aqfer

Aqfer Marketing Data Platform

Flexible. Scalable. Modern.
Ruthlessly Efficient Data Processing

Aqfer Marketing Data Platform (aMDP) is the proprietary data storage framework within the Aqfer Marketing Data Platform-as-a-Service (MDPaaS). aMDP is an elastic Software-as-a-Service (SaaS) product – based on a service-oriented architecture (SOA) – that integrates advertising, marketing, prospect, and customer data from commercial, enterprise, and client first-party data sources. Aqfer is the first company in the world to develop and deliver a marketing data platform to the market.

aMDP is built to be the primary data repository for all forms of marketing-related data. It ingests, collates, resolves, analyzes, and stores large volumes of marketing-related data. It then re-factors that data so it can be rapidly distributed to any and all advertising and marketing technology solutions for batch or near-real-time analysis, segmentation, visualization, and activation. The integration of all client data sources into aMDP, along with Aqfer's provided tools and APIs, significantly eases tasks such as analytics, media planning, attribution, activation, machine learning, and audience analysis.

Thanks to its proprietary Avro + Parquet data structure, aMDP is able to quickly, efficiently, and cost-effectively process billions of data records at a time. Unlike similar competitors in the marketing data management space whose platforms only provide aggregated data, it is able to harmonize, ingest and collate first-, second- and third-party log-level data (geo-location, cookie IDs, transactions, time stamps, URLs) in addition to non-aggregated data. aMDP can process billions of data records in under 60 seconds and organize them into audience-centric segments at-scale faster and more efficiently than any other marketing data management solution on the market.

Additionally, a key differentiator between aMDP and similar solutions is that all data stored within the solution is done so in the users' own virtual private cloud. Essentially, aMDP is ClientMDP, giving the user complete control over the data that passes into and from it at all times. aMDP is deployed and hosted on leading cloud services providers (CSPs) like Amazon Web Services and Google Cloud and allows for customization at each layer in the solution, including the Data, Services, and Application Layers.

aMDP can be licensed as a standalone product or packaged with additional MDPaaS products and modules to power enhanced use cases such as identity resolution and identity graphing.

aqfer

aMDP Components and Architecture

Aqfer Marketing Data Platform has three major architectural components: aMDP Lakehouse, aMDP Processing Engine, and aMDP Data Marts.

aMDP Lakehouse

A data lakehouse is a data management architecture that combines the flexibility and cost-efficiency of a data lake with the data management and structure features of a data warehouse into one package. This makes it possible to store both structured and unstructured data in a single repository and access it using a variety of downstream activation tools and platforms.

The aMDP Lakehouse is the system of record for storage of marketing data for a solutions provider or brand. It contains granular, detailed data that has been harmonized in the standard Aqfer Lakehouse schema that is optimized for projection to various marketing and advertising environments. The data in the aMDP Lakehouse is generally immutable (i.e., fixed) and so does not change over time. Data records are stored in a row-oriented format, specifically Apache AVRO.

aMDP Processing Engine

The aMDP Processing Engine handles the large-scale processing of data ingested into the aMDP Lakehouse, materializing aMDP Data Marts (see below), and distributing data to downstream applications and data stores.

It is optimized for modern CSPs, taking advantage of the effectively limitless cross-sectional bandwidth of their object stores and handling the complex non-linear data processing (collation, joins, shuffles, sorts, and high cardinality aggregation) typical in AdTech and Martech big data applications. The aMDP Processing Engine runs on Kubernetes clusters on multiple CSPs. Its worker nodes that access data protection-controlled data can operate within a compute-forward architecture or a data-forward architecture (more on this below).

The aMDP Processing Engine executes instructions that may be customized using its "low code" solution based on YAML instructions that are developed and tested in the aMDP Sandbox testing environment. Low code solutions are beneficial for solutions providers who need their deployments to be customizable and repeatable because the YAML files can be version-controlled and saved as generic vendor or industry configurations and then forked for end-customer-specific requirements.

aqfer

Data mapping in aMDP can be configured via translation tables that allow business analysts to maintain long lists of field and value translations or sub-tenant-specific translations. In addition, the aMDP Processing Engine supports customization via Python plug-ins that operate on individual records that handle unique or complex requirements (for example, sub-event or nested data processing) in a highly-optimized fashion.

aMDP Data Marts

aMDP Data Marts are data environments created to answer specific analytic problems and use cases for the data stored in aMDP Lakehouse. Clients can customize aMDP Data Marts to their own schema, or use Aqfer's standard schema.

