

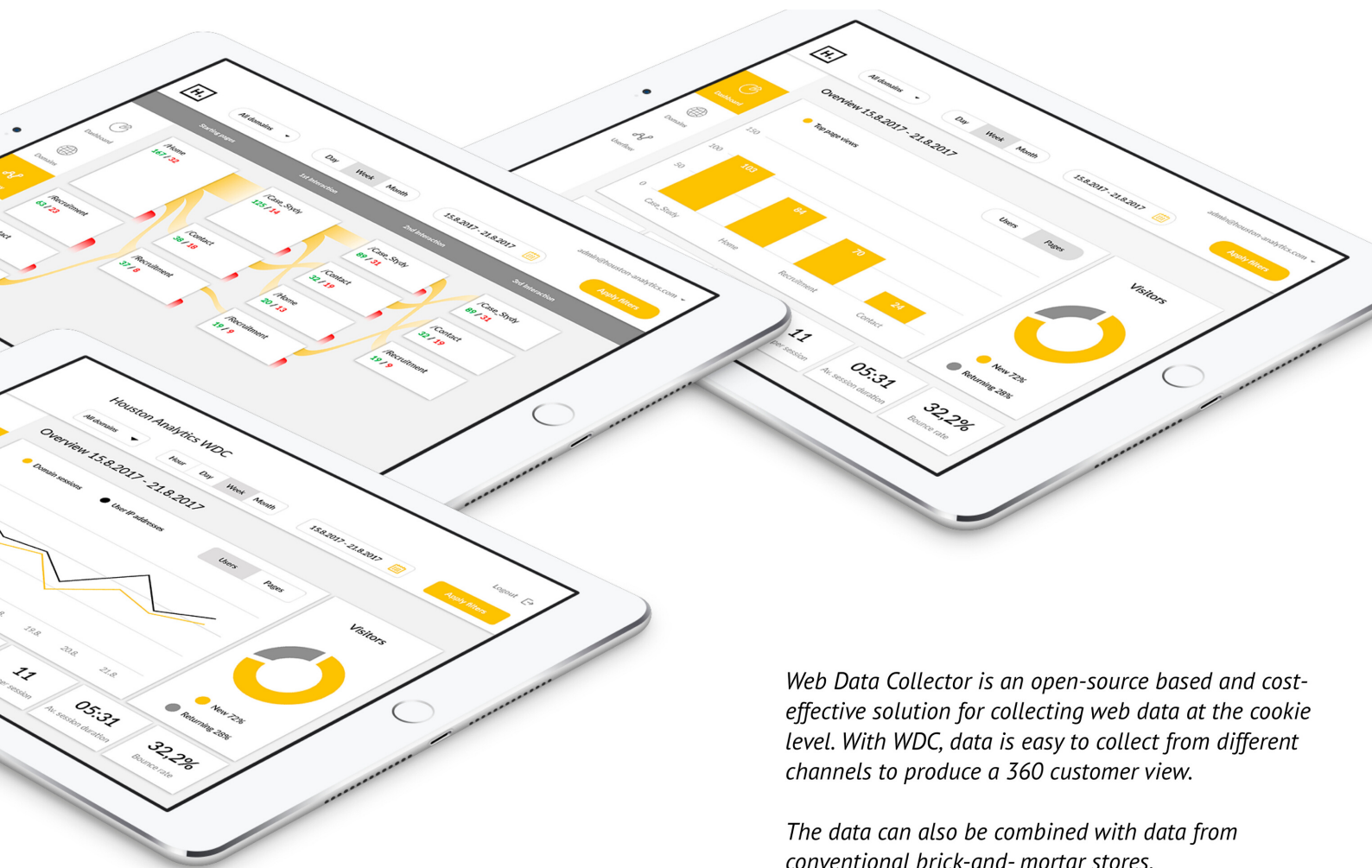


HOUSTON ANALYTICS

# WEB DATA COLLECTOR

Data sheet





*Web Data Collector is an open-source based and cost-effective solution for collecting web data at the cookie level. With WDC, data is easy to collect from different channels to produce a 360 customer view.*

*The data can also be combined with data from conventional brick-and-mortar stores.*

## WDC SOLUTION ARCHITECTURE

The Houston Analytics Web Data Collector (WDC) is a cloud-based SaaS offering for capturing data from multiple channels with a tag-based implementation. It is based on the Snowplow open-source web analytics platform. Snowplow is a complete web analytics platform that allows businesses to capture, store and analyze granular customer-specific event-level data. It is designed to gather all kinds of web data in different formats in a fast and comprehensive way, enabling a cost effective, personalized customer dialogue. WDC is created by analysts to analysts as a mean to achieve a 360 view of customers and prospects, enabling analytics and real-time interaction for personalized service.

Much effort has been put in to make the product easy and quick to install, configure, run and maintain, resulting in record time deployment into production.

Data privacy (e.g. GDPR), transparency and ownership requirements have been a core design requirement from start. With WDC the customer owns the data, and the data collection technology (end user facing technology) is open source for maximum transparency.



# WDC UPDATES AND VERSION CONTROL

Being a SaaS product, all upgrades are included in the monthly fee and can, if otherwise not agreed, be rolled out without customer intervention. Significant upgrades that may require customer actions will be announced two months in advance.

