



DATAIKU

BIG DATA – BUSINESS INTELLIGENCE – MACHINE LEARNING



1. ¿QUÉ ES DATAIKU?

Dataiku DSS es una herramienta colaborativa de tratamiento de datos donde pueden **trabajar conjuntamente** científicos de datos, ingenieros de datos y analistas de datos. A su vez tiene paneles dedicados a la monitorización de los equipos para ayudar a los gerentes de proyecto a controlar la actividad de su equipo.

Dataiku ha sido nombrada como **líder en el Cuadrante Mágico de Gartner**.

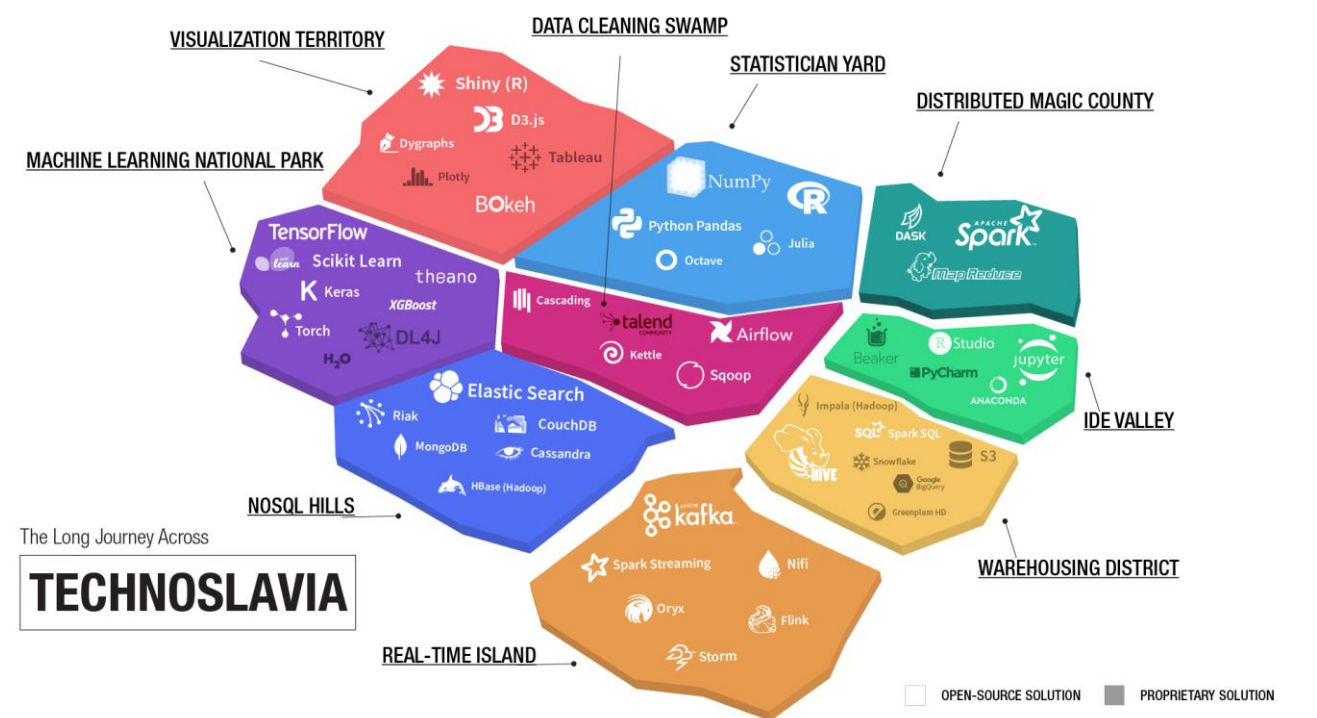
Figure 1. Magic Quadrant for Data Science and Machine Learning Platforms



Source: Gartner (February 2020)

Ofrece una versión gratuita de prueba y varias versiones de pago que mejoran significativamente sus funcionalidades y su capacidad de compartir los proyectos con otras personas (Información de las versiones ofrecidas: [pinche aquí](#)). Fijan los precios de las licencias en base a la cotización.

Dataiku se apoya en las siguientes tecnologías:



2. DESPLIEGUE

Dataiku DSS está diseñada para desplegarse como **servidor Linux** y no puede ejecutarse de forma nativa en Windows. Puedes conectarte al servidor desde el navegador de tu equipo, independientemente de su sistema operativo.

Ofrece la posibilidad de desplegarlo como contenedor utilizando Docker o Kubernetes.



3. ¿QUÉ OFRECE?

