![snowflake®]

# LEVERAGING DATA AS A
## STRATEGIC ASSET WITH
## SNOWFLAKE

**The Platform for Your Federal Data Strategy 2020 Action Plan**

CHAMPION
GUIDES

**EBOOK**

# TABLE OF CONTENTS

# INTRODUCTION: WHAT ARE THE FEDERAL DATA STRATEGY AND THE 2020 ACTION ITEMS?

The Federal Data Strategy, defined by the President's Management Agenda[1] (PMA), defines a long-term vision for modernizing the federal government. It outlines three key drivers of transformation:

- **Modern Information Technology**
- **Data, Accountability, and Transparency**
- **People: Workforce for the 21st Century**

The PMA stipulates that data, accountability, and transparency initiatives must provide the tools to deliver visibly better results to the public and improve accountability to taxpayers for sound fiscal stewardship and mission results. Investments in policy, people, processes, and platforms are key elements of this transformation and require cross-agency cooperation to ensure an integrated Federal Data Strategy[2] that encompasses all relevant governance, standards, infrastructure, and commercialization challenges of operating in a data-driven world.

The Office of Management and Budget approved and released the Federal Data Strategy's first-year action plan in December 2019, marking a milestone in how agencies are expected to implement the strategy within their organizations, and in collaboration across the government, in 2020.

The Federal Data Strategy provides a common set of data principles and best practices in implementing data innovations that drive more value for the public. Annual action plans will identify and prioritize practice-related steps for a given year, along with targeted timeframes and responsible entities.

The 2020 Action Plan establishes a solid foundation that will support implementation of the strategy over the next decade. Specifically, the plan identifies initial actions for agencies that are essential for establishing processes, building capacity, and aligning existing efforts to better leverage data as a strategic asset.

# WHAT IS THE
# SNOWFLAKE DATA CLOUD?

The Data Cloud is a global network of data supported by Snowflake's platform, where you can unite siloed data, easily discover and securely share governed data within and across agencies, and execute diverse analytic workloads. Snowflake's founders started from scratch and built a platform that would harness the immense power of the cloud. Whether you're a business, government, or technology professional, you can get the performance, flexibility, and near-infinite scalability to easily load, integrate, analyze, and securely share your data. Snowflake is a fully-managed service that's simple to use, but it can power a near-unlimited number of concurrent workloads. From a single platform, Snowflake enables your most critical data workloads, including data warehousing, data lakes, data engineering, data science, data application development, and secure data sharing.

Snowflake provides one platform with a revolutionary architecture for near-unlimited data opportunities:

## MULTI-CLUSTER SHARED DATA ARCHITECTURE

Snowflake's unique architecture logically separates but natively integrates storage, computing, and services. You can enable virtually all your users and data workloads to access a single copy of your data without impacting performance.

## GLOBAL DATA SOLUTION

Snowflake abstracts the complexity of underlying cloud infrastructures, allowing you to run your data solution seamlessly across multiple clouds and regions for a consistent experience.

## SECURE COLLABORATION

Snowflake enables you to easily create your own data exchange to collaborate securely across your organization's ecosystem. You can access additional shared datasets and data services via Snowflake Data Marketplace and connect with thousands of Snowflake data providers that comprise the greater Data Cloud.

# WHY IS SNOWFLAKE RELEVANT TO THE FEDERAL DATA STRATEGY?

The Federal Data Strategy provides a guide that helps agencies leverage the value of data by focusing on the following:

- Building a culture that values data and promotes public use
- Governing, managing, and protecting data
- Promoting efficient and appropriate data use

To implement the annual action plan initiatives effectively, agencies should consider a comprehensive platform that enables them to plan, execute, and increasingly leverage data as a strategic asset.[3]

## ONE PLATFORM, ONE COPY OF DATA, MANY WORKLOADS

Snowflake delivers a built-for-the-cloud architecture that enables disparate teams to work on the same data—structured and semi-structured—for a wide range of workloads across clouds.

## NEAR-UNLIMITED PERFORMANCE AND SCALE

Snowflake scales to virtually any volume of data and any number of users and concurrent workloads, without degrading performance.

## SECURE AND GOVERNED ACCESS TO ALL DATA

Snowflake empowers users to discover, share, and consume shared data with other agencies and departments, citizens, partners, and beyond.