By default, they are stored in Apache Parquet format, a columnar organization that is optimized for fast query access. The Parquet format allows a variety of query engines to access them in a cloud service provider (CSP) object store, such as Amazon Web Service Athena, Google Cloud Platform BigQuery, and Microsft Azure Synapse.

Additionally, clients can also create Named Queries from within Aqfer Portal for testing, troubleshooting, or quick access during data management. Named Queries allows users to define and save SQL queries that can be executed on-demand or on a user-defined schedule. Named Queries provide an abstraction layer over the underlying data, allowing users to create business-friendly views of their data without having to worry about the data's schema or location. .

Data Mart tables may be materialized to optimize specific use cases like multi-touch attribution, audience statistical analysis and selection, or advertising campaign optimization. In addition, Data Marts may be projected into query engines with their own proprietary storage, such as Snowflake, AWS RedShift, Google BigQuery, or Teradata.

aMDP Deployment Architecture

In aMDP's standard deployment architecture, Aqfer provides clients with their own secure virtual private cloud (available on the CSP of their choice) wherein aMDP is deployed. In this model, compute and storage are separate functions that take place within the Aqfer Control Plane.

Additionally, Aqfer also offers Compute-Forward or Data-Forward deployment models (see below). Regardless of the client's preferred deployment model, Aqfer cannot access the client's virtual private cloud without authorization, never has access to the client's data at any time, and will never scrape the client's data.

aqfer

Aqfer Control Plane

1

Client-controlled virtual private clouds isolate data at rest and can be configured for subtenancy with multiple sub-clients with full data encryption.

2

Aqfer's SOC2-certified processing separates storage from compute to improve performance and security as all data in this environment is destroyed after processing.

Client-Controlled Cloud Sub-Account (1)

Client Base MDP Deployment

Sub-Tenant Storage Environment

Sub-Tenant A

Sub-Tenant B

Aqfer Production Environment (2)

Client Base MDP Processing Cluster

Sub-Tenant A Processing Cluster

Sub-Tenant B Processing Cluster

(3)

Ecosystem of Distribution (4)

3

Metadata attached to each record includes audit trails to show when data has been collected, processed, or distributed to external sources

4

Rules-based policies can be set up to match the appropriate privacy consent frameworks in various regions around the world
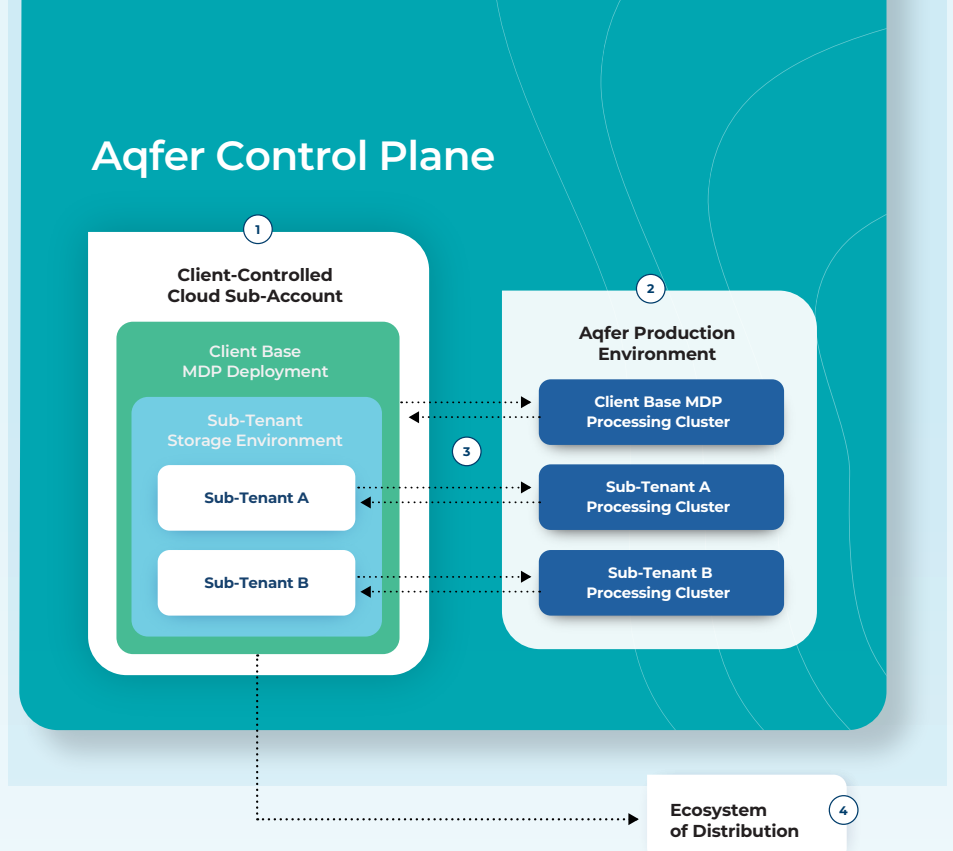
Compute-Forward Model

