

# Case Study

Ad Reporting for an Interactive TV Gaming Channel

ANKHOR Intelligence GmbH

[www.ankhor.com](http://www.ankhor.com)

[info@ankhor.com](mailto:info@ankhor.com)

## Case Study: Ad Reporting for an Interactive TV Gaming Channel

### Abstract

ANKHOR FlowSheet was successfully used to build an ad reporting solution for a customer offering an interactive television gaming channel. The result is a fast and cost-effective implementation for analyzing and visualizing unstructured and semi-structured data in the form of log files comprised of more than 10 million lines of text per day, and connecting this with information retrieved from several other types of data sources (SQL, SOAP). ANKHOR FlowSheet's intuitive and visual approach provides the necessary flexibility required by the customer to adapt to steadily changing requirements. Our customer is finally able to get the complete picture of the channel's usage and provide meaningful reports to their advertisers.

### Challenge

Our customer is the provider of an interactive gaming channel that offers casual games to cable TV subscribers regardless of their set top box technology. The game's video and audio is rendered on servers in a data center, which receive the user's remote control input via a backchannel, and then send it over an IP network and cable to the client. The service displays video ads while users are browsing the game catalog, before and after gameplay sessions and between levels within a game session. Advertisements can also be interactive, in the form of questionnaires or data collection. Since users can create a profile, their demographics (player name, birth year, gender, zip code) are known and used to select the type of advertisements to present, allowing for precise targeting.

The game servers constantly register which games are played by each unique user during the course of a session, as well how long each game is played. In addition, each time an ad is displayed, the system remembers which ad was shown to that user and for how long. For interactive ads, the user's responses are recorded, too.

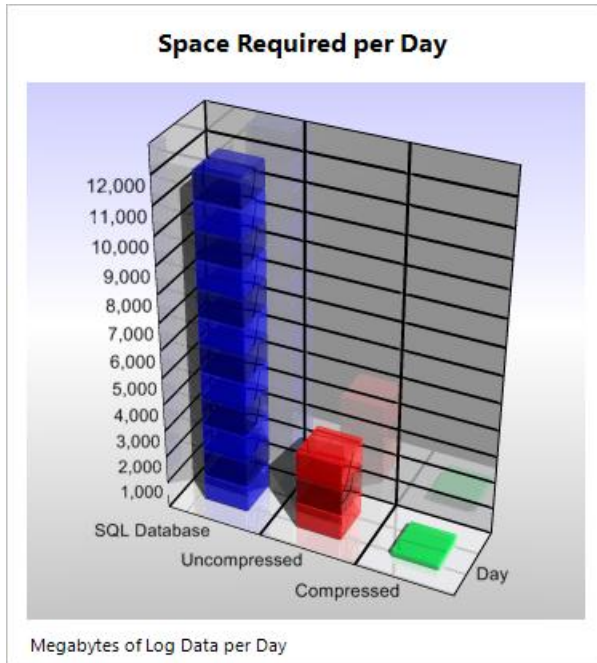
The above information is stored as lines of text in a set of semi-structured, varying format log files. Due to the game channel's popularity, these log files grow rapidly in size and number. The total of all log data to be processed amounts to about 240 Megabytes per 10.000 games played on each day.

The challenge consisted of the following customer requirements:

- Get detailed insight into the usage of the games channel: determine when and what games are played by each group of users (by age, gender, time of day, season, etc.) and for how long. Knowing which games users prefer greatly helps identify the type of content that will make the channel as attractive as possible.
- Analyze and visualize the ad impressions: determine which ads were shown to each unique user based on their demographics, when it was shown, and for how long.
- Produce meaningful reports for the advertisers to demonstrate that the desired target groups were reached.
- Provide documentation for the purpose of billing to the advertisers.
- Provide a means to export and further process the consolidated data extracted from the various data sources (log files and SQL databases).
- Keep the required space for storing the collected usage data as small as possible.
- Get a near real time picture of the channel by extracting, processing and analyzing the data as quickly as possible.
- Present the results in a visually compelling but understandable way, as a viewable dashboard using a standard web browser, which generates reports as printable PDF documents.

