Annals of "Dunarea de Jos" University of Galati
Fascicle I. Economics and Applied Informatics
Years XXXI – nº2/2025
ISSN-L 1584-0409 ISSN-Online 2344-441X
www.eia.feaa.ugal.ro

DOI https://doi.org/10.35219/eai15840409514

Transforming the Metaverse into Data: Evaluating Consumer Behavior with ETL

Maria Cristina Enache[★]

ARTICLE INFO

Article history:
Received July 7, 2025
Accepted August 12, 2025
Available online September 2025
JEL Classification
D12, L86, D83

Keywords:
metaverse, artificial intelligence

ABSTRACT

The metaverse has transformed the way we interact with digital environments, and the analysis of user behavior in these 3D environments is becoming increasingly important for business, education, and other fields. While the metaverse offers sensational opportunities for immersion and interaction, understanding consumer behavior and evaluating the effectiveness of virtual experiences require sophisticated data analysis tools. Here, ETL (Extract, Transform, Load) plays an essential role.

1. Introduction

Over the past several years, the notion of the metaverse has transitioned from a fictional idea to the near future, altering how consumers socialize, work, and more importantly, shop. Virtual reality and augmented reality platforms create immersive environments in which brands can define distinctive and interactive experiences for consumers. The metaverse has found its applications in various domains, revolutionizing the way we engage with technology and each other. In video games, it offers an entirely new dimension of experience for players by allowing complete immersion. One of the most populated metaverses in the gaming industry is Roblox, which has evolved past gaming since its release in 2006, providing potential marketing and eCommerce opportunities for retailers. Therefore, the metaverse is the next frontier of online commerce, taking traditional e-commerce and rendering it in three-dimensional, interactive, and continuous environments. With this quantum leap comes a pressing problem: how do we understand and study consumer behavior in these complex virtual worlds? In the metaverse, unlike a standard website, consumers can navigate, physically (virtually) touch products as a real-world activity, attend events, and interact with other avatars. As a result, disorienting amounts of unstructured data is produced ranging from space coordinates to movements, all the way to object movements and even virtual transactions; anything without a particular processing method can be categorized as digital noise. This is where the Extract, Transform, Load (ETL) comes into play – ETL can capture data from multiple metaverse sources, organize it into helpful forms, then ETL will populate the data into your analysis system where you can work with that data and interpret data into meaningful numbers. The aim of this article is to explore the ways ETL is vital to the process of "from metaverse to metrics"; that is, turning consumer interaction in the 3D universe into actionable business insight. We will consider what forms of potential data can be ingested, how it is translated, what can be quantified, and how it can be leveraged to make marketing, design, and strategy decisions in immersive digital space.

2. Understanding the Metaverse

The metaverse refers to a group of connected, persistent, and shared virtual worlds where users communicate with each other and with virtual objects through avatars. Although the concept has been popularized by science fiction, the metaverse is taking tangible forms today with platforms such as Decentraland, Roblox, Meta Horizon Worlds, or Spatial. They facilitate the hosting of events, the construction of shops, the exploration of themed worlds, and live social interactions within an immersive 3D. The defining characteristic of the metaverse is immersion—no longer do experiences get passively received on a screen, but instead lived in a virtual world where users navigate, talk, "touch" objects, and make near-instinctual decisions. This kind of freedom of movement and interaction enables a dynamic setting, very different from traditional websites or mobile applications. For brands and marketers, the metaverse holds out the prospect of creating extremely immersive marketing experiences: virtual showrooms, interactive product launches, gamified campaigns, or partnerships with digital influencers.

★ Dunarea de Jos University of Galati Romania. E-mail addresses: mpodoleanu@ugal.ro (M. C. Enache).

