Implement SQL in Engineering Workflows

Gerald Britton
Pluralsight Author

@GeraldBritton www.linkedin.com/in/geraldbritton

Stepping Through Data Ingestion

Extraction

Validation

Transformation

Aggregation

Enrichment

Loading

ETL vs ELT

Extract, Transform, Load	VS	Extract, Load, Transform
Extraction		Extraction
Validation		Loading
Transformation		Validation
Aggregation and Enrichment		Transformation
Loading		Aggregation and Enrichment

```
DECLARE @xml TABLE (xmldata XML)

INSERT INTO @xml (xmldata)
  SELECT *
  FROM OPENROWSET(
    BULK '/tmp/Pluralsight.xml',
         SINGLE_BLOB)
    AS xmlData;

SELECT xmldata.value(
   '(/ConfigArchive/ProfileName)[1]',
        'varchar(max)')
    AS ProfileName
FROM @xml
```

- Create a table to hold the xml data
- Load xml data from document
- Extract XML element from loaded document

```
CREATE TABLE json_external_table (
    id NUMBER,
    json_content CLOB
)
ORGANIZATION EXTERNAL (
    TYPE ORACLE_LOADER
    DEFAULT DIRECTORY data_dir
    ACCESS PARAMETERS (
        RECORDS DELIMITED BY '\n'
        FIELDS (id INTEGER EXTERNAL,
            json_content CHAR(4000))
        LOCATION ('json_file.json')
    )
);

SELECT * FROM json_external_table;
```

◀ **Create an Oracle external table**

◀ **Define the format and fields**

◀ **Setup the path to the file**

◀ **Read from the external table**

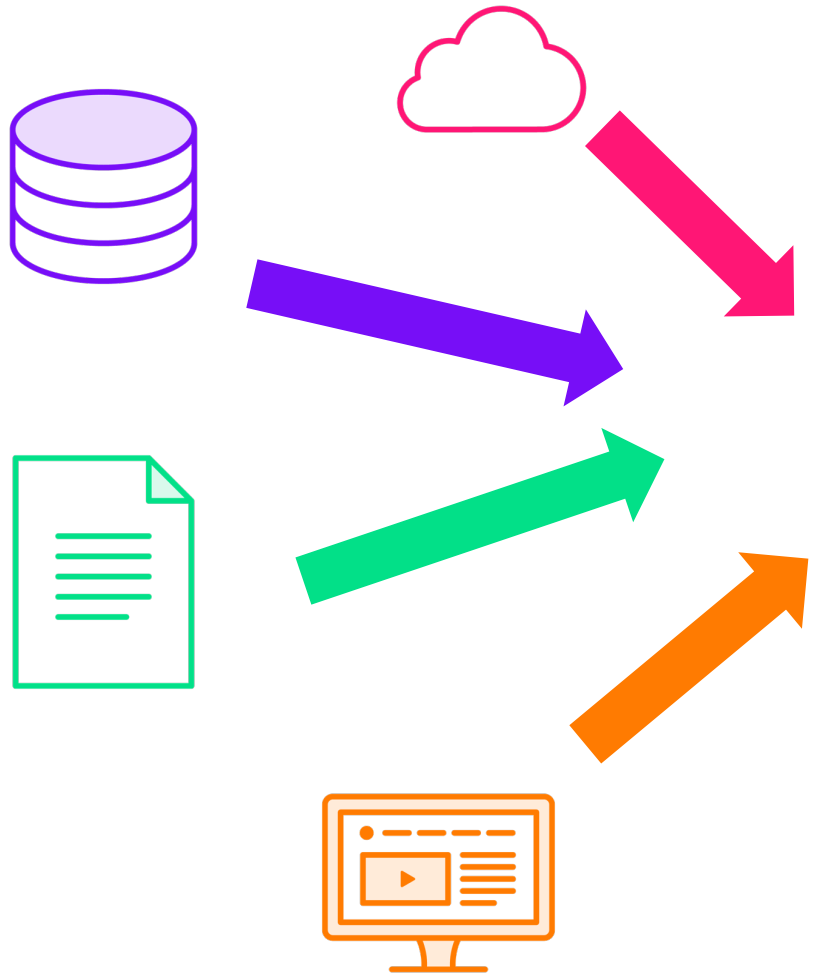
```sql
SELECT TRY_CAST('20230229' AS date);

SELECT *
FROM your_table
WHERE REGEXP_LIKE(your_column,
        '^[0-9]+$');

SELECT '123' AS my_value,
    CASE WHEN is_numeric('123')
        THEN 'Valid'
        ELSE 'Invalid'
    END AS is_valid;

SELECT CAST('123' AS INTEGER)
    AS converted_value;

INSERT INTO enriched_data
SELECT s.*, c.custname
    FROM source_data AS s
    JOIN cust_data AS c
        ON s.cust_id = c.cust_id;
```

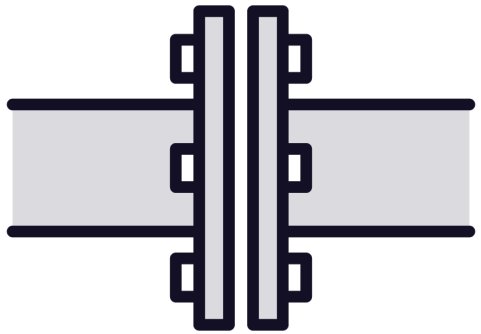
- SQL Server TRY_CAST and TRY_CONVERT
- Oracle REGEXP_LIKE
- PostgreSQL is_numeric
- ANSI SQL CAST
- Enriching data with customer name

Automating Data Ingestion

Solutions for enterprise-ready data pipelines.

Data Pipelines





ETL / ELT

Key Functions and Capabilities of ETL Systems

- Workflow automation
- Data quality
- Orchestration
- Real-Time
- Metadata management
- Monitoring and Performance

Top ETL Systems

- Informatica Cloud Data Integration
- IBM Infosphere DataStage
- Oracle Data Integrator
- Apache Airflow
- Microsoft SQL Server Integration Services
- Talend Data Fabric
- AWS Glue and Data Factory
- Azure Data Factory