



Introduction

Keboola Connection

A new generation platform for fast and transparent data integration in the cloud.

Unique ability to **instantly** integrate, blend and enrich **thousands** of data sources

Agile deployment of **data science** applications; fast iteration

10x faster implementation than the best existing solutions available

Keboola Connection Overview

- All your business data in one platform — ETL (Extract, Transform, Load)
- Easy framework for pulling data from over **170,000 data sources**
- Automated, **transparent** and completely **secure**
- Appstore for data science **applications** ready to deploy; apps can be created by 3rd parties
- **Scalability** on AWS cloud infrastructure
- **Data democratisation**, giving access to the right people to the right data at the right time

Keboola Connection Value

- Connecting your data without worrying about maintenance; everything easily accessible on demand in a unified format.
- Plugging in new data sources is fast and easy; no programming is required.
- Planning and monitoring sequences of tasks of any complexity. You don't have to touch a thing.
- Marketplace for data science application in case you don't have a full time data scientist.
- Using the right muscle for each job. From MySQL to Redshift and SQL to R, we cover you. Pay for what you really use. #goCloud
- Delivering data and insights across the organisation, targeted to fit everyone's business needs and goals.

Trends That Matter

30%

of enterprise access to broadly based big data will be via intermediary data broker services, serving **context** to business decisions by 2017.

Gartner [2015]

Context

35%

shortfall of data-savvy workers needed to make sense of big data by 2018.

McKinsey [2011]

Technology

60%

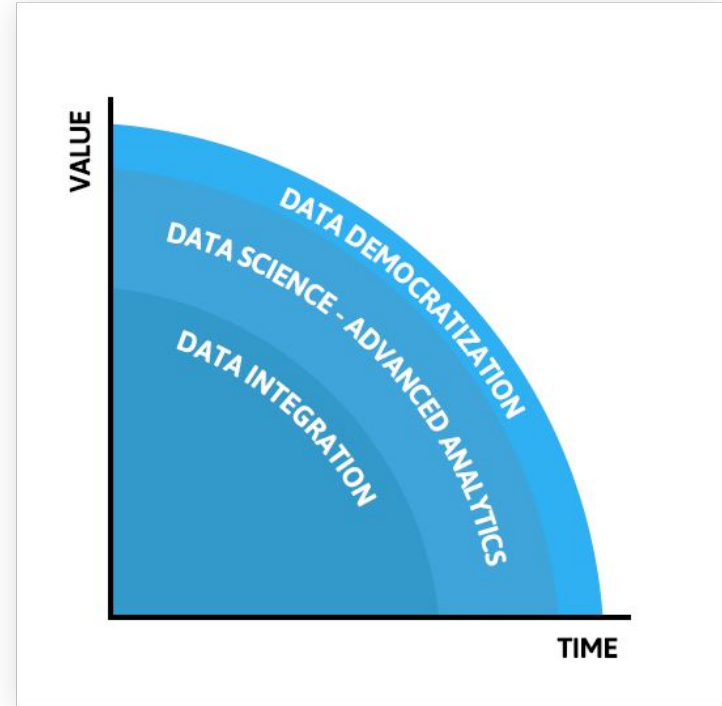
increase in operating **margin** of an average retailer if using big data to its full potential.

McKinsey's [2015]

Efficiency

Data Enablement Maturity

- 1) **Data integration** – organisations struggle to bring all their meaningful business data under one roof. **Consolidation** typically takes up to **85%** of their analytical capability.
- 2) **Advanced analytics** – these companies have cracked their basic operational, 'static' reporting and are working on **predictive** modeling, advanced segmentation, leveraging **machine learning** and similarly evolved practices.
- 3) **Data democratisation efforts** – these companies truly understand what being **#datadriven** means and distribute information vertically for all levels of employees to leverage and create value.



Challenges We Solve

1

Connecting Your Data

Our open platform makes it easy to quickly connect relevant data from many sources and provide context.

2

Reducing Cost of Data Infrastructure

Our automated solution removes the burden of updating your home grown or legacy ETL, so you can focus on solving your business issues.

3

Advanced Analytics

We enable organisations to quickly deploy data science applications to get new value from your data faster.

Client Problems We Solve

Marketing Director

Data Overload - Dealing with many data sources and data types; consolidating everything for our final reporting takes full time resources.

Not Enough Time - New campaigns start before we are finished evaluating the last one; it's time consuming.

Offline vs. Online - How do we find the link?

Attribution - How do we allocate budgets?

