Progress DataDirect + Google BigQuery: Accelerate Your Data Warehousing and BI Initiatives

November 21, 2019

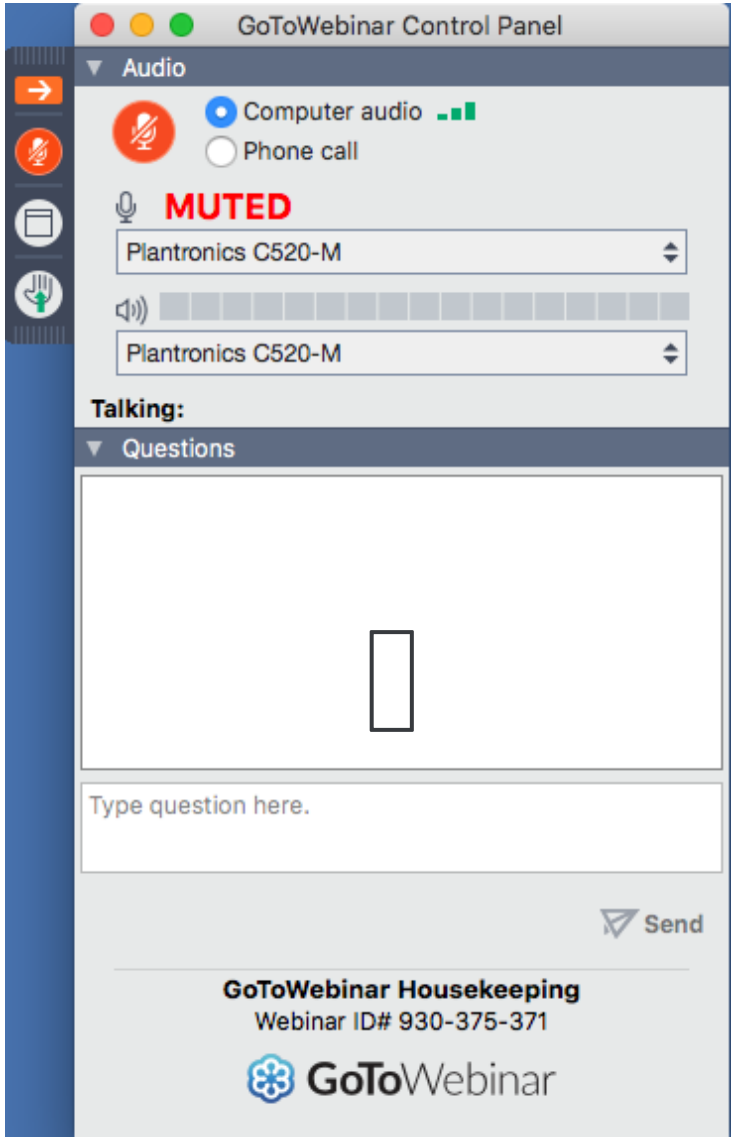
Dipak Patel

Dir of Product Management

James Goodfellow

Product Marketing Manager

GoToWebinar: Q&A

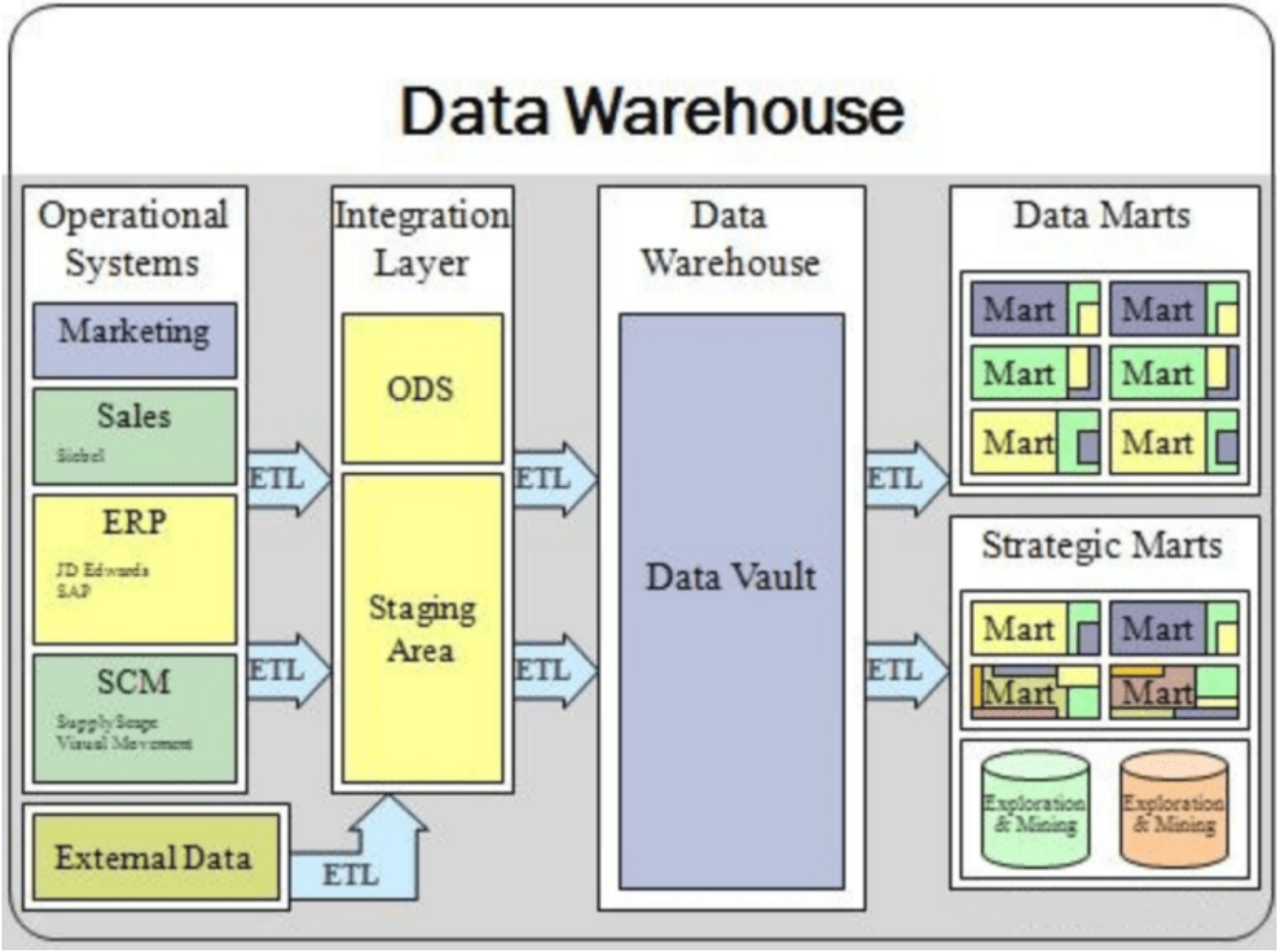


- Submit your questions & comments using the Questions panel at any time during the webinar