Yet with this complexity comes also the difficulty of tracking and analyzing what precisely happens within these spaces: where do visitors stop, what products do they consider, how much time do they spend in a zone, what triggers an action to buy? The resolution to such questions demands an infrastructure that will collect, interpret, and structure behavioral data – that is to say, a well-defined ETL process. The metaverse is a collection of shared, persistent, interconnected virtual worlds, where individuals interact with each other and virtual objects through avatars. Named by science fiction, the metaverse is taking tangible forms today in the likes of platforms like Decentraland, Roblox, Meta Horizon Worlds, or Spatial. The platform provides for hosting events, creating stores, visiting themed worlds, and live social interactions in an immersive 3D environment. The key feature of the metaverse is immersion—no longer is experience passively received on a screen, but lived in a virtual world in which one moves about, speaks, "touches" objects, and makes near-intuitive choices. The active process of moving about and interacting spawns an active ecosystem, miles removed from static websites or mobile apps.

For brands and marketers, the metaverse promises the potential for extremely tailored marketing experiences: virtual shops, experience-based product introductions, game-based campaigns, or partnerships with virtual influencers. Yet so complex is it that it also presents the challenge of being able to monitor and measure what is actually happening within them: where one stops, what one reads, how long one stays in a location, what triggers a purchase action? The answers to these and other such questions require some infrastructure to collect, analyze and model behaviour data – i.e., a properly organized ETL process.

3. Consumer Behavior in 3D Environments

Consumer activity within three-dimensional spaces is fundamentally distinct from that experienced in traditional e-commerce. Instead of browsing web pages and menus, metaverse shoppers move through virtual spaces, actively explore, engage with objects and with other people, and purchase decisions are influenced by multisensory and social experiences. This shift requires a novel manner of understanding and interpretation of consumer behavior. In the world of 3D space, the customer is not just a "click" in a sales funnel, but an active participant in an ecosystem. He is able to:

- ✧ Shop in an online store and "test" products with avatars;
- ✧ Participate in real-time events (e.g., the start of the collection, concerts promoted by the brand);
- ✧ Play with objects through gestures (lifting, examining, placing in the cart);
- ✧ Collaborate with other users in public zones or receive real-time recommendations.

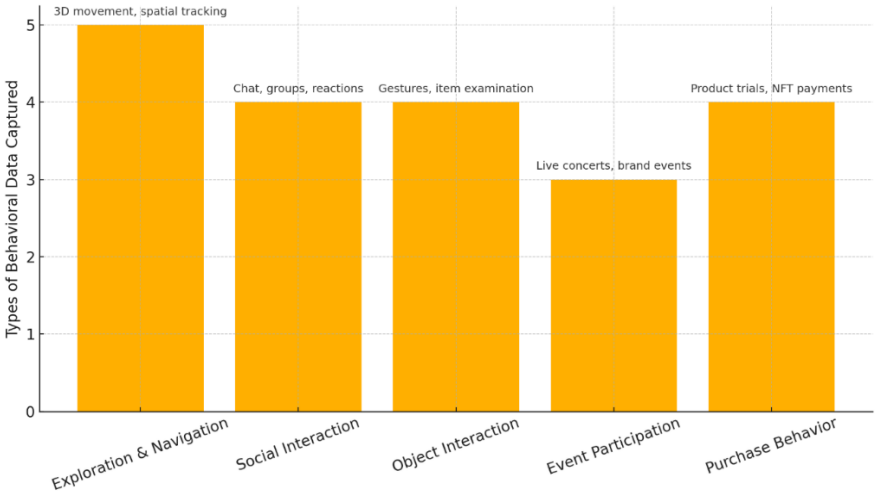

Figure 1. Consumer Behavior Dimensions in 3D Spaces

All of these activities generate vast amounts of behavioral data:

- ✧ The user's location and path in space (3D tracking);
- ✧ The time spent within a space or near an item;
- ✧ Interactions (clicks, gestures, voice, responses);
- ✧ Virtual or real-world purchases, including NFT payments;
- ✧ Social interaction (communication, groups, response to stimuli).

Relative to behavior on a conventional site, where behavior is discrete and trackable (e.g., clicking on an item, adding to cart), metaverse behavior is fluid and situation-dependent. Data collection and sharing hence become far more complex, requiring robust integration, aggregation, and analysis processes. Therefore, a correct understanding of metaverse consumer behavior includes not only the creation of data but also a methodical and scalable means of processing it—where the ETL process is a badly needed application.

