Overview

Hybrid Cloud architectures and workflows within enterprises are being held back.

While the problem to easily provision VM's and storage on premise or in the cloud has been solved and is available with only a few clicks, the vast majority of an enterprise's data is unstructured file and object and this key data is siloed in the location it was created and there is no orchestration or management between the two. This needs to change.

This blog will explore the requirements of a new product category needed for Hybrid Cloud architectures: "**Data Orchestration Products based on Superna's Data Plane 2.0 Platform** ".

What are the factors to consider?

1. Files and objects need to coexist transparently
2. Security should be applied consistently regardless of whether files or objects are used
3. Location tracking (where is my data now?)
4. Transparency (I don't need to know anything about protocols)
5. Duplicate data (the copy and versions problem)
6. High Availability access to files and object data
7. Copy, move , sync operations between files and objects should be the same workflow
8. Criteria based data deletion
9. North South Data orchestration (on premise to cloud)
10. East West Data orchestration (cloud to cloud or on premise DC to DC)
11. Data life cycle (old data needs cost less as it's value to the business decreases)

What IT architecture issues exist with Hybrid Cloud today?

1. Business processes and workflows are limited to on premise or Cloud. Data created on premise stays on premise and data created in the cloud stays in the cloud.
2. Business processes and end user departments do not have access to the data they need in the cloud or on premise due to different data access authorization, different access protocols and

network security.

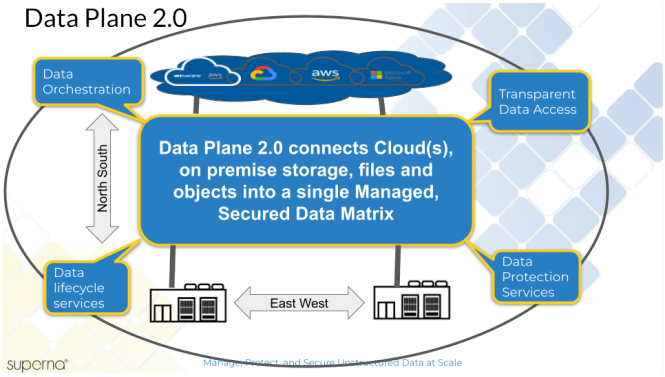
3. Network security has been architected to secure the on premise network from the Internet. This has been the case for decades but hybrid cloud needs to allow this.
4. Inefficient resource allocation and data silos that do not serve the business units due to organizational structures where Cloud IT departments were created separately from on premise IT even though compute, storage, network, and security departments already existed.

What does a Data First Strategy look like?

Businesses create, process and analyze data to create value for the business. Business users should not need to know where the data is stored or have any knowledge of the "protocols" to access, secure, protect, backup and restore data they create.

I want to introduce the future of Hybrid Cloud Architectures, "**Data Plane 2.0**". The goal of Data Plane 2.0 is to create a transparency layer that connects siloed file and object data from the Cloud with on premise storage systems. The Data Plane has 4 dimensions, **Data Orchestration** (East West data and North South data movement), **Transparent Data Access**, common **Data Protection Services** and **Data Lifecycle Services**.

Let's look at each dimension in more detail.



Data Orchestration

1. **Data movement** means files move into object stores and objects move to file systems. Data can move between file systems or object stores in the cloud or on premise. **North South** - This means copy, move, sync data from Cloud → on premise and/or on premise → Cloud. This unidirectional relationship is called a data pipeline. **East West** - This means copy, move, sync data between on

United States
225 Cedar Hill Street, Suite 200
Marlborough, Massachusetts 01752

premise locations and is configured as unidirectional relationships.

2. **Use cases for Data Orchestration**
    a. Disaster recovery
    b. Archiving cold data from file to object
    c. Security - make copies for cyber protection
    d. Backup file to file, file to object (3-2-1 backup strategy)
    e. Data aggregation for analysis (ML/AI or batch processing, transformation (video rendering, billing systems)
    f. IOT Data aggregation from edge locations to central data centers
    g. Triggered or scheduled data movement based on user actions (open file, save file, read, filet etc, user location, user id, data age, data type
3. **Cost Based Policies**
    a. The cloud providers have ingress and egress fees which may play a factor in decisions to orchestrate data movement. These policies would be designed to block or allow data movement into or out of the cloud based on cost thresholds.

Transparent Data Access

1. **Security Transparency** - The goal for this dimension is to sync the security between the file and object to allow consistent enforcement of data access and mapping file user access permissions to object data and vice versa
2. **Data Access** - Users should not need to understand protocols.
   If files are synced to cloud object storage, the access to this data should be presented over SMB/NFS with the same permissions the data had on premise.
   Object data in the cloud synced to on premise needs to be presented over SMB or NFS with security permissions flowing from Cloud to on premise SMB/NFS.
3. **File and Object Transparency** - Data needs to maintain key metadata regardless of where it was created, example owner, group, permissions, date stamps (created, modified, accessed). Without metadata the data itself loses it's context and policies that depend on metadata will fail.
4. **Location Transparency -** Users do not need to know if data is a file or an object or where it is located when searching for data.