Dataiku DSS es una herramienta All-in-One, que aborda todo el proceso de transformación de datos desde su preparación hasta su análisis, enfocada en actividades de IA y machine learning.

Hace énfasis en no tener la necesidad de utilizar código para sus distintos procesos.



Conectividad

Se puede utilizar una gran cantidad de fuente de datos, desde ficheros de datos hasta bases de datos analíticas MPP de forma ya implementada.

Con la posibilidad de conectarse a casi cualquier fuente de datos gracias a la utilización de plugins. Estos plugins se pueden crear usando R o Python para conectarse a cualquier API, base de datos o fichero. También se pueden obtener ya implementados en el repositorio de la propia comunidad.

New dataset

Files Upload your files Server's Filesystem Files in folder	Hadoop <i>Hadoop connection is not enabled on your DSS instance. Please contact your administrator</i>	SQL MySQL PostgreSQL Vertica Amazon Redshift Snowflake IBM Netezza Other SQL Greenplum Teradata Oracle MS SQL Server SAP Hana Google BigQuery Athena	Cloud Storages Amazon S3 Azure Blob Storage Google Cloud Storage FTP SFTP SCP HTTP HTTP (with cache)
NoSQL MongoDB Cassandra ElasticSearch	DSS Managed dataset Folder Dataset from another project Metrics Internal stats Editable	Import existing Choose connection to import from Import from catalog	More dataset options? Browse plugins

Preparación de datos

Con **Dataiku** se puede transformar y preparar datos de forma intuitiva con una interfaz simple, viniendo integrados potentes procesos de transformación de datos.

También ofrece la posibilidad de crear tus propias transformaciones mediante código (Python, R, SQL, Shell, Hive, Impala, Pig, Spark SQL, Spark Scala, PySpark, SparkR o Sparklyr).

Estas transformaciones se pueden guardar y compartir.

Visual recipes	Code recipes
Sync	Python
Prepare	R
Sample/Filter	SQL
Group	Shell
Distinct	Hive
Window	Impala
Join with...	Pig
Split	Spark SQL
Top N	Spark Scala
Sort	PySpark
Pivot	Spark R
Stack	

Aunque incorpora buenos procesos de transformación de datos no incluye características como diccionarios de datos, gestión de metadatos, gobernanza de datos... Como sí nos ofrecen otras herramientas como Talend o Pentaho Data Integration, entre otras.

En este video ([pinche aquí](#)) tenemos una muestra de lo que nos puede ofrecer la herramienta.

Machine Learning

Esta herramienta facilita la implementación de ML al modelo de datos y de manera visual, aprovechando las bibliotecas de Scikit-Learn, MLlib, XGboost. Puedes importar cualquier biblioteca externa a través de una API o crear, entrenar y desplegar tu propio modelo de ML utilizando Python o R (incluye soporte para notebooks). Además, puedes reutilizar tus modelos de entrenamiento ya creados previamente.

- **Aprendizaje automático**

La propia herramienta genera de forma automática el modelo de entrenamiento y sus características adaptándose a los tipos de datos que utilizamos. Compara decenas de algoritmos y muestra los que mejor se adaptan al modelo para que nosotros seleccionemos que algoritmo utilizar.

- **Deep learning**

Dataiku te permite definir la arquitectura y personalizar los ajustes de entrenamiento de tu modelo. También maneja imágenes de forma automática, incluidas sus características. Incluye soporte para Keras, Tensorflow backend, integración con Tensorboard y puedes entrenar tu modelo aprovechando la GPU de tu máquina.

En este campo te ofrece de manera similar lo mismo que DataRobot o Alteryx pero soportando a la vez potentes transformaciones de datos y visualizaciones de forma intuitiva.

En estos dos vídeos se muestra un ejemplo de integración de ML al modelo de datos:

[Vídeo 1](#)

[Vídeo 2](#)