4. The Role of ETL in Data Analysis

In the metaverse, where every user action can trigger one digital event or another information comes together quickly, into many forms and from many sources. The trick is to construct, clean, and tune these data sources into a systematic approach so that they can be useful for business action and decision-making. This is where the ETL steps come into play – Extract, Transform, Load, and is the foundation for every data analysis pipeline you use today.

Extract (Data Extraction) is the first step of the process where we collect data from various sources: metaverse platforms (Decentraland, Roblox, Horizon Worlds, etc.), VR hardware, log files, user databases, or APIs streaming real-time events. These sources can have heterogeneous data - e.g., 3D coordinates in cartographic form, timestamps that are in non-local time zones, or interaction data captured in nested JSON formats. An expert ETL has to tame this mess and extract a complete and consistent extraction.

Transform (Data Transformation) is a step of critical significance in determining sense from raw data. Transformation may include:

- ✧ Cleaning: removal of erroneous, incomplete, or redundant data.
- ✧ Normalization: transforming data into a shared format (e.g., all distances in meters, all interactions in equivalent events).
- ✧ Aggregation: summing events at the session, user, or location level.
- ✧ Enrichment: supplying useful metadata – i.e., correlating a coordinate to an "area of interest" (store, stage, social space). Transformation is where the data is rendered contextual and analytical and is ready to render meaningful metrics.

Load (Data Loading) is the final stage, the processed data is moved to an analytical system or central store (data warehouse, data lake, dashboard). Examples: Google BigQuery, Amazon Redshift, Snowflake, or NoSQL databases MongoDB. They can be accessed by visualization tools (e.g., Tableau, Power BI), analysis engines, or machine learning algorithms from here.

5. Metrics for Evaluating Consumer Engagement

Assessing the behavior of consumers in the metaverse cannot be achieved through observing their actions in a vague way. It is important to introduce certain metrics that quantify the level of user activity in a measurable and business-critical form. The ETL process is important to extract and transform data into a framework of actionable measures that can inform marketing efforts and enhance user experience.

Second, we'll examine some of the most relevant measures that can be calculated to ascertain consumer activity in 3D environments:

- Active time vs. total time - Possibly one of the most important measures is active time users spend in an environment. It is measured by calculating the difference between the moment the user enters the environment and the moment they leave it. It may be compared with the overall available time (e.g., duration of a shopping trip or an event) to provide a certain measure of the size of active engagement. A high active time is an indicator of an engaging experience, and a low active time may indicate a less engaging experience.
- Interaction frequency per session - Another important criterion is the number of times an individual user will interact with objects within the virtual world. Examples include interaction with products (e.g., "view," "add to cart"), interaction with interactive events (e.g., likes, comments), to social activity (e.g., messages, inviting other users to locations). Frequent interaction per session suggests an rich user experience, with consumers engaged and interested in exploration.
- 3D conversion rate - Within a metaverse, the conversion rate can be not only finished purchases (buying a physical or digital good) but also other actions that achieve a goal, like:
  - Finishing a learning session.
  - Purchasing an NFT collectible item.
  - Virtual event registration. It is one of the important measures to decide on the success of marketing and sales, to measure consumers' behavior in virtual purchasing.
- Social interaction and peer-to-peer communication - The second important measure is social interaction – namely, how many times and in which way users interact with other members of the community. The usage can be:
  - Voice or text communication with other users.
  - Participation in social group or group activity.
  - Reactions to actions of other consumers (e.g., comments, giving away or exchanging impressions). Social interaction involves strong emotional involvement and serves as a prominent indicator when analyzing user loyalty and creating a community feeling.
- Paths of movement and heatmaps - Paths of movement and heatmaps are visual representations that show where consumers are spending most of their time in a virtual space and where they are returning most frequently. These metrics can be utilized to quantify points of interest in a virtual store, theme park, or

exhibition, and spatial design insights they offer are very valuable. Reading through these can help brands get their virtual locations correct and improve the shopping or interaction experience.

