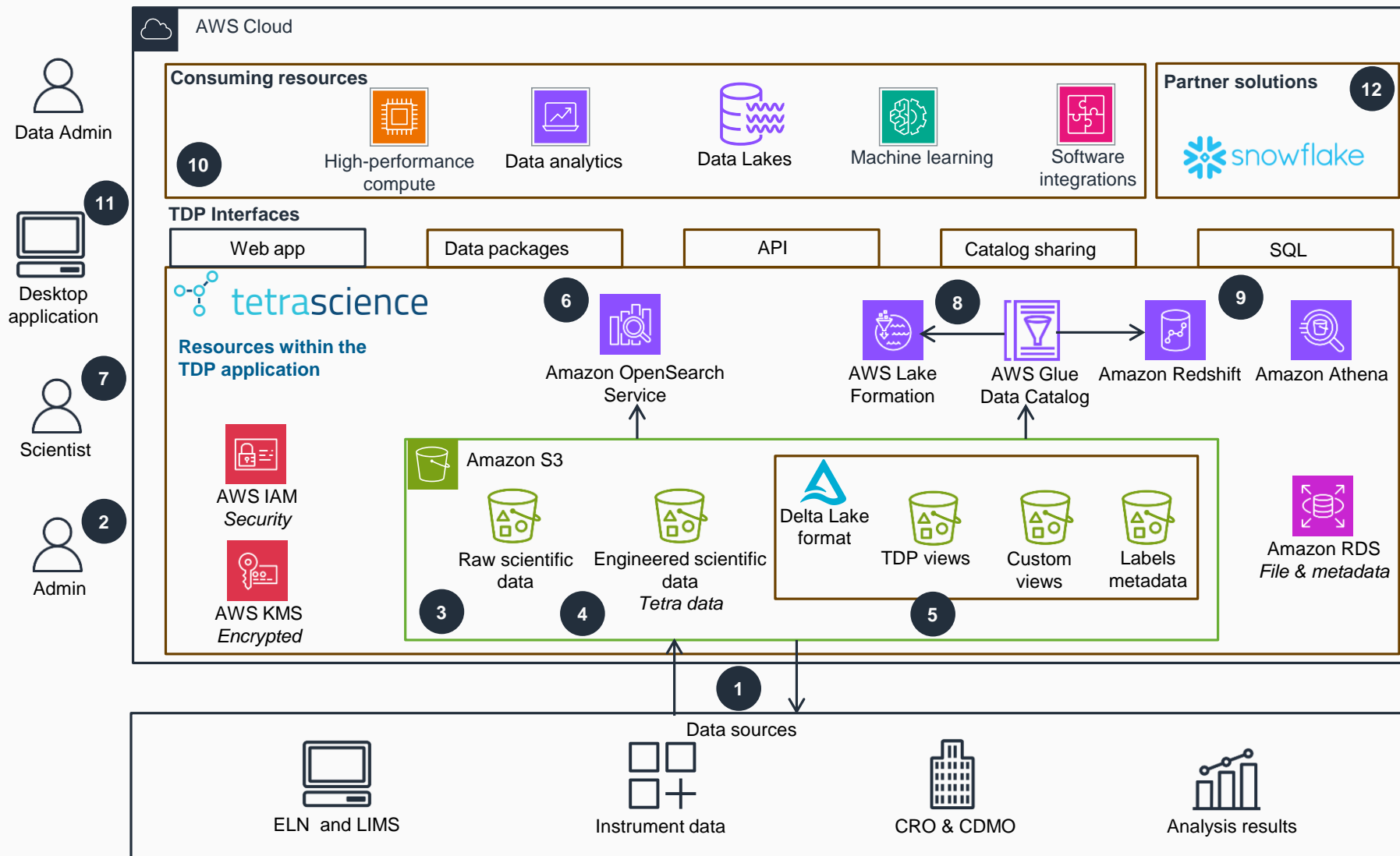


Guidance for Integrating the TetraScience Tetra Data Platform on AWS

This architecture diagram shows various ways you can integrate the Tetra Data Platform (TDP) with your own data lakes on AWS. This slide outlines steps 1-6. For steps 7-12, refer to the next slide.

