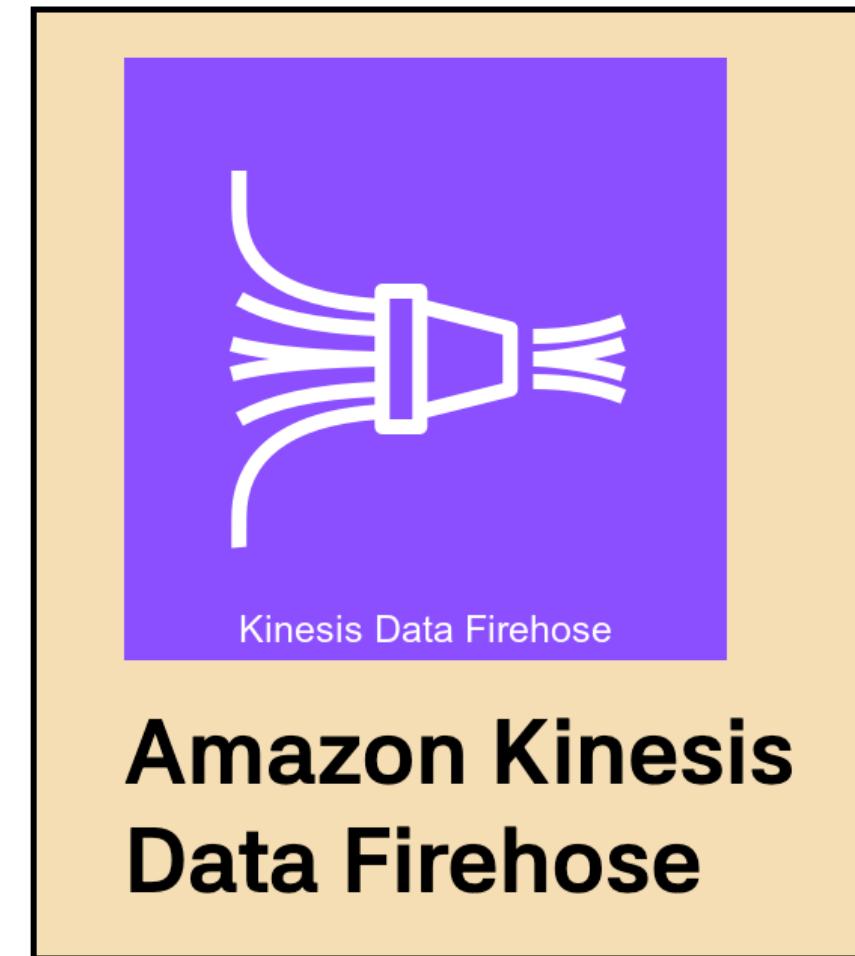




# Amazon Kinesis Data Firehose

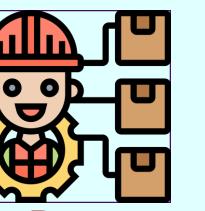
# Table of Contents



- 1 What is Kinesis Data Firehose
- 2 Key Concepts
- 3 Data Flow for S3
- 4 Data Flow for RedShift
- 5 Data Flow for OpenSearch
- 6 Data Flow for Splunk
- 7 Sources - Direct Put
- 8 Destinations
- 9 Transform source data in Amazon Data Firehose

# What is Kinesis Data Firehose

Data Sources



Data  
Producers

1. Generate Streaming Data
2. Send Data to Firehose

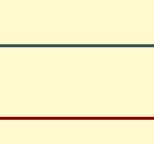
Streaming Data

## Third-Party & Custom Destinations

### Third-Party Services



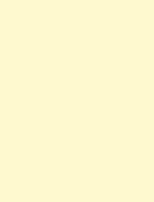
### Amazon Data Firehose



4. Return  
Transformed Data

3. Transform  
Data (Optional)

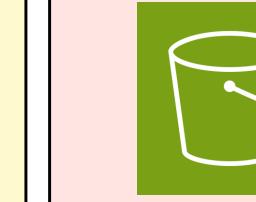
Data  
Transformation



Data  
Transformation

### AWS Service Destinations

5. Deliver to S3



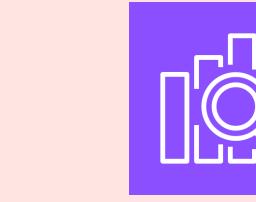
Amazon S3

6. Deliver to Redshift



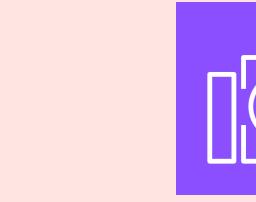
Amazon Redshift

7. Deliver to OpenSearch



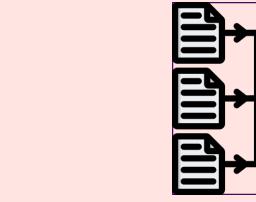
OpenSearch Service

8. Deliver to OpenSearch Serverless



OpenSearch Serverless

9. Deliver to Iceberg Tables

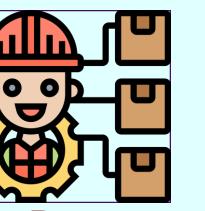


Apache  
Iceberg Tables

1. Fully managed streaming data delivery service: Delivers to Amazon S3, Delivers to Amazon Redshift, Delivers to Amazon OpenSearch Service, Delivers to Amazon OpenSearch Serverless, Delivers to Splunk, Delivers to Apache Iceberg Tables, Delivers to custom HTTP endpoints, Delivers to third-party providers (Datadog, Dynatrace, LogicMonitor, MongoDB, New Relic, Coralogix, Elastic)

