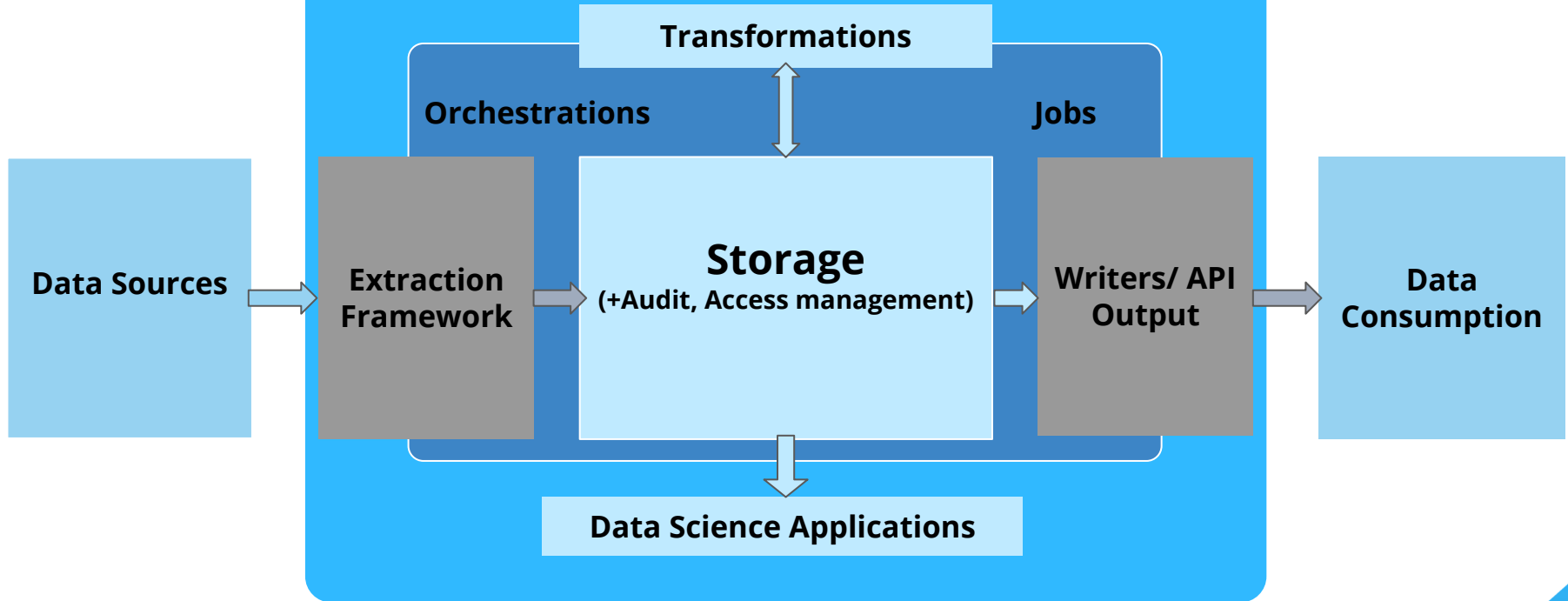




Under the Hood



Keboola Connection



EXTRACTORS

Marketing

Google Analytics, Mailchimp, Silverpop, Adwords, Omniture, Stat, Qualtrics and more



Social

Facebook, Twitter, Instagram, Youtube and more

Sales / Finance

Salesforce, Magento, SAP, Zendesk, Shopify, MS Dynamics, Recurly and more

Flat files

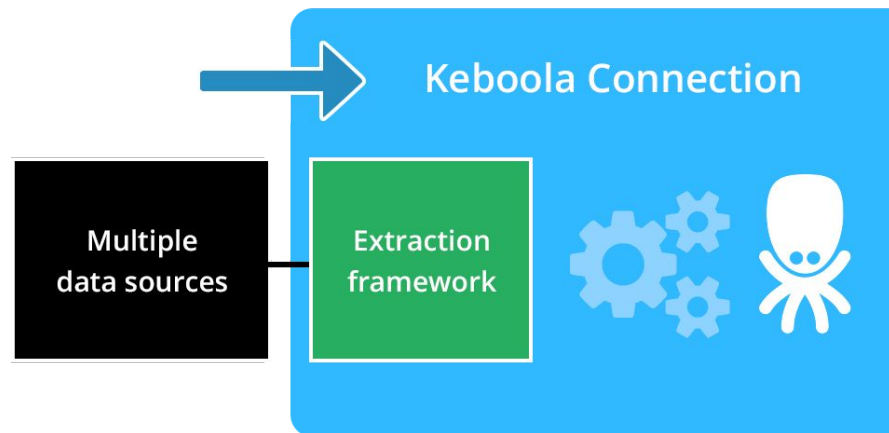
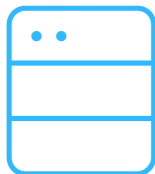
CSV, XLS