The Evolution of Data Warehousing

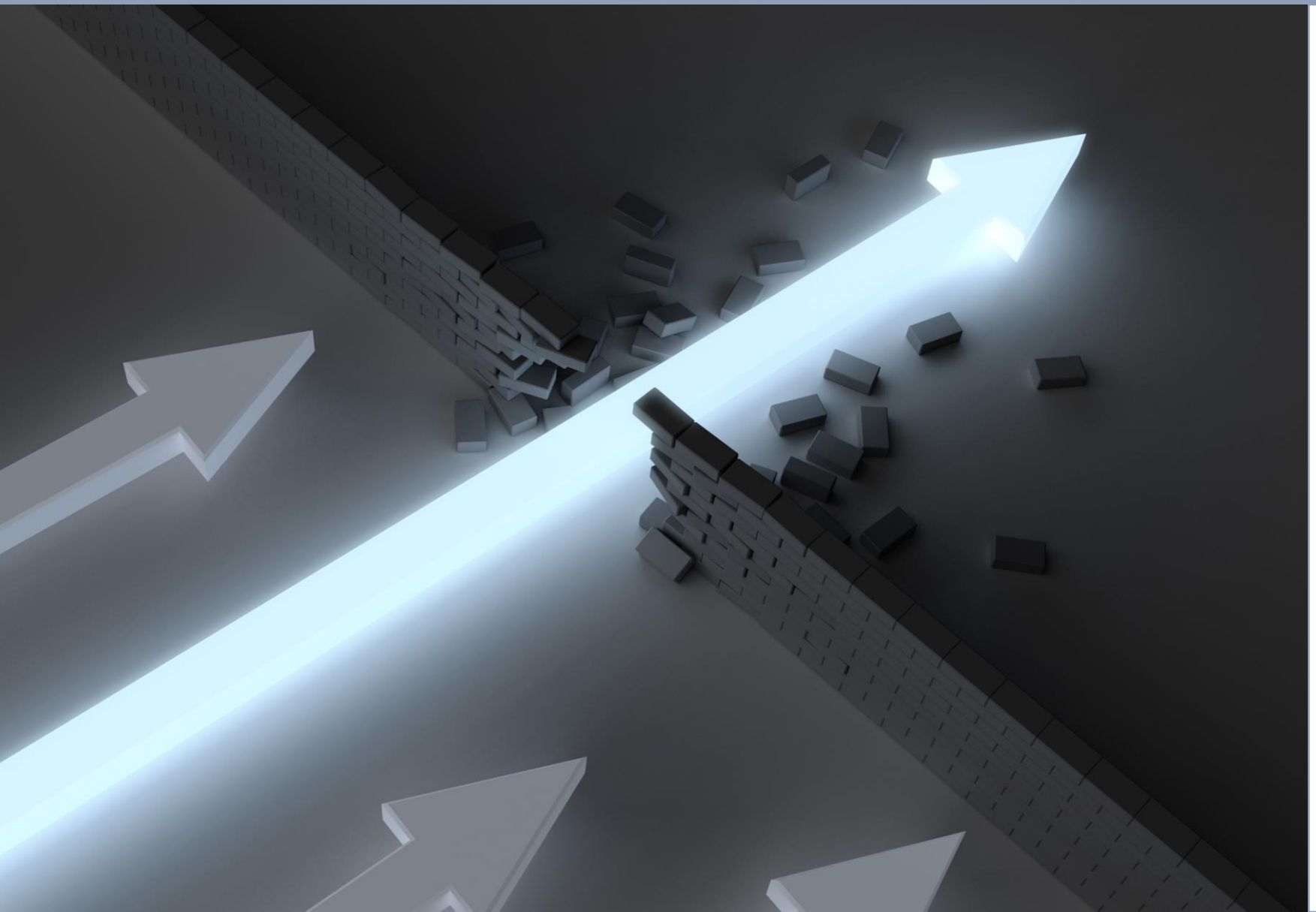
- Limited sphere of applications, tools and data marts
- On-premises, appliance-based warehouse architecture
- Connectivity generally wire-based using ODBC and JDBC.