Visualización de datos

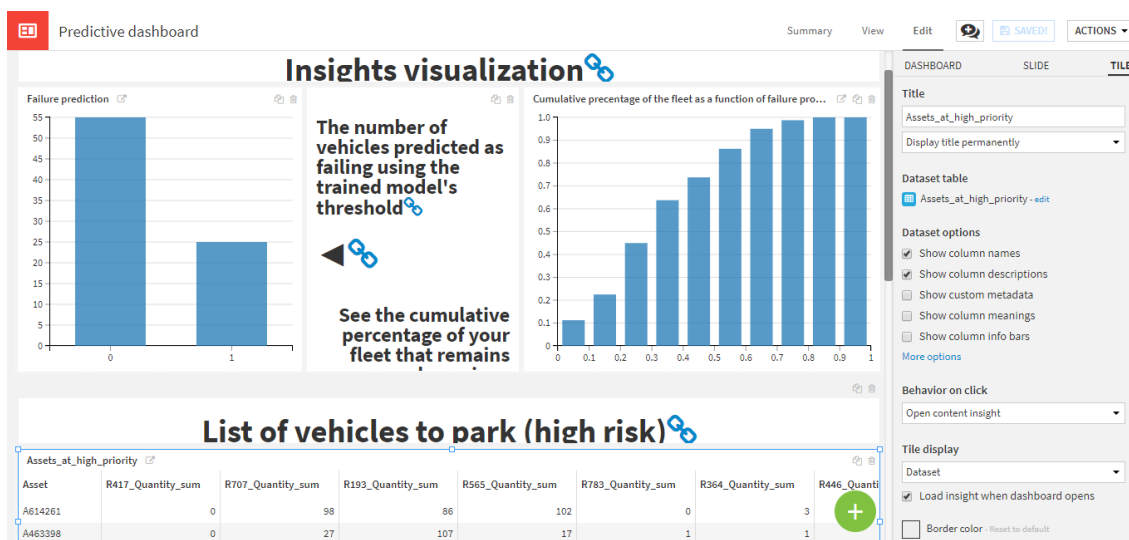
Dataiku señala las inconsistencias de los datos con los que estás trabajando y genera informes de manera automática.

Te permite generar multitud de gráficas y mostrar reportes integrando soluciones de Tableau o Qlik.

Tiene la posibilidad de ejecutar los análisis de datos en motores de Spark, Hadoop o SQL.

Y puedes desarrollar tus propias aplicaciones de visualización utilizando HTML, Javascript, Bokeh (Python) o Shiny (R). (Incluyendo notebooks)

Estas visualizaciones o reportes están diseñados principalmente a ayudar en los procesos de transformación de datos o machine learning pero no para ofrecer soluciones finales como sí lo pueden hacer herramientas de Microsoft (PowerBI, SSRS), Pentaho (PDA, PRD) o LinceBI.



Compartir conocimiento

Dispone de un repositorio donde se puede compartir conocimiento y añadir documentación para los miembros del equipo, esto también ayuda a la incorporación de nuevos miembros al equipo. Además, ofrece la posibilidad de mantener discusiones sobre los temas a tratar.

Control de versiones

Cada acción en el sistema se versiona y registra a través de un repositorio Git integrado lo que permite hacer un seguimiento de la progresión del proyecto y volver a versiones anteriores.

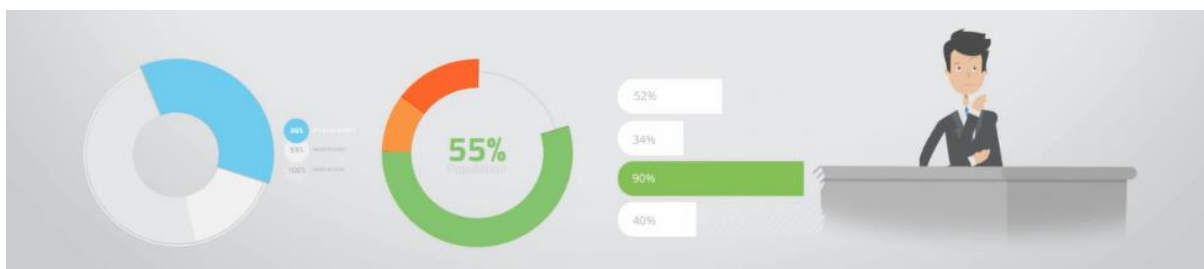
Seguridad

A nivel de seguridad se puede crear distintos tipos de usuarios con distintos accesos y permisos.

4. DESCRIPCIÓN DE STRATEBI

En Stratebi ofrecemos **gran cantidad de soluciones analíticas** por una compañía de **rápido crecimiento**, innovando en las áreas tecnológicas de mayor desarrollo en la actualidad: **Business Intelligence, Big Data y Social Intelligence**, muchas de ellas, basadas en soluciones **Open Source**.