Database

SQL & NoSQL / Cloud & On-prem

DWH / Storage

Oracle, SAP, Teradata, AWS, Informix, Dropbox, Drive, FTP

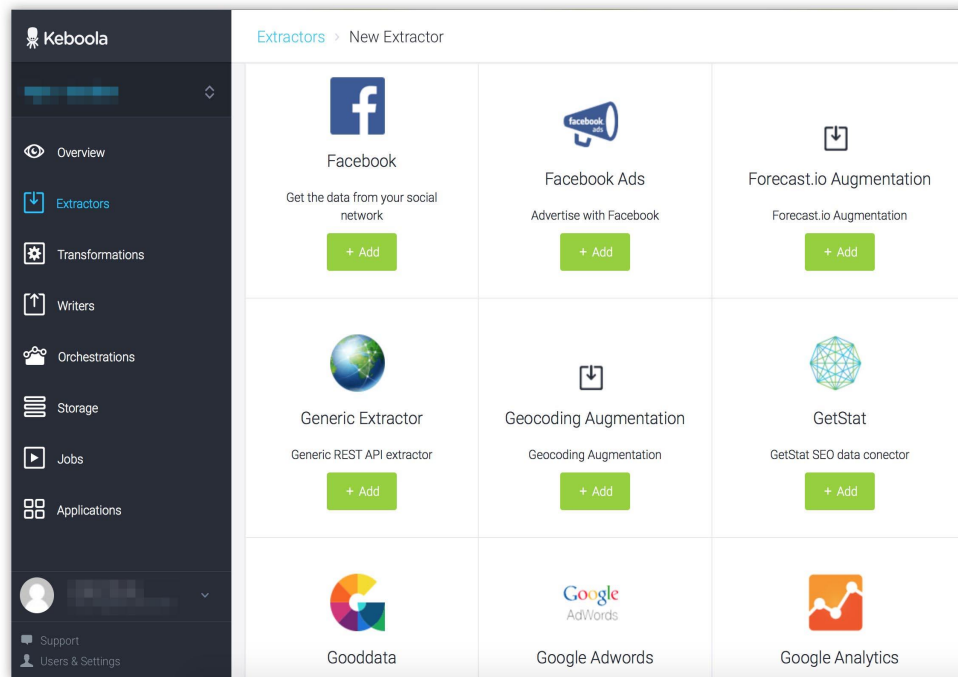


Over 170,000 data sources at your fingertips, instantly!

