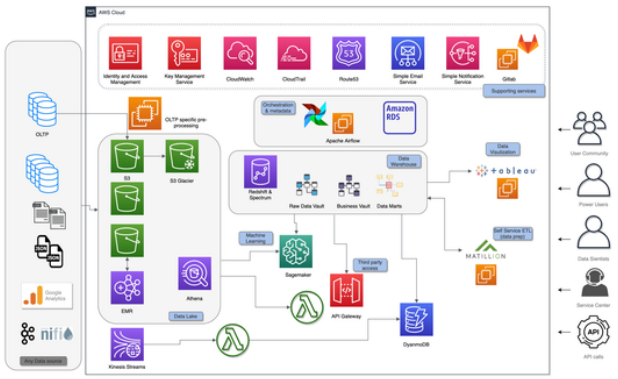
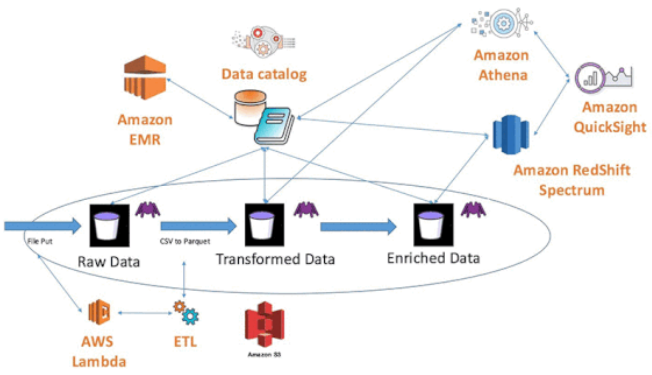
Sample Architecture and Use Case Diagrams

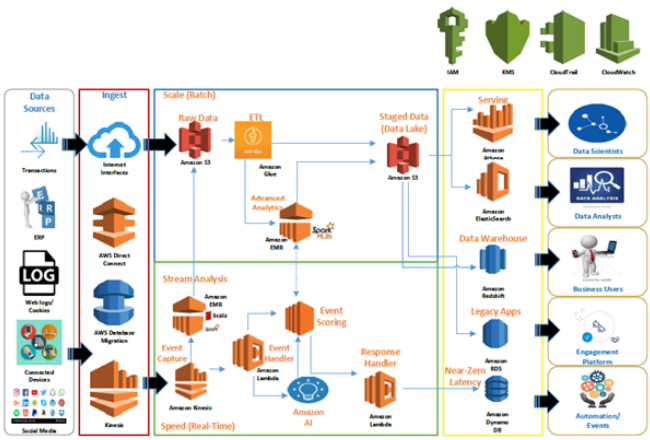
This Big Data Use Cases Link is a great source of looking at various AWS data lake use cases covering:

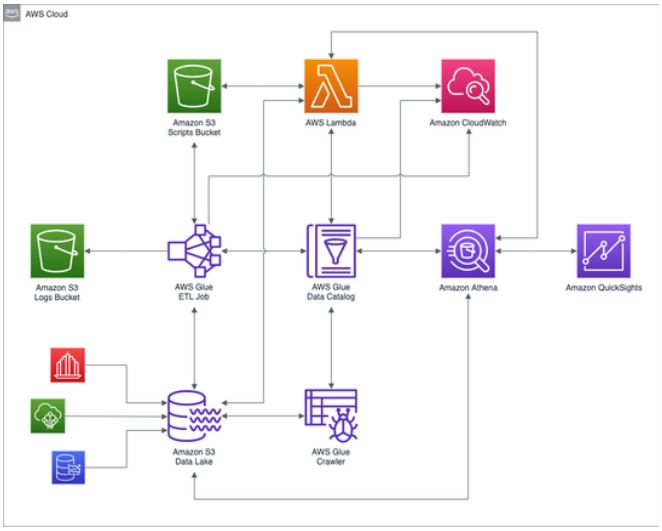
- On-demand big data analytics
- Clickstream analysis
- Event-driven extract, transform and load (ETL)
- Smart Applications
- Data Warehousing