Client Problems We Solve



Head of IT

Data Infrastructure – Managing our own ETL system is not efficient. When one of our key systems changes API, we have to update our entire system. Rewriting scripts and process workarounds is not sustainable.

Distribution – Matching the right people with the the right data is very difficult. We are mainly demand driven.

Human Factor – Nobody is 100 % reliable and mistakes happen. The whole data management and flow are not transparent enough and are error-prone.

Client Problems We Solve

Business Analyst / Data Scientist

Disjointed Data - Rather than developing a new hypothesis, we spend time organising disjointed data.

Data Hygiene - Lots of time is being wasted on data quality issues. Fixing this never ends.

Iteration Speed - Preparing and implementing new algorithms is ineffective and their iteration is time consuming.

Use Cases

Keboola in Retail

Challenge

A **retailer** with a chain of stores and a growing e-commerce platform needed insights into the **purchase behaviour** of their customers.

- disconnected data in warehouse and siloed cloud based systems
- lack of 'data science' skills in house

Big Goal

Drive more revenue & increase **Customer Lifetime Value** (customer lifetime value)

Data Sources

- loyalty management system
- point of sales system in store
- point of sales e-commerce

Solution

Required data is integrated in the **Keboola Connection** ETL, where it is cleaned and prepared for further analysis.

Custom or pre-built R algorithm is deployed to run advanced RFM segmentation to be able to **predict a next purchase** and plan more **effective campaigns**.

Result:
Increase in purchasing frequency and higher CLV

Use Cases

Keboola in Omni-Channel Marketing

Challenge

A client from the financial sector with several business units generates new business through mix of offline activities and online campaigns.

Their marketing department didn't have **clarity** on the influence of different channels on attribution beyond the point of the first **conversion**.

- multiple disconnected channels and online reporting systems
- complexity of ownership of different data sources

Big Goal

Complete **accountability** in marketing efforts

Transparency from customer acquisition to their life value

Ability to identify most profitable customer groups by origin

Data Sources

- **Google Analytics**
- **Facebook Ads & Insights**
- **Google Adwords**
- **Salesforce CRM**
- **Offline campaigns (events)**

Solution

Required data is integrated in the Keboola Connection ETL, where it is cleaned, connected and prepared for further analysis.

Advanced transformation and **R** algorithms are used to map the entire consumer journey from the first **advertisement interaction** to **conversion** to **repeated purchases**.

Result:
Visibility on customer groups by CLV
(customer lifetime value)
Clarity on effect and symbiosis of certain channels

Unique Value Proposition

Data **integration** of thousands of online and offline data sources with **unprecedented speed**.

Leveraging easy to scale.
Pay what you use.
Cloud based platform.

Network of specialized **partners** around the world within the reach of the Keboola **ecosystem**; specialized solutions from marketing **attribution** to retail analytics or natural language processing.

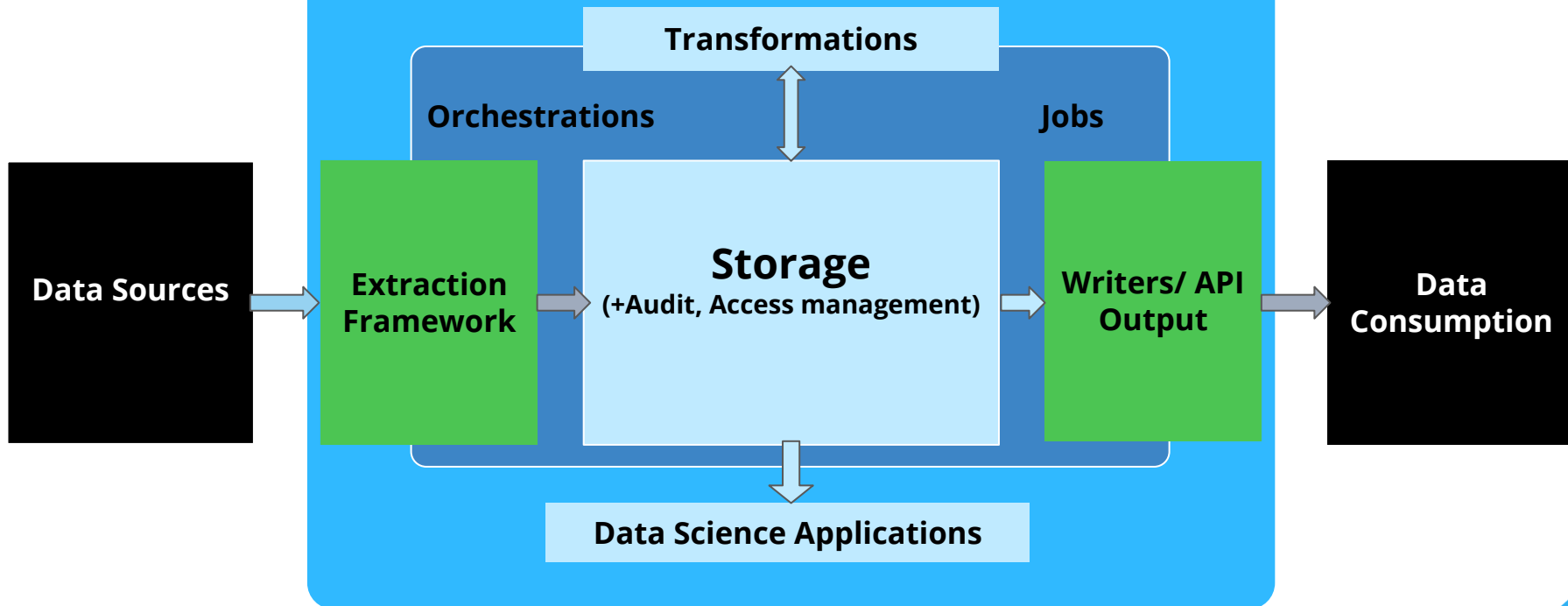
No vendor lock in. We value **freedom** and believe that customers should always be able to take their data away if they want to. So we introduced **Data Take Out**.



