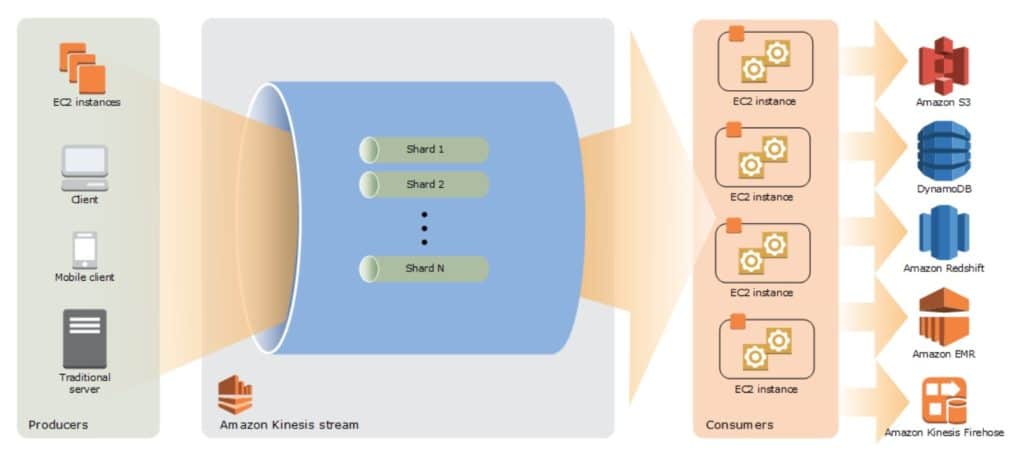
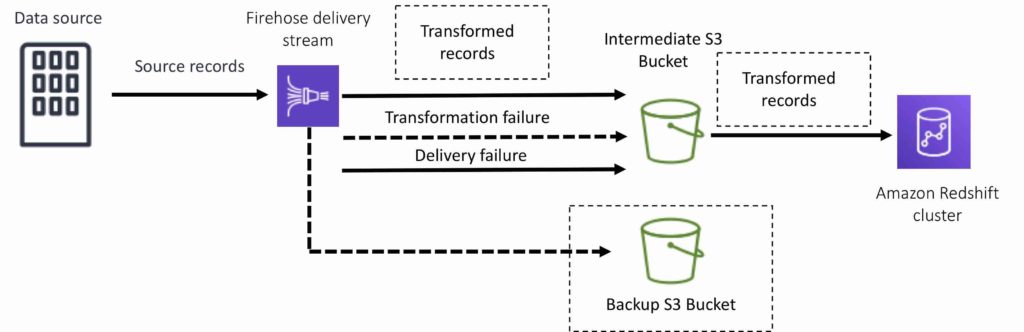
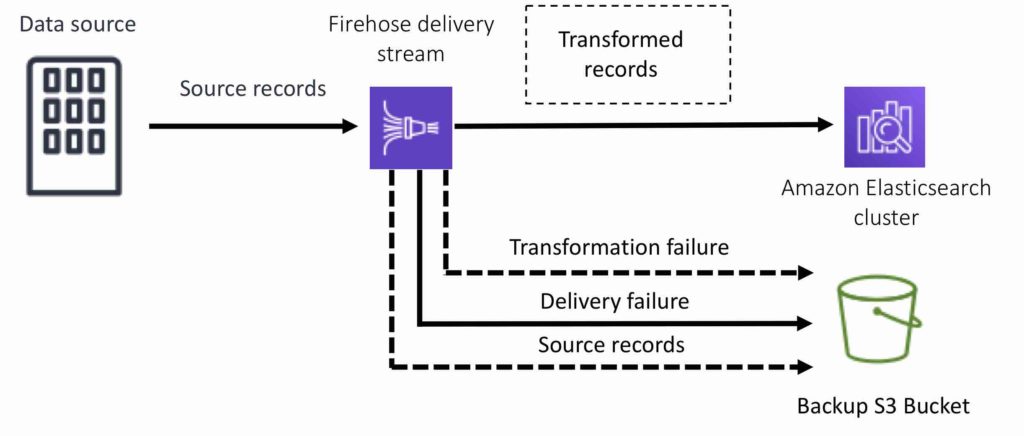
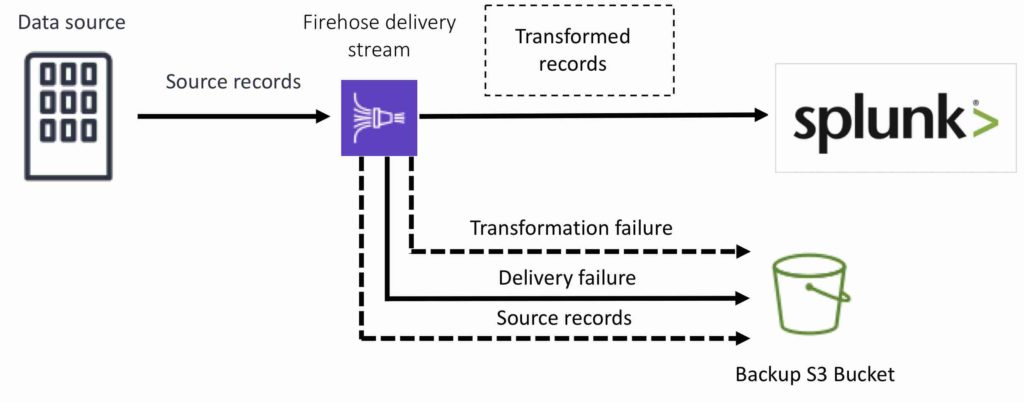
**Kinesis**

* Kinesis makes it easy to collect, process and analyze real-time, streaming data so you can get timely insights and react quickly to new information
* Collection of services for processing streams of various data
* Data is processed in ‘shards’ – with each shard able to ingest 1000 records per second
* There is a default limit of 500 shards, but you can request an increase to unlimited shards
* A record consists of a partition key, sequence number, and data blob(up to 1MB)
* Transient data store – default retention of 24 hours, but can be configured for up to 7 days
* There are four types of Kinesis service and these are detailed below
* **Kinesis Video Streams**
  + Kinesis Video Streams makes it easy to securely stream video from connected devices to AWS for analytics, machine learning(ML) and other processing
  + Durably stores, encrypts, and indexes video data streams, and allows access to data through easy-to-use APIs
  + Producers provide data streams
  + **Stores data for 24 hours by default, up to 7 days**
  + Stires data in shards – 5 transaction per second for reads, up to a max read rate of 2MB per second and 1000 records per second for writes up a max of 1MB per second
  + Consumers receive and process data
  + Can have multiple shards in a stream
  + Supports encryption at rest with server-side encryption(KMS) with a customer master key
* **Kinesis Data Steams**
  + Kinesis Data streams enables you to build custom applications that process or analyze streaming data for specialized needs
  + Kinesis Data Streams enables real-time processing of streaming big data
  + Kinesis Data Streams is useful for rapidly moving data off data producers and then continuously processing the data
  + Kinesis Data Streams **stores data** for later processing by applications(key difference with Firehose with delivers data directly to AWS services)
  + Common use cases include:
    - Accelerated log and data feed intake
    - Real-time metrics and reporting
    - Real-time data analytics
    - Complex stream processing
  + Diagram concepts
    - Producers continually push data to Kinesis Data Streams
    - Consumers process the data in real time
    - Consumers can store their results using an AWS service such as DynamoDB, Redshift, or S3
    - Kinesis Streams applications are consumers that run on EC2 instances
    - Shards are uniquely identified groups or data records in a stream
    - Records are the data units stored in a Kinesis Stream
    - 
    - A producer creates the data that makes up the stream