A Traditional Data Warehouse Approach



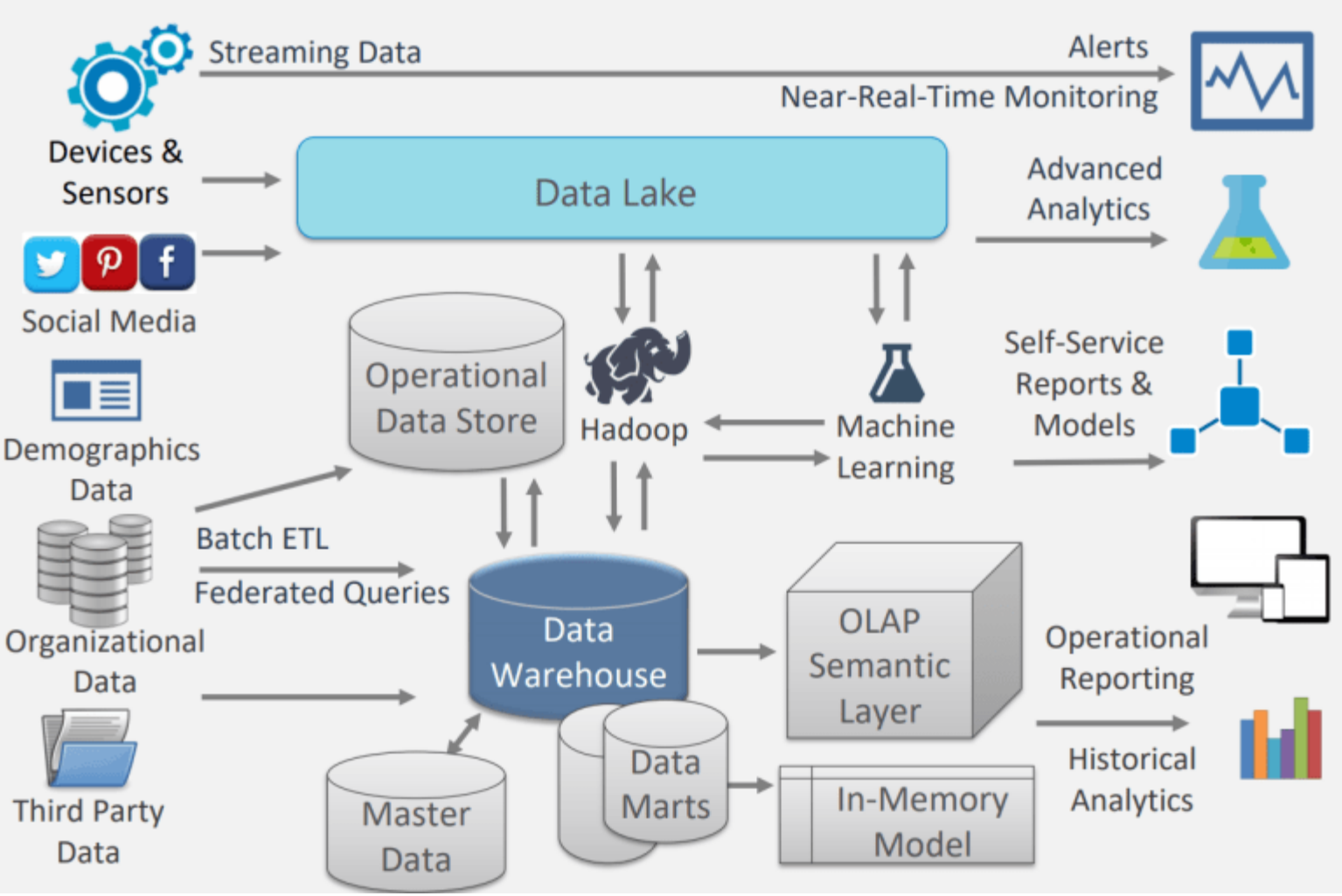
*courtesy of SQLHammer.com

Emergence of the Cloud Data Warehouse



- Modernize
- Scalable
- Cost-Effective
- Require Far Less Time & Effort

The Cloud Approach to Data Warehousing

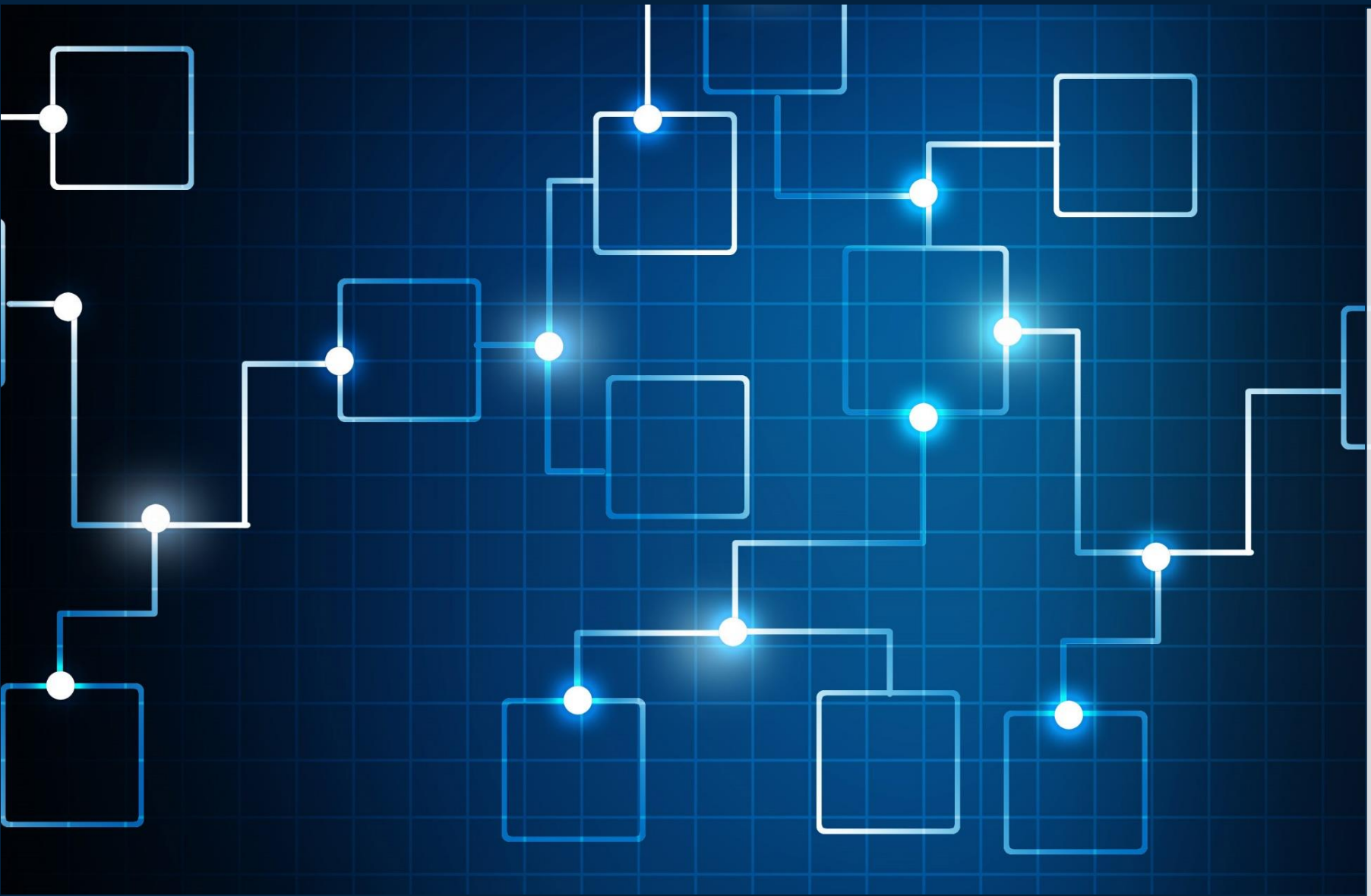


*courtesy of SQLHammer.com

Cloud Data Warehouse Challenges

- Speed of Data Access
- Query and Reporting Functionality
- Security

Progress DataDirect Driver for Google BigQuery



- Greater Connectivity and Functionality
- Faster Speed of Access
- Enhanced Security

Google BigQuery Driver in Action

Progress®

Use Cases for Google BigQuery JDBC and ODBC Drivers

Definition of JDBC

Java™ database connectivity (JDBC) is the JavaSoft specification of a standard application programming interface (API) that allows Java programs to access database management systems.

The JDBC API consists of a set of interfaces and classes written in the Java programming language.

Using these standard interfaces and classes, programmers can write applications that connect to databases, send queries written in structured query language (SQL), and process the results.

Since JDBC is a standard specification, one Java program that uses the JDBC API can connect to any database management system (DBMS), as long as a driver exists for that particular DBMS.

Definition of ODBC

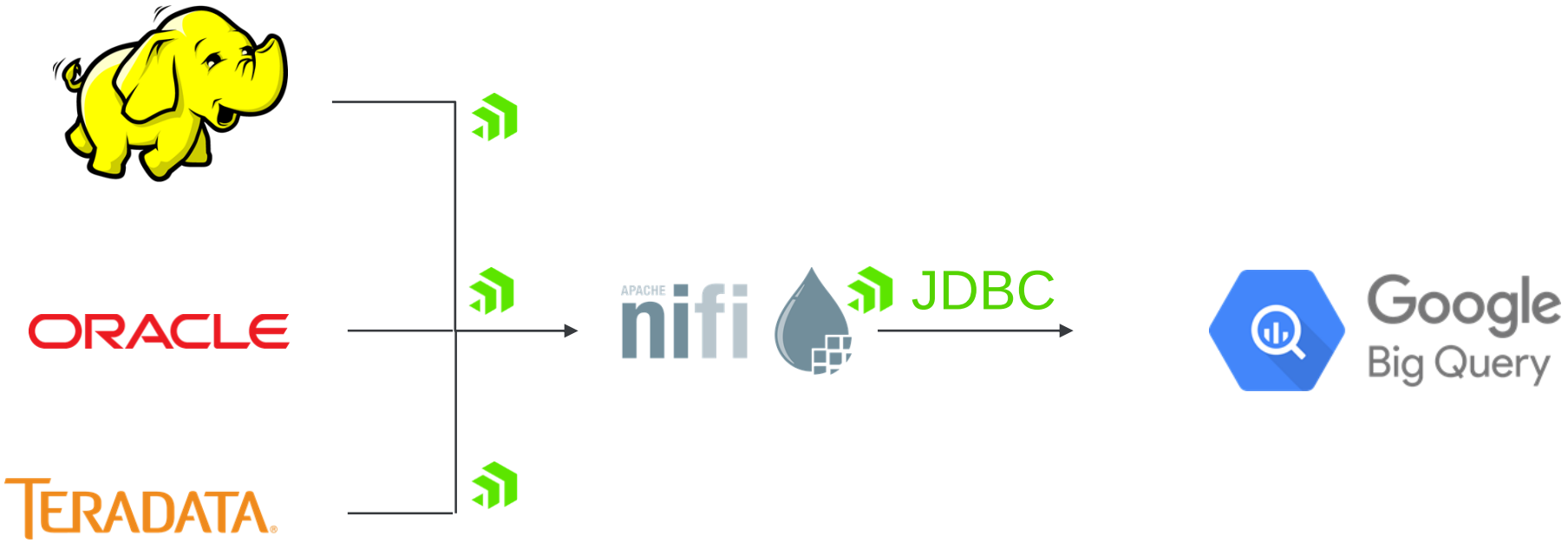
The Microsoft Open Database Connectivity (ODBC) interface is a C programming language interface that makes it possible for applications to access data from a variety of database management systems (DBMSs).