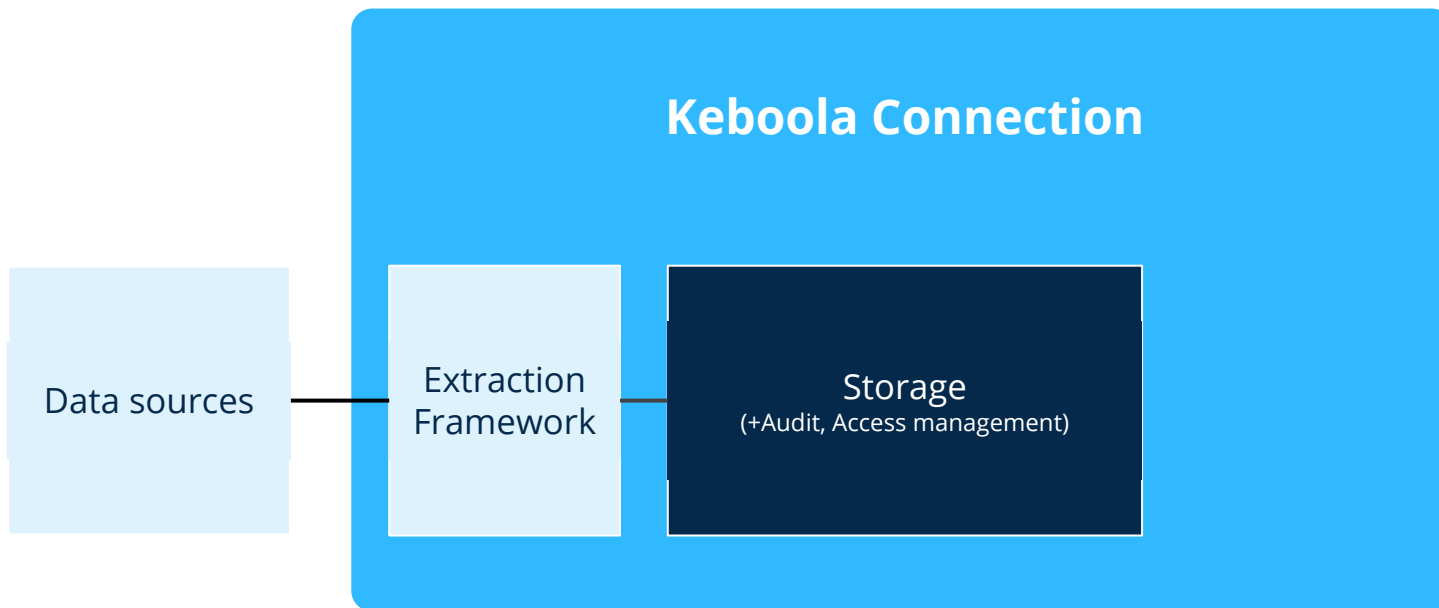
EXTRACTORS

With Keboola, you can instantly extract data from more than **170,000** different sources.

- Over **70 custom pre-built extractors** ranging from marketing applications to finance, with own UI; clients with development resources can also build their own extractor.
- Our 'golden bullet' is the **Generic REST Extractor**, which in itself is configurable for thousands more data sources in a **matter of hours!**



STORAGE



Secure data warehouse with a full rollback.

STORAGE

Storage API is the key part of the infrastructure. It is fully managed by Keboola and run in **Amazon S3**.

It mainly **receives** data from extractors, and **stores** it safely for further operations.

Data is then accessible for transformations, data blending and augmentation.

Internal tasks are conducted in an **isolated environment** for **security** purposes, and all activity is logged for audit at any point in time.

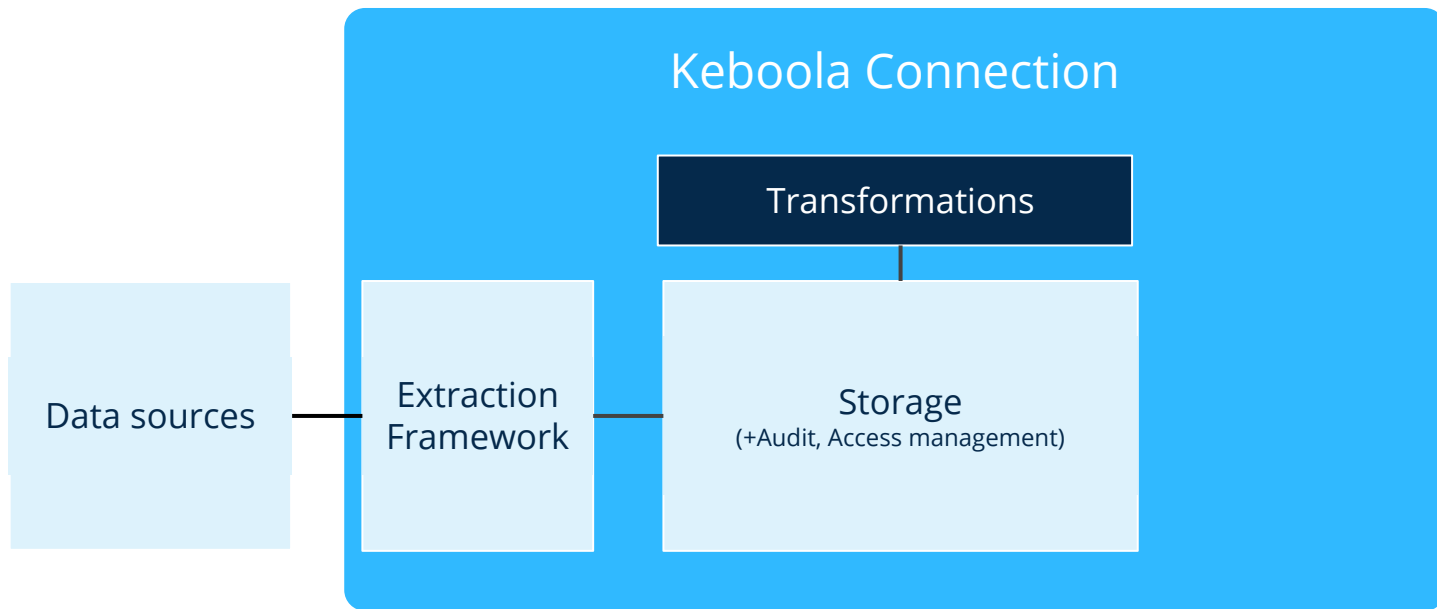
This log is visible and searchable to all users of the project and offers a full rollback function.