# What is Kinesis Data Firehose

Data Sources



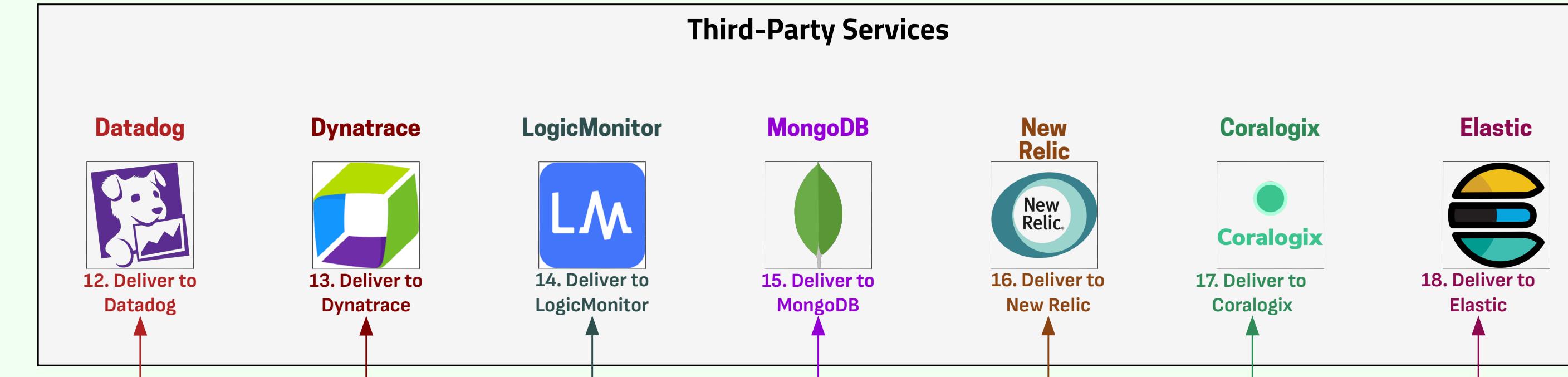
Data  
Producers

1. Generate Streaming Data
2. Send Data to Firehose

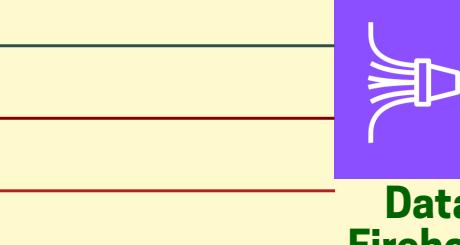
Streaming  
Data

## Third-Party & Custom Destinations

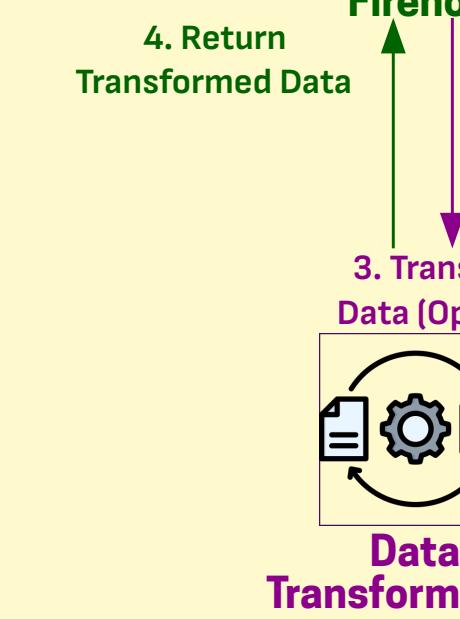
### Third-Party Services



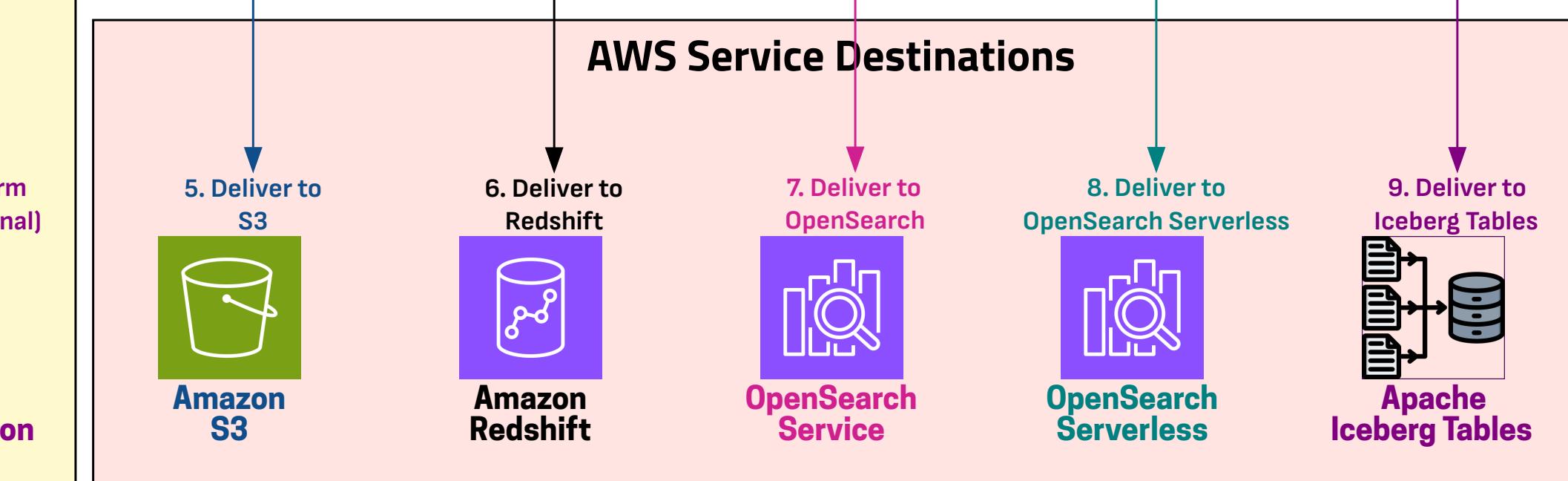
### Amazon Data Firehose



Data  
Firehose



### AWS Service Destinations

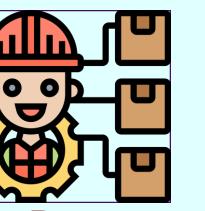


2. 🧩 No need to write applications

⟳ No resource management required

# What is Kinesis Data Firehose

Data Sources



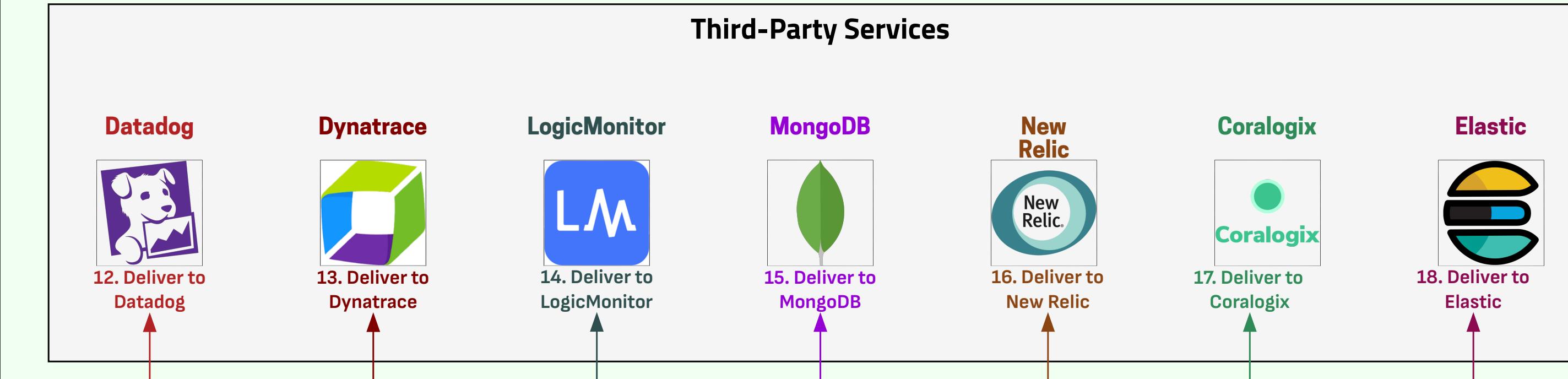
Data  
Producers

1. Generate Streaming Data
2. Send Data to Firehose

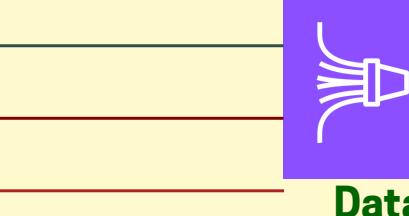


## Third-Party & Custom Destinations

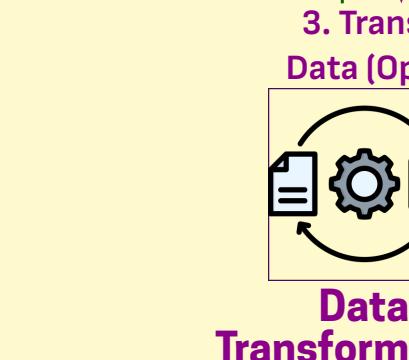
### Third-Party Services



### Amazon Data Firehose



4. Return  
Transformed Data

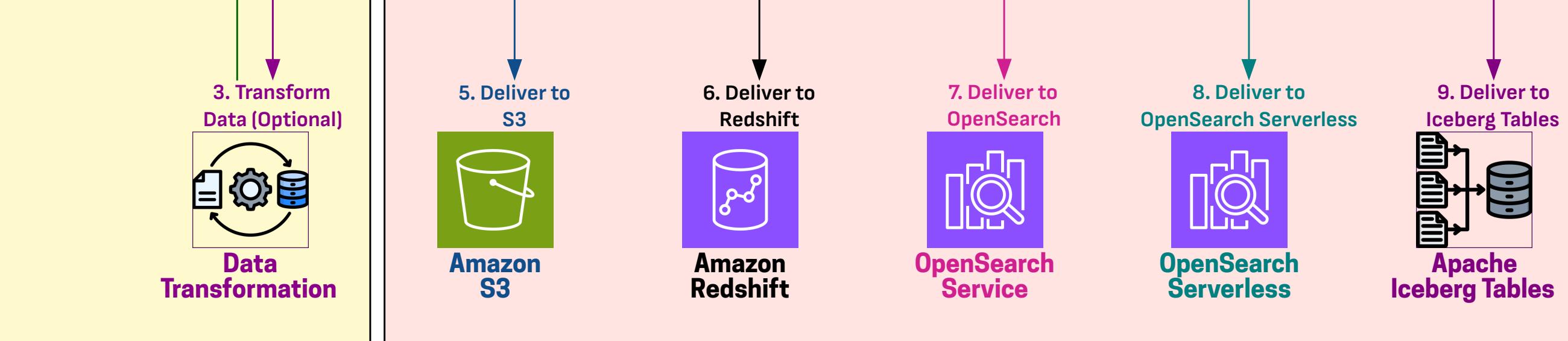


### 3. ⚙ Simple configuration

Configure data producers to send data

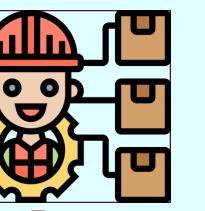
Automatic delivery to specified destination

