The Six Essential Functions of the Modern Data Platform

Modern

Modern data platforms are infamously fragmented and still proliferating. Feature sets are at once duplicative and bewilderingly diverse. In order to find the best solution for a business, we need a set of standard functions that every data platform should fulfill. We can then add specialty functions as needed to identify the platform that best meets our needs.

In a previous post, The Core Principles of a Modern Data Platform, we examined the core design principles to show what users should expect from any modern data platform worthy of the name. You'll see that those principles give shape to these essential functions.

This article will examine the essential functions of a modern data platform and how DataOS fits in. You will see how DataOS works with common data tools and, in many cases, can replace them.

DataOS

DataOS is the first true data operating system. It does for your data and data stack what your computer's operating system does for your files and applications. DataOS facilitates operations between apps, supports and enhances security, and helps your data stack perform better and more consistently. As a result, it interacts with the elements and functions of your data stack in ways that no other product does.

1. Data Ingestion

Everything starts with data ingestion: bringing data in from all the data sources you use into the storage systems you've designated for each source. This is the beginning of many data pipelines and workflows.

Modern Solution
DataOS uses a built-in implementation of Flare to ingest data. The process can be automated and set to occur on a schedule or triggered by events elsewhere in the data stack. The operating system can also automate ingestion by any standard tool.

2. Data Storage and Processing

The data storage and processing layer is fundamental to the modern data platform. This is also one of the most rapidly evolving functions. Currently, three architectures

are most common: warehouse, lake, and lakehouse.

Each architecture has its own set of tools to accomplish tasks.

- **Data Warehouse**: The oldest solution still in widespread use and still the gold standard for structured data.
- **Data Lake:** A solution developed for storing large volumes of unprocessed or unstructured data.
- **Lakehouse**: The newest storage and processing solution. As the name suggests, it combines many of the best features of a warehouse with those of a data lake.

Modern Solution
DataOS is storage-agnostic. Any sort of storage volume, from a Snowflake warehouse to the hard drive of a single desktop, is abstracted as a "depot" in DataOS. DataOS works with data from any depot without moving or copying the data.

3. Data Transformation

If you're using a data warehouse, then you're probably using a tool that leverages native SQL for transformation. The other common approach is using an orchestration engine coupled with custom transformation in a programming language like Python.

Modern Solution
DataOS can work with a transformation tool like dbt or Matillion, passing data to the tool for transformation, then passing the transformed data to another tool for analysis or to any depot in its network. Alternatively, the

built-in Flare can write and execute transformation jobs that run on Apache Spark. DataOS supports both batch and incremental workloads.

4. Modern Business Intelligence and Analytics

The latest BI and analytics tools are built to fit within a larger modern data platform. Generally, the goal is self-service data dashboards for individual users who want to manipulate and explore data rather than using static graphs and tables.

Modern Solution
DataOS can pass data to a visualization tool such as Tableau, or it can use its native Apache Atlas to build reports and dashboards customized to different stakeholders.

5. Data Catalogs and Governance

If there is any area where modern data platforms struggle, it is aiding discovery, enabling trust, and bringing context to data. One of the best ways to do this is with rich and editable metadata. In fact, metadata is becoming "big data" in its own right. The main approaches to this function are either open-source tools or proprietary tools developed in-house.

Modern Solution
DataOS has Apache Atlas built-in for building data catalogs and other tools such as Metis for adding and editing metadata. But as with most other cases in this article, DataOS can also work through other tools in your data stack.

6. Modern Data Privacy and Access Governance

Especially with the growth of regulatory frameworks for data privacy (e.g., HIPAA or GDPR), companies must manage privacy and access controls throughout their data stacks. Several tools are emerging to meet this need.

Modern Solution

DataOS handles privacy and security through its Attribute-Based Access Control (ABAC). This is one of its most powerful features — you can assign users their access level, as well as any special access privileges, through tags. You can tag a data table, row, or column with access requirements, and from then on users will be given only the data they are authorized to see, automatically and transparently at every touch point in the network.

Learn how DataOS delivers on all six essential functions of a modern data platform →

BY PAUL SCOTT

TheModern DataCompany

The Six Essential Functions of the Modern Data Platform