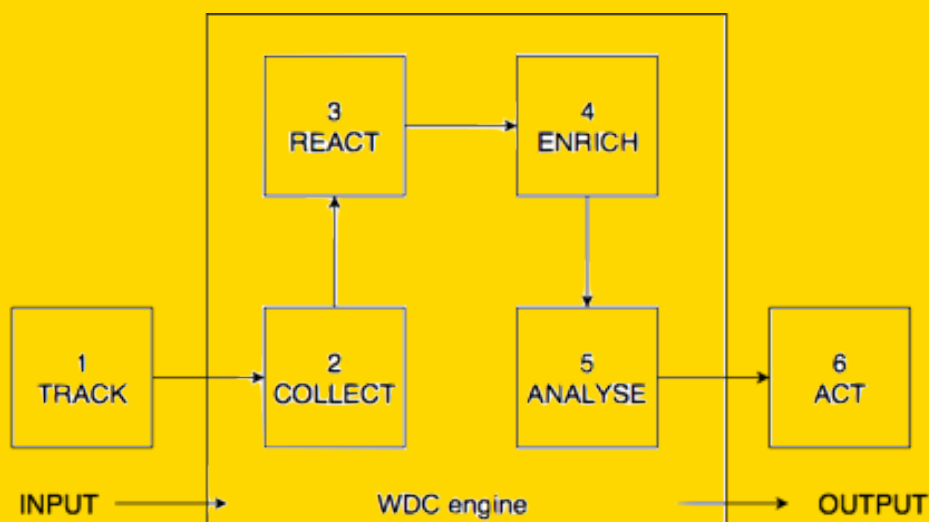
Backward compatibility in integrations is guaranteed.

## WDC SETUP AND ELEMENTS

WDC is a scalable data collecting framework that consists of 3 separate areas:

- Snowplow open-source data tracking framework on the frontend-
- Proprietary scalable cloud-based backend framework built on top of AWS infrastructure
- Raw data output and BI integration ready (e.g. Microsoft PowerBI)

The WDC consists of 6 service layers, that can be extended and/or configured for client use – or then used based on default settings out-of-the-box.



## 01 TRACK

The data tracking part of WDC is based on the Snowplow open-source framework <https://github.com/snowplow/snowplow>. The open nature of the tracking script and all the ready-made connectors ( iOS, Android, IoT, web and multiple programming language specific SDK's ... ) ensure transparency and extensive compatibility to different interfaces. Technical documentation of the data tracking: <https://github.com/snowplow/snowplow/wiki/Snowplow-technical-documentation>

## 02 COLLECT

Tracking pushes a data stream to the WDC data collector service. The collector is designed to scale automatically to ensure all data gets through and is controlled for quality. Option to stream raw data to customer premises is set up from the data collector. The data is available in near-real time (milliseconds).

## 03 REACT

Real time decision-making and triggers based on event stream data and rules can be set up in this part of WDC. This enables near real-time interactions with customers/users across all channels, web sites, videos, newsletters, social media, applications (like for instance CRM, MA etc).

## 04 ENRICH

Data enrichment such as IP address lookup and other business specific data can be injected into the data stream. Data stream analysis on data happens in this element. The data here is available in the millisecond to second levels.

## 05 ANALYZE

Collected data is normally available within one minute (AWS restriction), and in case local storage is used, this can be significantly faster. WDC provides out-of-the box functionality for web data collection, aggregation, enriching and visualization, as well as the fundamental ideology of forwarding the relevant share of web data for further processing in an Enterprise Data Warehouse like Teradata and/or other data architecture core components like Data Lakes/Hadoop etc. Standard WDC offers ODBC/JDBC links for integration with other systems.

## 06 ACT

WDC delivers raw data and support to BI systems (e.g. Microsoft PowerBI) to support management and relating decision-making. The data is owned by the entity licensing the WDC and it can be easily integrated into most reporting solutions and marketing automation systems.

## WDC CLOUD ARCHITECTURE ON AWS

The system is deployed in Amazon AWS EU. The following Amazon provided services have been used:

- Amazon EC2
- Amazon S3
- Amazon ELB
- Amazon Kinesis Stream
- Amazon Kinesis Firehose
- Amazon Redshift

# PRICING STRUCTURE AND COST BASE

Pricing is based on sessions. The definition of session is based on the Snowplow tracking cookie and its lifetime of 30 minutes.

SQL line of amount of sessions:

```
select app_id, DATE_PART(month, etl_tstamp), DATE_PART(year, etl_tstamp), count(distinct(domain_userid || domain_sessionidx)) from atomic.events
```

app\_id = business identifier  
atomic.events = WDC database

WDC pricing starts from 150€ / month (VAT 0%) for 50.000 monthly sessions.

WDC setup and configuration is sold separately if required.

## ACT / DATA OUTPUT

WDC delivers raw data and access rights for the customer to act on it. Customers are the data owners allowing them to use it in any way they want, e.g. display it in a BI solution. Ready-made BI templates are offered by Houston Analytics to ease the implementation costs.

**NOTE!** Customers are required to purchase the BI licenses separately, as they are not part of WDC SaaS offering. WDC templates for BI solutions are not part of WDC SaaS offering, and they are offered "as is".

## CONTACT

For further information, please contact:

[missions@houston-analytics.com](mailto:missions@houston-analytics.com)