- Retention Rate - Each virtual world has a most important measure that is the retention rate, which measures how many users continue to return to the environment after a period of time. For instance, a customer returning to look at the same virtual store or to participate in a second session of training is likely to become a repeat customer or recommend others to the experience. Retention is determined by the variety and attractiveness of the experience being offered, social and economic rewards, and personalization of the interaction.

6. The technology behind the ETL process in the Metaverse

The ETL (Extract, Transform, Load) process is such that the initial step, which is referred to as data extraction, is to gather data from a multitude of sources. These sources can include metaverse sites, such as virtual or augmented reality applications or education sites, where data is typically stored in log format or distributed database. Virtual socialization systems are also involved in the generation of a large portion of data from social interaction, e.g., messages, chat conversations, or peer-to-peer communication. Virtual commerce systems also constitute another significant source from where data regarding virtual buying, selling, and business activities of users are collected either from virtual stores or from economic transactions made by players. Additionally, data regarding physical activity in virtual environments is provided by user devices, i.e., head or hand motions, by tracking devices like VR headsets. Technically, the data extraction stage might use application programming interfaces (APIs) to extract the data from the metaverse platforms or web scraping tools in cases where platforms do not have these standardized interfaces.

The second stage of the ETL process involves data transformation, which involves cleaning, preprocessing, and normalizing the data that has been extracted so that it becomes ready for use in analytical purposes. Data cleaning refers to the removal of inaccuracies and discrepancies, and completion of missing data, and normalizing inconsistent or duplicate values, such as duplicated entry of the same action by a user. Normalizing the form refers to the standardization of data representation, e.g., converting recorded times from a standard plethora of time zones to their standardized UTC format, for the same purpose. Data aggregation is the procedure of arranging discrete data points, for example, visits or clicks, to obtain synthetic measures like cumulative time spent in the virtual world, total interactions, or participation level per session. Besides, derived variables may be established, like an engagement score computed based on the interactions, the time consumed in specific regions of the metaverse, and the feedback from users. To implement these transformations, technologies such as Apache Spark for processing of big data are used, Python (along with Pandas and NumPy libraries) to perform data cleaning and manipulation, and ETL tools such as Talend or Apache Nifi, which assist in integrating and transforming data from multiple sources.

The final operation in the ETL process is loading the data that has been transformed into a data warehouse or database for subsequent analysis. To analyze the data produced in the metaverse, it may be loaded into a cloud database system such as Amazon Redshift, Google BigQuery, or Snowflake, or in a custom-built data repository system. The architectures and technologies used in this stage are relational databases (SQL) and non-relational databases (NoSQL), which are used for the storage of structured and semi-structured data, i.e., transactional data, user interactions, and their activities. Data warehouses provide infrastructure requirements for storing and processing large amounts of information as well as enabling advanced analytical capabilities. Once the data are saved to the storage system, Business Intelligence tools such as Tableau or Power BI are used in order to generate reports and interactive charts that allow for a better comprehension of users' behavior and the conclusion of appropriate conclusions when making decisions. After the data has been extracted, transformed, and loaded, advanced analysis methods are applied to make forecasts and reveal hidden patterns in customer behavior. Machine learning algorithms, for example, such as the K-means clustering algorithm, can therefore be applied to group users based on their similar behavior, such as frequency of use or type of activities conducted within the virtual environment.

Sentiment analysis is also a crucial field, and it involves using natural language processing (NLP) techniques to examine the messages and comments by the users on the metaverse. Sentiment analysis allows for the measurement of general sentiment towards brands, products, or events occurring in the virtual world. Additionally, recommendation systems based on individualized collaborative filtering can be constructed that suggest relevant products or virtual activities for each user, on the basis of their own and other users' behavior with similar profiles. In support of ETL procedures and connected analyses in the metaverse, a series of particular technologies are used. Apache Kafka is used in processing real-time virtual environments' data streams, while AWS Lambda enables on-demand data processing as well as interaction with other cloud resources. Google Analytics and Mixpanel support tracking users' activities in virtual spaces, providing rich descriptions of interaction, navigation paths, or engagement levels. For the application of machine learning software, such as TensorFlow and Scikit-learn, is what is normally applied in an effort to analyze behavior from users and formulate recommendations aimed at improving experience in virtual environments.

