

# Cable Television Analytics: Advertisement Viewership Analysis

## AT A GLANCE

### Challenges

- Difficulty in creating accurate analysis models using Hive Query Language due to Hive's lack of data visualization
- Hive query execution time was too long and consumed excessive resources on the Hadoop cluster

### Solution

- Xcalar provides an interactive modeling experience to create a dataflow that extracts the required insights
- Information from various sources and formats is easily integrated into the dataflow
- Xcalar's optimized batch dataflow execution provided results in less than half the time, using one third of the compute resources

### Value

- 3X increase in analyst productivity through interactive, visual modeling
- 8X performance increase, when processing data at scale
- Higher accuracy by exposing, assessing, and resolving data anomalies at each step while modeling
- Ability to quickly create new models to extract more value from the data
- 10X cost savings over Hadoop/Hive approach

Xcalar is working with a leading telecommunications company that provides cable TV and broadband services to tens of millions of customers. This company's key driver for new investment in analytics is to grow their advertising revenue stream by improving their understanding of viewer behavior. They continually seek a deeper understanding of ad viewership to better target their ads for impactful delivery, whether at specific times of the day, or interspersed with specific types of content programming, to maximize the value of ad impressions for their sponsors.

The company's set-top boxes collect detailed information about which channels are being watched at specific times, down to single-second precision. Additional data from Nielsen, broadcasters, and other sources, provide needed context:

- Channels that broadcast programs in each country or region
- Time intervals when ads are displayed

This supplemental data came in a variety of formats. The challenge was in verifying data accuracy, combining it, and outputting actionable results that could be used to make informed programming adjustments, such as inserting ads at more strategic times of the day, aligning ad time slots adjacent to programming content tailored for the intended audience, or even shuffling the order of ads within an ad break.

This customer sought to gain key insights into viewer behavior by observing when a user changes the channel or fast-forwards during an ad. By logically dividing each ad into four sequential, equal-sized time slices, and observing which slice was playing at the time a viewer changed the channel or fast-forwarded, they realized they could build a histogram measuring the viewers' disinterest in a particular ad. Furthermore, they realized that if this information could be collected in near real-time, the company could maximize viewership by adjusting advertisement schedules immediately.

Using Xcalar Design Enterprise Edition, analysts combined the data required to develop these insights using a spreadsheet-like interface and point-and-click operations. Data arrived in various formats, including CSV, JSON, or XML, and resided either in NFS shared storage or on an HDFS cluster. Xcalar accessed the data in-place in the original format from these sources and created metadata that allows Xcalar Design to render data in spreadsheet format. Analyzing this data required the building of a dataflow that included checking for unexpected values, mis-typed strings, or missing data, and developing techniques to clean up these anomalies. After the analytics data was cleaned and prepared, the analysts joined the set-top box data with the programming data to determine the intersection between channel changes and ads playing at that time. To identify the advertisements that were being

aired at the viewer's location, analysts then joined additional datasets into the dataflow and correlated local channel numbers or antenna IDs with national or international programming networks.

Dataflows in Xcalar Design can be operationalized—deployed at big data scale—with the click of a button. Customizing a resulting dataflow allowed it to be run against any date range of viewership and programming data. Now, the company could run the analysis across dozens of cable channels, tens of millions of users, and a week's worth of data in just a few minutes using a four-node cluster of virtual machines running Xcalar Data Platform. A comparable Hadoop/Hive solution took more than twice as long to execute, using a cluster four times the size. Our customer also discovered that they had derived additional benefits which directly decrease their time to insight. Using Xcalar Data Platform, once a high-level algorithm is identified, it only takes a day or two to develop a high-quality dataflow, including the data prep steps. In addition, while developing dataflows, they discovered anomalies in the data, which previously had taken weeks or months to identify using Hive Query Language; many had been missed entirely.

Xcalar Data Platform processed the data over 8X faster than Hadoop/Hive. Meanwhile, the process efficiencies gained by using one data platform, from click to insight to scale, increased the speed of their analytics projects by 3X, which translates to a faster analytics cycle. This shorter time-to-insight resulted from Xcalar Design's easy data visualization capabilities, clear recognition of data anomalies, and immediate data model operationalization without the need to re-architect for scale.

## About Xcalar

Xcalar is an open and extensible analytics platform for the complete analytics pipeline, including data quality, virtual data warehousing, data science, and operationalization. Users interactively build dataflows with visual programming, SQL, and structured programming, and execute those dataflows at petabyte scale on unstructured, structured, and semi-structured data. Xcalar's enterprise-grade software scales to hundreds of nodes and thousands of users for both cloud and on-premises deployments. Its patented technologies deliver actionable insights with simplicity, speed, and scale.

## KEY FEATURES, PRODUCTS, AND SERVICES

### Key Features

- Dataflows can be re-used to rapidly prototype and productionize algorithms
- Visual programming, SQL, and Python for development flexibility
- Machine learning algorithms in TensorFlow can be seamlessly integrated in dataflows at any stage in the pipeline.

### Products

- Xcalar Data Platform Enterprise Edition
- Xcalar Design Enterprise Edition

### Services

- Product training
- Solution architecture and design
- Infrastructure setup, configuration, and monitoring in AWS environment
- User-defined function data import/export
- Transformation design and implementation
- Dataflow design and implementation
- Cluster sizing and performance tuning