aMDP's compute-forward architecture allows it to be deployed in the client's own virtual private cloud. This allows data processing to occur within the cloud environment of the client or brand rather than within Aqfer's systems. This gives clients control over the data processing location while still leveraging the full functionality of the platform. The compute-forward architecture can be deployed in one of three ways:

1. Aqfer's SOC2-compliant compute environment, which isolates client workloads

2. A vendor's multi-tenant compute environment in the client's virtual private cloud

3. A dedicated compute environment for an individual brand or secure data collaboration partnership in the client's own virtual private cloud

aqfer

Data-Forward Model

aMDP's data-forward architecture allows it to be deployed in a dedicated account for each client. These accounts can be part of Aqfer's billing organization, the client's organization, or a brand- or publisher-specific organization. In this model, data processing can be performed in Aqfer's SOC2 compliant trusted compute environment or in a dedicated account in the cloud organization of Aqfer's client or the brand/publisher itself. This allows data and computation to be deployed where the clients' needs dictate, including in the clients' own virtual private cloud if necessary.

How aMDP Operates

aMDP has four main stages/processes that control the flow of data into and within each client's environment: **Data Integration Service, Data Collation Service, Data Distribution Service,** and **Data Analytics Service** (more on each of these to follow). In order for any data-related action to take place during any and each of these processes, a Job must first be created and executed.

Jobs are used in Aqfer to run recurring tasks and processes in aMDP. The Jobs interface in Aqfer Portal allows users to create, view, configure, and manage jobs. There are two main environments for jobs: live production and aMDP Sandbox, each of which is accessed via Aqfer Portal.

Production jobs are the live jobs used by clients. aMDP Sandbox allows users to test and develop Jobs before promoting them to live production. This helps ensure live production Jobs are running as expected. Once tested, Jobs created in aMDP Sandbox can be promoted to production environments for live use. In the Jobs List in Aqfer Portal, users can view all live production and aMDP Sandbox jobs.

There are six core Job types performed within aMDP:

Mapper: These transform and enrich data by mapping input fields to output fields. The mapper configuration defines input schemas, output schemas, and data mapping logic.

Collator: These join and aggregate data sets together. The collator configuration specifies inputs, keys, join logic, aggregations, and output schema.

Analysis: These perform analytics and machine learning on the data. The analysis configuration sets up the analytic logic and modules.

aqfer

File Transfer: These move data files between storage systems. The file transfer configuration specifies source and destination paths.

Connector: These allow aMDP to Integrate with external services like CRM and marketing automation systems. The connector configuration sets up API credentials and endpoints.

Pipeline: These chain multiple jobs together into an orchestrated workflow, allowing for complex ETL and data processing workflows to be configured as code and automated. They configure and sequence various steps including importing data, mapping and transforming fields, joining data sets, and executing analytics. They act as the conductor to run each component Job in a predefined order.

Let's now look at what each of the four main stages of aMDP operate and what they control.

Data Integration Service

First-, second- and/or third-party marketing-related data is ingested from various source(s) via data channels via the **aMDP Data Integration Service** stage. This tooling includes Channel Verification, Archiving, Validation, Schema Standardization, Sub-Event Aggregation, and storage to the output schema. The Data Integration Service is used for both event-level data as well as for pre-aggregate and reference data.

