

D&A TRANSFORMATION TALES

Simplifying Historical Data Management with **Data Lakes**



BACKGROUND

Our client is a leading edutech company. They approached Aspire to build a data lake to streamline and maintain their historic data from over 60 years within a specific time frame. They realized they were facing a multitude of challenges with maintaining their historical data and current data in check with their legacy systems. The challenges they faced included:

- Inability to match and link student profile data with assessment data
- Implementing matching algorithms for confidential rules
- Redundant data from unclear target sources
- Inefficient current and history data management
- Corrupt, mis-formatted or incomplete data received from multiple sources



BACKGROUND

The requirement of the client included a solution that can help them:

- Enrich and maintain datasets
- Manage a single version of truth for reporting requirements
- Implement data catalog and organize data in logical groups that represent domains of information such as;
 - Data security/access controls
 - Data extraction/view showing datasets based on access control list
 - Schedule and monitor work load
 - Email notification of status to specific stakeholders

To overcome these challenges our data and analytics experts suggested streamlining their data challenges with Data Lakes that would enable the client to:

- Create and maintain the customer data in Iceberg table formats in data lake
- Enhance query performance while querying iceberg table data through Athena from Data Lake
- Include tags for meta data for grouping domain as well as for access controls
- Manage AWS IAM roles and permissions

