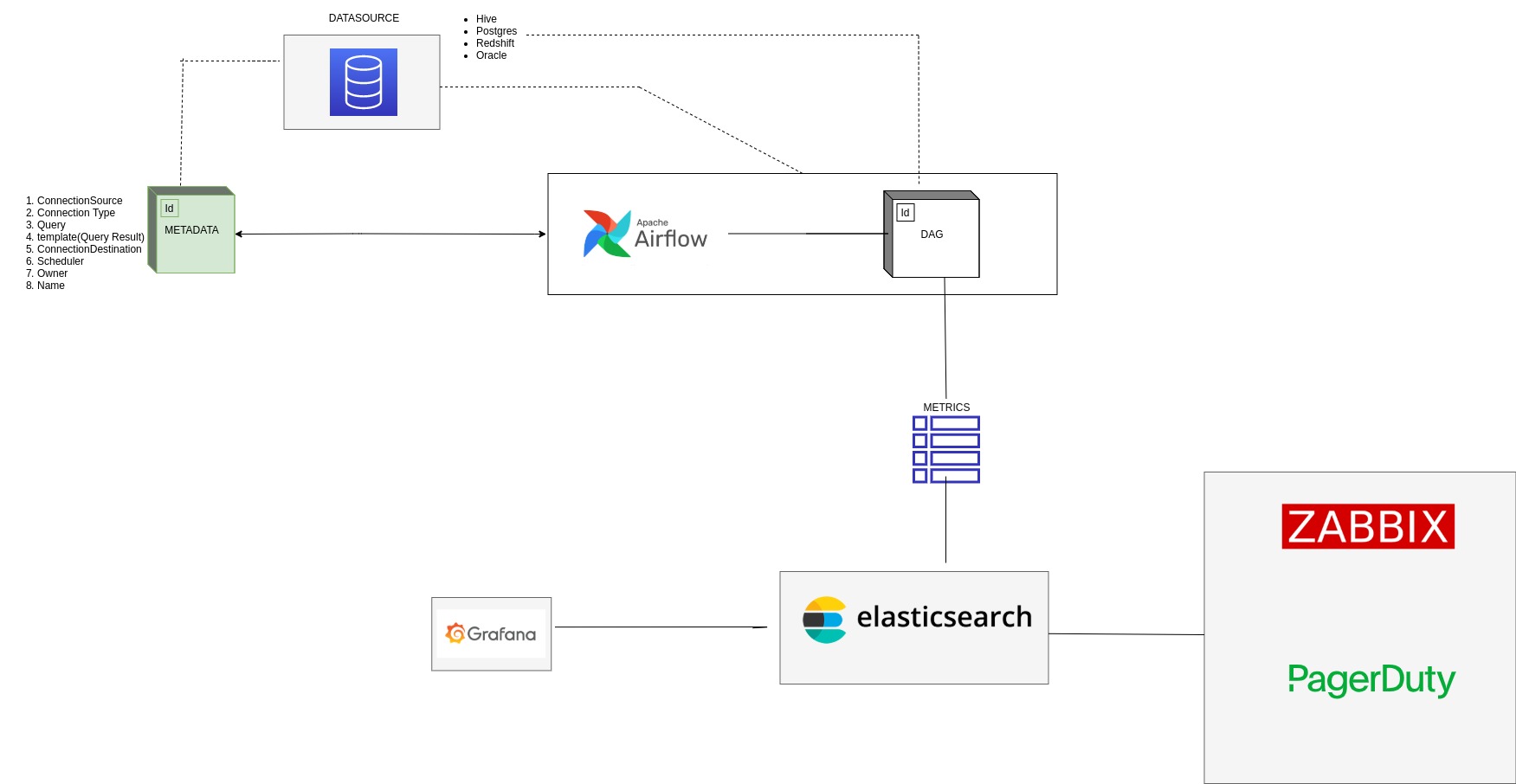
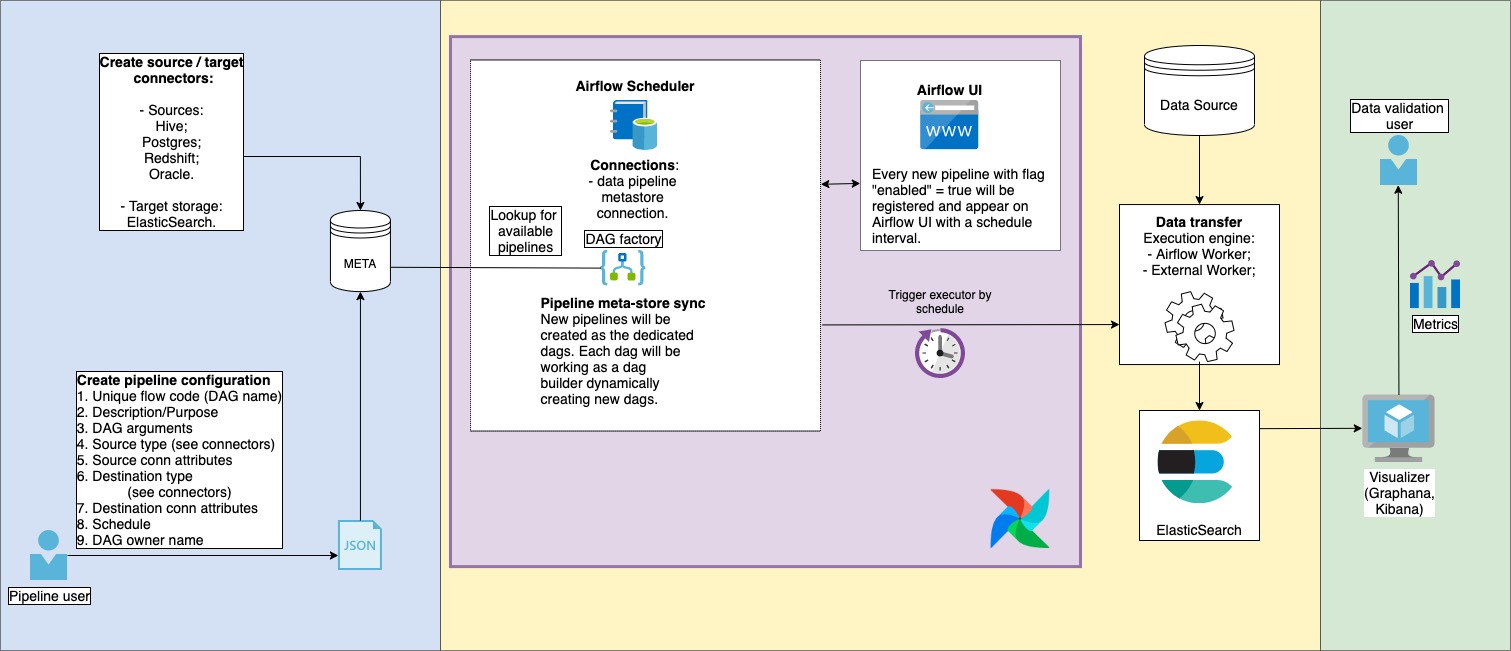
## Short Description