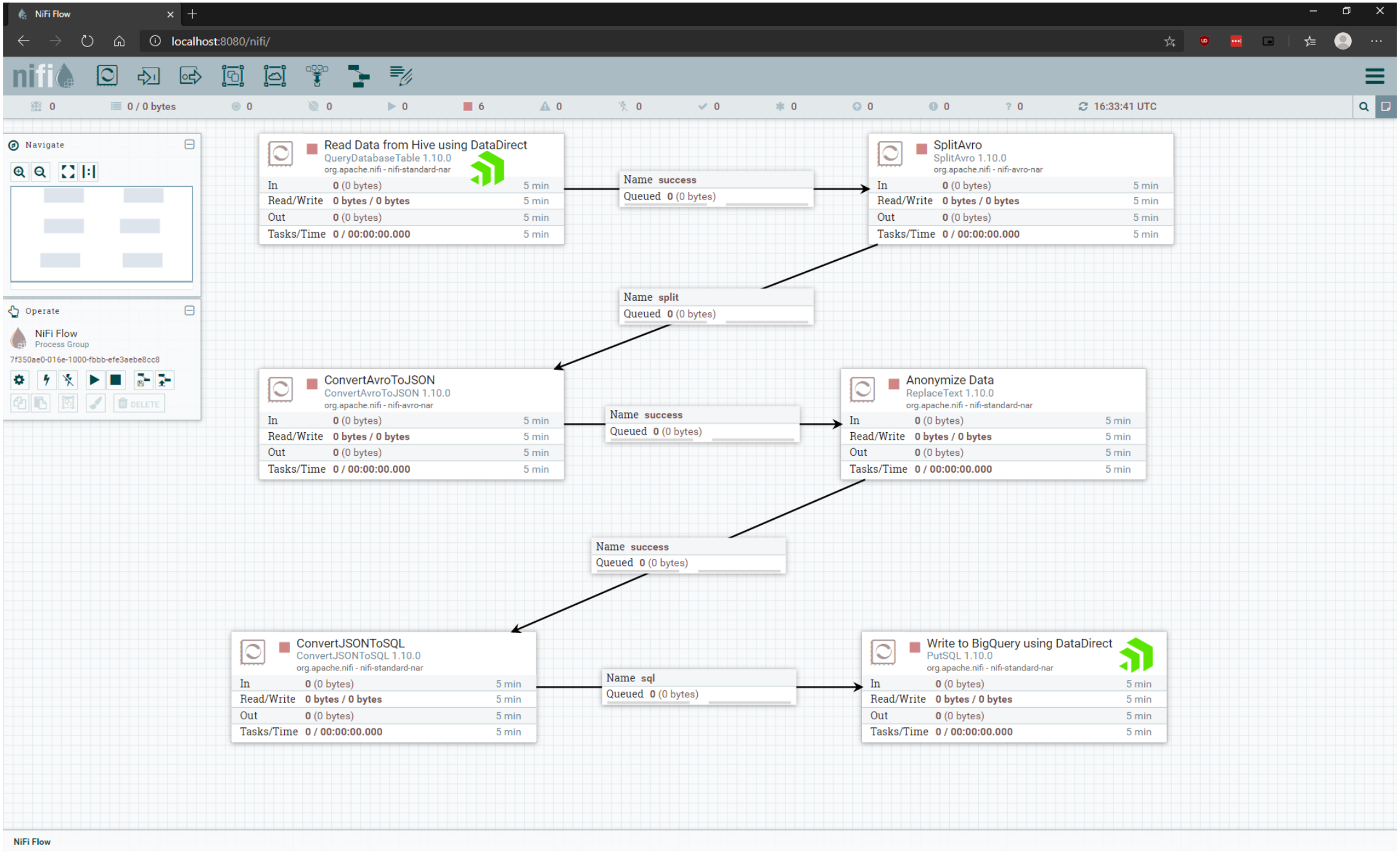
ODBC is a low-level, high-performance interface to connect to databases, send queries written in structured query language (SQL), and process the results.

The ODBC interface allows maximum interoperability—an application can access data in diverse DBMSs through a single interface.

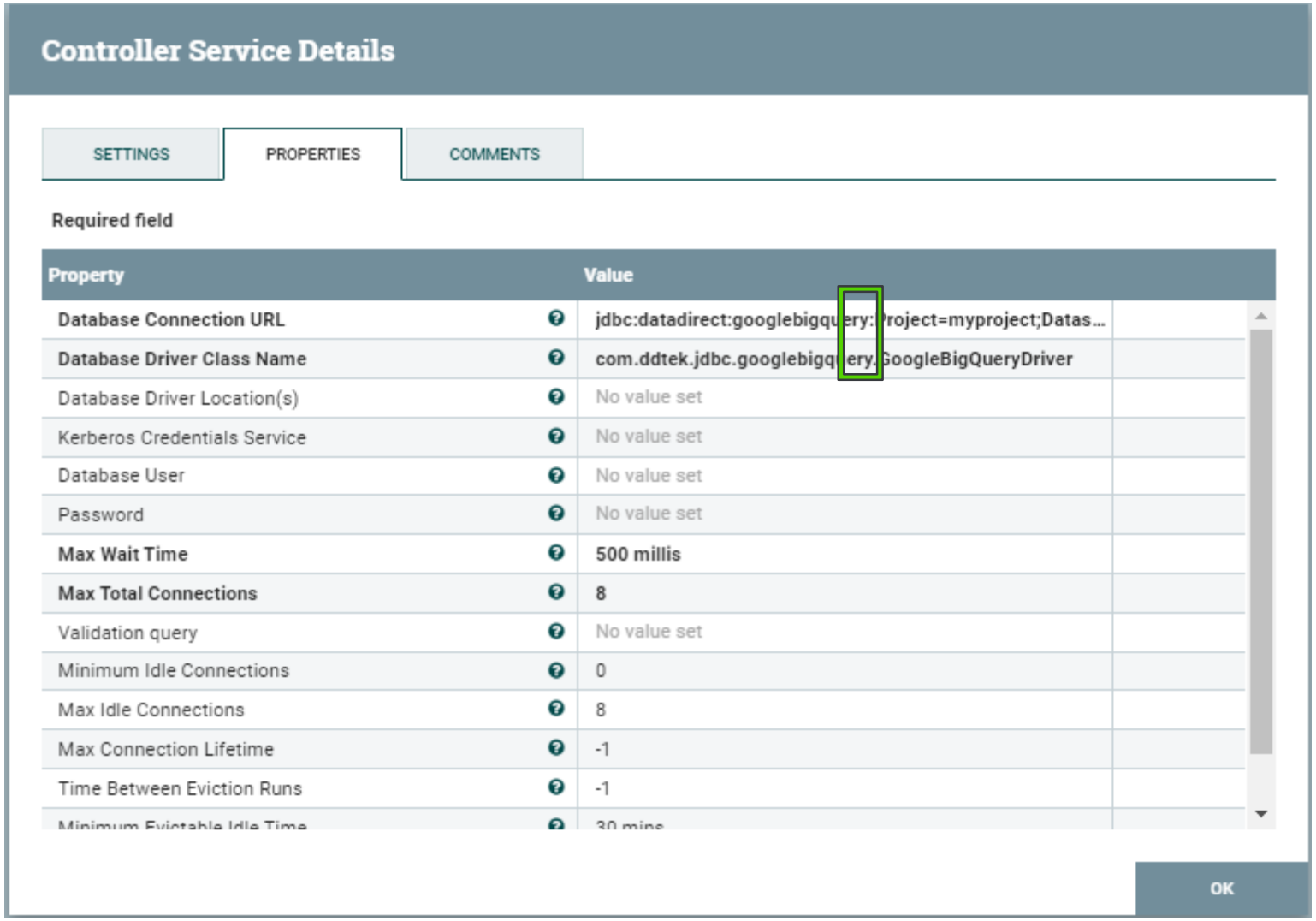
Moreover, that application will be independent of any DBMS from which it accesses data. Users of the application can add software components called drivers, which interface between an application and a specific DBMS