    - Producers can be used through the following:
      * Kindesis Streams API
      * Kinesis Producer Library(KPL)
      * Kinesis Agent
    - A record is the unit of data stored in a Kinesis data stream
    - A record is composed of a sequence number, partition key, and data blob
    - **By default, records of a stream are accessible for up to 24 hours from the time they are added to the stream(can be raised to 7 days by enabling extended data retention)**
    - A data blob is the data of interest your data producer adds to a data stream
    - The maximum size of a data blob(the data payload before Base64-encoding) within one record is 1 megabyte(MB)
    - A shard is the base throughput unit of Kinesis data stream
    - One shard provides a capacity of 1MB/sec data input and 2MB/sec data output
    - Each shard can support up to 1000 PUT records per second
    - A stream is composed of one or more shards
    - Consumers are the EC2 instances that anlyze the data received from a stream
    - Consumers are known as Kinesis Streams Applications
    - **When the data rate increases, add more shards to increase the size of the stream**
    - Remove shards when the data rate decreases
    - Partition keys are used to group data by shard within a stream
    - Kinesis Streams uses KMS master keys for encryption
    - To read from or write to an encrypted stream the producer and consumer applications must have permissions to access the master key
    - Kinesis Data Streams replicates synchronously across 3 AZs
* **Kinesis Data Firehose**
  + Kinesis Data Firehose is the easiest way to load streaming data into data stores and analytics tools
  + Captures, transforms and loads streaming data
  + Enables near real-time analytics with existing business intelligence tools and dashboards
  + Kinesis Data Streams can be used as the source(s) to Kinesis Data Firehose
  + **You can configure Kinesis Data Firehose to transform your data before delivering it**
  + With Kinesis Data Firehose you don’t need to write an application or manage resources
  + Firehose can batch, compress and encrypt data before loading it
  + Firehose synchronously replicates data across 3 Azs as it is transported to destinations
  + Each delivery stream stores data records for up to 24 hours
  + A source is wher eyour streaming data is continuously generated and captured
  + A delivery stream is the underlying entityt of Kinesis Data Firehose
  + A record is the datat of interest your data producer sends to a delivery stream
  + The maximum size of a record(before Base64-encoding) is 1000 KB
  + A destination is the data store where your data will be delivered
  + Firehose Destinations include