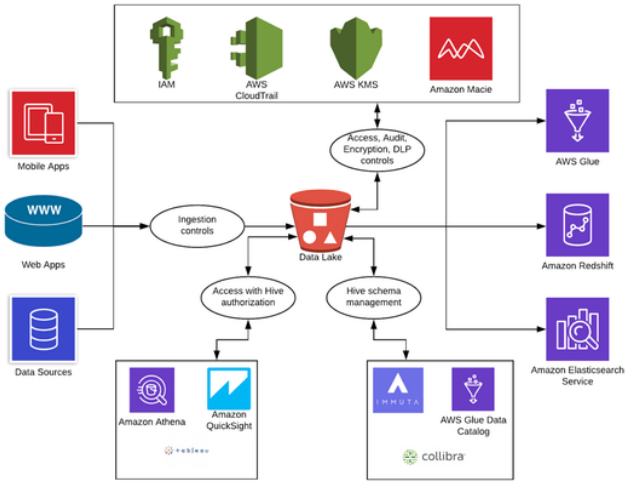
Below are various examples of architecture structures and use case diagrams for data lakes; small, medium and large.

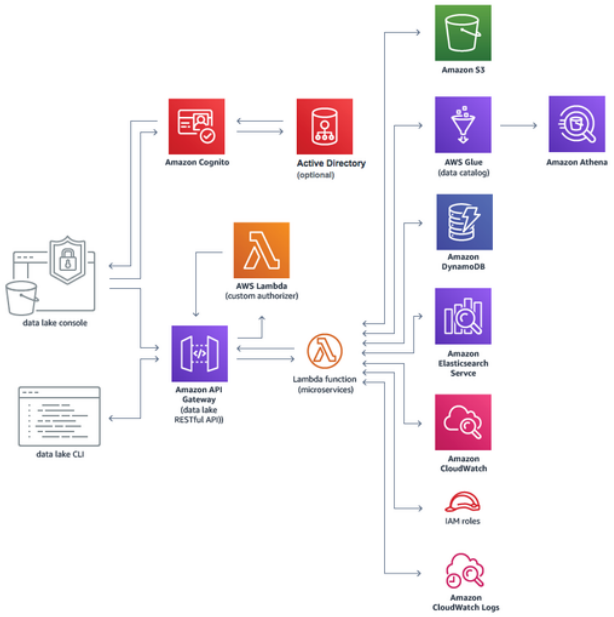


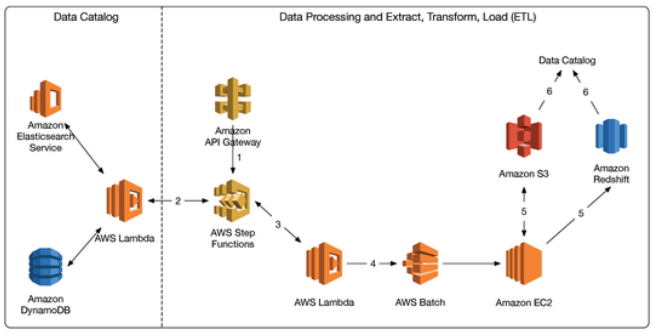


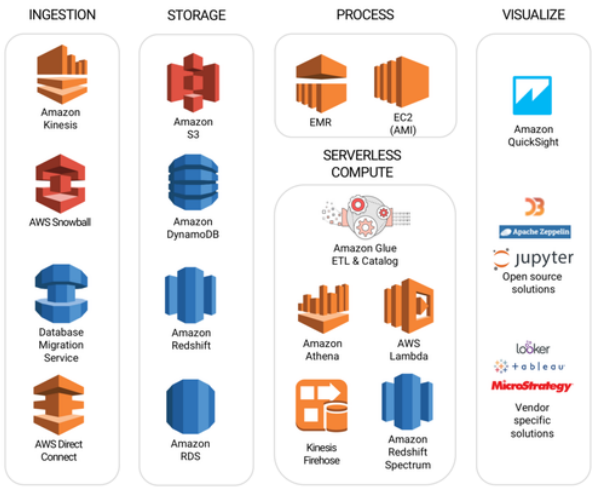




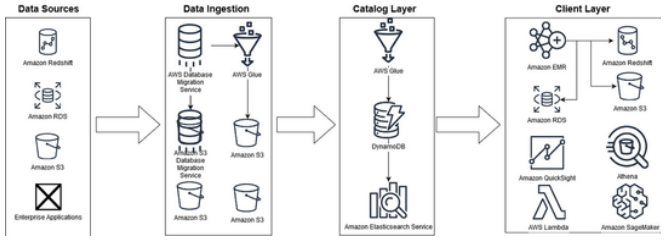


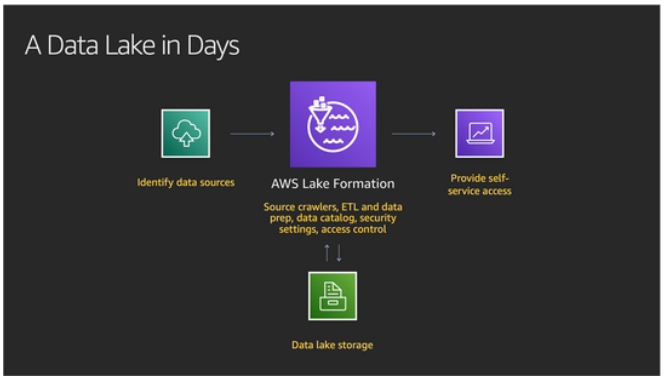

Amazon S3
Amazon Cognito
Active Directory
(optional)
AWS Glue
(data catalog)
Amazon Athena
Amazon DynamoDB
AWS Lambda
(custom authorizer)
data lake console
Amazon Elasticsearch Servce
Amazon API Gateway
(data lake RESTful API)
Lambda function
(microservices)
data lake CLI
Amazon CloudWatch
IAM roles
Amazon CloudWatch Logs