Batch Data With Apache NiFi

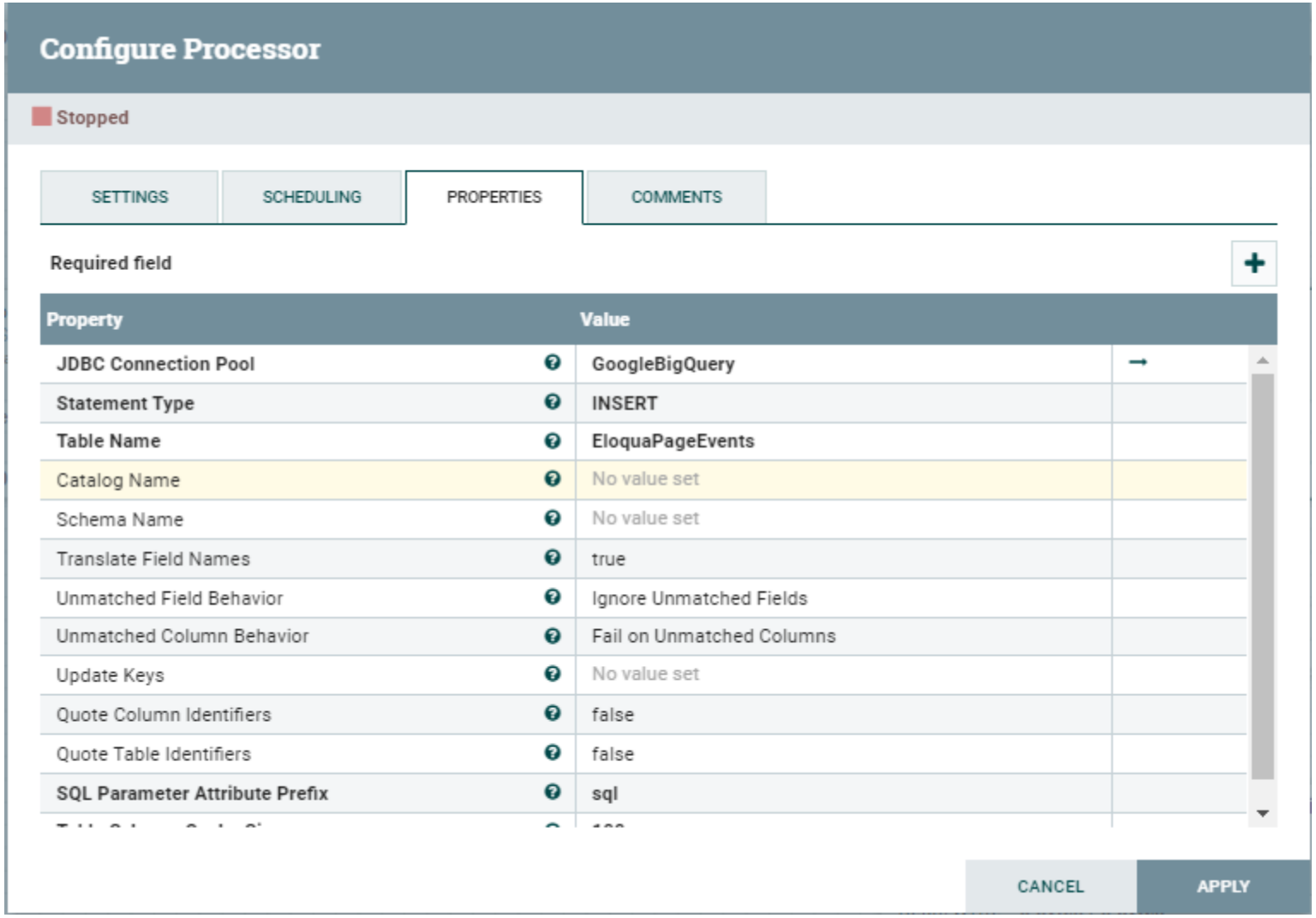




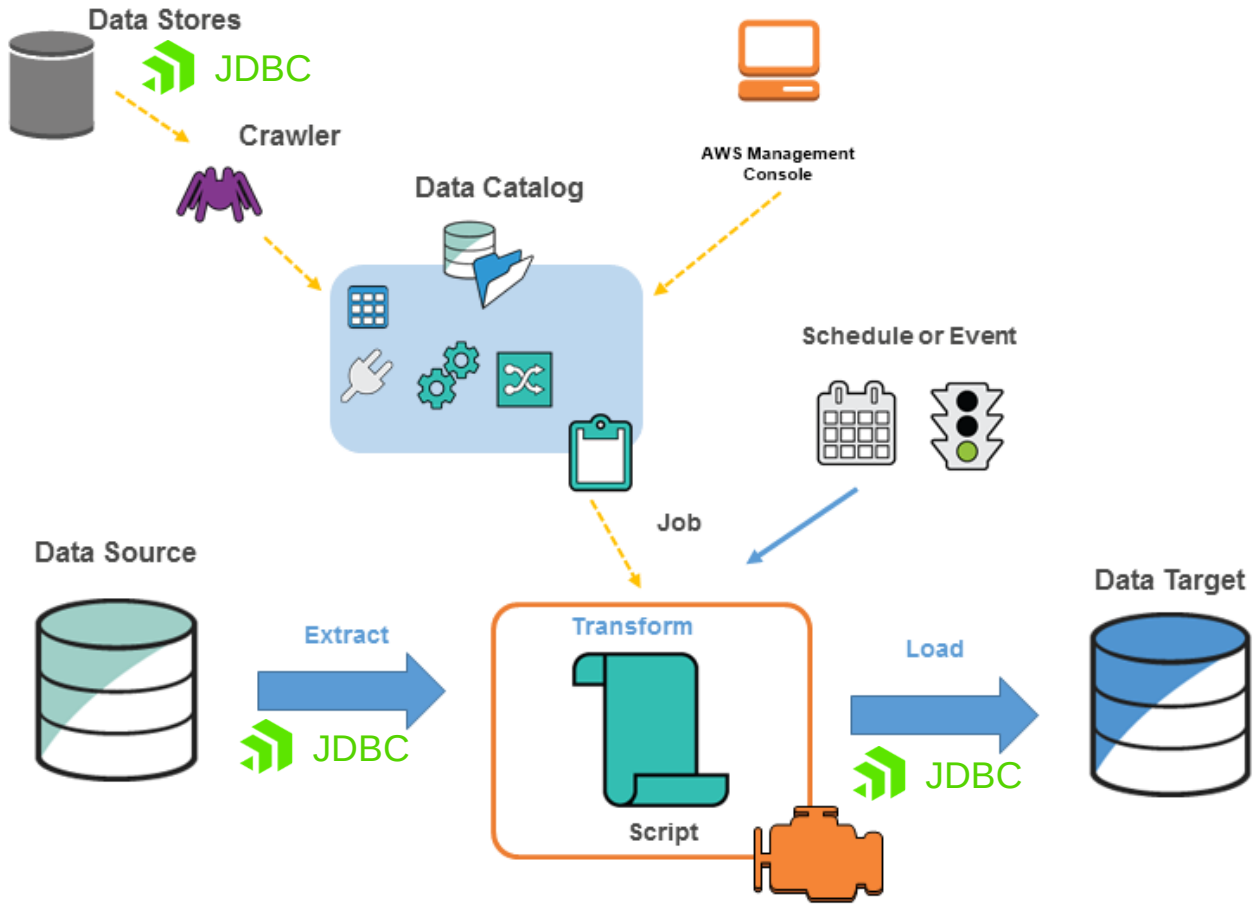
Configure Controller for GBQ



Configure Processor



AWS Glue – Similar to nifi



Progress® Corticon® Digital Decisioning: Read/Write Facts and Data

JDBC

REST/JSON API using ARC connector (read only)

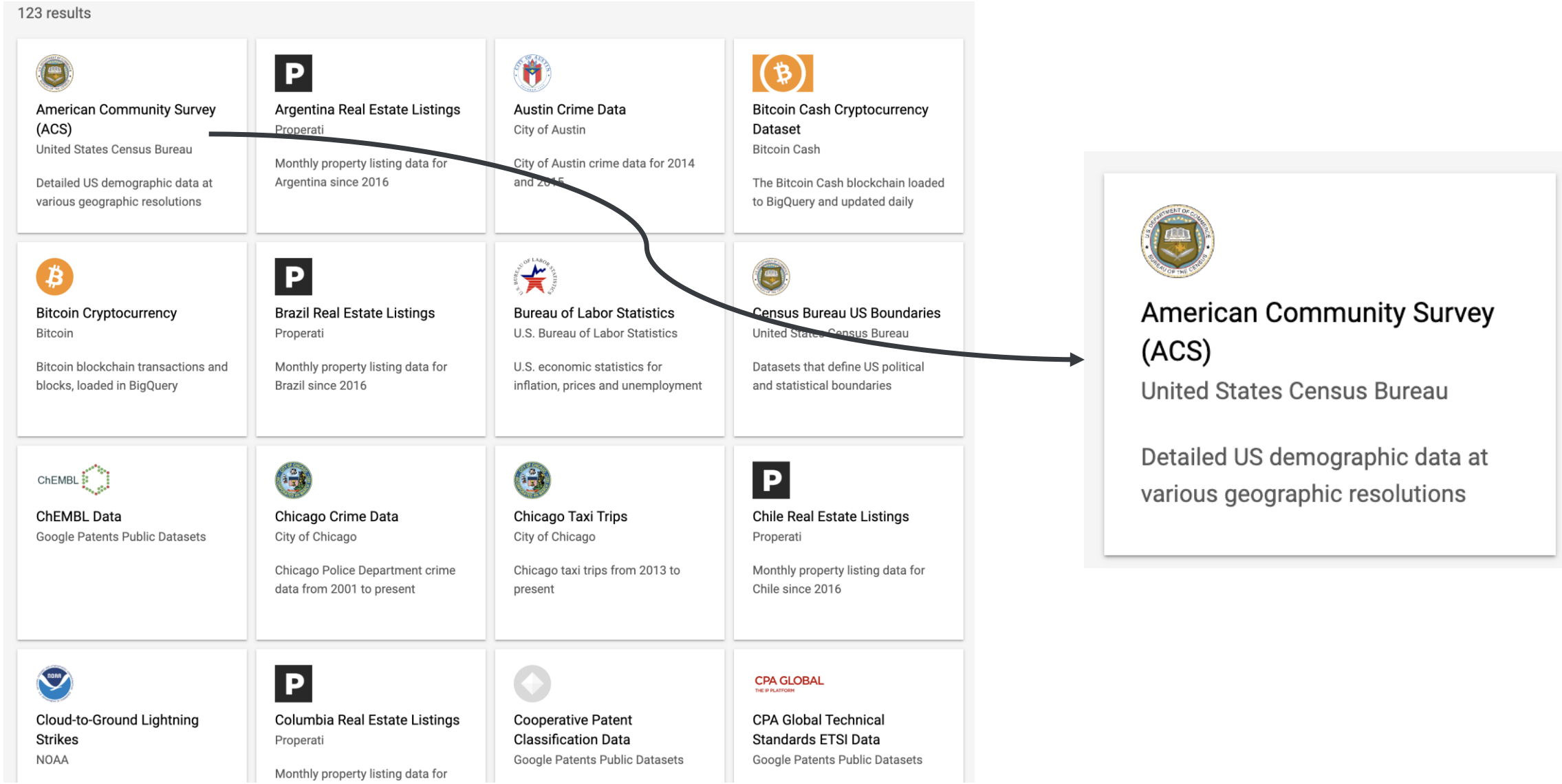
JDBC

MS DYNAMICS CRM using REST/OData connector (full CRUD)

JDBC

ANY DATABASE via DCI JDBC connectors (full CRUD)

Easily Access Public and Commercial Datasets



Accessing American Community Survey Data

What is the difference in median income in Brooklyn by Census tract from 2010 to 2017?

```
WITH acs_2017 AS (
  SELECT geo_id, median_income AS median_income_2017 FROM `bigquery-public-
data.census_bureau_acs.censustract_2017_5yr`
  WHERE geo_id LIKE '36047%' --Selecting Brooklyn
),
acs_2010 AS (
  SELECT geo_id, median_income AS median_income_2010 FROM `bigquery-public-
data.census_bureau_acs.censustract_2010_5yr`
  WHERE geo_id LIKE '36047%' --Selecting Brooklyn
),
acs_diff AS (
  SELECT a17.geo_id, a17.median_income_2017, a10.median_income_2010, (a17.median_income_2017  -
