Practical Guide to delivering sensitive data in AWS

Included in this guide are vetted ways to set up a data delivery solution in AWS under the known constraints within Koch Industries today but still satisfying sensitive data compliance/risk requirements.

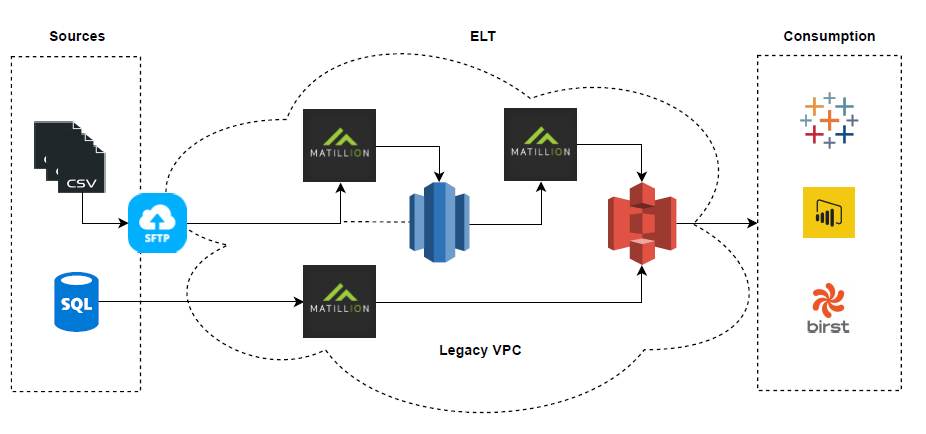
**All in Legacy/IAM User – Access Keys**This solution is delivered with all components hosted in Legacy.

Benefits:

* Completely accessible from on-prem
* Easy Integration with on-prem and other legacy applications

Overview

* Encryption enforced at all storage locations
* SSL enforced for all data in transit
* Flat files delivered via SFTP Gateway or Matillion to S3
* Data protected in S3 Buckets by bucket policy
* Matillion/Redshift access to secured S3 bucket via IAM User Access Key(s)
* Redshift database user access granted for consumption today; Transition to Federated Identity later