## COMPLIANT ACROSS REQUIRED STANDARDS

Snowflake meets the privacy and security regulations guidance for federal departments. It allows you to easily comply with regulatory standards and monitor your compliance on an ongoing basis. Snowflake is FedRAMP Authorized (Moderate) on AWS commercial and Microsoft Azure Government regions, and meets SOC 1 Type 2, SOC 2 Type 2, ISO 27001, FISMA Moderate, NIST 800-171, FIPS 140-2, ARS 3.1 PCI DSS, HITRUST, and HIPAA standards.[4]

## CROSS-REGION, CROSS-CLOUD

Snowflake allows you to distribute data across regions or cloud providers like AWS, Microsoft Azure, and Google Cloud Platform. You can mix and match clouds as needed.

## NEAR-ZERO MAINTENANCE, AS-A-SERVICE

Snowflake requires little maintenance compared to other solutions. Legacy database tasks, such as building, indexing, and managing partitioning schemes, become a thing of the past. Using Snowflake can reduce operational overhead while allowing your essential team members to focus on culture, governance, data assets, transparency, and resources.

This makes Snowflake uniquely suitable for supporting Federal Data Strategy initiatives within and across federal agencies and organizations.

Snowflake is
FedRAMP Authorized

# FEDERAL DATA STRATEGY
## ACTION ITEMS THAT SNOWFLAKE EMPOWERS

This section describes how Snowflake can enable organizations to successfully achieve the objectives of several Federal Data Strategy Action Items.

### ACTION 6: PUBLISH AND UPDATE DATA INVENTORIES

Snowflake's unique approach to data sharing allows customers to collaborate and share datasets seamlessly and securely across cloud regions and even cloud providers. Snowflake Secure Data Sharing eliminates the need for complex and insecure data extraction jobs, file-based exchanges, and complex ingestion routines by data consumers.

Snowflake offers three modes for sharing datasets between a producer account and one or more consumer accounts:

- **DIRECT SHARING:** A provider grants read-only access to any of its database objects to one or more consumer accounts. The consumers see only the tables and views that have been shared with them, and they can instantly begin executing SQL queries against the shared datasets. Use this mode when the number of consumers is small and the consumers are known ahead of time.

- **SNOWFLAKE DATA MARKETPLACE:** A Snowflake customer can become a producer account by publishing one or more datasets into the public Snowflake Data Marketplace. The marketplace is a web-based portal that enables consumers to browse and search for available datasets. With a single click, consumers can subscribe to a dataset and incorporate that data into their Snowflake environment. Snowflake functions as the administrator of Snowflake Data Marketplace. Use this mode to make data available to the greatest number of potential consumers.

- **PRIVATE DATA EXCHANGE:** A private data exchange is a web-based data portal experience accessible only within a single customer organization. Each Snowflake customer can administer a private data exchange. This allows them to decide which data inventories to publish and which subscription requests to approve. Use this mode when the publisher requires tighter governance policies over shared datasets.

Regardless of which sharing mode you choose, all data remains in the data provider's Snowflake database at all times. This is an excellent way to publish data to internal and external audiences easily and securely.

### ACTION 8: IMPROVE DATA AND MODEL RESOURCES FOR AI RESEARCH AND DEVELOPMENT

Snowflake's platform was designed from the ground up to support machine learning and AI-driven data science applications. Snowflake integrates with all the leading AI data science platforms, and it can function as the repository for both the raw data that feeds the models and the generated models themselves. Snowflake's support for semi-structured data formats allows it to function as a full-featured data lake, which complements its core data warehousing functionality.

Because Snowflake enables organizations to prepare data and build machine learning and AI models with ease, you don't need to worry about complex integrations or related expenses. Instead you can focus on creating value out of your AI initiatives.

## ACTION 9: IMPROVE FINANCIAL MANAGEMENT DATA STANDARDS

Snowflake's customers include leading financial services enterprises like Capital One, S&P Global, and FactSet. Meeting strict compliance and financial management standards is part of the core capabilities of Snowflake. Snowflake includes many features such as Dynamic Data Masking and end-to-end encryption for data in transit and at rest. You can fulfill modern governance, compliance, and reporting needs by tracking and measuring compliance, averting cyber threats and fraud in real time, and easily meeting regulatory requirements.

## ACTION 10: INTEGRATE GEOSPATIAL DATA PRACTICES INTO THE FEDERAL DATA ENTERPRISE

Geospatial data provides a powerful tool for organizations. The insights from analyzing geospatial data can increase resource efficiencies, reduce fraud, and eliminate risks. Snowflake provides full support for geospatial data types and functions and can include any of the following types of geospatial objects:

- Point
- MultiPoint
- LineString
- Polygon
- MultiPolygon
- GeometryCollection
- Feature
- FeatureCollection

Geospatial data comes in a wide range of formats. Snowflake currently supports the following formats as input and output:

- GeoJSON (default output format)
- Well-known text (WKT) and extended well-known text (EWKT)
- Well-known binary (WKB) and extended well-known binary (EWKB)

Snowflake's geospatial innovations are part of a broader ecosystem that includes Snowflake Data Marketplace and Snowflake Partner Network for easy access to an even wider set of features.