### AWS Service Destinations



# What is Kinesis Data Firehose

Data Sources



Data  
Producers

1. Generate Streaming Data
2. Send Data to Firehose

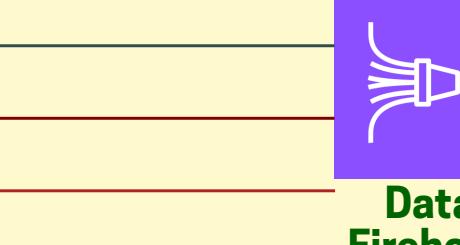
Streaming Data

## Third-Party & Custom Destinations

### Third-Party Services

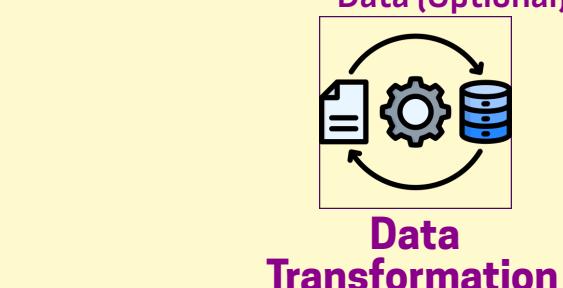


### Amazon Data Firehose



4. Return  
Transformed Data

3. Transform  
Data (Optional)



Data  
Transformation

### AWS Service Destinations

#### AWS Service Destinations

5. Deliver to S3



Amazon  
S3

6. Deliver to Redshift



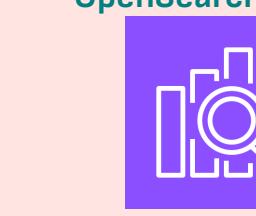
Amazon  
Redshift

7. Deliver to OpenSearch



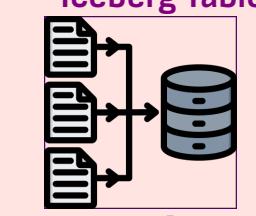
OpenSearch  
Service

8. Deliver to OpenSearch Serverless



OpenSearch  
Serverless

9. Deliver to Iceberg Tables

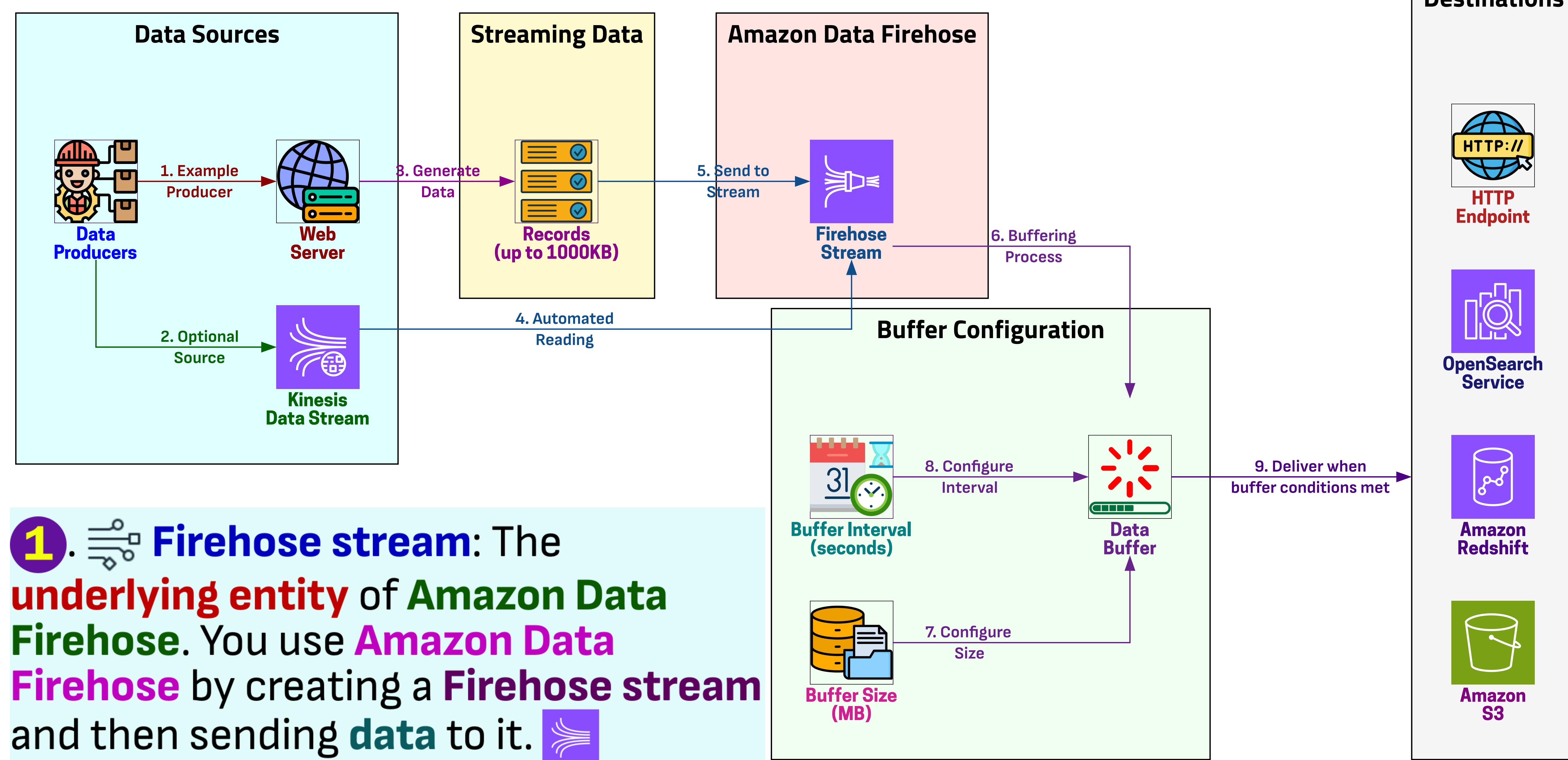


Apache  
Iceberg Tables

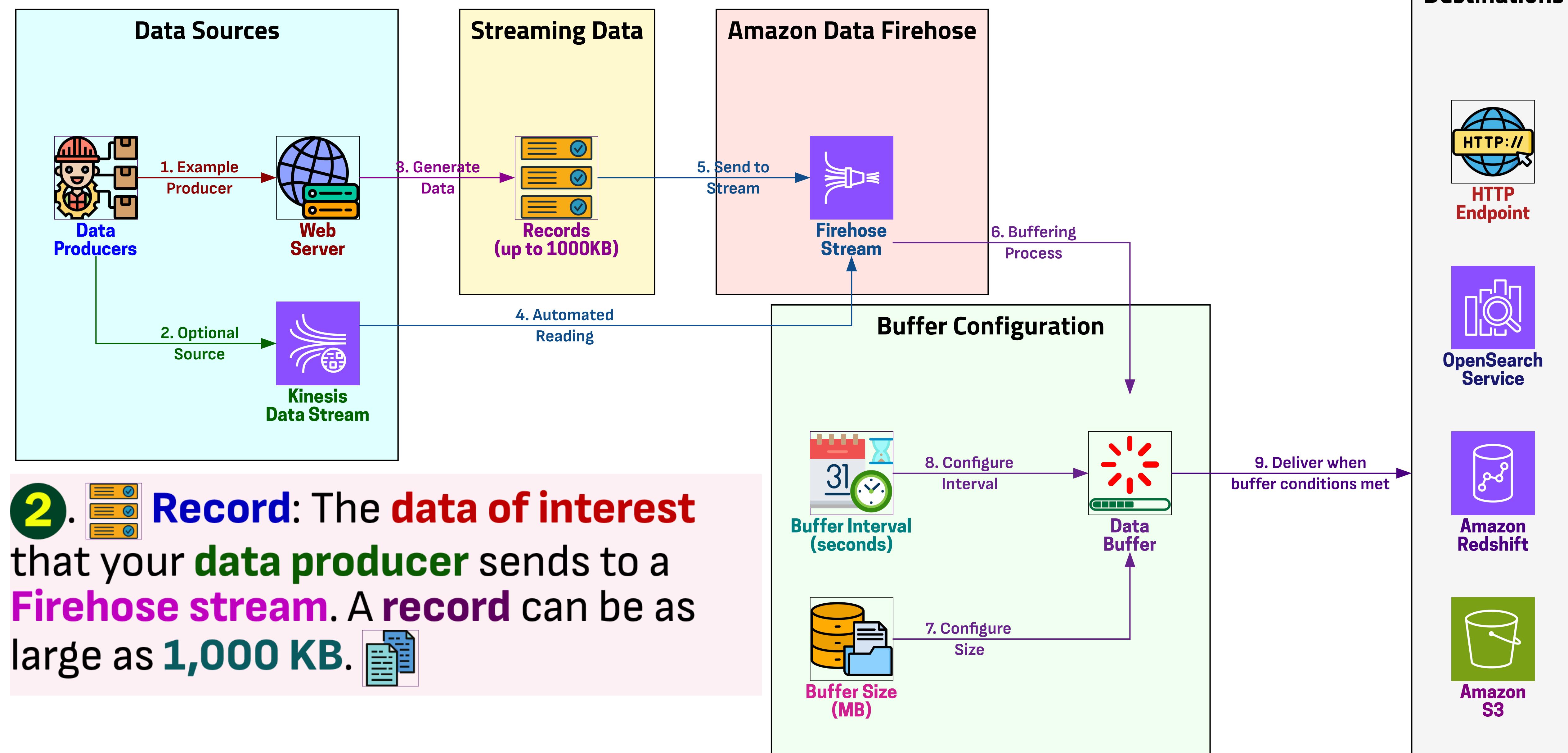
#### 4. Data transformation capabilities