Details

| Step | Notes | Resource Required (cited resource used) | Reference/Example |
| --- | --- | --- | --- |
| IAM User is created | This provides a way to lock down access to S3 buckets within the shared IAM accounts in Legacy. Requires a request to the AIS Architects, which is vetted then by the BizAspps Cloud ASA, then approved by the Cloud Champion, and finally created by AIS, who provides the Access Key for the new user. | -Cloud ASA (Tracy Kline)  -Cloud Champion (Chris Bauer, Sandesh Shetty )  -Koch Cloud Team/AIS (Gary Powers) | arn:aws:iam::106904787262:user/kbs-as-svc-user |
| S3 bucket creation | Dedicated S3 bucket should be created and must be locked down via bucket policy to the least privileges necessary for solution to work and remain in compliance. Anyone can create a bucket and bucket policy. Recommendation is to deny all access except to the IAM User created previously and to users who need console or programmatic access. | N/A  (However, help can be obtained from David Banks, Mike Worthington, Zach Slavin, or Patrick Purviance) | Secured S3 Bucket example  Bucket Name: [kbs-analytics-solutions-secuirty-poc-dev](https://s3.console.aws.amazon.com/s3/)  Bucket Policy:  {  "Version": "2012-10-17",  "Statement": [  {  "Sid": "userID definitions: AROAJBQUBQFKWMQ3XSZXY = KBS\_DevOps\_Engineer\_NonProd; AIDAJP5UORHEZ32OYJP6Y = KBS\_DevOps\_Engineer\_NonProd/kbs-as-svc-user;",  "Effect": "Deny",  "Principal": "\*",  "Action": "s3:\*",  "Resource": [  "arn:aws:s3:::kbs-analytics-solutions-security-poc",  "arn:aws:s3:::kbs-analytics-solutions-security-poc/\*"  ],  "Condition": {  "StringNotLike": {  "aws:userId": [  "AROAJBQUBQFKWMQ3XSZXY:JUNGY",  "AIDAJP5UORHEZ32OYJP6Y"  ]  }  }  }  ]  }  {  "Version": "2012-10-17",  "Statement": [  {  "Sid": "userID definitions: AROAJBQUBQFKWMQ3XSZXY = KBS\_DevOps\_Engineer\_NonProd; AIDAJP5UORHEZ32OYJP6Y = KBS\_DevOps\_Engineer\_NonProd/kbs-as-svc-user;",  "Effect": "Deny",  "Principal": "\*",  "Action": "s3:\*",  "Resource": "arn:aws:s3:::kbs-analytics-solutions-security-poc",  "Condition": {  "StringNotLike": {  "aws:userId": [  "AROAJBQUBQFKWMQ3XSZXY:JUNGY",  "AIDAJP5UORHEZ32OYJP6Y"  ]  }  }  }  ]  }  AIDAJP5UORHEZ32OYJP6Y is the IAM User account.  Notice use of Sid to document.  arn:aws:s3:::kbs-analytics-solutions-security-poc |
| Flat file delivery | Data is delivered via any number of methods to the S3 bucket using IAM User Access Key | N/A | We are using SFTP Gateway to deliver data to S3. Full writeup on this is available separately. |
| Redshift Cluster creation | Redshift cluster created and needs one or more roles to get access to access other AWS resources. However, the sensitive data is not accessed by Redshift role assumption in this architecture, so a basic AWS managed role would suffice. | -Koch Cloud Team/AIS (Zach Slavin, Gary Powers)  -KBS Database Services (Scott Stanley) | Redshift Cluster  AWS Account: kbs-nonprod (106904787262)  Cluster Name: kbs-bia-risk-analytics  We created a dedicated role, but again, it is only providing access to non-sensitive data:  rn:aws:iam::106904787262:role/kbs-bia-riskmgmt-redshift-role  Redshift Security Group is open to all traffic via inbound/outbound rules due to being in Legacy  SSL is enforced by Redshift Parameter Group |
| Matillion ETL for Redshift creation | Matillion instance purchased from AWS Marketplace and requires a member of the Power Users role in Legacy (KBS Database Services has this) to purchase and stand up. Matillion also requires a role to access other AWS resources. This role can be created by anyone, but the policies required to be attached must be attached by Koch Cloud Team or AIS. | -Koch Cloud Team/AIS (Zach Slavin, Gary Powers)  -KBS Database Services (Scott Stanley) | Matillion Instance  AWS Account: kbs-nonprod (106904787262)  Host site: <http://ip-10-225-234-62.ec2.internal/>  Role:  arn:aws:iam::106904787262:role/Matillion\_ETL\_Role  Attached Policies:  arn:aws:iam::106904787262:policy/KBS-BIA-S3FullAccess  arn:aws:iam::106904787262:policy/KBS-BIA-Matillion-Policy |
| Extract/Load Processing | Data is extracted from on-prem data source or from S3 bucket (using IAM User Access Key) and loaded raw into preferred AWS database platform | -KBS Database Services (Scott Stanley)  -KBS BIA Team (Patrick Purviance) | Accomplished with Matillion, hosted in the Legacy kbs-nonprod account, which provides access to Koch on-prem and other cloud based data sources/destinations. Matillion can then be used to access secured sensitive data by way of Access Key.  See Matillion instance details above. |
| Transformation Processing | Data is transformed once loaded into preferred AWS database platform | N/A  (However, help can be obtained from Patrick Purviance) | Accomplished with Matillion, hosted in the Legacy kbs-nonprod account, which provides access to Koch on-prem and other cloud based data sources/destinations.  See Matillion instance details above. |
| Data Consumption | Data is now in a database platform in a Legacy account which has on-prem connectivity, so this is no different than traditional data consumption methods | N/A | Tested connecting to Redshift instance, kbs-bia-risk-analytics in kbs-nonprod (106904787262), with database login. Federated Identity authentication expected to be available soon.  Applications tested:   * Matillion * SQL Workbench * Tableau Desktop * Power BI Desktop * Excel (ODBC DSN) \* * Access DB (ODBC DSN) \*   \* Requires [32-bit ODBC Redshift Driver](http://docs.aws.amazon.com/redshift/latest/mgmt/install-odbc-driver-windows.html) |