    - S3
    - Redshift
    - Elasticsearch Service
    - Splunk
  + Producers provide data streams
  + No shards, fully automated
  + Can encrypt data with an existing AWS KMS key
  + Server-side encrytpion can be used if Kinesis Streams is used as the data source
  + Firehose can invoke an AWS Lambda function to transform incoming data before delivering it to a destination
  + For S3 destinations, streaming data is delivered to your S3 bucket. If data transofrmation is enabled, you can optionally back up source data to another S3 bucket
  + 
  + For Redshift destinations, streaming data is delivered to your S3 bucket first. Kinesis Data Firehose then issues an Amazon Redshift COPY command to load data from your S3 bucket to your Redshift cluster. If data transformation is enabled, you can optionally back up source data to another S3 bucket
  + 
  + For Elasticsearch destinations, streaming data is delivered to your ES cluster, and it can optoinally be backed up to your S3 bucket concurrently:
  + 
  + For Splunk destinations streaming data is delivered to Splunk, and it can optionally be backed up to your S3 bucket concurrently:
  + 
  + Kinesis Data Analytics
    - Kinesis Data Analytics is the easiest way to process and analyze real-time, streaming data
    - Can use standard SQL queries to process Kinesis data streams
    - Provides real-time analysis
    - Use cases:
      * Generate time-series analytics
      * Feed real-time dashboards
      * Create real-time alerts and notifications
    - Quickly author and run powerful SQL code against streaming sources
    - Can ingest data from Kinesis Streams and Kinesis Firehose
    - Output to S3, Redshift, ElasticSearch and Kinesis Data Streams
    - Sits over Kinesis Data Streams and Kinesis Data Firehose
    - A Kinesis Data Analytics application consists of three components:
      * Input – the streaming source for your application
      * Application code – a series of SQL statements that processi nput and produce output
      * Output- One or more in-appliaction streams to hold immediate results
    - Kinesis Data Analytics supports two types of inputs: streaming data sources and reference data sources:
      * A streaming data source is continuously generated data that is read into your application for processing
      * A reference data source is static data that your application uses to enrich data coming in from streaming sources
    - Can configure destinations to persist the results
    - Supports Kinesis Streams and Kinesis Firehose(S3, RedShift, ElasticSearch) as destinations
    - IAM can be used to provide Kinesis Analytics with permissions to read records from sources and write to desitinations