These data channels are associated to a specific data source with a specific set of credentials, both for ingestion and distribution. aMDP comes standard with a number of pre-existing data channels in the form of commercial partner integrations, relieving clients of managing configuration details (e.g., column mapping or API call semantics) Client-specific, custom enterprise connections that require unique mapping jobs can be developed by Aqfer in 14 days or less in most cases.

There are several stages that take place as part of the Data Integration Service:

1. The Validation stage is responsible for curation of the data from incoming Data Channels at the row/record and field/attribute level. Every record is examined to ensure it is compliant with the expected schema. Attribute values are also checked to make sure they are of the correct type.

2. The Mapping stage maps the source data schema for a data channel to your client schema in two steps for data channels for commercial partner integrations or a single step for enterprise sources. More complex mappings and transformations can be implemented via a client-supplied plug-in for the Mapping stage using Scala or Python code.

aqfer

3. The Enhancement stage allows the adornment or modification of detail records with additional data. For example, user agent strings might be parsed into browser and OS fields, or URLs might be decoded into classifications that represent their activity. These enhancements can be controlled by a combination of a client-controlled YAML configuration file and client-supplied plug-in logic.

4. The Sub-Event Aggregation stage groups sub-events such as in-page actions, in-app actions, clicks and click-through conversions, video engagement events, etc., together with the logical data subject business event that engendered the sub-events. The output of the Sub-Event Aggregation stage is stored in the aMDP Lakehouse as transient data and can be useful for many types of near real-time analysis until it is replaced by Interim Collation data (see below).

aMDP Collation Service

Once data is ingested, it is then stored in the aMDP Lakehouse, where it can then be processed via the aMDP Collation Service within the aMDP Processing Engine.

The aMDP Collation Service provides the tooling required to collate data in a multitude of scenarios, including: intra-day and daily event collation, data augmentation, ID translation, and entity key aggregation. Client-specific customizations are available for virtually every stage within the Data Collation Service.

There are several stages that take place as part of the aMDP Collation Service:

1. The Interim Collation stage physically organizes the events by data subject key using sort and merge techniques. The data subject (entity type) is specified in an Aqfer controlled, client-specific YAML file, which also controls the frequency of Interim Collation (by default, hourly), and can be modified upon request by filing a customer support ticket. In addition, Interim Collation performs sub-event aggregation for late-arriving sub-events and parent events that arrive after the Data Integration Service process is complete for their associated parent event or sub-events.

2. The Augmentation stage allows adornment or modification of data subject data after the events have been collated. Typically, Augmentation involves substituting source user identifiers for the clients' user identifiers and appending additional user or customer attributes to the data. User ID translation is controlled by configuration files managed by Aqfer operations, but clients can specify how to do their specific data augmentation.

   The data emitted from the Augmentation stage is stored in the aMDP Lakehouse as transient data, superseding the transient data emitted by the Sub-Event Aggregation Stage until it is replaced by Final Collation data (see below). This creates temporary fact tables in the aMDP Data Mart so intra-day events can be analyzed prior to Final Collation.

aqfer

3. The Final Collation stage combines the collated data from the Augmentation stage and intelligently merges the data so that data subject key-organized AVRO data structures in the aMDP Lakehouse are populated. Additionally, final sub-event aggregation takes place. It also populates the fact tables in the aMDP Data Mart and removes transient data from the Sub-event Aggregation and Augmentation stages.

Data Distribution Service

Following processing during the aMDP Collation Service, the processed and enriched data is then ready for distribution. The Data Distribution Service provides the tooling for moving data from aMDP to virtually any endpoint destination desired by the client, such as aMDP Data Marts, files in S3 or Secure File Transfer Protocol, and API endpoints. The service handles distributing data in the appropriate format and via the required protocols for each endpoint, which provides flexibility to distribute to a variety of partners and systems.

Distribution is handled in a secure and privacy-compliant manner through encryption and access controls. The distribution process is automated through workflows and APIs, which enables real-time distribution when new data is available rather than via batch transfer.