The screenshot displays the Keboola Storage API interface. On the left is a dark sidebar with navigation links: Overview, Extractors, Transformations, Writers, Orchestrations, Storage (highlighted), Jobs, Applications, and Users & Settings. The main panel is titled 'Storage' and shows a list of buckets. The selected bucket is 'in.c-ex-api-generic-tm1', which contains a 'Cubes' sub-entry. The 'Cubes' entry is expanded, showing a list of views: results, views, views_Columns, views_Rows, and views_Titles. The 'views' entry is selected, showing a table of data. The table has columns for ID, Created, Primary key, Last import, Last change, Rows count, and Data size. The data is as follows:

ID	Created	Primary key	Last import	Last change	Rows count	Data size
in.c-ex-api-generic-tm1.Cubes	2015-08-25 15:04	Not set	2015-09-07 21:40	2015-09-07 21:40	0	32.00KB

Below the table is a section titled 'Latest imports' with a bar chart showing the number of imports over time. The chart has a y-axis from 0.0 to 0.9 and a series of bars representing imports at different times.

TRANSFORMATIONS



Clean, join, unify and transform data.

TRANSFORMATIONS

Transformation is essentially **data manipulation**: cleaning, joining and unifying prepared in **SQL** and **R**.

In Keboola, you can use the right tool for each task. Available transformation backends are: **MySQL**, **Redshift** and **Snowflake**.

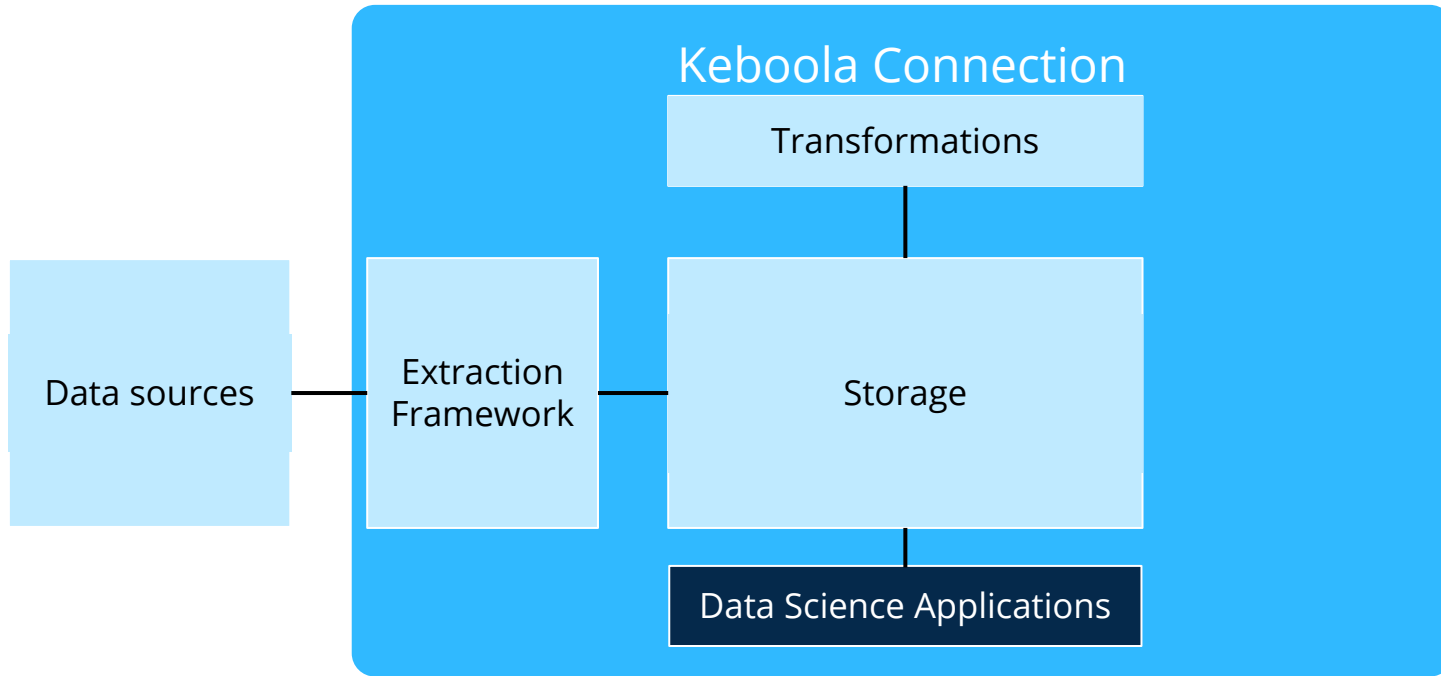
Sandbox is instantly available for fast iterations of your query for both, MySQL and Redshift backends.