Data monitoring system with Airflow DAG





## Benefits

* Creation of open source, easy to incorporate cloud agnostic Data Monitoring System that may be used out-of-the-box

## Scenarios

* Catch data anomalies at validation step not when business logic fails
* Centralized alerts about absent or late data

{

**"pipelines"**: [

{

**"code"**: **"postgres-tableA-monitoring"**, **- уникальное значение pipeline'а;**

**"project"**: **"SomeMigration | some-migration"**, **- назанвание проекта. используется в наименовании дага (указывается без пробелов)**;

**"flow"**: **"tableA | inventory-distribution"**, **- описание цели/задачи. используется в наименовании дага (указывается без пробелов)**;

**"description"**: **""**, **- будет доступно внутри дага**;

**"arguments"**: { **- аргументы для настройки задачи**;

**"dag"**: {

**"start-date"**: **""**,

**"schedule"**: **"0 \* \* \* \*" - расписание дага**

}

},

**"source"**: {

**"type"**: **"jdbc"**,

**"connection"**: {

**"declaration"**: **"explicit"**, **- флаг указания на принцип использования подключения. explicit - явное указание значений**, **implicit - указание через Airflow connection**;

**"url"**: **"jdbc:mysql://hostname/database"**, **- для случае когда подключение производится внутри агента**

**"driver"**: **"org.mysql.Driver"**,

**"username"**: **"dbuser"**,

**"passwordKeyPath"**: **"/some/key/store/path" - при использовании KeyStore**;

},

**"sqlQuery"**: **"SELECT \* FROM schemaA.tableB where datetimeField >= {{ ds }}" - используется jinja и Airflow macros**;

},

**"target"**: {

**"type"**: **"elasticsearch"**,

**"connection"**: {

**"declaration"**: **"explicit"**,

**"url"**: **"https://hostname"**,

**"index"**: **"inventory-distribution"**,

**"type"**: **"inventory"**,

**"login"**: **"dbuser"**,

**"passwordKeyPath"**: **"/some/key/store/path" - при использовании KeyStore**;

}

}

},

{

**"code"**: **"hive-tableA-monitoring"**,

**"project"**: **"SomeMigration"**,

**"flow"**: **"tableB | inventory-subordinat-X"**,

**"description"**: **""**, **- будет доступно внутри дага**;

**"arguments"**: {

**"start-date"**: **""**,

**"schedule"**: **"0 \* \* \* \*"**

},

**"source"**: {

**"type"**: **"hive"**,

**"connection"**: {

**"declaration"**: **"implicit"**,

**"id"**: **"hive-connection-id-A" - для случаев когда атрибуты указаны в Airflow connection**;

},

**"sqlQuery"**: **"SELECT \* FROM schemaA.tableB where datetimeField >= {{ ds }}" - используется jinja и Airflow macros**;

},

**"target"**: {

**"type"**: **"elasticsearch"**,

**"connection"**: {

**"declaration"**: **"implicit"**,

**"id"**: **"elasticsearch-inventory-id"**

}

}

}

]

}