Glue

Taufiqurrakhman Nur Hidayat, S.Kom., M.Cs.

Crawlers

- Add information about your crawler
  - Add Crawlers
- Specify crawler source type
  - Data stores
- Add a data store
  - Choose a data store : S3
  - Crawl data in : Specified path in my account
  - Include path :
    - Choose S3 path : CSV File
- Add another data store
  - No

Identity and Access Management (IAM)

- AWSGlueServiceRole, plus access to your data stores

- Roles
- Create role
- Select type of trusted entity
  - AWS service
- Choose the service that will use this role
  - Glue
- Attach permissions policies
  - AWSGlueServiceRole
  - AmazonS3FullAccess
- Review
  - Role name

Crawlers

- Choose an IAM role
  - Choose an existing IAM role
  - IAM role
- Create a schedule for this crawler
  - Run on demand
- Configure the crawler's output
  - Add database
- Finish

Jobs

- Add job
- Configure the job properties
  - Name
  - IAM role
  - Type
    - Spark
  - Glue version
    - Spark 2.4, Python 3 (Glue version 1.0)
  - This job runs
    - A proposed script generated by AWS Glue
- Run job

Jobs

- Script file name
- S3 path where the script is stored
- Temporary directory

- Choose a data source

- Choose a data target
  - Create tables in your data target
  - Data store
  - Format

- Map the source columns to target columns.
  - Save job and edit script

Jobs

Taufiqurrakhman Nur Hidayat, S.Kom., M.Cs.