Transformations can be saved as **templates** and reused across multiple projects.

The screenshot displays the Keboola web interface for managing transformations. The left sidebar contains navigation links: Overview, Extractors, Transformations (highlighted), Writers, Orchestrations, Storage, Jobs, and Applications. The main content area is titled 'Transformations > Bucket Pairing > Transformation main'. It features two sections: 'Input Mapping' and 'Output Mapping'. The 'Input Mapping' section lists four input buckets with their sizes and names, each mapped to a specific output bucket. The 'Output Mapping' section shows one output bucket. Below these sections is a 'Queries' section with a text editor containing a SQL query. The query is as follows:

```
create table traffic as
select
  pages.`id`,
  pages.`source`,
  pages.`medium`,
  pages.`landingPagePath`,
  pages.`bounceRate`,
  pages.`pageviews`;
```


DATA SCIENCE APPLICATIONS



Don't have an in-house data scientist? Fret not!

DATA SCIENCE APPLICATIONS

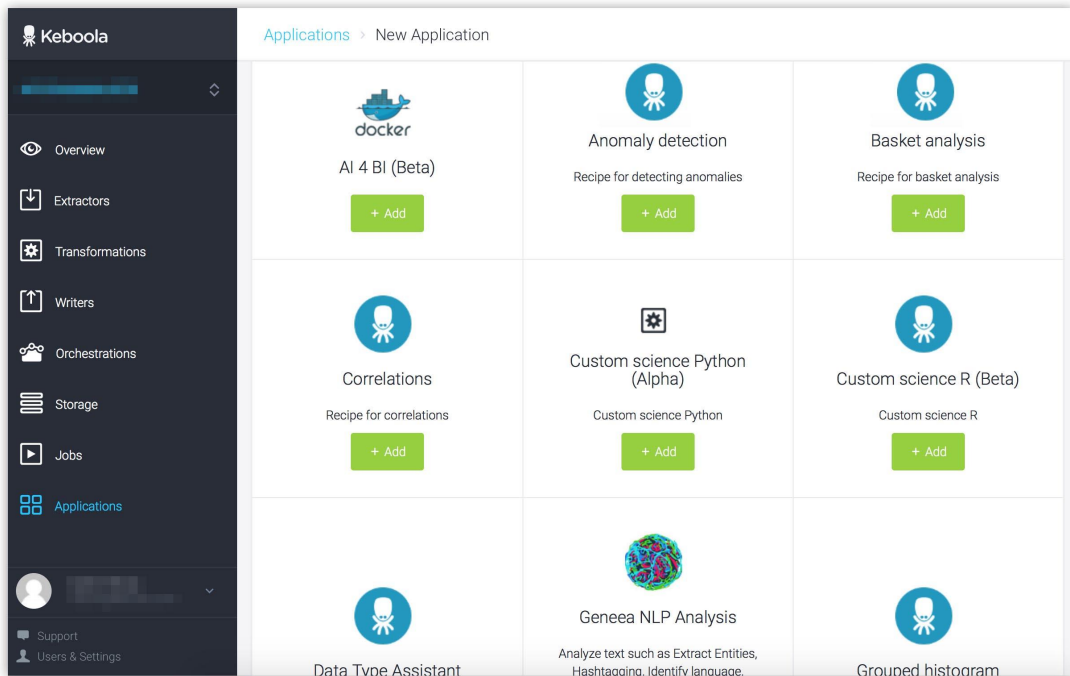
See our App Store with **'ready to deploy'** data science applications that can be released within a couple of clicks and are **easy to configure**.