a10.median_income_2010) AS median_income_diff, geo.tract_geom AS tract_geom FROM acs_2017 a17
  JOIN acs_2010 a10 ON a17.geo_id = a10.geo_id JOIN `bigquery-public-
data.geo_census_tracts.census_tracts_new_york`geo
    ON a17.geo_id = geo.geo_id
)
SELECT * FROM acs_diff WHERE median_income_diff IS NOT NULL
```

20.8 MB in 2 seconds for just 1/100th of a cent

Google BigQuery Console

Census Bureau ACS: What is the difference in median income in Brooklyn by Census tract from 2010 to 2017?

```
1  WITH acs_2017 AS (
2    SELECT geo_id, median_income AS median_income_2017
3    FROM `bigquery-public-data.census_bureau_acs.censustract_2017_5yr`
4    WHERE geo_id LIKE '36047%' --Selecting Brooklyn
5  ),
6
7  acs_2010 AS (
8    SELECT geo_id, median_income AS median_income_2010
9    FROM `bigquery-public-data.census_bureau_acs.censustract_2010_5yr`
10   WHERE geo_id LIKE '36047%' --Selecting Brooklyn
11 ),
12
13 acs_diff AS (
14   SELECT
15     a17.geo_id,
16     a17.median_income_2017,
```

Run | Save query | Save view | Schedule query | More

This query will process 20.8 MB when run.