Además, somos **Partners Certificados en Microsoft PowerBI y Vertica**, con gran número de proyectos con ambas tecnologías

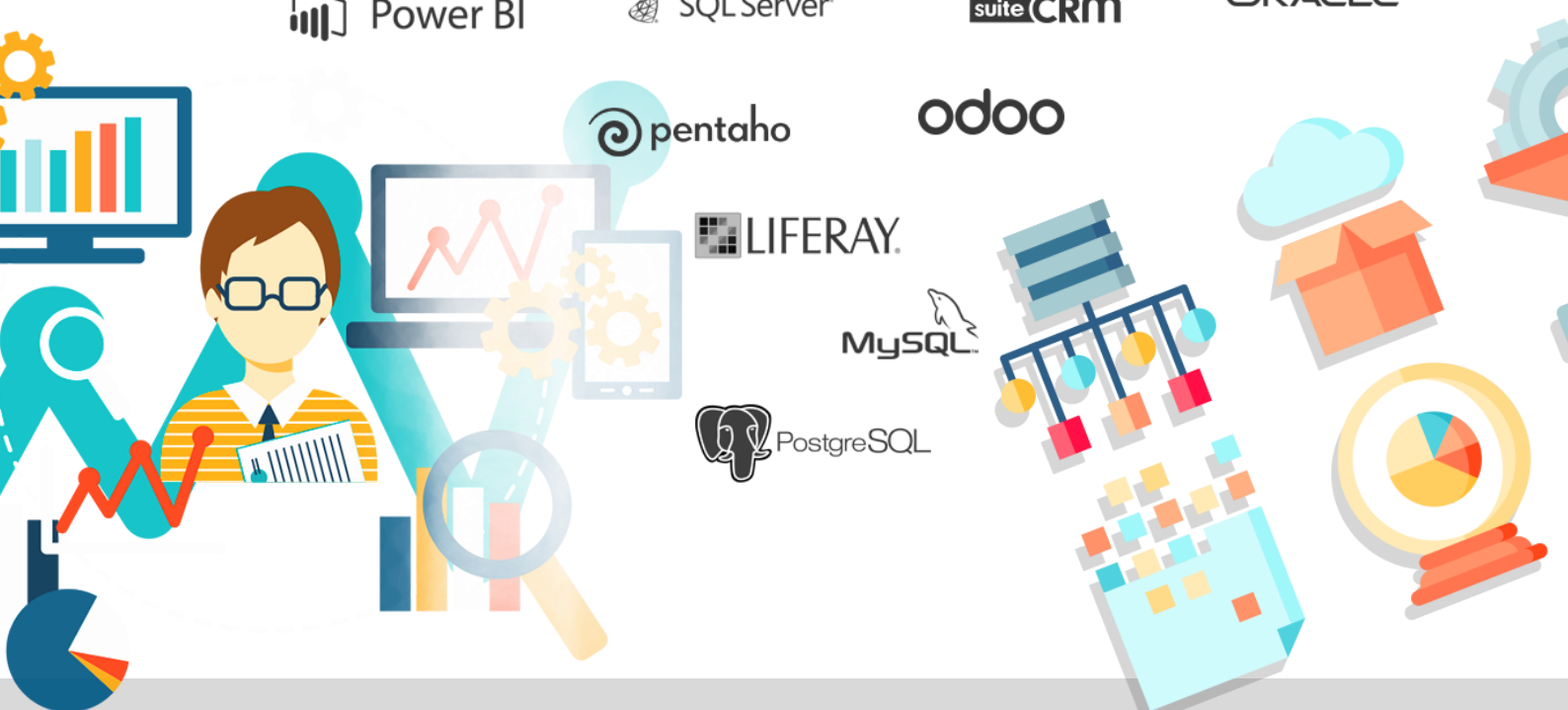


Desarrollamos nuevas soluciones analíticas basadas en Open Source, para la generación de Cuadros de Mando en tiempo real, con tecnologías IoT para SmartCities, machine learning, etc...