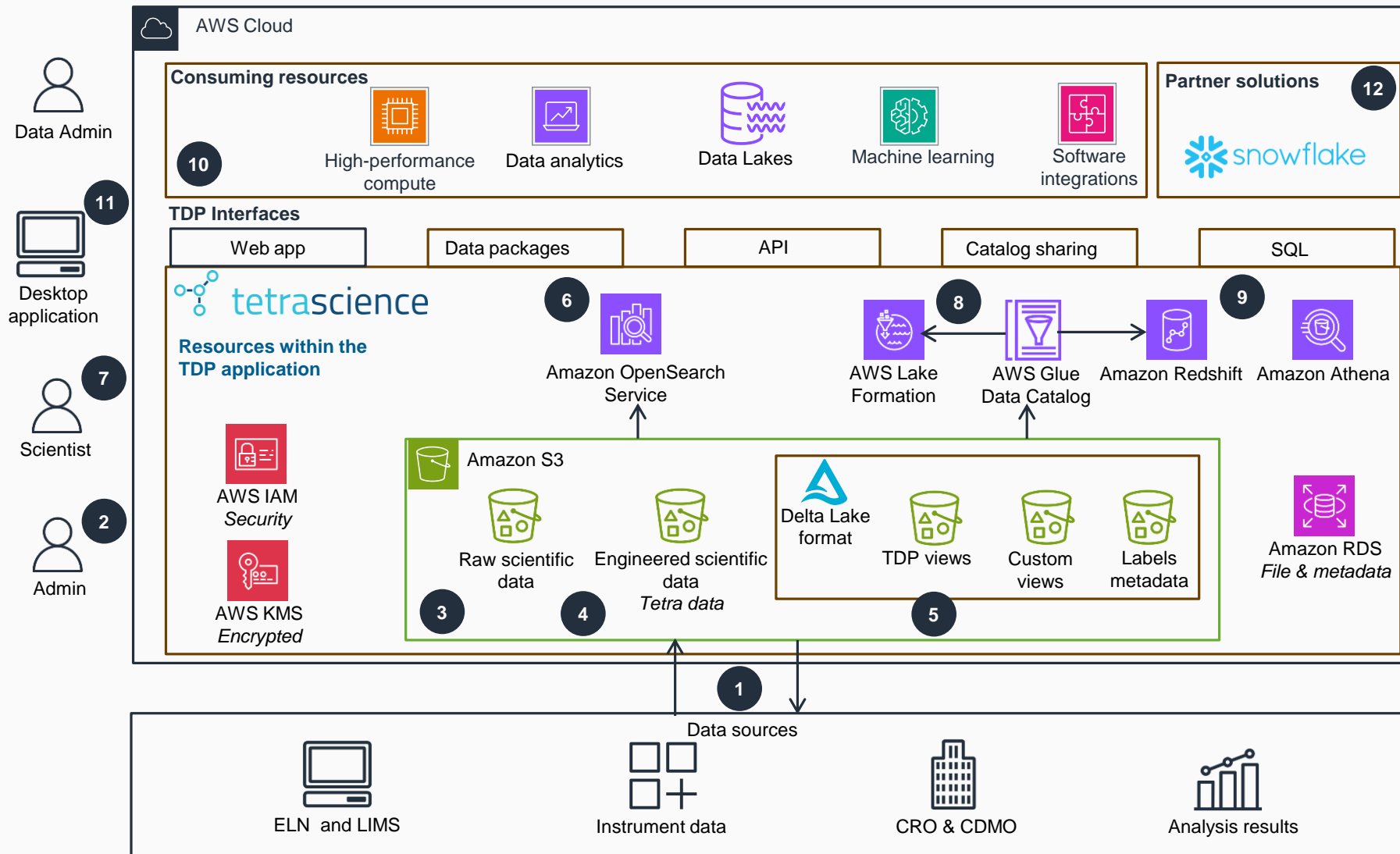


- 1 The Tetra Data Platform (TDP) connects your instruments and software systems, such as an electronic lab notebook (ELN) and a laboratory inventory management system (LIMS). It also connects with partner organizations, although the individual connector components required for that are not shown here. TDP can also connect with contract research organizations (CROs) and contract manufacturing and development organizations (CDMOs).
- 2 The Administrator uses a web application to set monitoring paths and extract metadata, including time, sample, user, and assay data before uploading raw data to TDP. Authentication utilizes **AWS Identity and Access Management (IAM)** roles and credentials.
- 3 TDP collects, enhances, and stores raw data in an **Amazon Simple Storage Service (Amazon S3)** bucket. Metadata resides in **Amazon Relational Database Service (Amazon RDS)** and replicates to the **Amazon S3** metadata bucket. **AWS Key Management Service (AWS KMS)** encrypts all data in TDP.
- 4 TDP converts raw data into engineered scientific data and stores it in the Tetra Data **Amazon S3** bucket.
- 5 Engineered scientific data is converted into tables managed by **AWS Glue** in an open-source Delta Lake format.
- 6 The **Amazon OpenSearch Service** maintains a search index for data.



Guidance for Integrating the TetraScience Tetra Data Platform on AWS

Steps 7-12



- Scientists log in to the web application to search instruments, experiments, and assay data.
- The catalog sharing interface allows Data Administrators to utilize the **AWS Glue Data Catalog** with **AWS Lake Formation** for access controls over Tetra Data tables, supporting secure data sharing between Regions and accounts.
- The SQL interface through **Amazon Redshift Spectrum** and **Amazon Athena** provides interactive querying, analysis, and processing of Tetra Data stored in **Amazon S3** and cataloged in **AWS Glue**. **Athena** supports Delta Lake tables that are registered with **AWS Glue**.
- Scientists analyze the engineered scientific data (Tetra Data) with high-performance computing, data analytics, data lakes, business intelligence, and machine learning. This consumption flows from **Athena**, **Amazon Redshift**, or the table catalog exposed by **Lake Formation**. TDP also provides software integrations using APIs and webhooks.
- Locally, scientists create graphs, author reports, and conduct performance analysis using data packages.
- Data Administrators integrate partner solutions and software as a service (SaaS) products into TDP by provisioning API access keys and/or natively supported integrations such as Snowflake.