Transform data before delivery

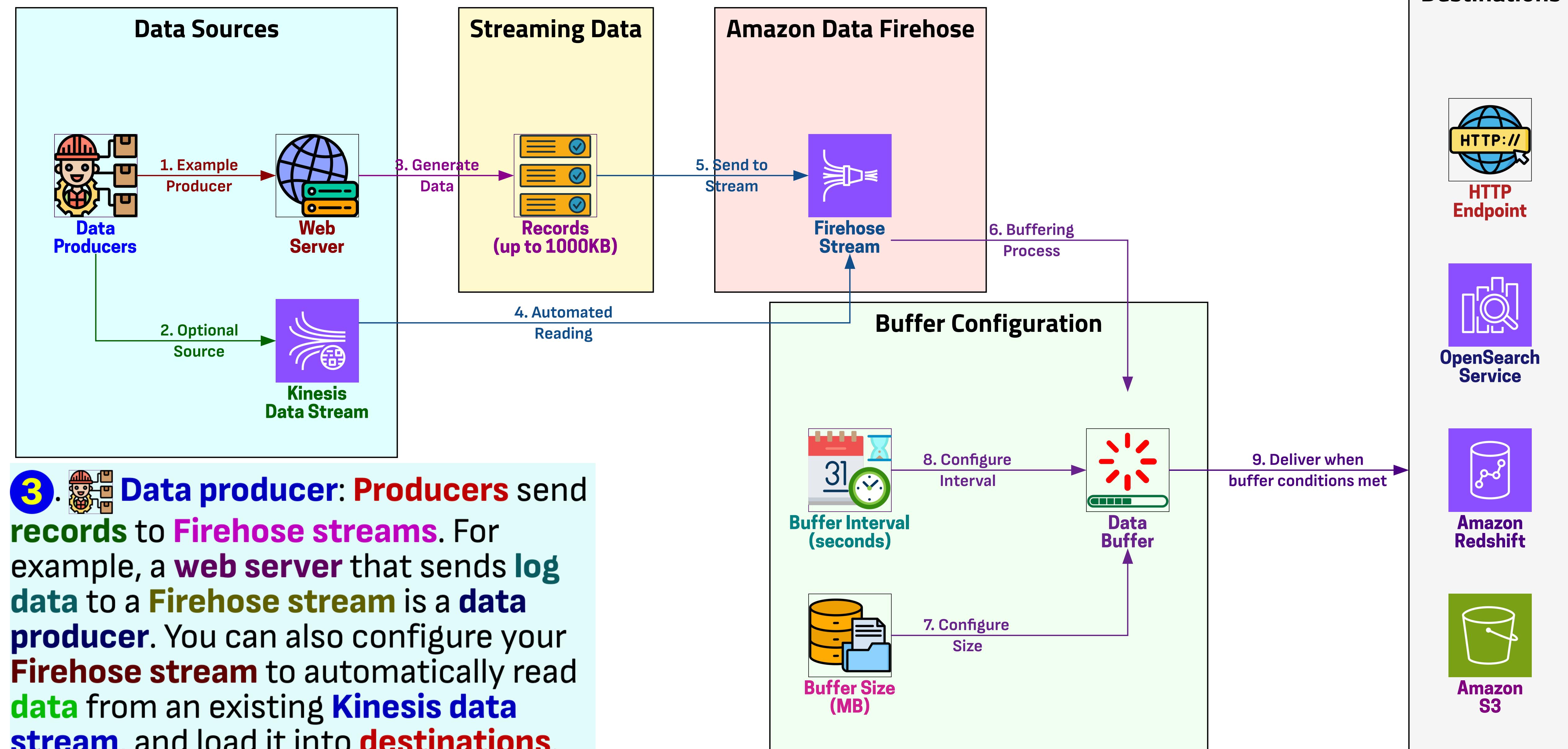
# Key Concepts



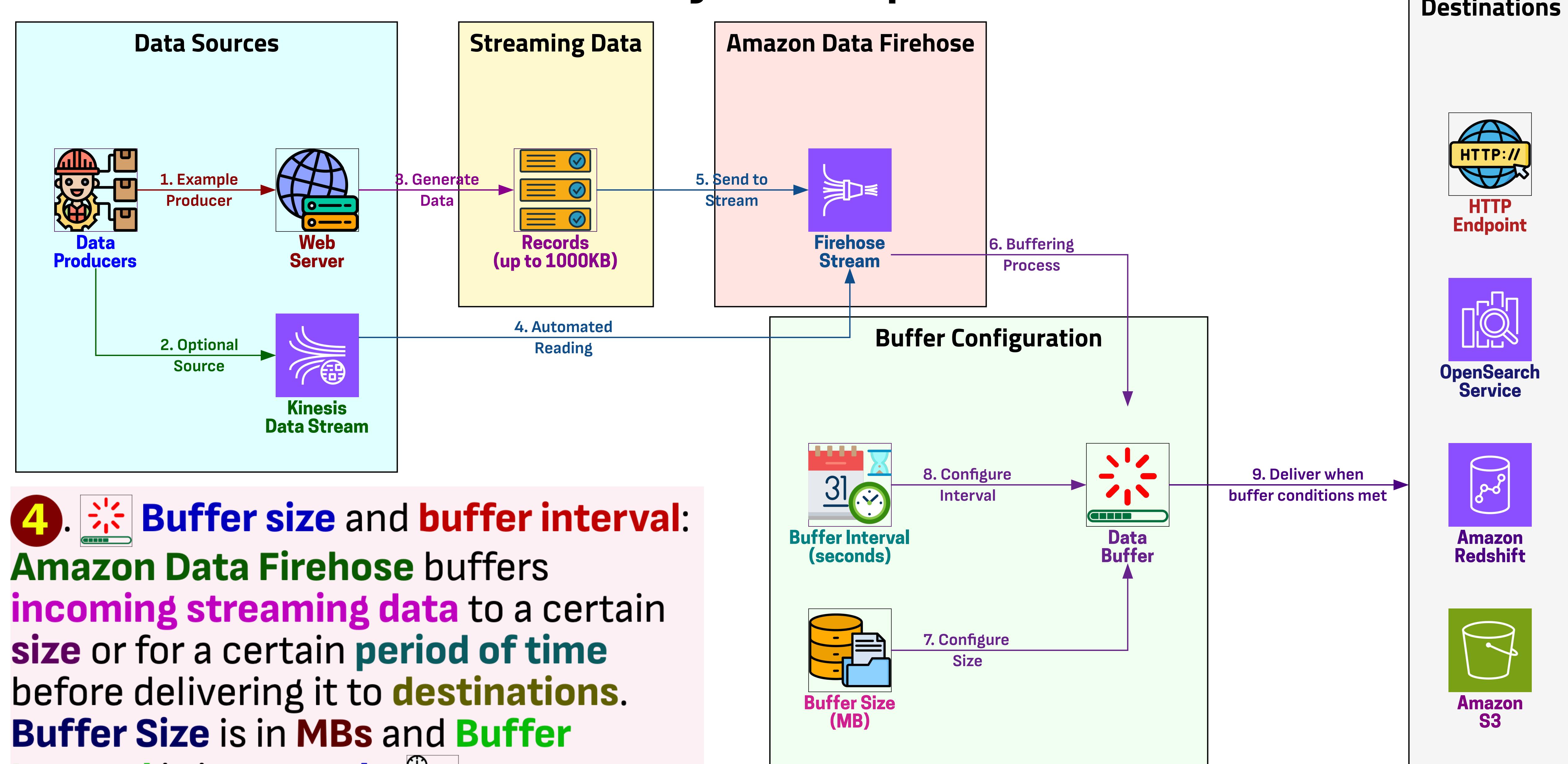
# Key Concepts



# Key Concepts

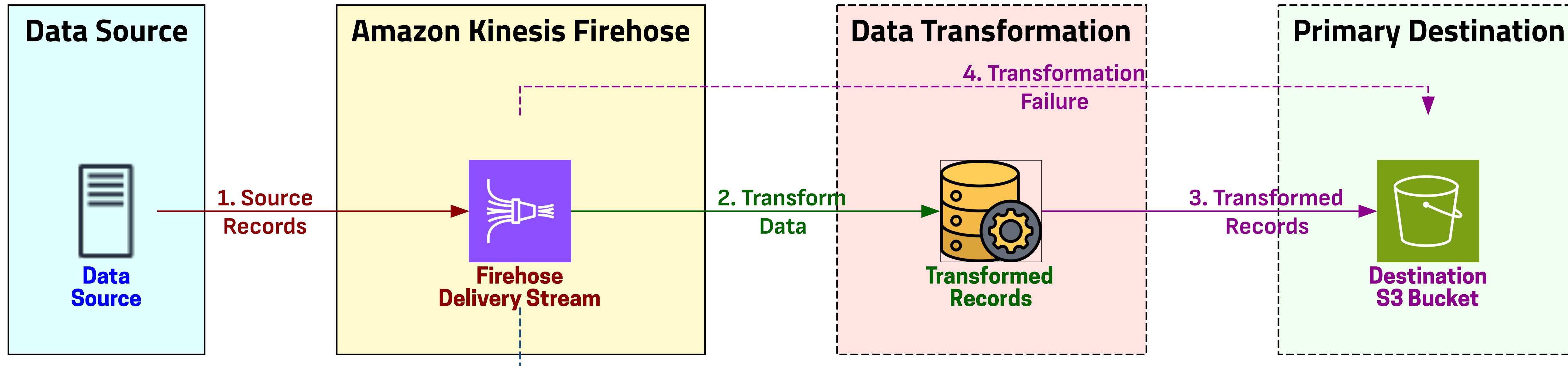


# Key Concepts

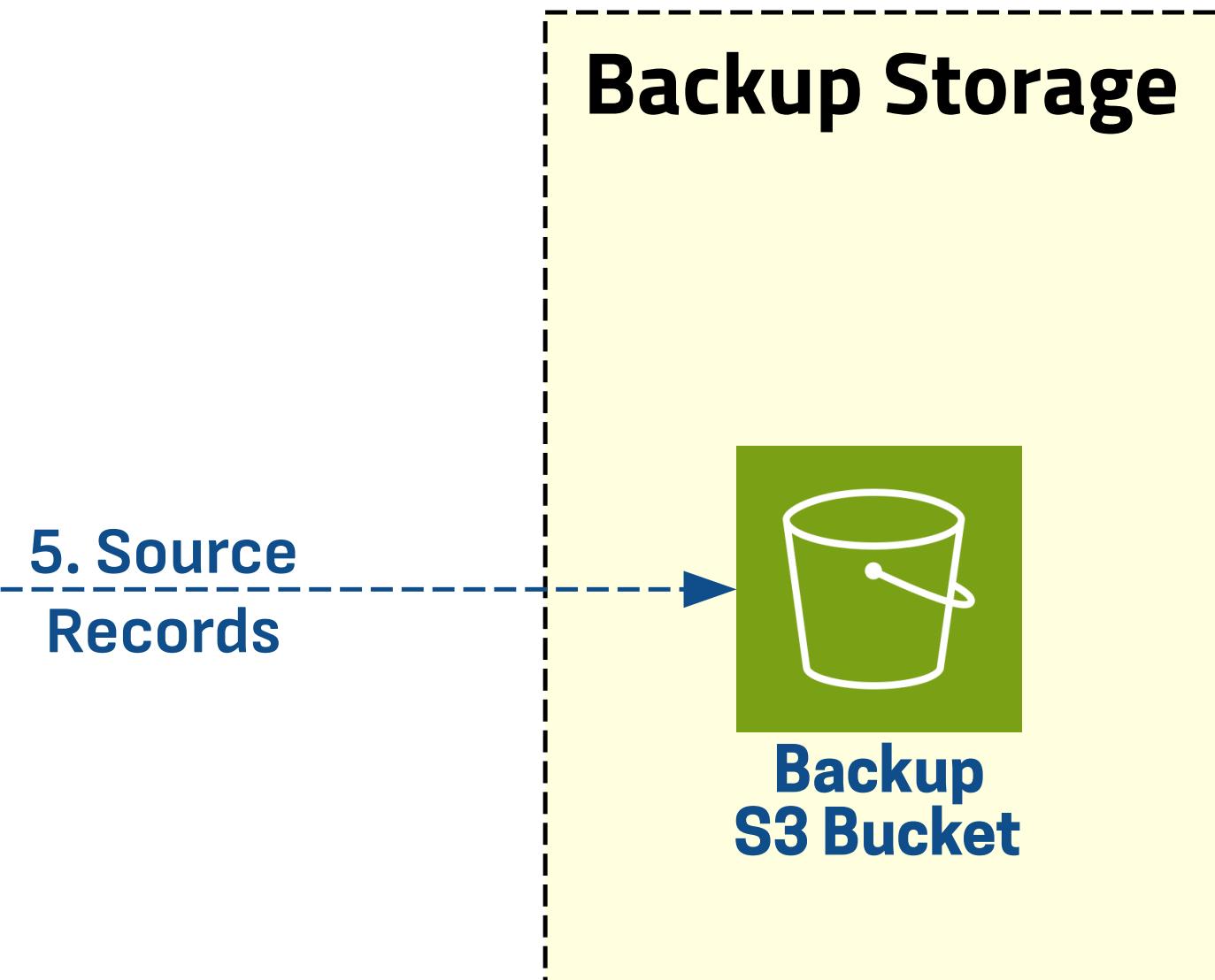


4.  **Buffer size** and **buffer interval**:  
Amazon Data Firehose buffers incoming streaming data to a certain size or for a certain