Process unstructured data using 3rd party applications and technology such as NLP from Geneea.

Let us prepare **custom algorithms** for you, or use **your own solution** thanks to the help of a **docker** image.

Examples of available apps:

- Next order prediction
- Segmentation
- Anomaly detection
- Basket analysis
- Grouped histogram (Video [example](#))
- and much more



APPLICATIONS - BASKET ANALYSIS

Discover what goods customers buy together.

Understand what combinations of goods to bundle.

In Keboola Connection, we have a ready-made set of applications revolving around **purchasing patterns** in R language. It comes with description of data format required. The app is ready to deploy instantly and is suitable for most common use cases.

If your business has very specific needs, deploy your own R code, or reach out to our team and network of specialised partners to help you build personalised one.

Most suitable use cases are in **Retail, E-commerce shops, CRM and Sales in B2B organisations.**



WRITERS



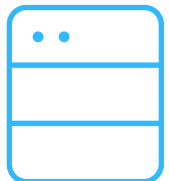
Business Intelligence

Tableau, GoodData, YellowFin, SAS
Visual Analytics, Chartio, Periscope,
Looker, etc.



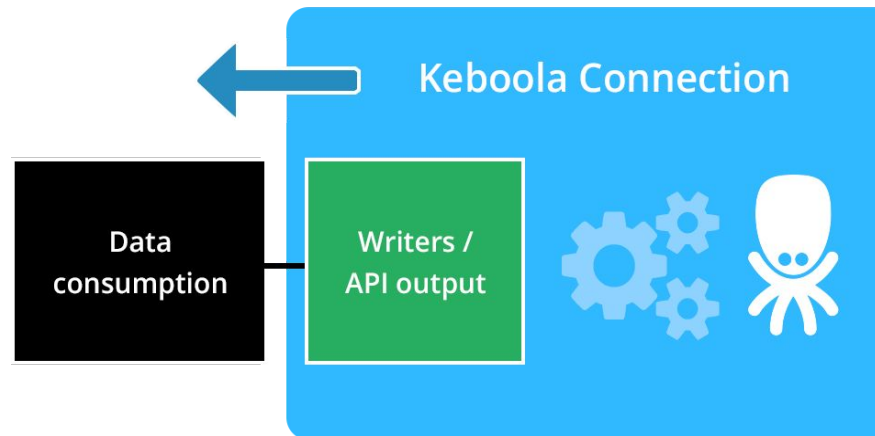
Marketing Solutions

Facebook, Adwords custom audiences,
Marketo, Mailchimp, Silverpop,
Trading desks, etc.



DWH / DB / Storage

Oracle, SAP, Teradata, AWS, Informix
Dropbox, Drive, FTP, Redshift, Elastic,
BigQuery, etc.