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

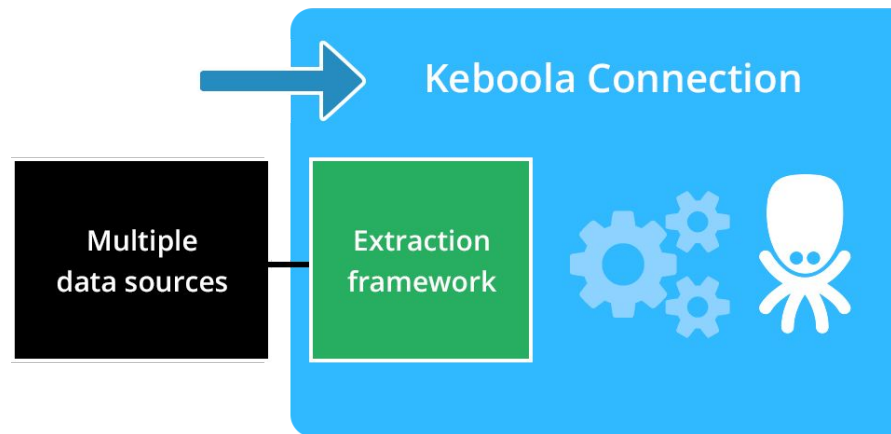
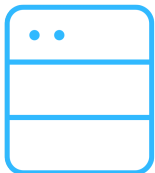
CSV, XLS