## ACTION 11: DEVELOP A REPOSITORY OF FEDERAL ENTERPRISE DATA RESOURCES

You can leverage Snowflake as the data warehouse, data lake, or data exchange, or you can participate in Snowflake Data Marketplace for your repository of federal data resources.

**The data warehouse** delivers a single, governed source for virtually all your data, including JSON and XML, that you can query immediately with near-unlimited and cost-effective storage.

Snowflake also provides a flexible solution to enable or enhance your data lake strategy. Use Snowflake as your **data lake** or integrate Snowflake with your existing data lake, taking advantage of a cloud-built architecture that meets your unique needs. Mix and match components of data lake design patterns and unleash the full potential of your data.

Snowflake enables organizations to easily establish a private **data exchange** for user, citizen, and partner ecosystems where data is discoverable, accessible, and secure.

**Snowflake Data Marketplace** (for publically available datasets) gives access to live and ready-to-query data from your data ecosystem to data scientists, BI and analytics professionals, and everyone who desires data-driven decision-making. Your data ecosystem can include other government organizations and partners, and potentially thousands of data providers and data service providers.

## ACTION 13: DEVELOP A CURATED DATA SKILLS CATALOG

Snowflake is relational and supports ANSI SQL and ACID transactions. It offers built-in functions and SQL extensions for traversing, flattening, and nesting of semi-structured data, with support for popular formats such as JSON and Avro.

This allows you to modernize your data platform while leveraging your existing employees and resource pool, with no new training needed. In addition, Snowflake is cloud-agnostic across Amazon Web Services (AWS), Microsoft Azure and Google Cloud Platform (GCP), which greatly simplifies the overall technology stack. Your team can work with one platform, regardless of the data source or cloud infrastructure provider technologies. And Snowflake is provided as a service, allowing you to focus on curating and developing analytics skill sets instead of managing databases and infrastructure.

## ACTION 15: DEVELOP A DATA PROTECTION TOOLKIT

The Snowflake platform was built from the ground up with security in mind. Snowflake's architecture, implementation, and operation are designed to protect data in transit and at rest against both current and evolving security threats. Snowflake was built from the ground up to deliver end-to-end data security for all users. It follows best-in-class, standards-based practices for the controls and processes that secure it. As part of its overall security framework, it leverages NIST 800-53 and the CIS Critical Security Controls, a set of controls created by a broad consortium of international security experts to identify the security functions that are effective against real-world threats. Snowflake comprises a multi-layered security architecture to protect data and access to that data. This security architecture is complemented by the monitoring, alerts, controls, and processes that are part of Snowflake's comprehensive security framework.

Snowflake is a multi-tenant service that implements isolation at multiple levels. It runs inside a virtual private cloud (VPC), a logically isolated network section within AWS, Microsoft Azure, or Google Cloud Platform. The dedicated subnet, along with the implementation of security groups, enables Snowflake to isolate and limit access to its internal components. Snowflake provides additional security features to support HIPAA, PCI DSS, and other compliance requirements. In addition, Snowflake has the tools necessary to ensure data privacy, governance, and security. Snowflake is FedRAMP Authorized (Moderate). You can control access privileges at a granular level to support specific regulatory requirements that will mitigate identified risks in your data management approach for multi-tenant environments.

The key Snowflake capabilities that enable data protection are:

### Authentication

Snowflake employs robust authentication mechanisms, and every request to Snowflake must be authenticated.

### Authorization

Snowflake provides a sophisticated, role-based access control (RBAC) authorization framework to ensure data and information can be accessed or operated on only by authorized users. In Snowflake's access control model, users are assigned one or more roles, each of which can be assigned different access privileges.

### Encryption Throughout

In Snowflake, all customer data is always encrypted when it is stored on disk, and data is encrypted when it's moved into a staging location for loading into Snowflake. Data is also encrypted when it is stored within a database object in Snowflake, when it is cached within a virtual warehouse, and when Snowflake stores a query result.

### Data encryption and key management

Snowflake uses strong AES 256-bit encryption with a hierarchical key model rooted in a cluster of hardware security modules. Each customer account has a separate key hierarchy of account-level, table-level, and file-level keys. Snowflake automatically rotates account and table keys on a regular basis. Data encryption and key management are entirely transparent to customers and require no configuration or management.