7. ETL Pipelines as Consumer Insight Infrastructure

The runaway expansion of the commercial metaverse—immersive, interactive 3D environments in which users can navigate, interact with, and purchase digital or physical-world goods—has generated a record volume of behavioral data. Product interaction and avatar movement, purchase decisions, and social influence patterns are all such real-time data that must be harvested, normalized, and analyzed on the fly. At its core lies the ETL (Extract, Transform, Load) pipeline, the basic data infrastructure enabling the multimodal data integration from multiple virtual worlds and real-time decision-making along with longitudinal analytics for metaverse businesses.

Figure 2 describes the three fundamental stages of an ETL (Extract, Transform, Load) pipeline, adapted to interactive 3D environments in the metaverse. Each stage is essential to collect, process, and use behavioral data generated by users in virtual spaces.

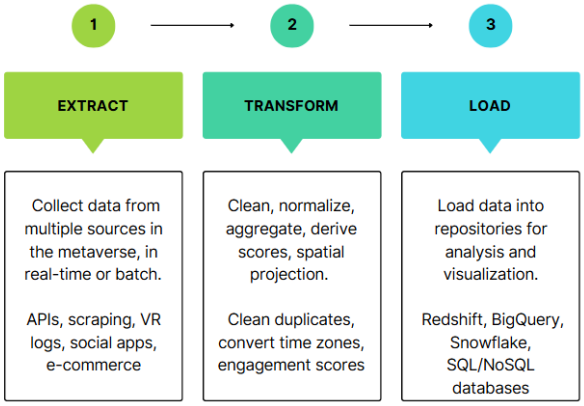

Figure 2. Assessing the behavior of consumers in the metaverse

The main types of behavioral data that can be collected from users during their activities in the metaverse are shown in the figure below. These sources are extremely varied and reflect the complexity of human interaction in virtual reality. It includes both action data (navigation, gestures), social (chats), commercial (shopping), contextual (layouts, promotions) and biometric (physical movements or sensory responses). In practice, all of this data provides a "complete picture" of how users interact with the virtual world.

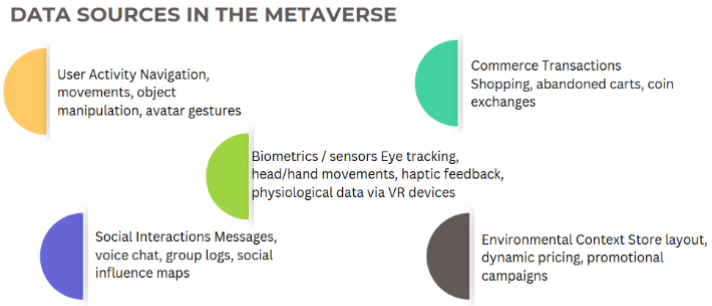

Figure 3. Data sources in the metaverse

By means of computational analysis of behavioral data, businesses can achieve advanced consumer segmentation by grouping users based on interaction behavior, frequency, and navigation patterns. Real-time virtual environment and content realignment according to user activity and preference facilitate dynamic personalization. Aggregating overall interaction data like heatmaps and session flows facilitates user experience (UX) optimization in the form of iteratively improving spatial layout and interface logic. Analyzing campaign performance is also facilitated by ETL pipelines to make possible correlating immersive ad campaigns with quantifiable engagement and purchase behavior.

Automated recommendation platforms are also enhanced by possessing structured interaction data to allow meaningful products and experiences to be provided by machine learning algorithms. In addition to this, user disengagement signals in early stages can be identified by churn prediction models to initiate proactive retention. Sentiment analysis and feedback extraction by natural language processing of user-generated content can offer insights into the perception of a brand and emotional involvement.

Being a core objective of privacy and compliance, pipelines enable data protection norms through encryption and governance structures. Lastly, such functions are combined to facilitate revenue maximization by identifying top-performing elements of the virtual commerce infrastructure and aligning them with consumer behavior cues.

Figure 4 highlights the key applications and strategic objectives of ETL pipelines within the commercial metaverse.