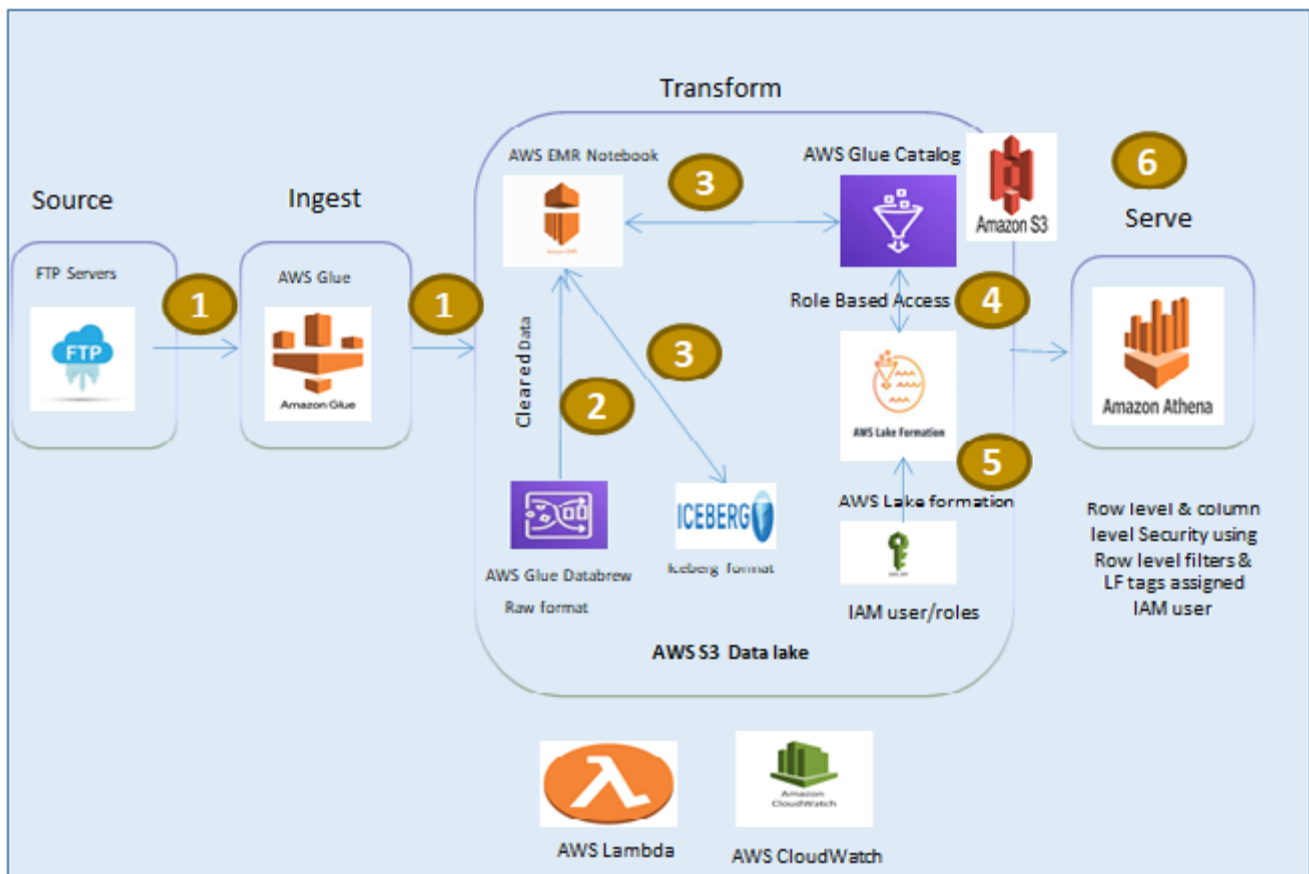




THE SOLUTION

After taking the client's needs and challenges into consideration, Aspire's Data and Analytics team curated a well-crafted plan to build and deploy Data Lake that would help the client process, store, clean, share, analyze, model and monetize their datasets using BI and ML solutions. The solution involved:

- Ingestion of all data sources using ETL glue Jobs
- Storage of ingested Data in the S3 bucket as and when extracted from source
- Creation of Glue Databrew Recipes to match profile algorithm
- Saving cleaned data output in S3 bucket
- Creation of required IAM roles
- Registration of S3 iceberg table location for Lake Formation
- Using EMR Spark Engine Database and Iceberg tables for cleansing files and storing it in S3 location
- Creation of Glue Data catalog based on underlined tables metadata
- Creating LF tags & Data filters for Lake formation to be attached to the underlying Glue catalog Iceberg Database/tables/ columns for grouping domain information & data classification
- Attaching IAM users with selected Glue catalog iceberg tables with respective LF Tags/data filters to restrict data access control
- Using Athena query editor to connect & query Iceberg tables with access control





TECHNOLOGY SNAPSHOT

- Enterprise Cloud Platform – AWS Cloud Platform,
- AWS Glue job & catalog
- AWS EMR
- AWS S3
- AWS Lambda
- AWS Cloudwatch
- AWS Athena
- AWS Glue Databrew
- AWS Data Lake formation



BUSINESS BENEFITS

- Enhanced data quality and consistency.
- Data Lake formation helps to classify data and implement role-based access control
- Amazon Athena implemented to query the various format S3 storage data with access control
- Row and cell-level security implemented to protect sensitive data like Personal Identifiable Information (PII)
- Table definition, physical location and business relevant attributes are stored as central repository and grouped by business



```
override fun onPageSelected(position: Int) {
    selectedTab = position
}
})
mTabLayout.getTabAt(selectedTab)
mTabLayout.tabMode = TabLayout.MODE_FIXED
mTabLayout.tabGravity = TabLayout.GRAVITY_FILL
mTabLayout.setupWithViewPager(mViewPager)

private fun displayViewPager(packageResult: ArrayList<Package>, bundleResult: Bundle,
mViewPager.adapter = PackagePagerAdapter(
    supportFragmentManager,
    packageResult,
    bundleResult)
Utility.tabsFontChanges(context=this, packageName)

private fun initToolbar() {
    Activity
```



Aspire Systems is a global technology services firm serving as a trusted technology partner for our customers. We work with some of the world's most innovative enterprises and independent software vendors, helping them leverage technology and outsourcing in our specific areas of expertise. Our core philosophy of "Attention. Always." communicates our belief in lavishing care and attention on our customer and employees.

For more info contact: info@aspire.com or visit www.aspiresys.com

USA
+ 1 630 368 0970

SINGAPORE
+65 3163 3050

INDIA
+91 44 6740 4000

BELGIUM
+ 32 3 204 1942

NETHERLANDS
+ 31 (0)30 800 92 16

POLAND
+48 58 732 77 71

MEXICO
+52 222 980 0115