### Data protection and recovery through retention and backups

Snowflake was designed from the ground up to be a continuously available cloud service that is resilient to failures to prevent customer disruption and data loss. Its continuous data protection (CDP) capabilities provide easy self-service recovery from accidental errors, system failures, and malicious acts.

### Dynamic Data Masking

Snowflake provides Dynamic Data Masking, which allows a designated administrator to create and apply column-level masking policies. These policies can be defined specifically to restrict access to data in the columns (of tables or views) on which the policy is applied.

# CONCLUSION: WHY SNOWFLAKE IS IDEAL FOR YOUR FEDERAL DATA STRATEGY INITIATIVE

Snowflake was created to help organizations break free from the limitations of conventional software solutions. The patented multi-cluster shared data architecture easily and securely enables a wide variety of workloads—data warehouses, data lakes, data pipelines, and data exchanges—and many types of BI, data science, and data analytics applications. In addition, the platform easily loads, integrates, and analyzes all types of structured and semi-structured data inside a unified repository that seamlessly operates across clouds and across regions, while supporting these workloads.



**Google Cloud Platform**   **Amazon Web Services**   **Microsoft Azure**
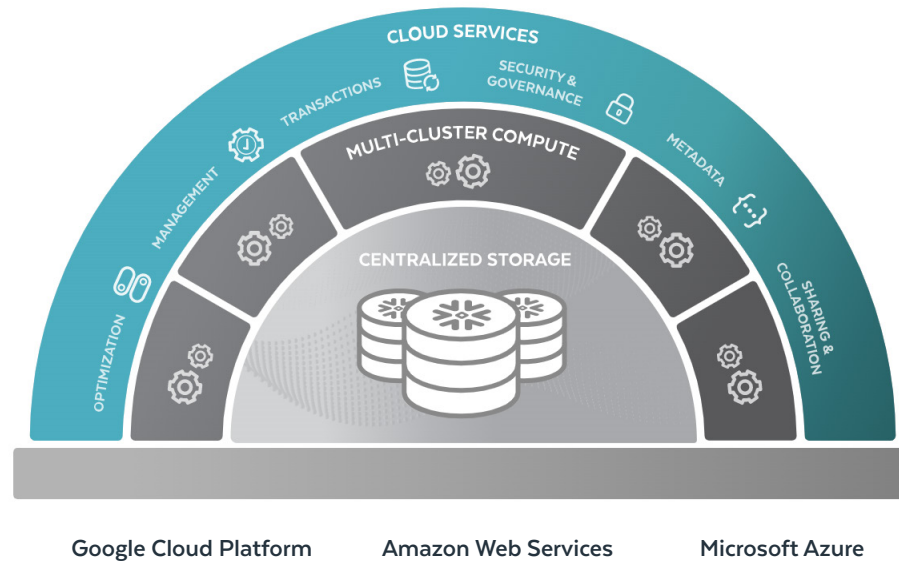
With Snowflake, all of the departmental units of your organization, other agencies, external partners, and citizens can collaborate by seamlessly and securely sharing data to make timely, data-driven decisions. Snowflake provides a simple, powerful, and flexible foundation that enables you to shift your focus from managing a sprawl of disparate infrastructure to deriving insights from virtually all your data, by all your users.

Leveraging data as a strategic asset to achieve your business objectives has never been more straightforward. Meeting the goals of the 2020 Action Items for your Federal Data Strategy initiative can be easier, and your initiatives can potentially be more performant, optimized and efficient, secure and compliant with Snowflake as your comprehensive platform.

## ABOUT SNOWFLAKE

Snowflake delivers the Data Cloud—a global network where thousands of organizations mobilize data with near-unlimited scale, concurrency, and performance. Inside the Data Cloud, organizations unite their siloed data, easily discover and securely share governed data, and execute diverse analytic workloads. Wherever data or users live, Snowflake delivers a single and seamless experience across multiple public clouds. Snowflake's platform is the engine that powers and provides access to the Data Cloud, creating a solution for data warehousing, data lakes, data engineering, data science, data application development, and data sharing. Join Snowflake customers, partners, and data providers already taking their businesses to new frontiers in the Data Cloud.

**snowflake.com/federal**

Snowflake is FedRAMP Authorized

## CITATIONS

[1] whitehouse.gov/wp-content/uploads/2018/04/ThePresidentsManagementAgenda.pdf

[2] strategy.data.gov/assets/docs/2020-federal-data-strategy-framework.pdf

[3] performance.gov/CAP/leveragingdata

[4] Specific certifications can vary across different Snowflake Service Editions