Query results

SAVE RESULTS | EXPLORE WITH DATA STUDIO

Query complete (2.0 sec elapsed, 20.8 MB processed)

Job information | Results | JSON | Execution details

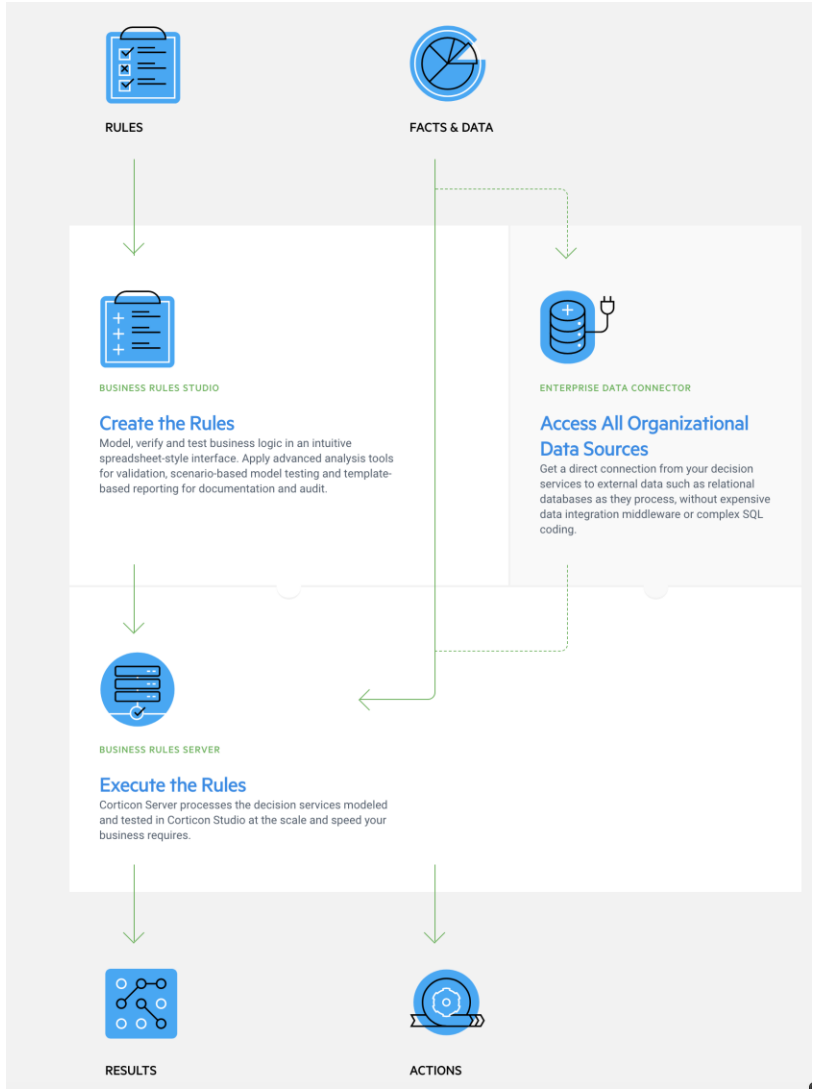
Some cell values are very long and the display is truncated to the first 1024 characters to improve browser performance. If full values are necessary, try lowering the number of rows per page before clicking "Show full values".

Show full values

Row	geo_id	median_income_2017	median_income_2010	median_income_diff	tract_geom
1	36047000100	88238.0	85396.0	2842.0	POLYGON((-73.9975 40.699653, -73.997395 40.699572, -73.997348 40.69957, -73.99 -73.996946 40.699618, -73.996817 40.699546, -73.996759 40.69948, -73.996708 40. 40.699163, -73.996001 40.699007, -73.995122 40.698767, -73.994292 40.698535, -7 -73.992124 40.697224, -73.991365 40.69701, -73.991407 40.697777, -73.991409 40. 40.699845, -73.990872 40.700066, -73.990787 40.70039, -73.990819 40.700554, -73. -73.99129 40.701049, -73.991676 40.701352, -73.992101 40.701684, -73.992207 40. 40.701844, -73.99236 40.701879, -73.992612 40.702061, -73.992768 40.70215, -73.9

Rows per page: 100 | 1 - 100 of 748 | First page | Last page

Progress Corticon Architecture



BigQuery

Interesting Questions Your Apps Can Ask

- What countries have the longest life expectancy?
- Which countries have the largest proportion of their population under 25?
  Over 40% of the world's population is under 25 and greater than 50% of the world's population is under 30!
- Which countries are seeing the largest net migration
- What is the current global picture of PM10 concentration?
  PM10 (small particulate matter of 10 microns or less in diameter) is known to cause disease and cancer, contributing to an estimated 3 million premature deaths worldwide per year in 2012.
- What's the average age of first marriages for females around the world?
- What's the average response time for each type of dispatched vehicle?
- Many, many, more

Commercial Datasets

AccuWeather

AccuWeather offers min/max temperatures, precipitation, and snowfall to enable users to track climate trends that influence consumer behavior.

Learn more

Explore via BigQuery

Dow Jones

Dow Jones DNA offers historical News Snapshots including Dow Jones Intelligent Indexes (subject, region, and industry taxonomies) which cover approx 19.5M companies in 28 languages.

Learn more

Explore via BigQuery

HouseCanary

HouseCanary offers historical home price indices going back up to 40 years, home price indices forecasted up to 3 years, current residential housing values and current housing rental values.