Database

SQL & NoSQL / Cloud & On-prem

DWH / Storage

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

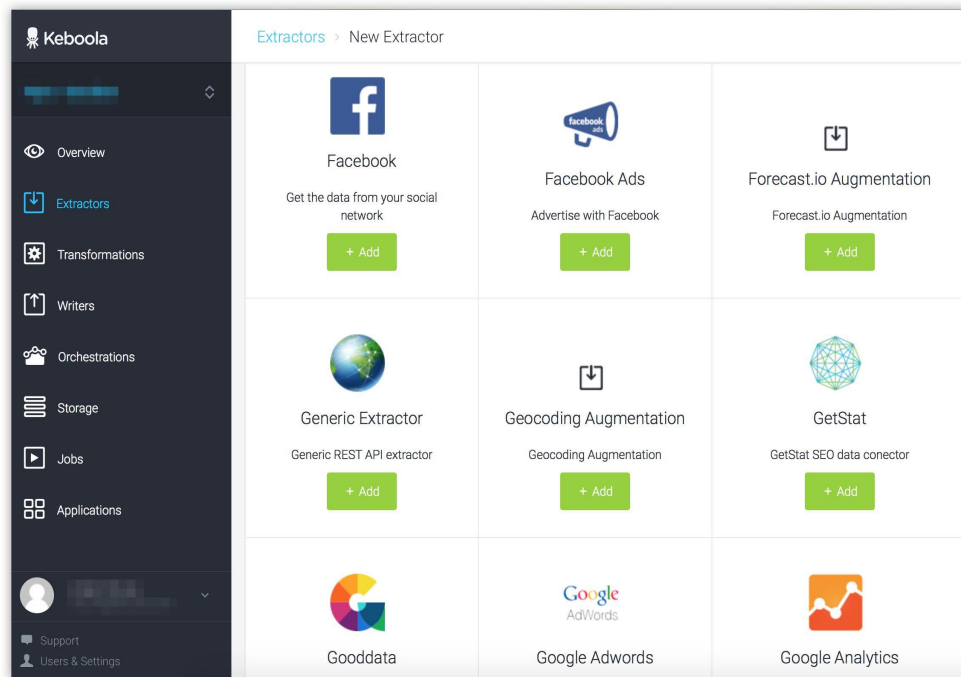


Over 170,000 data sources at your fingertips, instantly!

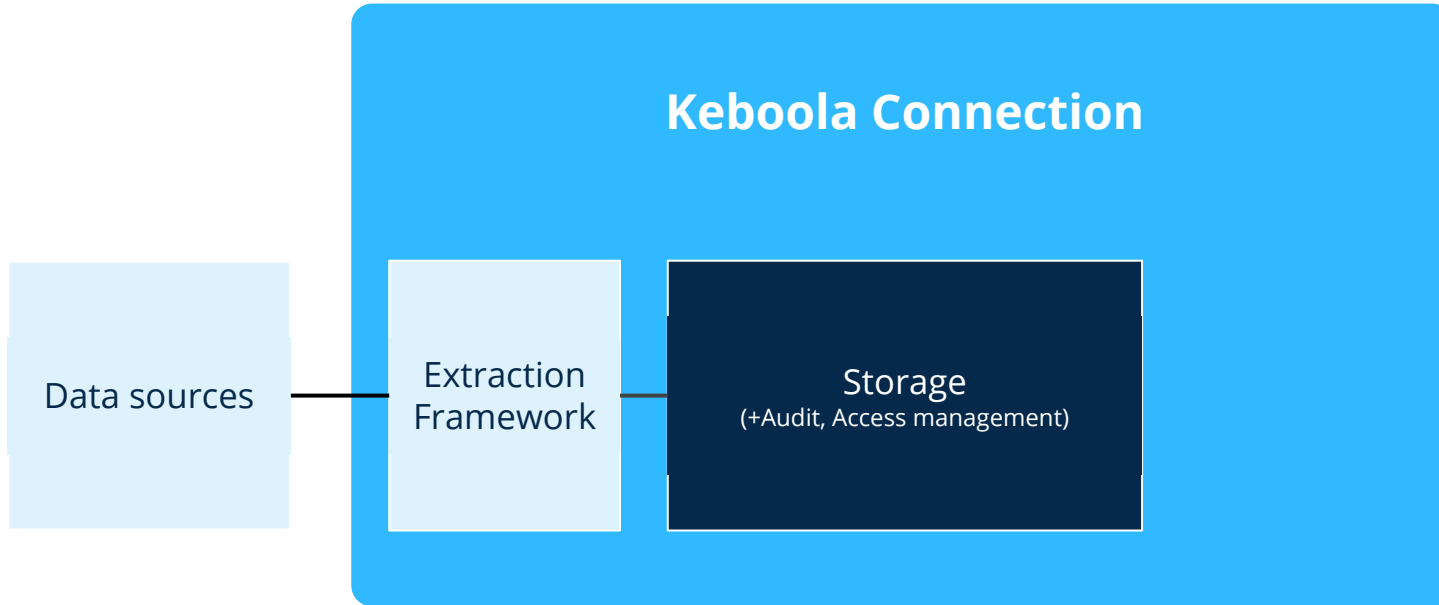
EXTRACTORS

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.