period of time before delivering it to destinations. **Buffer Size** is in **MBs** and **Buffer Interval** is in **seconds**. 

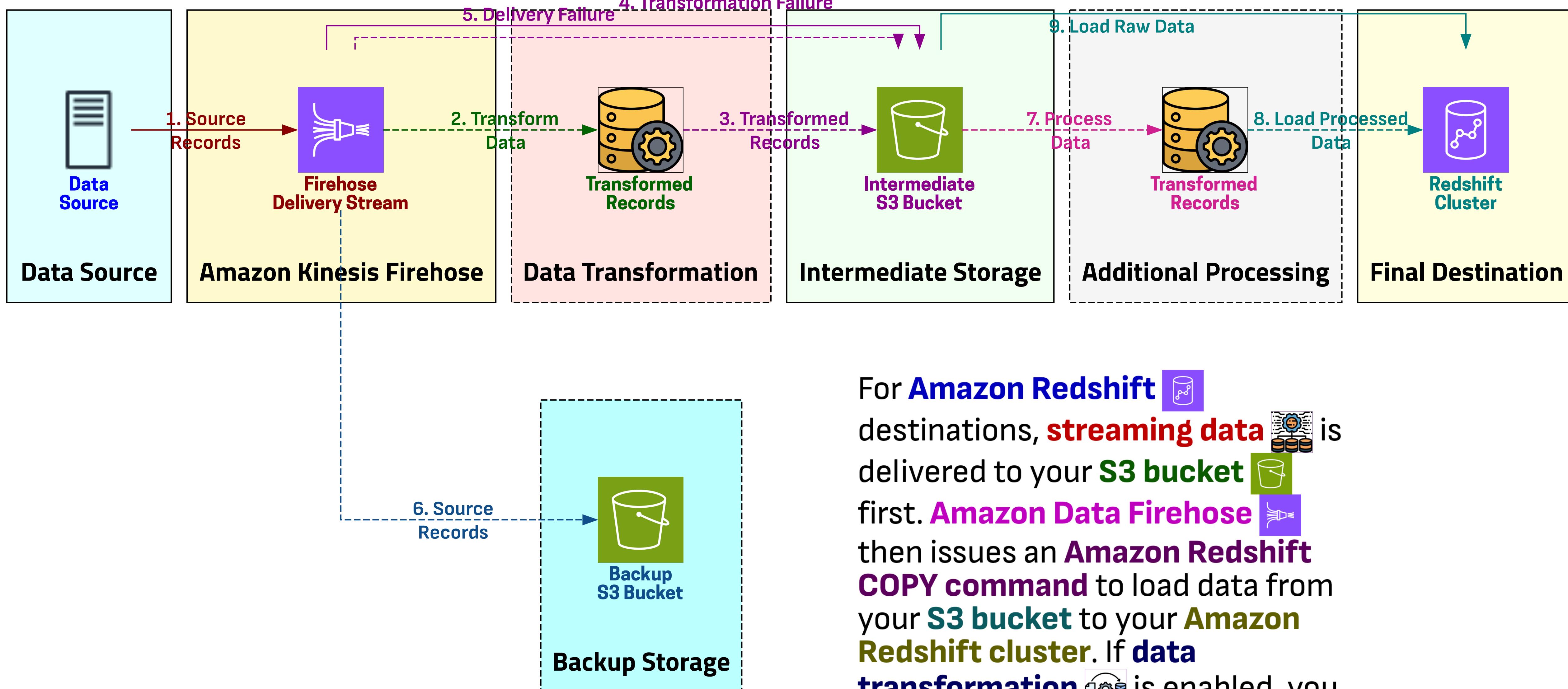
# Data Flow for S3



For **Amazon S3** destinations, **streaming data** is delivered to your **S3 bucket** . If **data transformation** is enabled, you can optionally **back up source data** to another **Amazon S3 bucket** .

