DMS Overview

Last updated by | Miroslav Plese | Jul 17, 2023 at 5:36 PM PDT

Overview

Data Movement service (DMS) supports high-scale hydration of analytical data into Dataverse entities. Using DMS APIs, 1P apps can move their data from Managed Lake into Dataverse tables at a high rate while ensuring execution reliability and workload prioritization and balancing across orgs. This service is part of the Insights App Platform offering and will be available natively to 1P apps through Dataverse APIs.

Primary driver for this project has been the need to modernize and consolidate existing high volume data hydration and movement scenarios over a single stack that would utilize emergent high-performing xMultiple DV APIs, Core Services for better reliability and BCDR support, and maximize use of available DV processing resources to increase overall throughput. In the past major Dynamics implementation partners faced difficulties ensuring reliable high-volume data loads leading to CRIs and escalations. Our tests of DV table hydration performance have shown up to 8x improvement over other technologies, which ultimately will benefit Microsoft with lower cost to execute data movement, and 1P app teams by reducing the time required to synchronize different data storage systems.

Why use DMS?

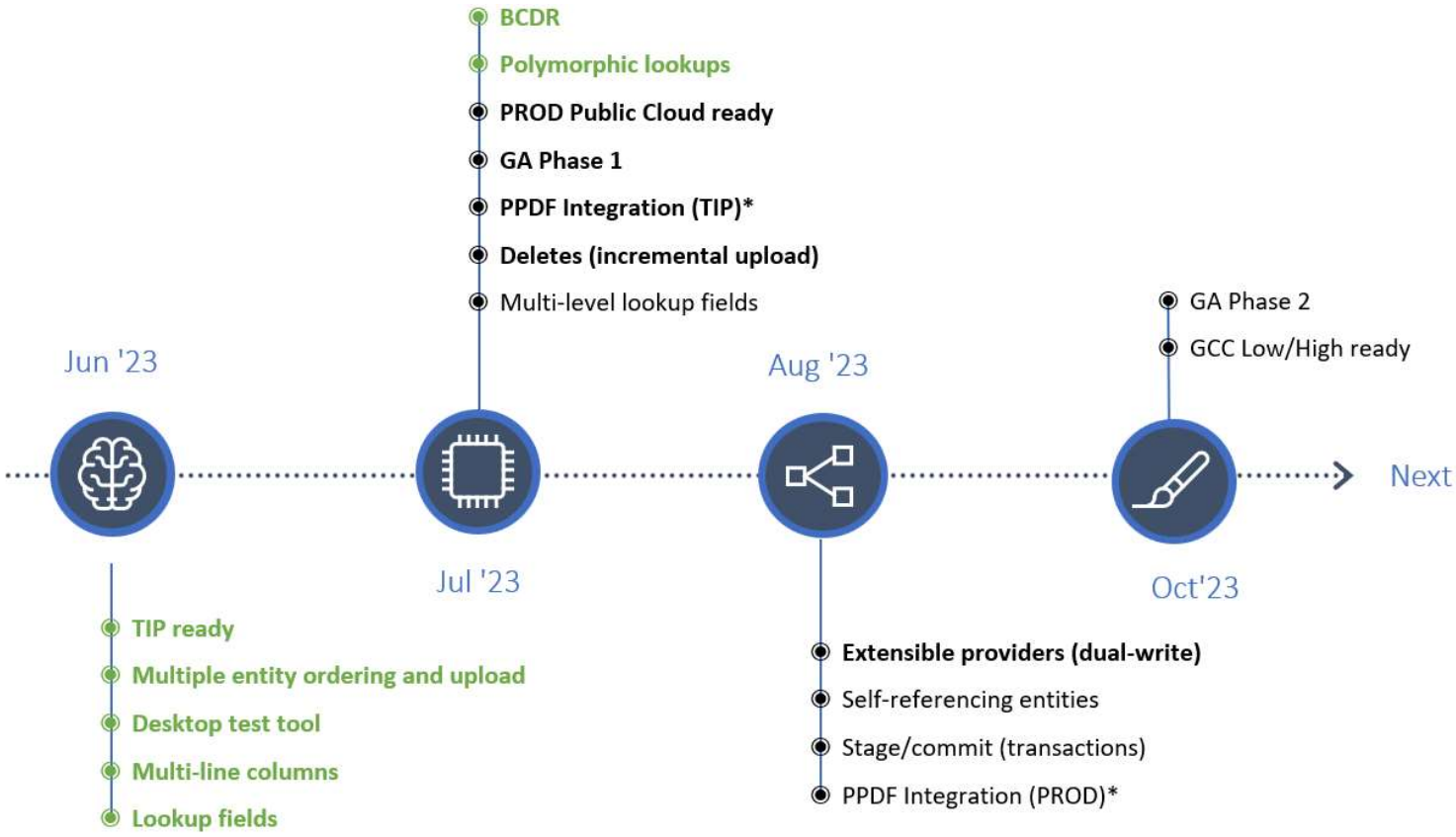
- Data Movement Service supports **hydration** of single or multiple (relationship-ordered) CDM entities from MDL to SQL or Elastic DV entities and ensures DV table data cleanup post-completion. It can perform either alternate or primary key based upserts at the destination table and supports lookup fields too.
- DMS is built on **Dataverse constructs** and provides APIs for managing/monitoring workloads and uses standard behaviors such as role management, API throttling and others. It can easily integrate with MEF to build E2E analytical data processing pipelines.
- The service was designed ground-up with **reliability** and **flexibility** goals in mind. It monitors worker health, automatically recovers on transient issues using checkpointing, ensures fairness, avoids noisy neighbor problems and is also fully compliant with BCDR.
- DMS is highly **optimized** and **performant**, adapting its processing strategy based on available DV scale-out targets. It minimizes network I/O through batch record processing and maximizes use of available processing threads in an org. It also supports incremental data upload scenarios.
- **Lower operational costs** – DMS run-time metadata is kept on Power Platform Metadata Store (PPMS) which is a shared storage and uses Core Services for the compute layer. With the higher efficiency and reliability it brings, partners can benefit from an overall reduction of their bills and less maintenance and customer support costs.

Feature details

- Runs on Core Services infra, which provides high availability, scaling, and region availability for Dataverse services.
- Built as a flexible system that can scale throughput to align with Dataverse org scale-out. If DV org tier is upgraded, DMS perf will improve proportionally based on degree-of-parallelism hint from DV and API throughput.
- Easy integration with MEF to build complete analytical data processing pipelines.
- Incremental processing support through Create and Delete requests.

- Built-in resilience – performs durable tasks execution that can recover transient failures from a prior checkpoint. Automatic retries on transient failures – handle Dataverse HTTP status 429 with retry-after. DMS runs active-active in paired regions offering proper disaster recovery along with better performance.
- Standard Dataflow refresh requests from different Dataverse orgs will be processed with fairness, i.e. one DMS job cannot use up all the resources and starve others, which avoids noisy neighbor problem.
- Leverages new high-perf Dataverse batch Upsert APIs (xMultiple). Automatically switches between xMultiple and single CRUD APIs for entities not supporting the xMultiple APIs.
- Performs batching of SQL transactions to reduce network chattiness.
- Can handle a high number of concurrent standard Dataflow requests from multiple Dataverse orgs.
- DMS system works for every org size – Lite or Basic.
- Provides standard DV tables and APIs for processing requests and managing status, with the ability to retrieve and monitor status in real-time.
- Designed grounds-up with a divide-and-conquer strategy – uses Spark-like job stages and tasks management over a pool of background workers.

Timeline



(green - completed, bold - in progress)