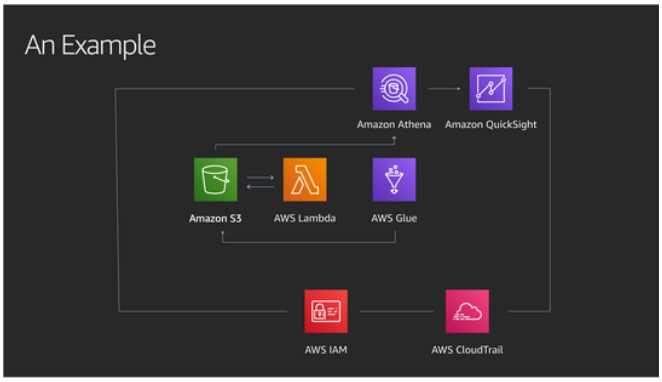

Data Catalog
Data Processing and Extract, Transform, Load (ETL)
Amazon Elasticsearch Service
Amazon API Gateway
Data Catalog
Amazon S3
Amazon Redshift
AWS Lambda
AWS Step Functions
Amazon DynamoDB
AWS Lambda
AWS Batch
Amazon EC2

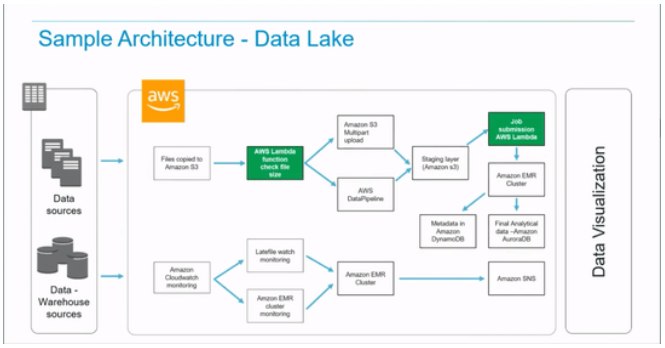

INGESTION
STORAGE
PROCESS
VISUALIZE
Amazon Kinesis
AWS Snowball
Database Migration Service
AWS Direct Connect
Amazon S3
Amazon DynamoDB
Amazon Redshift
Amazon RDS
EMR
EC2 (AMI)
SERVERLESS COMPUTE
Amazon Glue ETL & Catalog
Amazon Athena
AWS Lambda
Kinesis Firehose
Amazon Redshift Spectrum
Amazon QuickSight
Apache Zeppelin
jupyter
Open source solutions
looker
tableau
MicroStrategy
Vendor specific solutions