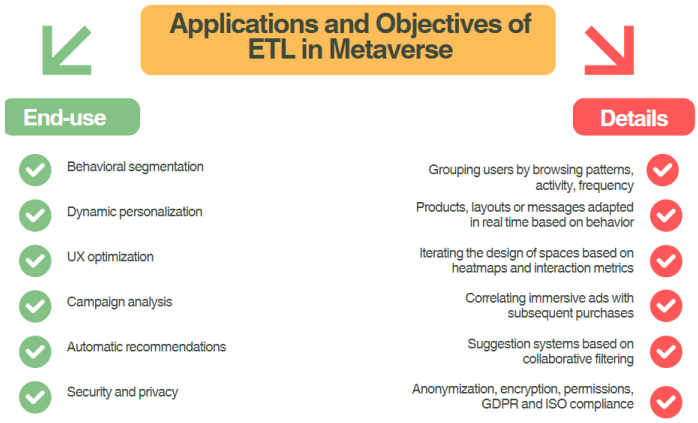

Figure 4. Key applications and strategic objectives of ETL pipelines within the commercial metaverse

8. Conclusion

ETL streamlines the process by which unstructured data, from an assortment of discrete locations and in mixed formats, is extracted, transformed, and loaded into an optimized and consumable system of analysis. Whether used in the realms of e-commerce, education, or otherwise, the ETL process is utilized to extract applicable measures out of that permit precise measurement of user interaction and optimization of their experience. As companies and educational institutions in the metaverse keep pouring funds into data-intensive tech and analytics, we would expect a huge surge in virtual experiences being tailored to individuals. Using measurements such as active time, conversion rate, social interaction, and user feedback, they can easily and efficiently change marketing strategies or improve learning processes. Additionally, ETL technology will continue to enable predictive analytics that will propel the creation of virtual platforms and help construct more attractive and better-performing environments. In the long term, the use of ETL tools will enable companies to deeply understand consumer behavior, not only at an individual level but also at a collective level, allowing them to create virtual experiences that perfectly align with their needs and desires. For education, this could mean creating more interactive and personalized learning environments that improve students' academic performance while also increasing retention and engagement. Lastly, ETL is not just a technical process, but actually a bridge between metaverse information and strategic decision-making that shapes the fate of digital business and learning. With the help of advanced data analysis techniques, companies and schools can make the metaverse a destination for engagement, but also for an implicit knowledge of human behavior.

References

1. *Ganac, C. G. (2018) - Investigating consumer optimum stimulation level and exploratory online buying behavior, DLSU Business & Economics Review, 28(1), 67–85*
2. *Kaur et al. (2023) - Consumer behavior in the metaverse, Journal of Consumer Behaviour. 23. 10.1002/cb.2298 https://www.researchgate.net/publication/376933554_Consumer_behavior_in_the_metaverse*
3. *Yang & Chattopadhyay (2024) – Understanding consumer behavior in the metaverse" (în Handbook of Qualitative Research Methods in Marketing) https://hbhi.jhu.edu/publications/understanding-consumer-behavior-metaverse*
4. *Sun et al. (2022) –Big Data Meets Metaverse: A Survey (arXiv) https://arxiv.org/abs/2210.16282*
5. *Nitin Rane & Choudhary (2023) – "Metaverse Marketing Strategies: Enhancing Customer Experience and Analyzing Consumer Behavior Through Leading-edge Metaverse Technologies…" (SSRN) https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4624199*
6. *Hadi, R., Melumad, S., & Park, E. S. (2023). The Metaverse: A new digital frontier for consumer behavior. Journal of Consumer Psychology. https://doi.org/10.1002/jcpy.1356*
7. *Hollensen, S., Kotler, P.,(2022). Metaverse – The new marketing universe. Journal of Business Strategy, 44(3), 119–125.https://doi.org/10.1108/JBS-01-2022-0014*
8. *William Villegas et. all, Educational Advances in the Metaverse: Boosting Learning Through Virtual and Augmented Reality and Artificial Intelligence, January 2024, ww.researchgate.net/journal/IEEE-Access-2169-3536, PP(99):1-1, DOI: 10.1109/ACCESS.2024.3393776*