In addition, the following factors had to be taken into consideration:

- The data needs to be analyzed on a per-session basis. However, the session log entries are intermingled, and sessions can reach over the boundaries of individual log files.
- In addition to the log files, secondary data like player-to-household mappings or billing information need to be retrieved from other sources like SQL databases and custom SOAP APIs.
- For performance and cost reasons, the log file data should not be expanded and stored in a SQL database, as this would increase the data volume and read time for processing by 40 times compared to ZIP-compressing instead:



### The Solution

ANKHOR proposed the implementation of a completely integrated solution based entirely on ANKHOR FlowSheet’s data flow graph principle. This provided the following benefits:

- The same tool and execution platform is used for automatic condensation, analysis and presentation of the data. The original source data is condensed, removing irrelevant data by building, slicing and dicing in-memory OLAP data cubes. There are no manual steps in the process – the condensing is done automatically in the background at regular intervals. The resulting data visualization is rendered on demand, on a local machine or a remote instance of the ANKHOR HTTP Server, and is accessible by a standard web browser.
- Fast adaptation to new requirements from the customer, cable operators and advertisers thanks to the visual editing of data flow graphs and intuitive usage of the ANKHOR FlowSheet Professional development tool.
- Built-in FlowSheet operators provide direct access to all other required data sources, like SQL databases containing detailed information about households, users and advertisements, and third-party components accessible via the SOAP protocol.
- Direct processing on the compressed log files using built-in ZIP decompression operators. Due to the

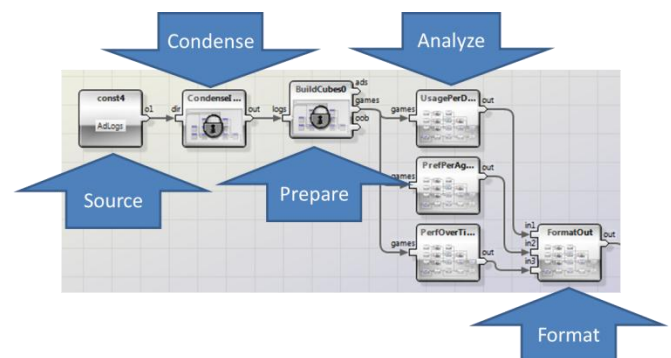
enormous reduction of the data size, archiving the complete historic data (e.g. for one or two years) comes at a very low cost compared to SQL databases.

- Fast processing time due to in-memory building and processing of OLAP data cubes and incremental update of the cubes with new data. The current data cubes are backed up in compressed cache files. If necessary, they can be rebuilt from the original compressed log data.
- Extremely efficient use of multi-core server hardware due to the inherently parallel execution of data flow graphs.

### Components

The core element of the solution is the Condense operator:

- Stores session data that overlaps the boundaries of individual log files and data checkpoints.
- Detects the arrival of new log files and creates new checkpoints.
- Condenses the source data into per-hour records.
- It can run independently on any system that has access to the compressed source data, providing results for download by the later processing stages. Alternatively, it can be integrated into an all-in-one FlowSheet running on the ANKHOR HTTP Server.



The condensing stage is followed by the Build Cube operator, which combines records for a single hour and builds the required OLAP data cube structures for further analysis into a common and consistent data view. This operator also manages the cube caching.

The analysis stage uses the data cubes provided by the previous stage. Several different kinds of analyses can be executed on the data in parallel or based on the user's selection of a specific FlowSheet in the ANKHOR dashboard. New types of analysis can be added easily thanks to ANKHOR's interactive visual editing solutions. They all share the same consistent view on the data provided by the Build Cube operator. It is even possible to download the cubes from a remote location and execute the analysis (and further stages) on a local machine. This is great for implementing personal, customized views on one consistent set of data.