Data Protection Services

1. **Disaster Recovery** - Data needs to be replicated to the Cloud or another on premise site or both for DR. DR requires synchronized data access permissions to ensure consistent security access to the data. Hybrid Cloud architectures and workflows complicate protection of data for DR and Security policies
2. **Continuous Real Time Threat Detection -** Real Time Threat Detection must operate against file or object data transparently and must offer: Ransomware detection, mass delete detection, data loss prevention, content aware security policies, insider threat and zero day threat detection behavior based detection capabilities.
3. **Threat detection automation** - The Data Plane **must** comply with the cybersecurity best practices with full end to end automation and self learn (Detect, Identify, Protect, Respond, Recover).
4. **Coordinated Threat Detection** - On premise and cloud threat detection needs to be coordinated.
    a. This means the Data Plane needs to be cyber threat aware and avoid data movement when threats are detected in the Cloud or on premise locations.
    b. The Data Plane is focused on data but security consists of network, compute, and endpoint systems. The Data Plane must **publish** threat levels to security systems and **subscribe** to threat levels from security systems. The Data Plane must take automated actions to protect data when notified of externally detected threats.
5. **Intrinsic Cybervault** - A secure **offline** copy of data must be maintained with **intelligent defense**. This is the ability to block replication to the vault when the Data Plane detects a threat anywhere in the Data plane (Cloud or On Premise).
6. **Who did what when?** File system auditing is a mandatory requirement for all compliance regulations. Object systems have no auditing features on premise or in the Cloud. The Data Plane must provide full auditing capabilities to plug this gap in object security.
7. **Dynamic Data Classification** - This requires content aware security that makes decisions based on the

United States
225 Cedar Hill Street, Suite 200
Marlborough, Massachusetts 01752

contents for the data. PHI, HIPPIA, PCI and many other compliance standards require content search within files/objects to classify the data and make decisions on how the data should be secured, stored or block data movement.

8. **Encryption at rest or inflight** - Policies on how data should be stored requires metadata properties attached to the data itself. The Data Plane needs to ensure storage and replication policies are factored in for example encryption in flight and at rest should be applied by the System.
9. **Smart Transfer** - This means the Data Plane should not allow data to move east,west or north, south if a threat is detected. This should be a base function that prevents compromised data from moving through the data matrix.

Data Lifecycle Services

1. **Archive Services** - As Data ages its value to the business decreases over time. A core requirement for the Data Plane is to understand access patterns and data age with the goal of data orchestration to move data to a storage device and media that matches the long term storage costs of the data. The Data Plane should suggest archive policies to the administrator to apply.
2. **Dynamic Performance Management** - Active data requires storage that provides optimized performance for each workflow. The Data Plane must be able to identify workflows that have performance impacts, identify top workloads, identify performance anomalies and root cause performance issues based on behavior analysis of all workloads. It must be a learning based system that baselines and learns based on the environment it is deployed in covering both on premise and cloud storage. **Context aware performance** means the user identity, the data type is factored into the automated policies
3. **Automated Archive Policies** - Cloud and on premise storage offer different long term storage requirements with similar immutable options for legal compliance. The Data Plane needs to apply automated archive policies across file and object storage regardless of where the data is stored.
4. **Data Location Tracking** requires the Data Plane to track data

location with an index to provide searchability for users and administrators. Data moves must be tracked for both user initiated or policy based data movement

5. **Cost Management** - This means the Data Plane must understand the underlying cost of storage, usage of storage and provide storage tier aware reporting. The cloud is inherently based on pay for consumption but on premise is not. This requires on premise storage to catch up and provide showback and chargeback at granular levels for business units to understand the true cost of data. The Data Plane must bridge this gap between Cloud and on premise file and object reporting for **cost management and storage usage accountability** to the business units.
6. **Audit Trail** - Data Orchestration is triggered based on end user actions or lifecycle policies or security policies or DR policies. The Data Plane must audit all data movement within the Data Plane for **compliance and debugging**. A single source of truth is required to quickly locate data and understand the history of its movement east, west, north and south.
7. **Content Indexing** - This fundamental value of policy based decision making based on the contents of file has been out of reach for Enterprise customers. The future of hybrid cloud architectures needs to support content aware policy driven data protection and security.

Is Data Plane 2.0 available now?

Data Plane 1.0 is available now with many of the key requirements available in Superna products that manage, protect and secure unstructured data at scale.

Superna's roadmap will enable the first Data Plane 2.0 Data Orchestration platform to fully realize the benefits of hybrid cloud architectures. A single platform with on demand applications covering all 4 dimensions of Data Plane 2.0, online upgrade with an always on Data Plane that can be upgraded without loss of any of its services.

Individual functions may exist in various vendor products, but without a vision of what the future looks like you will end up with siloed data and a business can never fully realize the value and benefits of the hybrid cloud architecture.

United States
225 Cedar Hill Street, Suite 200
Marlborough, Massachusetts 01752

Stay tuned for more updates on Superna Data Plane 2.0 product features and announcements.

About the Author
Andrew MacKay - President & CTO of Superna

Superna Data Plane 1.0 Products

1. DR Edition
2. Ransomware Defender
3. Enterprise Airgap for File and Object
4. Easy Auditor
5. Golden Copy
6. AnyCopy
7. Search & Recover
8. Smart Archiver
9. Performance Auditor

United States
225 Cedar Hill Street, Suite 200
Marlborough, Massachusetts 01752