

# Cable Television Analytics: Advertisement Viewership Analysis

## CHALLENGES

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

## SOLUTION

- The interactive modeling experience provided by Xcalar Design greatly simplifies the process of creating a dataflow and extracting meaningful insights
- The dataflow integrates data from various sources in many different formats
- The optimized execution generates results in less than half the time, using one-third of the compute resources

## BENEFITS WITH XCALAR

- **3X increase** in analyst productivity through interactive, visual modeling
- **8X performance increase** over Hadoop/Hive, when processing data at scale
- Higher accuracy by exposing, assessing, and resolving data anomalies at every step
- Faster creation of models/algorithms to extract more value from the data
- **10X cost savings** over Hadoop/Hive approach

## KEY FEATURES

- Dataflows can be re-used to prototype and operationalize the algorithms rapidly
- Visual programming, SQL, and Python together provide tremendous development flexibility
- Machine learning algorithms from TensorFlow, SparkML, H2O.ai, etc. can be seamlessly integrated into the dataflows at any stage of the data pipeline

## Introduction

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

## Challenges

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

- The list of channels and their programs that are aired in each country and region
- The specific times when each ad is displayed

This supplemental data comes in a variety of formats. The challenge was in verifying the data accuracy, combining it, and outputting the 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 the programming content tailored for the intended audience, or even shuffling the order of ads within an ad break.

## Xcalar Solution

This customer sought to gain key insights into viewer behavior by observing when a viewer 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 is playing at the time a viewer changes the channel or fast-forwards, they realized they could build a histogram measuring the viewers' disinterest in each 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, 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 its original format and generated metadata that allowed Xcalar Design to render it in a spreadsheet format. Analyzing this data required the building of a dataflow that included checking for unexpected values, mistyped strings, or missing data, and developing techniques to clean

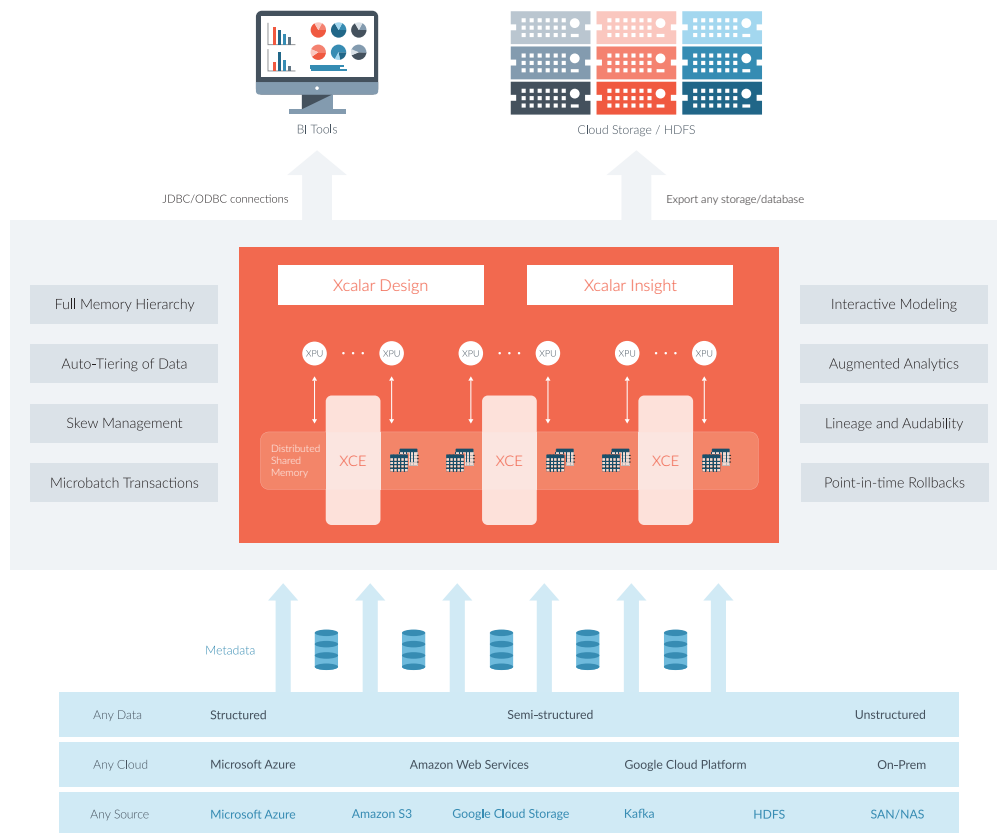
up these anomalies. After the analytics data was cleaned and prepared, analysts joined the set-top box data with the programming data to determine the intersection between channel changes and ads playing at that time. To identify the advertisements that were being aired at the viewers' locations, analysts 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. The ability to parameterize dataflows allows them to be run against any date range of viewership and programming data. Now, the company can run the analyses 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 - an 8x performance improvement.

Using Xcalar Design's powerful SQL, visual programming and structured programming interfaces, 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, working immersively with the entire raw source data. Anomalies in the data are quickly discovered and business rules easily applied for resolving integrity constraint violations. This process had previously taken weeks or months to identify and resolve using Hive Query Language; many anomalies had been missed entirely.

Overall, the process efficiency gained by using this single end-to-end data platform from click to insight to scale has increased the speed of their analytics projects by 3x.



#### About Xcalar

Xcalar is a scale-out platform for data processing applications and operationalizing ML. The platform is open and extensible, and suitable for developing and operationalizing business logic. Xcalar's use cases include virtual data warehousing to enable BI tools to query real-time transactionally consistent data, operationalizing ML algorithms at cloud-scale, as well as simplifying data transformation and quality processes. Users use a versatile IDE to interactively build dataflows using SQL, visual programming, and structured programming, and execute them at petabyte scale. Xcalar's enterprise-grade software scales linearly to hundreds of nodes and thousands of users for public/private cloud and hybrid deployments.