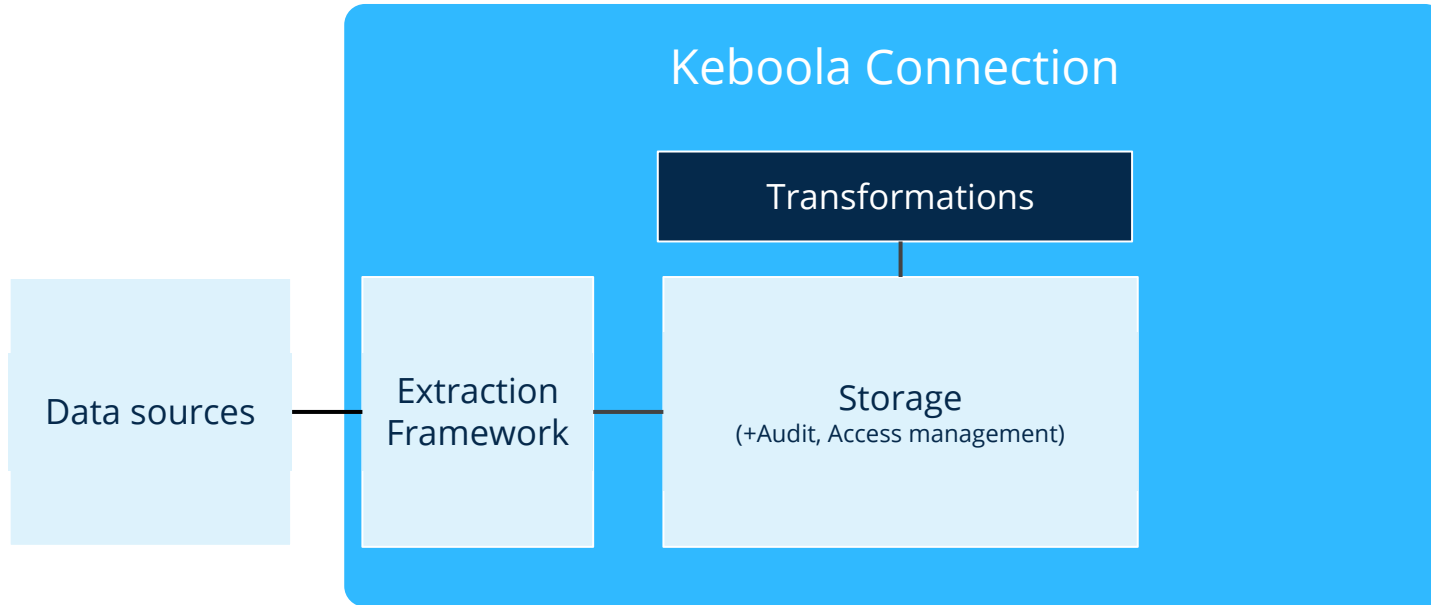
This log is visible and searchable to all users of the project and offers a full rollback function.

The screenshot displays the Keboola Storage API interface. On the left is a dark sidebar with the Keboola logo and a navigation menu including Overview, Extractors, Transformations, Writers, Orchestrations, Storage (highlighted), Jobs, Applications, and Users & Settings. The main content area is titled 'Storage' and features a search bar and a 'BUCKETS' section with a '+−' icon. A list of buckets is shown, with 'in.c-ex-api-generic-tm1' selected and expanded to show 'Cubes', 'results', 'views', 'views_Columns', 'views_Rows', and 'views_Titles'. The 'Cubes' sub-section is further expanded to show a list of cubes including 'in.c-ex-api-impactradiu...', 'in.c-ex-api-vend-vend', 'in.c-ex-api-zendesken...', 'in.c-ex-google-drive-ad...', 'in.c-ex-silverpop-secu...', 'in.c-ex-twitter-vojtiskovo', 'in.c-main', 'in.c-pigeon-stress-test', 'in.c-storage', 'out.c-main', 'sys.c-doctool', 'sys.c-ex-db', 'sys.c-ex-google-analytics', and 'sys.c-ex-google-drive'. The right-hand pane shows the 'Overview' tab for the selected cube 'in.c-ex-api-generic-tm1 / Cubes'. It includes tabs for Overview, Events, Data sample, Snapshots, and Graph. Below the tabs is an 'Actions' dropdown. A table displays the cube's metadata:

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

Below the table is a 'Latest imports' section with a small bar chart showing import activity over time.

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.