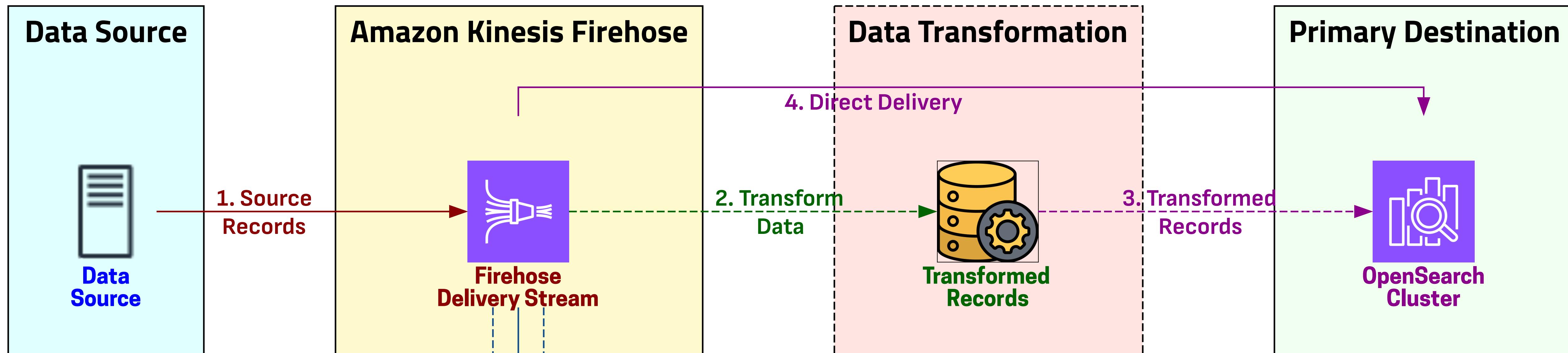


# Data Flow for RedShift

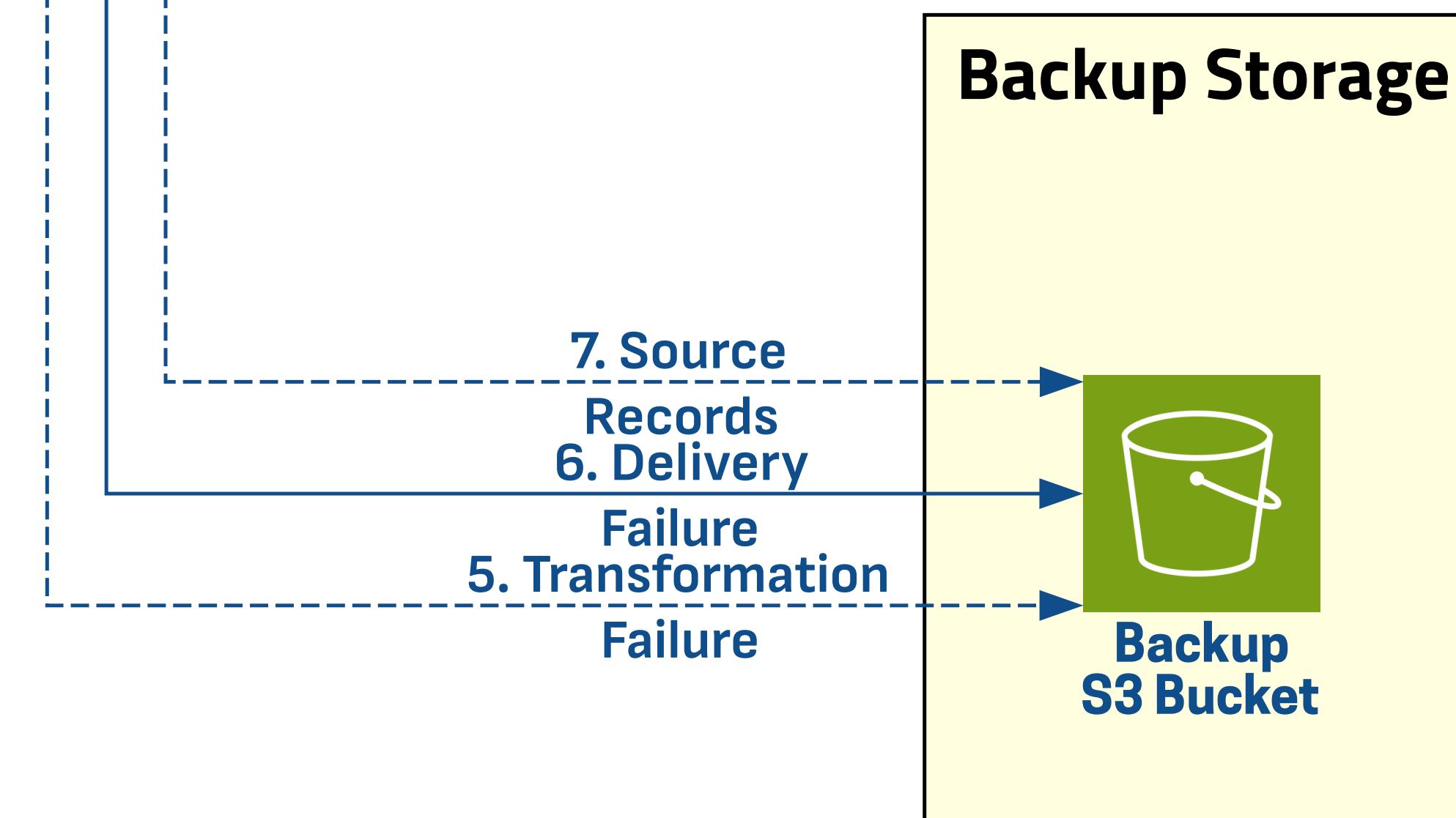


For **Amazon Redshift** destinations, **streaming data** is delivered to your **S3 bucket** first. **Amazon Data Firehose** then issues an **Amazon Redshift COPY command** to load data from your **S3 bucket** to your **Amazon Redshift cluster**. If **data transformation** is enabled, you can optionally **back up source data** to another **Amazon S3 bucket** .