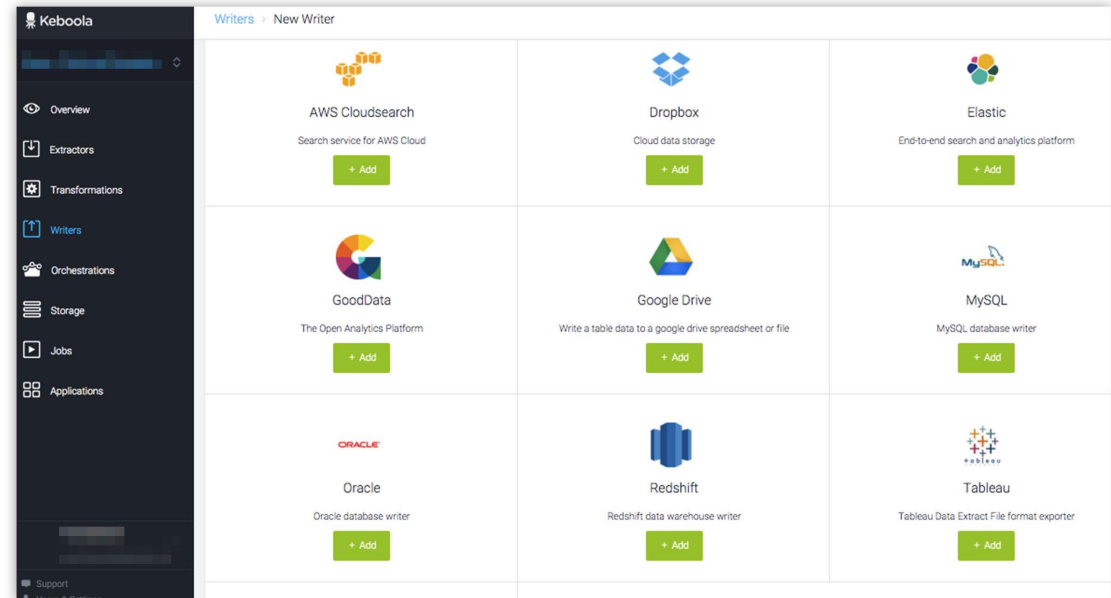
Dashboards are not the only way of consuming data.

WRITERS

Writers are components that take transformed and processed data, help **identify and describe** it and load it into its final destination.

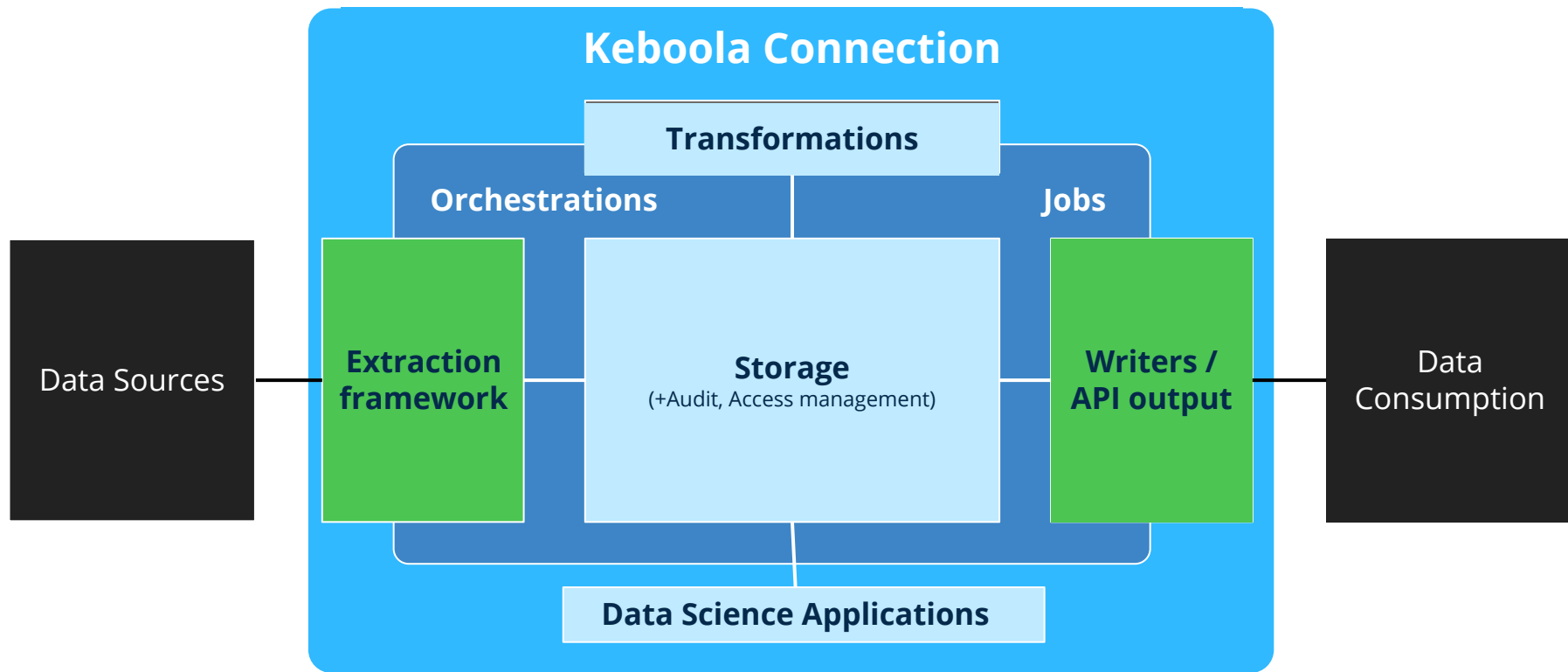
The most typical use case is feeding data into a **Business Intelligence** system. We partner with most of the top market players and integrate your data into their systems.