Data Analytics Service

The aMDP Data Analytics Service provides the tooling to create and maintain aMDP Data Marts. aMDP Data Marts are application-specific data stores generated by aMDP for use in marketing analytics and applications that contain aggregated and summarized views of the detailed data stored in the aMDP Lakehouse. Users can run customizable Named Queries on aMDP Data Marts using Aqfer Portal, or use Amazon Athena, Google BigQuery, or Microsoft Synapse for external data querying.

aMDP Data Marts store aggregate data in Apache Parquet format, making them optimized for the fast queries needed for modern marketing analytics use cases. The results of these queries can be used in downstream activations such as audience builds, attribution analysis, and to feed dashboards and reporting tools (or other data visualizations).

aqfer

Data Governance and Privacy Compliance

With aMDP, all marketing-related data gets managed in its source form (as permitted under governance guidelines and compliance mandates). aMDP's transactional (time of import) governance approach allows users to apply any and all necessary data governance and consumer privacy rules/requirements as part of the data import process. This ensures that no unauthorized or unlawful data is stored and used downstream by the client. Most data management solutions can only apply data governance and consumer privacy rules and requirements after data has been imported into the solution.

This approach ensures compliance with a variety of current consumer privacy regulations, such as GDPR, CCPA, and 'Right to Be Forgotten', and makes aMDP ready for any and all future policy/regulation changes (including 'Cookiepocalypse'). Clients can also plug a consent management platform into aMDP to further manage consent and permissioning.

The default aMDP Lakehouse schema includes the concept of Policy Effectivity, which attaches metadata to data subject-related data about the privacy and other policies that govern the use of the data. This can then be used to enforce policy restrictions at time of access or provide an audit trail of policies and permissions applicable to a data subject's information.

The aMDP framework provides controls to ensure that security best practices are followed (e.g., SOC 2), and to aid with regulatory compliance related to consumer privacy (GDPR, CCPA, etc.). To accomplish this, each aMDP Lakehouse is provisioned within a dedicated and isolated environment. In the case of AWS, for example, this occurs in an AWS account, which owns the storage, access controls, encryption keys, etc. This account is allocated to the owner of the data (the client), who maintains control and ownership of it.

aMDP services access this account and its resources via secured credentials that are stored within a hardware security module (HSM) or other similar mechanism. Your Aqfer account team (engineering and technical operations staff) do not have direct access to any client credentials that are stored in this way. Applications and people, where required, are granted access to "use" the credentials without ever having the ability to view them.

The aMDP Data Mart, by default, is generally the only system that can be queried by end-user applications. As a privacy and security feature, the default configuration is to mask data subject keys in the aMDP Data Mart using a secret one-way hash so that they are useful for longitudinal analysis but cannot be directly used for addressable marketing or tied to a user agent or device.

aqfer

Conclusion

Aqfer Marketing Data Platform (aMDP) revolutionizes marketing data management as a proprietary framework within the Aqfer Marketing Data Platform-as-a-Service. Operating on a Software-as-a-Service (SaaS) model with a service-oriented architecture, aMDP stands out for its elastic capabilities and versatility in integrating advertising, marketing, and customer data from diverse sources.

aMDP serves as a comprehensive repository for marketing-related data, adeptly handling ingestion, collation, and analysis of vast datasets. Notably, its Avro + Parquet data structure ensures rapid processing of billions of records, surpassing competitors in speed and flexibility. The platform's unique ability to harmonize log-level data from various sources enhances analytics, attribution, and audience analysis.

The standout feature is aMDP's commitment to user control and security — all data resides within users' virtual private clouds, ensuring unparalleled privacy. Deployed on leading cloud service providers, aMDP offers customization across data layers, reinforcing its adaptability.

The three core architectural components — aMDP Lakehouse, aMDP Processing Engine, and aMDP Data Marts — form a robust foundation for scalable data management. The data lakehouse architecture combines data lake flexibility with warehouse structure, facilitating the storage of structured and unstructured data.

aMDP's deployment models, Compute-Forward and Data-Forward, provide strategic choices for clients, allowing control over data processing locations. Operational stages, from Data Integration to Data Analytics Services, showcase meticulous control and automation, ensuring efficient data flow.

Crucially, aMDP prioritizes data governance and privacy compliance, employing a transactional governance approach during data import. This approach ensures compliance with regulations like GDPR and CCPA, setting a standard for privacy and security.

In essence, Aqfer Marketing Data Platform is a pioneering solution in cost-effective, scalable marketing data management. Its feature-rich architecture, user control, and commitment to governance underscore its leadership in the evolving landscape of marketing data platforms.

aqfer

To learn more,
visit us at www.aqfer.com
or contact us at info@aqfer.com