Learn more

Explore via BigQuery

Remine

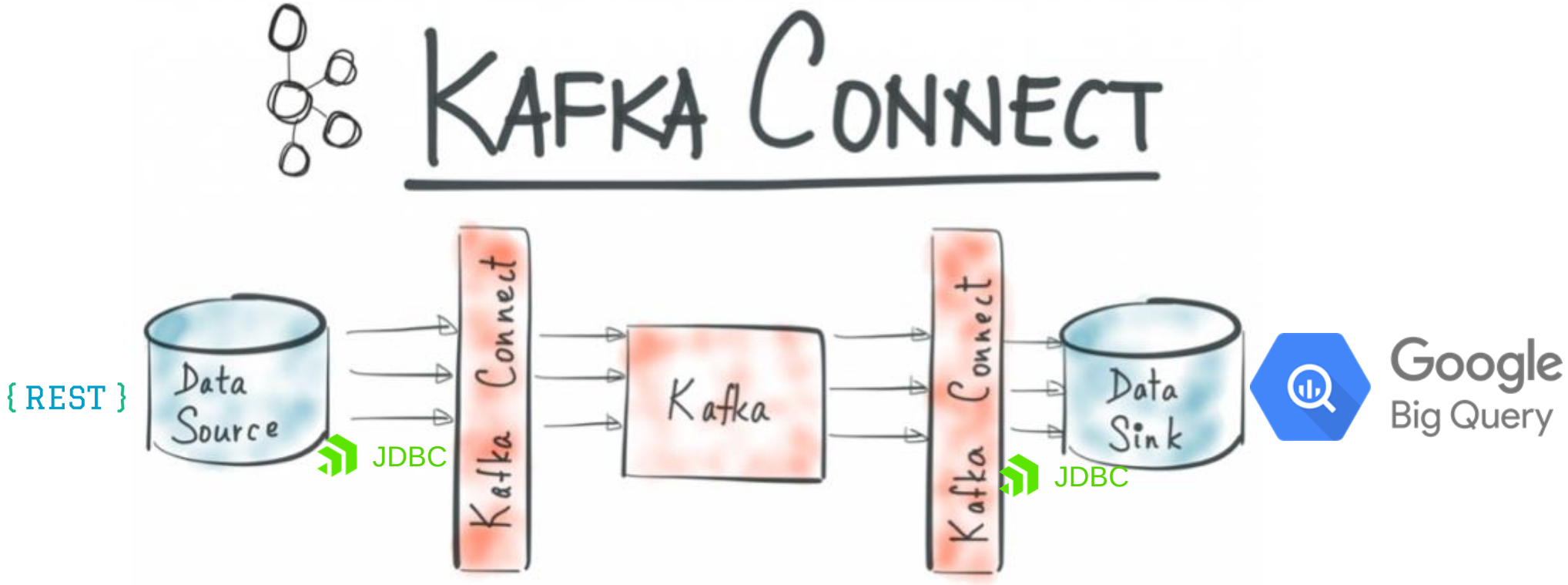
Remine offers property intelligence when single family homes will be going on sale within the United States.

Streaming Data Ingestion

Key Features of Kafka

- **Scalability**
- **High-Volume**
- **Data Transformations**
- **Fault Tolerance**
- **Reliability**
- **Durability**
- **Performance**
- **Zero Downtime**
- **Extensibility**
  - There are many ways by which applications can plug in and make use of Kafka. In addition, Kafka offers ways to write new connectors as needed.
- **Replication**

Kafka Connect for JDBC



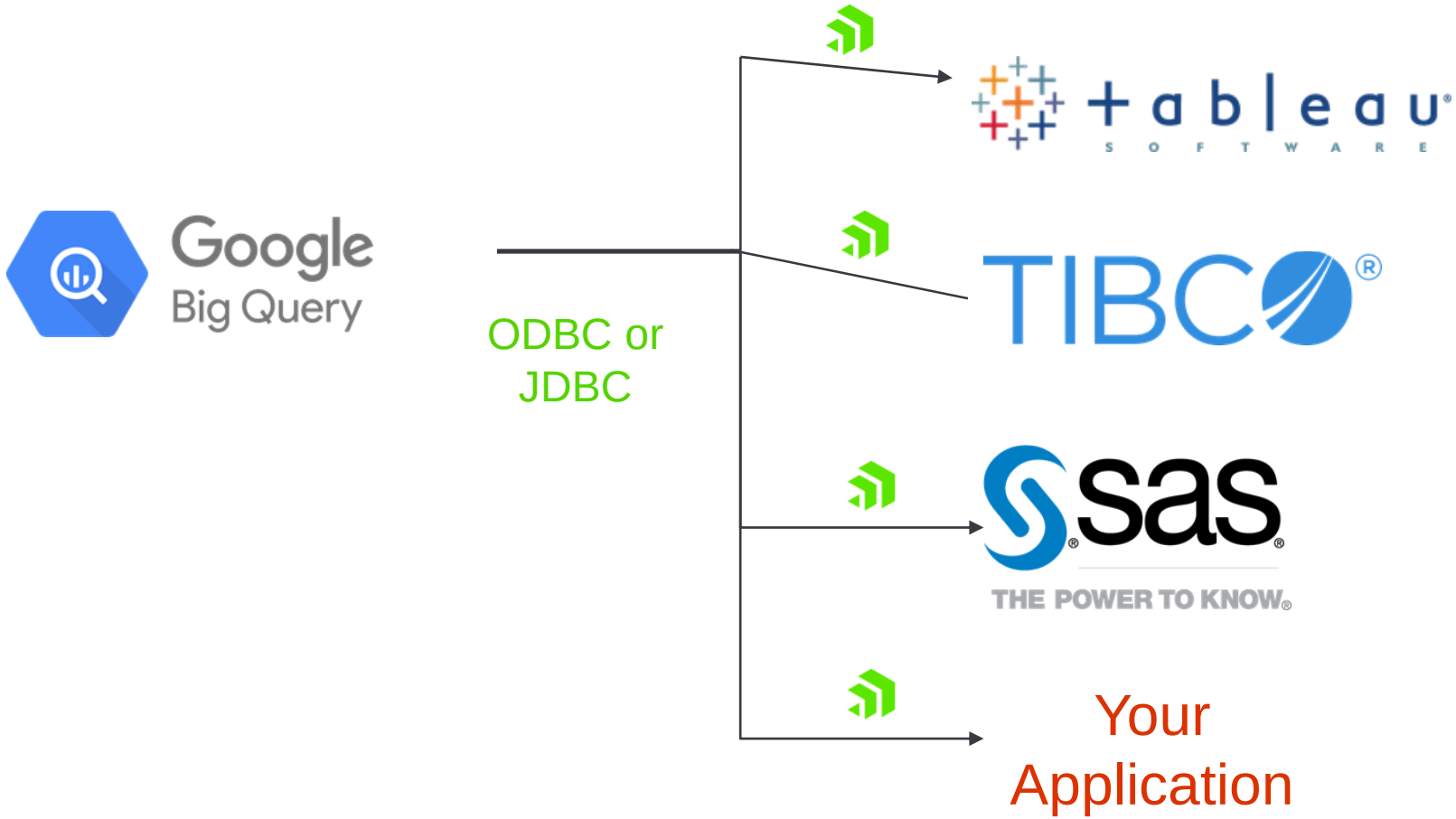
Configuring Kafka Connect

It is driven purely by configuration files, providing an easy integration point for developers

```
{
  "name": "jdbc_source_GoogleBigQuery_foobar_01",
  "config": {
        "connector.class": "io.confluent.connect.jdbc.JdbcSourceConnector”
        "connection.url": "jdbc:datadirect:googlebigquery:schemaMap=<Path to Schema Map File>;project
=<projectid>;dataset=<dataset>;accessToken=<accesstoken>;RefreshToken=<Refreshtoken>;clientid=<client
id>;clientsecret=<clientsecret>;schemaset=project:dataset",
        "table.whitelist": "foobar",
        "mode": "incrementing",
        "incrementing.column.name": "foobar_pk”
         }

}
```

Accessing GBQ from 3rd Party Application



Examples, can be a BI tool, Programming Languages for ML or Strategic OEM Partner's tool etc.,

Unleash the Power of GBQ from your Apps

- Plug-and-play with any JDBC or ODBC compliant application
- High performing, efficient, secure, and reliable
- Extend your application to leverage a data warehouse with lots of readily available data and AI/ML with very little effort

Progress®

Questions?


Progress®