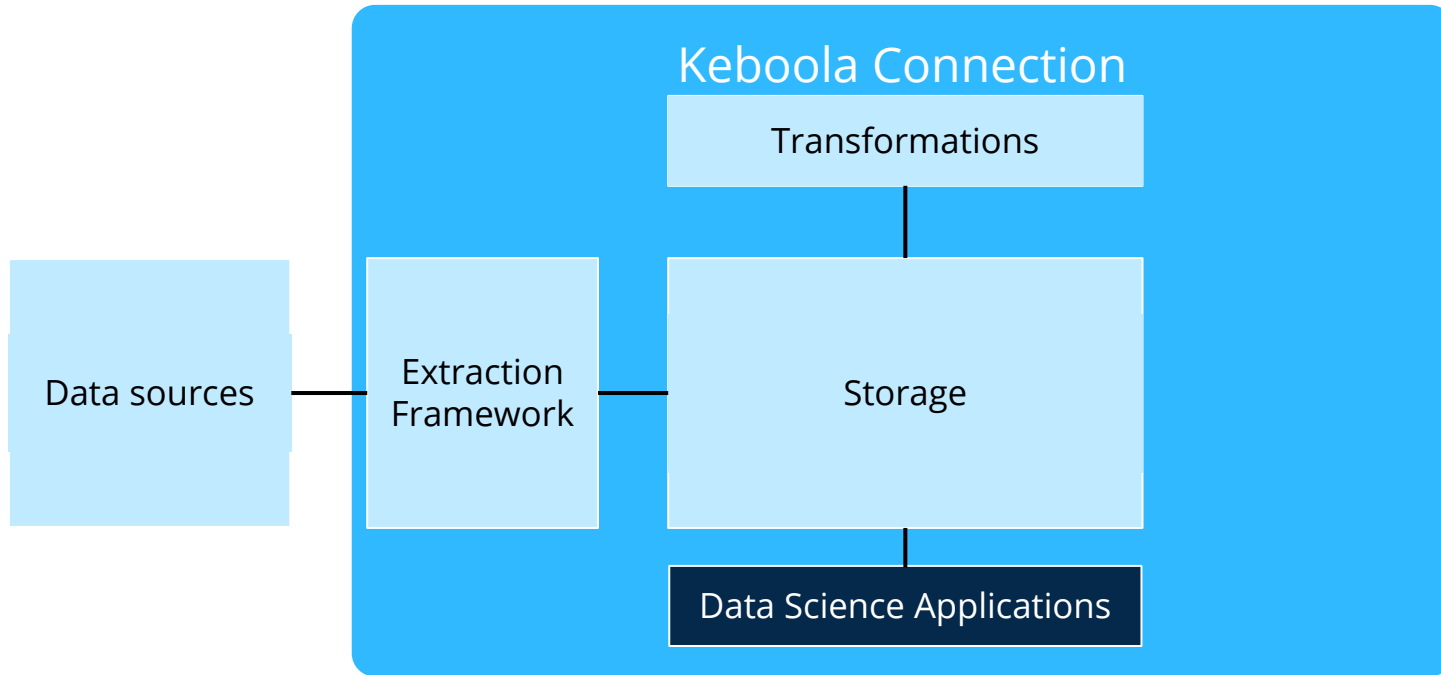
Transformations can be saved as **templates** and reused across multiple projects.

The screenshot displays the Keboola web interface for configuring a transformation. The left sidebar contains navigation options: Overview, Extractors, Transformations (highlighted), Writers, Orchestrations, Storage, Jobs, and Applications. The main content area is titled 'Transformation main' and is divided into three sections:

- Input Mapping:** Lists four input sources, each with a size and a 'mysql' backend:
 - 1.79 MB: in.c-ex-facebook-main.posts → in.posts
 - 393.22 kB: in.c-main.url_pairs → in.url_pairs
 - 7.37 MB: in.c-ex-google-analytics-main.pages → in.pages
 - 65.54 kB: out.c-main.pages → in.fb_pages
- Output Mapping:** Shows one output destination:
 - traffic → 5.03 MB mysql out.c-main.traffic
- Queries:** A text editor containing a SQL query to create a table and select data from a source table.

```
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

The screenshot displays the Keboola 'Applications' interface. On the left is a dark sidebar with navigation options: Overview, Extractors, Transformations, Writers, Orchestrations, Storage, Jobs, Applications (highlighted), Support, and Users & Settings. The main area is titled 'Applications > New Application' and contains a grid of application cards. Each card features a logo, a title, a description, and a '+ Add' button.

Application Name	Description	Status
docker	AI 4 BI (Beta)	+ Add
Anomaly detection	Recipe for detecting anomalies	+ Add
Basket analysis	Recipe for basket analysis	+ Add
Correlations	Recipe for correlations	+ Add
Custom science Python (Alpha)	Custom science Python	+ Add
Custom science R (Beta)	Custom science R	+ Add
Geneea NLP Analysis	Analyze text such as Extract Entities, Hashtaocina, Identify language.	
Data Type Assistant		
Grouped histogram		

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



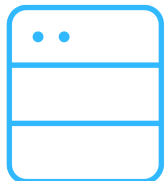
Business Intelligence

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



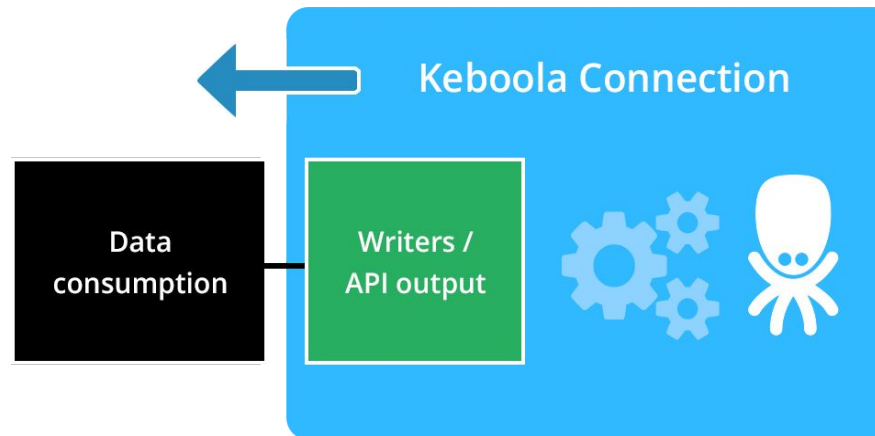
Marketing Solutions

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.



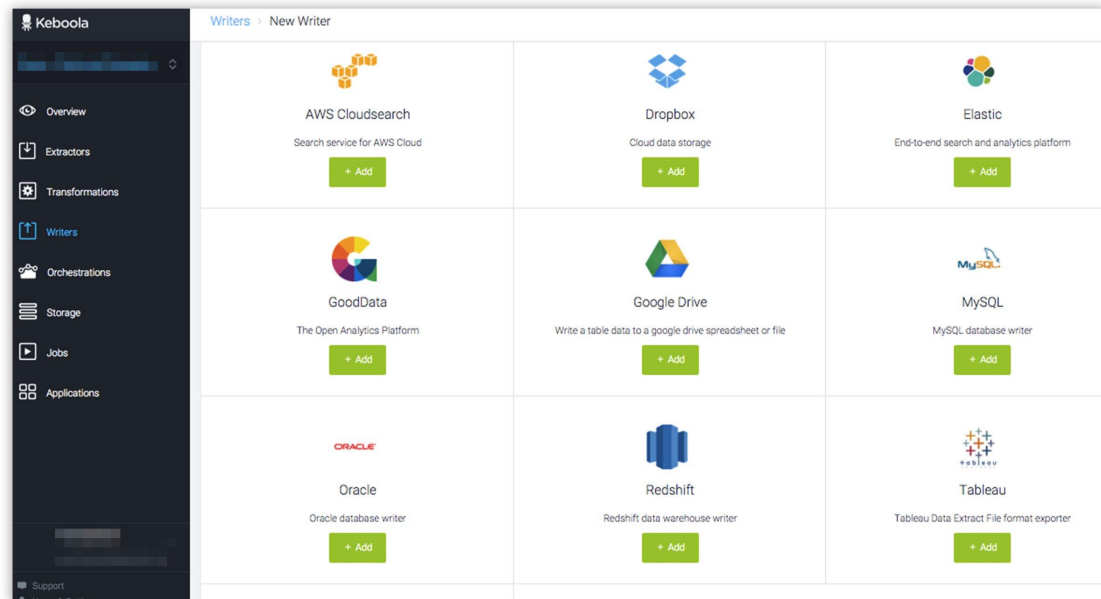
Dashboards are not the only way of consuming data.

WRITERS

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

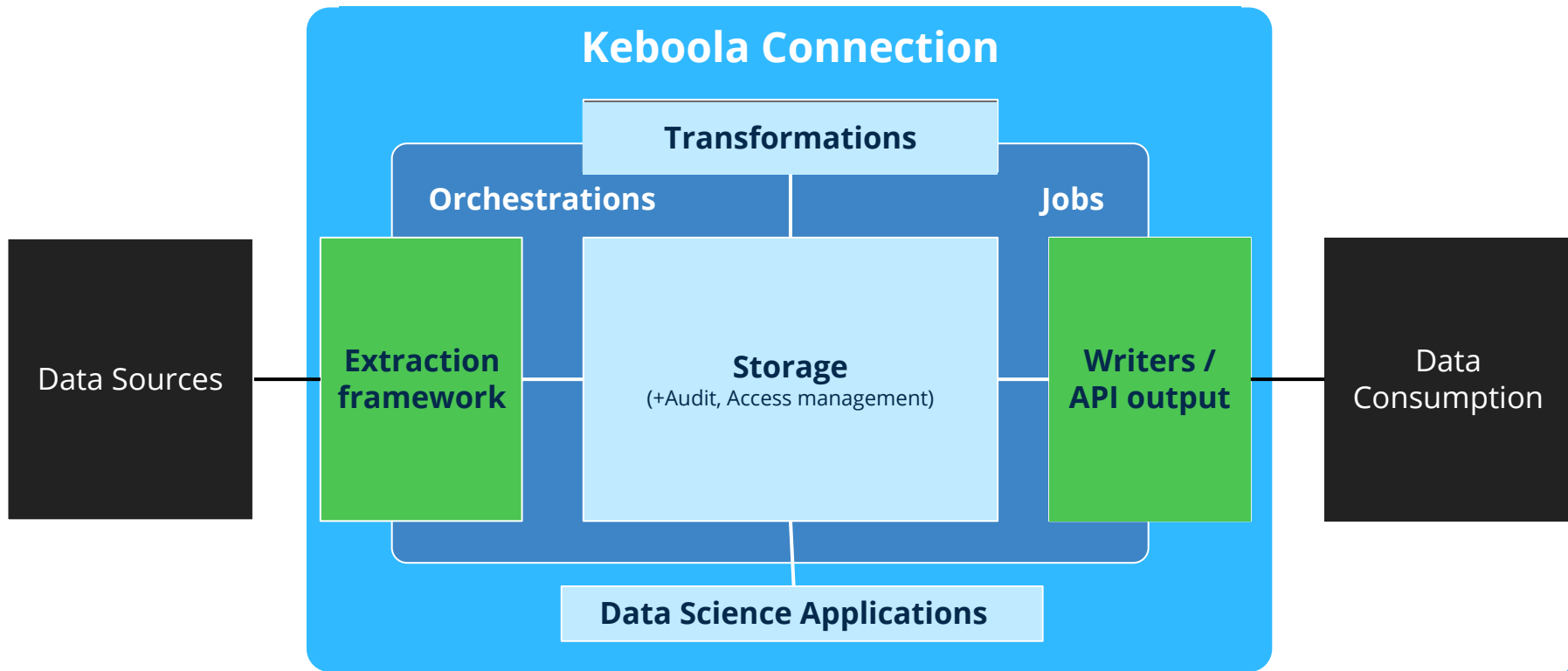
The most typical use case is feeding data into a **Business Intelligence** system. We partner with most of the top market players and integrate your data into their systems.

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



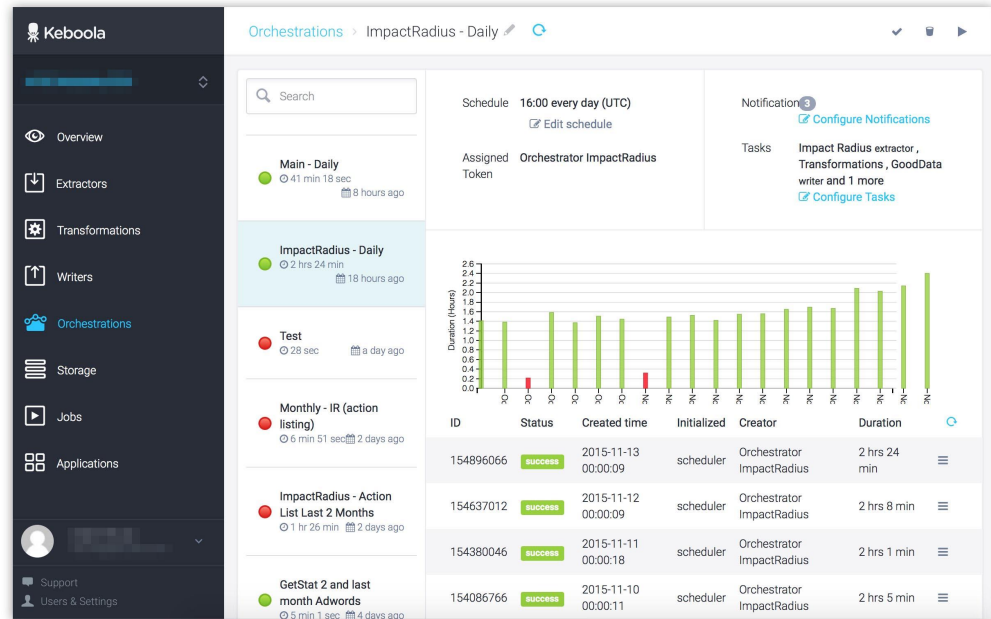
Fully automated data flow with security and transparency.

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