We also understand that beyond **dashboards** there is greater value in reusing data for efforts like **re-marketing** or **campaign planning**, so we don't limit users by where and how they can send their data.



[See for yourself how easy it is to configure a writer for Tableau.](#)

ORCHESTRATIONS & JOBS



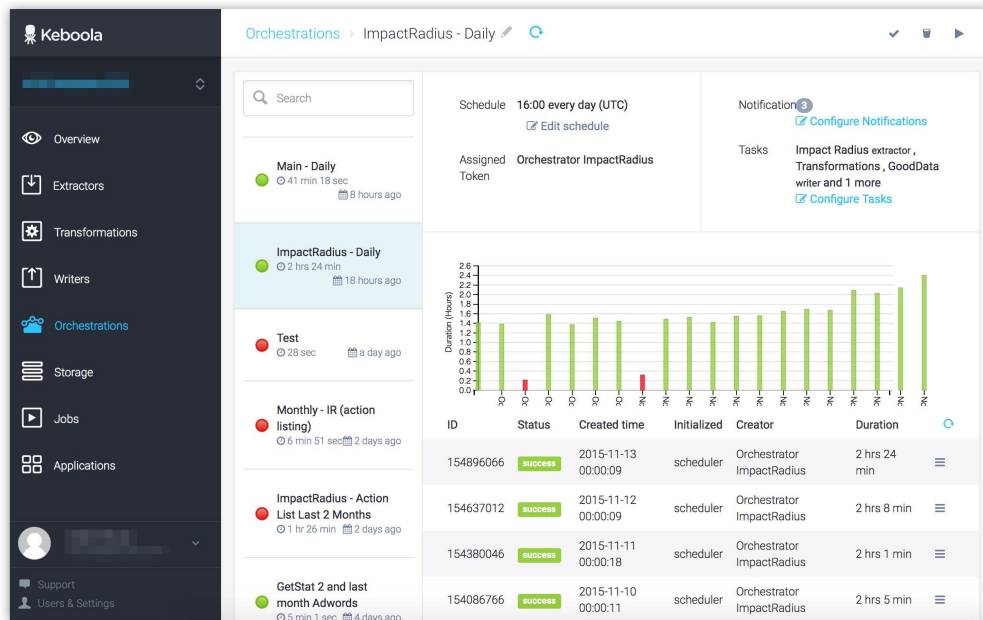
Fully automated data flow with security and transparency.

ORCHESTRATION & JOBS

Orchestration takes care of the automation of the whole sequence.

Parallel processing ensures time efficiency.

Jobs keep track of all tasks and help oversee performance of individual components. It means that **transparent security mechanisms** are in place throughout the whole process.



SECURITY COMPLIANCE

Security of data and the whole environment where corporate data travel and rest **is the core concern**.

Keboola is fundamentally built and operated on top of Amazon AWS and other 3rd party services such as PaperTrail (logging), PagerDuty (alerting) or NewRelic (monitoring). Keboola inherits from Amazon important security characteristics for data storage, encryption, access control, archiving and others. Detailed description of underlying Amazon security concerns are well documented in the Amazon and KBC whitepapers:

- [aws amazon security whitepaper](#)
- [aws amazon security in context of Singapore Privacy Considerations](#)
- [Keboola Connection security whitepaper](#)





Pavel.bulowski@keboola.com