In the final stage, the results of the analysis are formatted for output to the interactive dashboard, printing or PDF export. An extensive library of easily adaptable visualization methods is available to provide compelling results.

### A Quote from the Customer

"As an interactive television gaming channel with incredible usage and advertising response rates, we found that it was incredibly difficult to collect and analyze data for our TV channel. Other solutions allowed us to analyze pieces and parts but did not give us the whole picture. With Ankhor we were able to use a single tool (the ANKHOR Flowsheet) to collect the source web server log data, analyze the data and generate reports. We were also able to easily slice and dice our data to see it from different viewpoints, without programmer involvement. Our distribution and advertising partners consistently complemented us on our solid, detailed reports and were impressed with the level of data we were able to share with them. But the best part was that we were able to show the actual usage which led to more distribution deals. ANKHOR was invaluable in helping us grow our business."

Sangita Verma  
Founder and CEO  
TAG Networks, Inc.



### The Special ANKHOR Approach

The ANKHOR based solution takes a fresh new approach to achieve the goals set by the customer compared to "traditional" Business Intelligence solutions:

- **Adaptable to changing requirements:** ANKHOR's visual approach to creating solutions and its natural support of component reuse significantly simplifies and accelerates the development process, driving cost down.
- **Provisioning of data cubes covering specific aspects:** Instead of a data mart, the solution provides OLAP data cubes for download or export which offer a specific view on the original data (for example: specific data cubes with a focus on a time period, a specific application or a particular advertiser). These cubes can be used for further analysis using ANKHOR or other means.
- **Simplified structure for keeping data:** All historic data is archived in ZIP compressed log files, avoiding the expensive maintenance of a data warehouse. New customer requirements calling for a different view on the original data are no problem: The FlowSheet simply builds new OLAP data cubes, meeting the new requirements from the original set of data. No information is lost, as would be the case when condensing the original data before it is archived.
- **Fast incremental processing:** After processing the source data and building the OLAP data cubes for the first time, new data sets are added to the cubes in incremental steps, significantly reducing the processing time. 10 million new lines in the log files require less than 3 minutes to be processed, including the accesses to other data sources required to build the cubes.
- **Support for clustered processing:** By splitting the source data by deployment site, month or query type, the processing load is distributed over several servers running the ANKHOR execution engine. Since it is compressed, the source data can be fully replicated on all servers of the computing cluster.
- **Reuse existing channels for harvesting source data:** The existing infrastructure for collecting log files was reused. This avoided development cost and the need to qualify and deploy new versions of the game channel's server software.

## Screen Shots

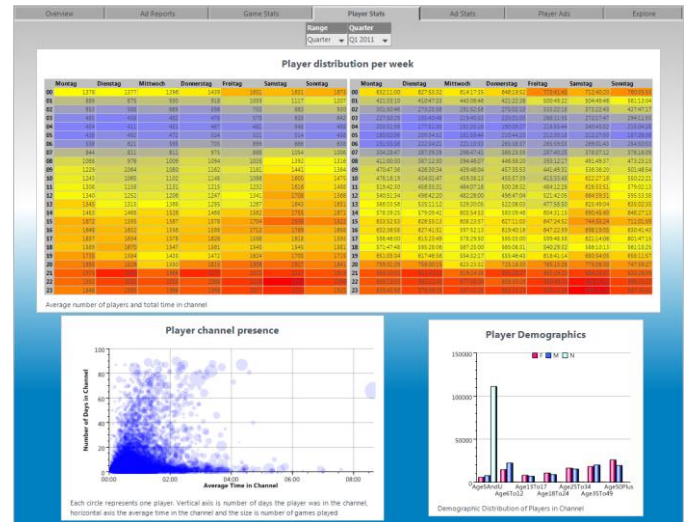
In the following we present a few Dashboard screen shots of the actual solution, based on fictitious data.