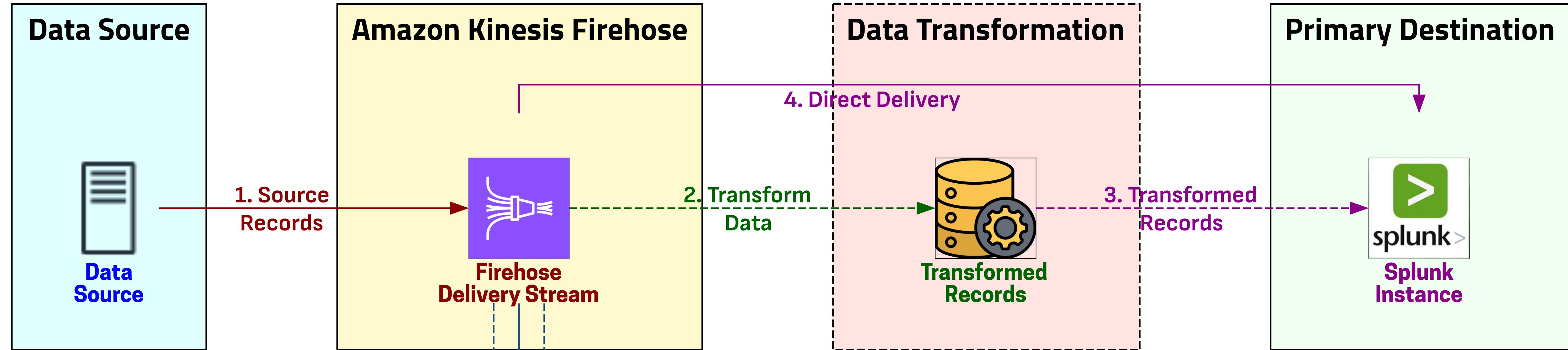
# Data Flow for OpenSearch



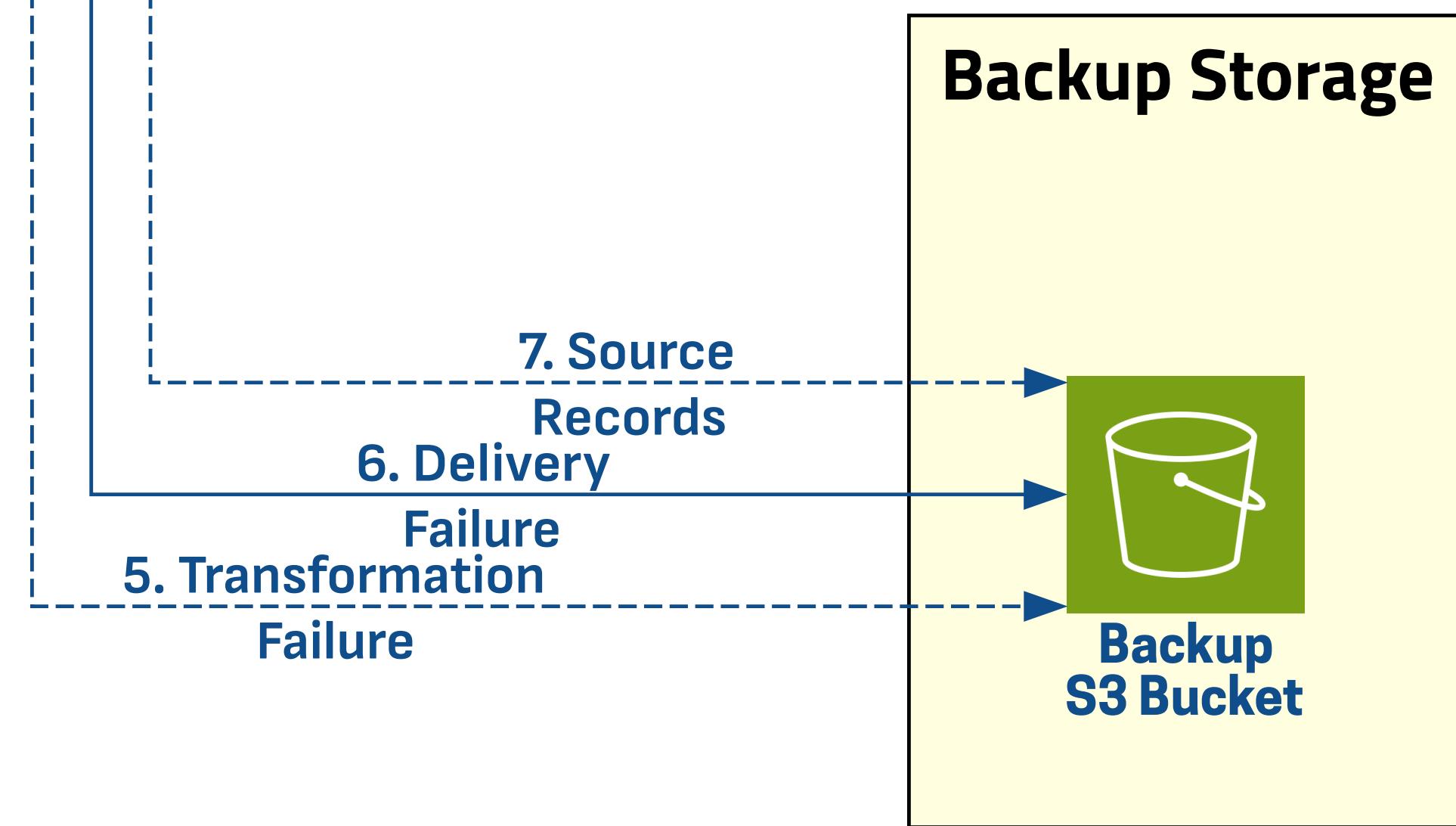
For **OpenSearch Service** destinations, **streaming data** is delivered to your **OpenSearch Service cluster**, and it can optionally be **backed up** to your **S3 bucket** concurrently.



# Data Flow for Splunk



For **Splunk** destinations, **streaming data** is delivered to **Splunk**, and it can optionally be **backed up** to your **S3 bucket** concurrently.



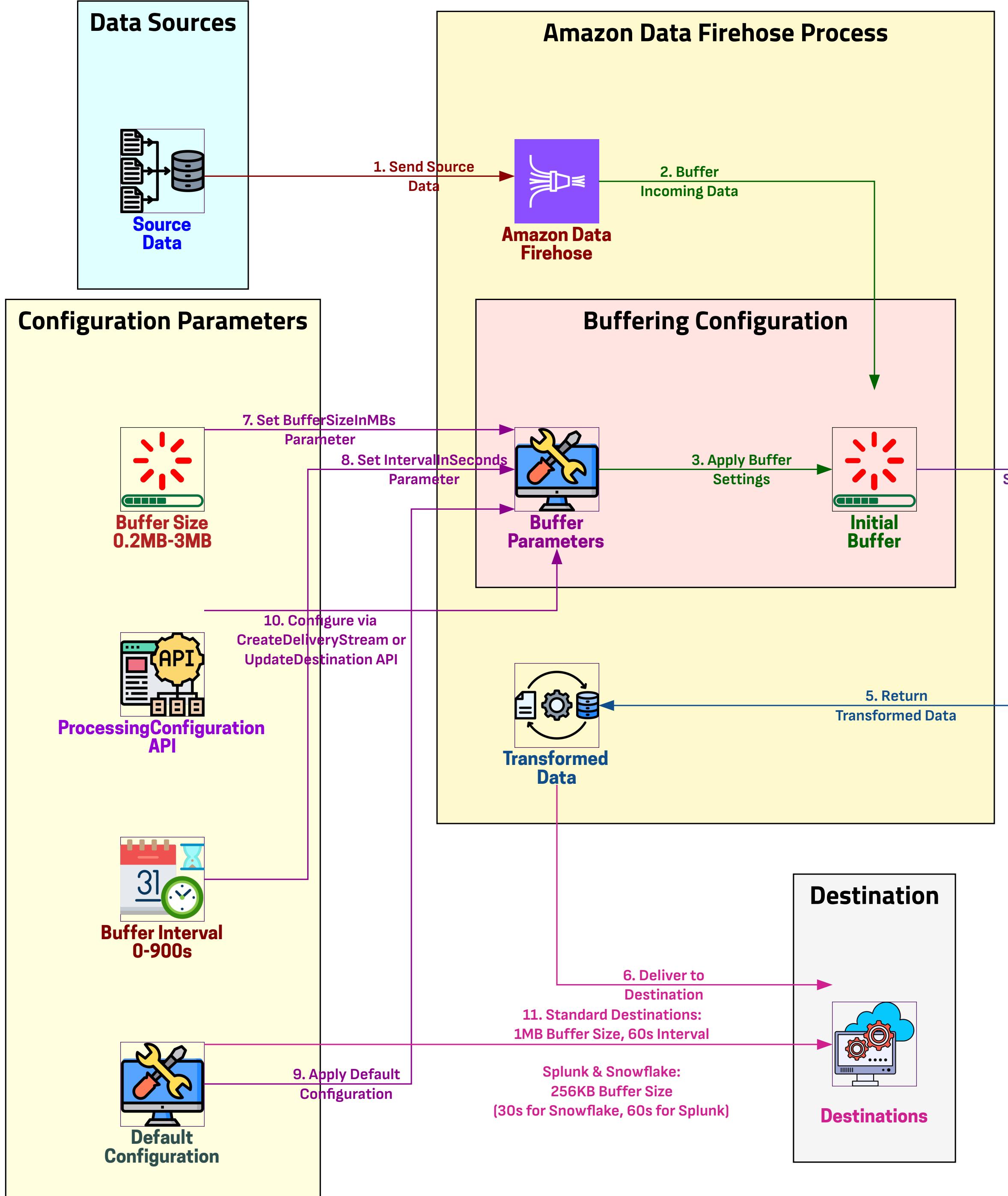
# Sources - Direct Put

1.  AWS SDK	9.  Amazon SNS	17.  Amazon Elasticache Redis SLOWLOG
2.  AWS Lambda	10.  AWS WAF web ACL logs	18.  Kinesis Agent (linux)
3.  AWS CloudWatch Logs	11.  Amazon API Gateway - Access logs	19.  Kinesis Tap (windows)
4.  AWS CloudWatch Events	12.  Amazon Pinpoint	20.  Fluentbit
5.  AWS Cloud Metric Streams	13.  Amazon MSK Broker Logs	21.  Fluentd
6.  AWS IoT	14.  Amazon Route 53 Resolver query logs	22.  Apache Nifi
7.  AWS Eventbridge	15.  AWS Network Firewall Alerts Logs	23.  Snowflake
8.  Amazon Simple Email Service	16.  AWS Network Firewall Flow Logs	

