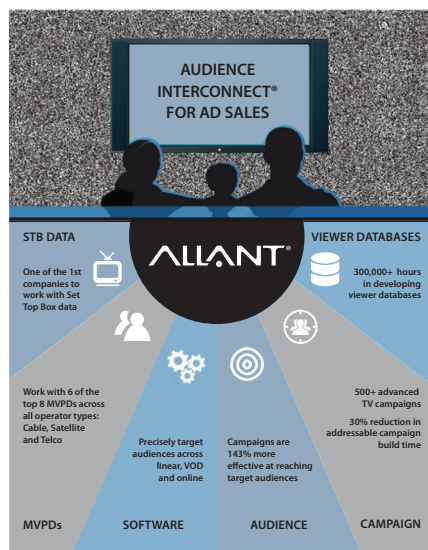


# Data-Driven Products at The Allant Group

## Background and Business Problem

Allant is a leader in advanced TV advertising software and data products as well as data and analytic driven marketing and advertising services. These products and services are designed to help ad buyers and sellers benefit from audience based marketing and advertising. Allant delivers to ad sellers, ad buyers, and marketers a distinctive understanding of their audiences through advanced data synthesis, insightful analytic products and services, and the industry's leading TV and online video solution, enabling them to optimize returns on their advertising investments across all media channels.



Allant seeks to provide new solutions to the market by expanding their offerings to include high value advertising and marketing products. They desired to build a differentiated, market-leading advertising solution to enable more valuable media buying and provide an expanded portfolio to their customers. To support this, they wanted to build an Integrated Analytic Environment to serve as a versatile and powerful foundation for data science, exploratory analysis, and reporting activities.

In other words, they needed a state-of-the-art, data-driven platform with extensible data integration, scalable data storage and processing capabilities, and rich analytics. Their new data platform had to support data ingest from various audience and activity acquired and public data sources—including de-identified ad impression and set-top box data as well as national demographic data from major providers. This data needed to be stored, transformed and made available for internal and external uses via a dynamic query access and data services layers.

Allant engaged SVDS to help build a new capability to enable a platform for audience keying and audience segmentation to support new, differentiated data products for Allant's growing SaaS business. We also built an integrated analytical environment to enable Allant's data science team to devise new insights and identify new analytical product offerings.

### *The Challenge:*

*Desired to expand data-driven product offerings*

*Required an integrated analytical stack to empower Allant's data science team*

*Needed high-performance, ad hoc analytics across extremely large database*

*Needed continuous integration across multiple data sources to yield high quality, granular, and comprehensive database*



## Solution

Silicon Valley Data Science and Allant jointly identified, organized, and prioritized the work efforts necessary for proof-of-concept execution for a highly flexible and scalable data platform utilizing open source tools that serves as the central store for the various data sources and analytics methods Allant wanted to use. We migrated Allant's Customer Data Integration (CDI)-Keying engine, built using Java Message Service (JMS) and Oracle, to an architecture centered on distributed data-management technologies with Cassandra as the data store. This was done by:

- using Spark to push the bulk of the functions originally executed at the "application layer" down to the distributed data layer;
- using Cassandra and Zookeeper to migrate in-process locking mechanisms to a more cost-effective, scalable system; and
- using Spark to create an high-throughput ETL pipeline for loading large datasets into Cassandra.

Allant's original customer recognition engine (employed to integrate customer profiles and their digital handles) was architected and designed to accommodate both batch and real-time inputs, with a common API to ensure unique creation of customer profiles. The common API is facilitated via a batch-to-single transaction interface (proprietary, custom software) that drops input records into JMS queues to perform the parsing, hygiene, matching, and keying sub-functions. This pipeline was re-implemented using Cassandra as a scalable, high-throughput data store, with Spark to achieve highly distributed function execution while maintaining fuzzy matching logic to preserve Allant's proprietary customer recognition rules. The resultant re-architected, re-platformed solution demonstrated a 20x throughput improvement while containing infrastructure costs.

SVDS also helped create tools and processes to help Allant's data science team perform robust analytics. We built an interactive analytics environment on top of Hadoop, Impala, Hive, and Spark using the open-source Jupyter notebook, and then enabled Allant's data scientists by helping them adopt tools like Python and Scikit-Learn. We helped Allant execute analytics at scale on the full dataset using the Hadoop cluster we built, instead of on data samples. We also helped them develop scikit-learn models in Spark or Hive. The result of this training was that Allant can now ingest big data into their analytic environment, extrapolate data on a national level, develop models and score a national audience universe and thus provide well-informed recommendations to their clients.

### *Our Approach:*

*Identified and prioritized data platform build activities to yield highest business value*

*Worked collaboratively with Allant's people as a blended engineering team*

*Built foundational components using agile development practices*

*Used proof-of-concept builds to test and uncover highest performing technologies, including open source options, proof-of-concept capabilities stack*

*Proceeded to execute full AMP and launched components into production*

### *New Capabilities:*

*Highly scalable audience segmentation platform*

*Data ingestion and processing 20x faster*

*New SaaS product that enables low-latency, ad hoc queries across entire datasets*

*Ability to build and deploy complex data science models at scale*