### Overview

On the first page, the Dashboard displays the number of players in the channel, the games played and ads that were viewed. The “last 24 hours” is a snapshot of the current channel usage. The information is updated once per hour.

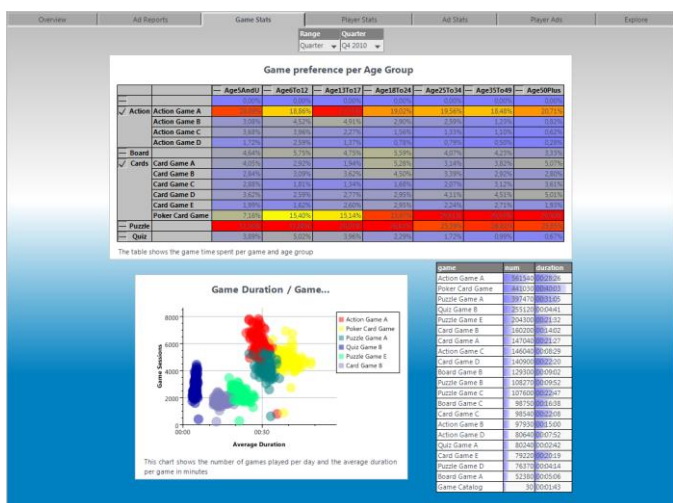


the size of the circle represents the number of games he or she played. The vertical axis corresponds to the number of days the player was in the channel, and the horizontal axis represents the average time spent in the channel. The chart on the right hand side shows the demographic distribution. The bar labeled "N" refers to users who did not disclose their gender.



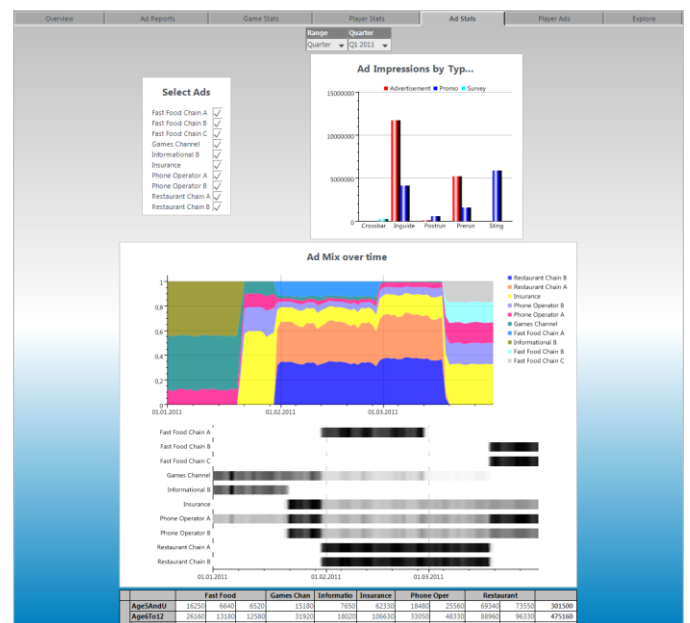
### Game Statistics

On this Dashboard tab, the user can extract detailed information on which games were most played during a given time period. A fold-out heat map displays what age group preferred what types of games. The games with the longest lasting appeal to players can be identified by setting the average game duration in relation to the number of game sessions.



### Ad Statistics

This tab provides detailed insight on specific ads presented to the players. The chart on the upper right displays the number of ad impressions based on the ad category (the “ad slot”). The “Ad Mix Over Time” chart is a visual representation of the percentage by which the ads booked by advertisers were displayed in relation to each other. The intensity chart below is a different representation for how often each ad was shown during the given period. In the table, detailed numbers for each ad can be inspected.



### Player Statistics

This view uses two heat maps to display which weekdays and hours received the highest average numbers of players and the highest total time spent in the channel. In the chart on the left side below, each circle represents one player, where