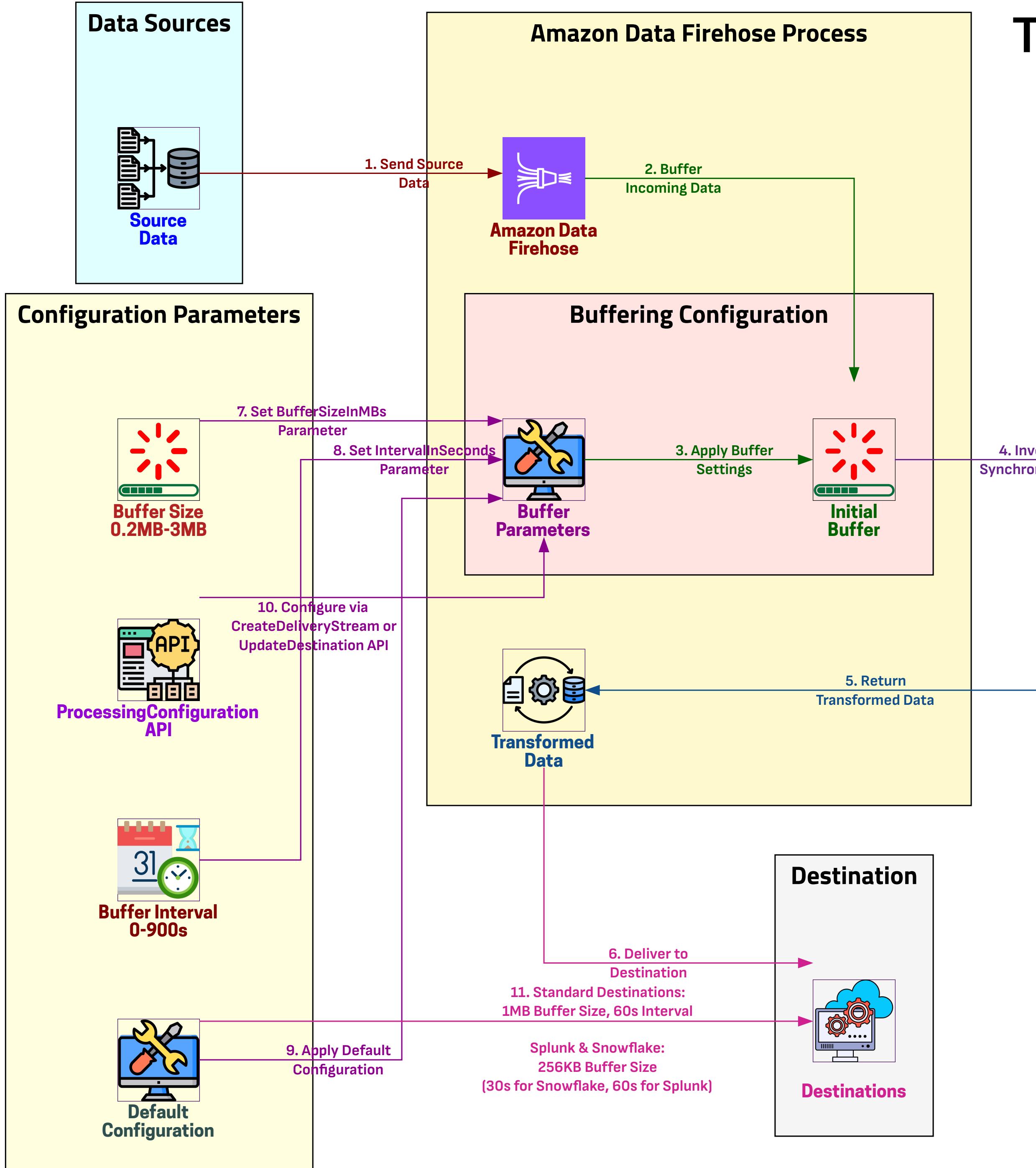
# Destinations

- |                                  |                                 |
|----------------------------------|---------------------------------|
| 1.  Amazon OpenSearch Service    | 11.  Honeycomb                  |
| 2.  Amazon OpenSearch Serverless | 12.  Logic Monitor              |
| 3.  Amazon Redshift              | 13.  Logz.io                    |
| 4.  Amazon S3                    | 14.  MongoDB Cloud              |
| 5.  Apache Iceberg Tables        | 15.  NEW New Relic              |
| 6.  Coralogix                    | 16.  Splunk                     |
| 7.  Datadog                      | 17.  Splunk Observability Cloud |
| 8.  Dynatrace                    | 18.  Sumo Logic                 |
| 9.  Elastic                      | 19.  Snowflake                  |
| 10.  HTTP Endpoint               |                                 |

# Transform source data in Amazon Data Firehose



# Transform source data in Amazon Data Firehose



6. 🔎 Special destination defaults

Splunk: 256 KB

Snowflake: 256 KB

8. ⏰ Default buffering interval

60 seconds for most destinations

7. ⏱ Lambda buffering interval range

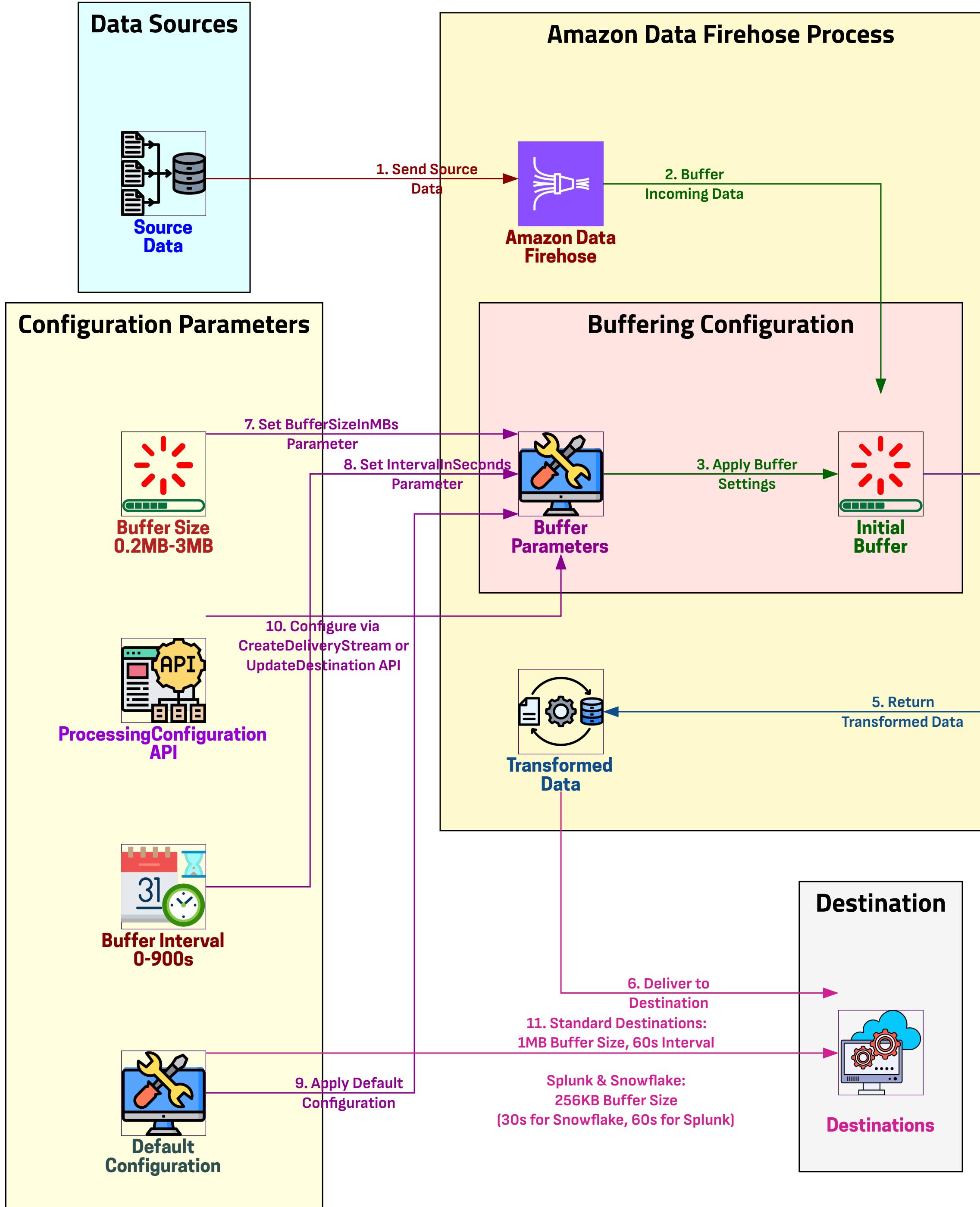
Minimum: 0 seconds

Maximum: 900 seconds

9. ⛅ Snowflake buffering interval

30 seconds default

# Transform source data in Amazon Data Firehose



## 10. ⚒ Adjust buffering parameters

Use API operations



CreateDeliveryStream

UpdateDestination

BufferSizeInMBs

IntervalInSeconds

## 11. 🔍 Lambda invocation

Synchronous invocation mode

Processes each buffered batch

## 12. 📦 Data flow

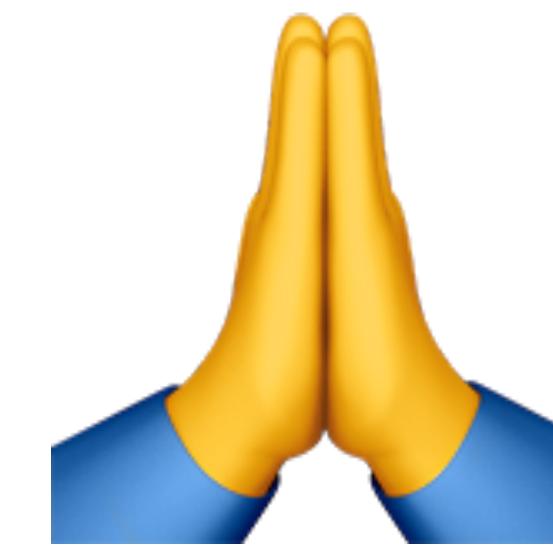
Lambda sends transformed data to Firehose

## 13. 🚚 Destination delivery

Triggered by buffering size

Or triggered by buffering interval

Whichever happens first



**Thanks  
for  
Watching**