# Lab instrument log acquisition and analytics

Build a data pipeline to automate lab instrument log ingestion, collate metadata, and transform and downstream to FDA compliant electronic laboratory notebooks. Eliminate multiple second-person reviews for a drug discovery and development process. Experience easy search capabilities and REST API support.

AWS Reference Architecture
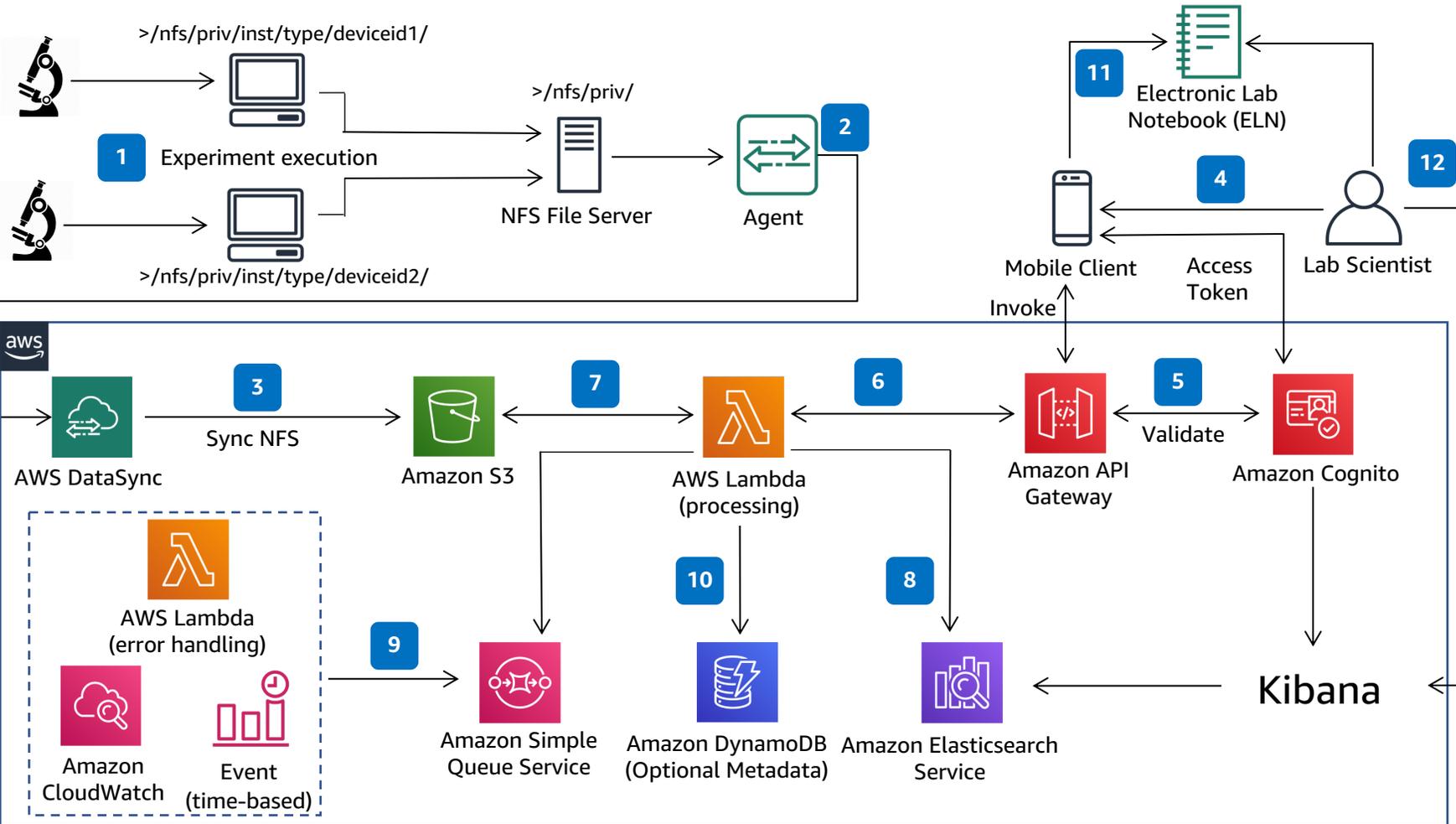
1. All instruments create the output log file in their respective network file system (NFS) shares on the machine attached to the instrument.

2. An agent (virtual machine) is deployed in the user environment, and is used to read data from their NFS storage.

3. An **AWS DataSync** task created in the customer's AWS account syncs data from NFS to a customer-owned **Amazon S3** bucket.

4. A lab scientist invokes a client application to look for an experiment output log file in **Amazon S3** for an associated experiment ID.

5. The client application gets a token from **Amazon Cognito** and invokes the API with the experiment ID and the **Amazon S3** file path (and possibly more IDs as needed from the client application).

6. **Amazon API Gateway** validates the token from **Amazon Cognito**, then calls the processing **AWS Lambda** function (one **Lambda** function for each type of log file), and returns response after the **Lambda** function runs.

7. The **Lambda** function reads the file from **Amazon S3**, and adds the experiment ID as a tag to the **Amazon S3** log file.

8. The **Lambda** function transforms the **S3** file, applies any business rules, converts the log to one or multiple JSON records, and pushes data to **Amazon Elasticsearch Service**.

9. If the processing **Lambda** function fails, it writes the **S3** file info to an error monitoring queue. A **Lambda** function based on a pre-defined timed event can read from the queue to reprocess or to inform support staff.

10. Optionally, a processing **Lambda** function maintains the **S3** file info with metadata (experiment IDs) in an **Amazon DynamoDB** table which can be consumed by other applications.

11. The processing **Lambda** function sends the JSON response to **Amazon API Gateway**, which is posted to an electronic lab notebook (ELN) via Mobile Client.

12. The lab scientist can access data in ELN, and can also discover and query data via Kibana for other experiments.