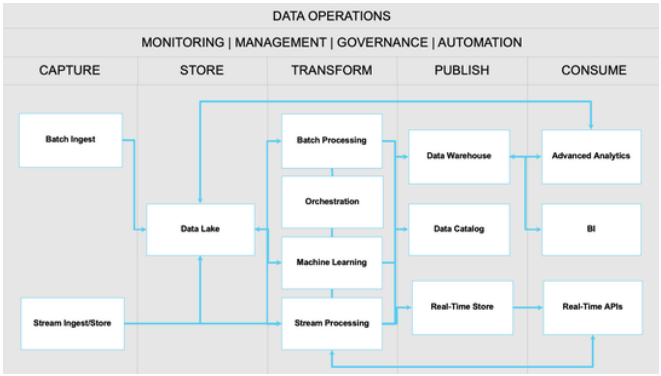

Data Sources
Collect & Process
Data Lake
Catalog & Load
Destination
Tools

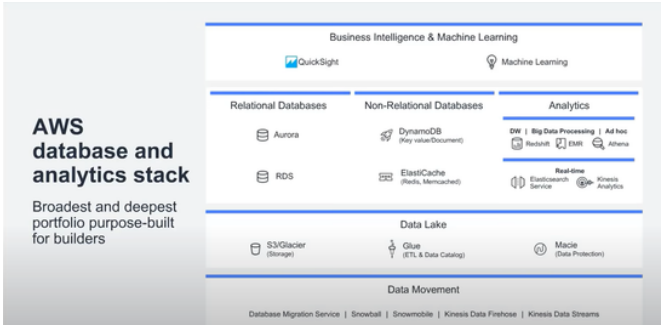

Data Sources
Amazon Redshift
Amazon RDS
Amazon S3
Enterprise Applications
Data Ingestion
AWS Database Migration Service
AWS Glue
Amazon S3
Catalog Layer
AWS Glue
DynamoDB
Amazon Elasticsearch Service
Client Layer
Amazon EMR
Amazon Redshift
Amazon RDS
Amazon S3
Amazon QuickSight
Athena
AWS Lambda
Amazon SageMaker


A Data Lake in Days
Identify data sources
AWS Lake Formation
Source crawlers, ETL and data prep, data catalog, security settings, access control
Provide self-service access
Data lake storage


An Example
Amazon Athena
Amazon QuickSight
Amazon S3
AWS Lambda
AWS Glue
AWS IAM
AWS CloudTrail


Sample Architecture - Data Lake
aws
Data sources
Data - Warehouse sources
Files copied to Amazon S3
AWS Lambda function check file size
Amazon S3 Multipart upload
AWS DataPipeline
Staging layer (Amazon s3)
Job submission AWS Lambda
Amazon EMR Cluster
Metadata in Amazon DynamoDB
Final Analytical data –Amazon AuroraDB
Amazon Cloudwatch monitoring
Lateflix watch monitoring
Amazon EMR cluster monitoring
Amazon EMR Cluster
Amazon SNS
Data Visualization


DATA OPERATIONS
MONITORING | MANAGEMENT | GOVERNANCE | AUTOMATION
CAPTURE
STORE
TRANSFORM
PUBLISH
CONSUME
Batch Ingest
Data Lake
Stream Ingest/Store
Batch Processing
Orchestration
Machine Learning
Stream Processing
Data Warehouse
Data Catalog
Real-Time Store
Advanced Analytics
BI
Real-Time APIs


AWS database and analytics stack
Broadest and deepest portfolio purpose-built for builders
Business Intelligence & Machine Learning
QuickSight
Machine Learning
Relational Databases
Aurora
RDS
Non-Relational Databases
DynamoDB
(Key value/Document)
ElastiCache
(Redis, Memcached)
Analytics
DW | Big Data Processing | Ad hoc
Redshift
EMR
Athena
Real-time
Elasticsearch Service
Kinesis Analytics
Data Lake
S3/Glacier
(Storage)
Glue
(ETL & Data Catalog)
Macie
(Data Protection)
Data Movement
Database Migration Service | Snowball | Snowmobile | Kinesis Data Firehose | Kinesis Data Streams