5. TECNOLOGÍAS

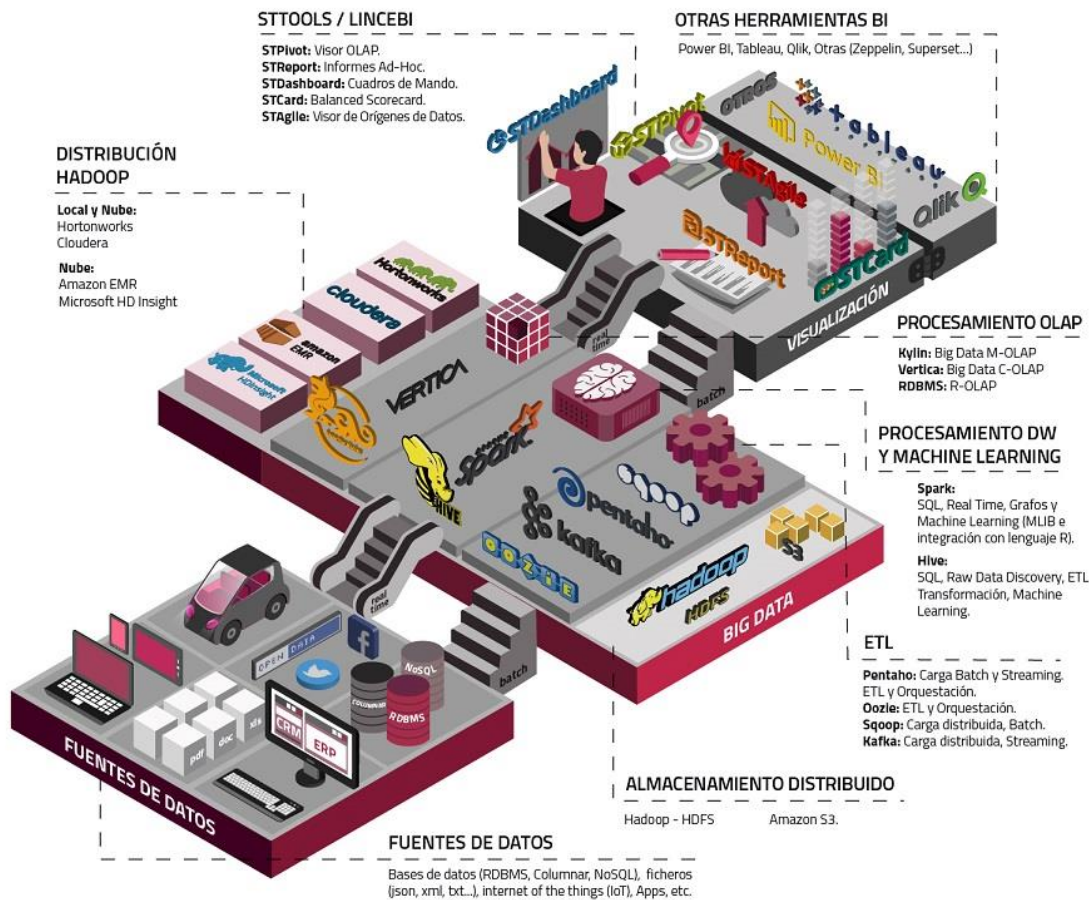
Recientemente, hemos sido nombrados Partners Certificados de Vertica, Talend, Microsoft, Snowflake, Kylligence, Pentaho, etc...



6. DESCRIPCIÓN DE LAS SOLUCIONES



Stratebi son los creadores de la solución LinceBI.com sobre la que podrás desarrollar e innovar.



7. INFORMACIÓN SOBRE STRATEBI



Stratebi es una empresa española, con sede en Madrid y oficinas en Barcelona, Alicante y Sevilla, creada por un grupo de profesionales con amplia experiencia en sistemas de información, soluciones tecnológicas y procesos relacionados con soluciones de Open Source y de inteligencia de Negocio.

Esta experiencia, adquirida durante la participación en proyectos estratégicos en compañías de reconocido prestigio a nivel internacional, se ha puesto a disposición de nuestros clientes a través de Stratebi.

Stratebi es la única empresa española que ha estado presente todos los Pentaho Developers celebrados en Europa (Mainz-Alemania, Barcelona, Lisboa, Roma, Amsterdam, Sintra, Amberes (2) Londres...), habiendo organizado el de Barcelona.

En Stratebi nos planteamos como **objetivo** dotar a las compañías e instituciones, de herramientas escalables y adaptadas a sus necesidades, que conformen una estrategia Business Intelligence capaz de rentabilizar la información disponible. Para ello, nos basamos en el desarrollo de soluciones de Inteligencia de Negocio, mediante tecnología Open Source.

Stratebi son **profesores y responsables de proyectos** del Master en Business Intelligence de la Universidad UOC, UCAM, EOI...

Los profesionales de Stratebi son los creadores y autores del primer weblog en español sobre el mundo del Business Intelligence, Data Warehouse, CRM, Dashboards, Scorecard y Open Source.

Stratebi es partner de las principales soluciones Analytics: Microsoft PowerBI, Talend, Pentaho, Vertica, Snowflake, Kyligence, Cloudera...

Todo Bi, se ha convertido en una referencia para el conocimiento y divulgación del Business Intelligence en español.

8. REFERENCIAS STRATEBI

Trabajamos en los principales sectores y con algunas de las compañías y organizaciones más importantes de España.

SECTOR PRIVADO

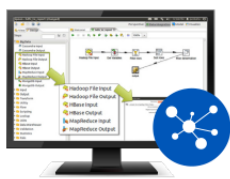


SECTOR PÚBLICO



9. EJEMPLOS DE DESARROLLOS ANALYTICS

A continuación se presentan **ejemplos de algunos screenshots** de cuadros de mando diseñados por Stratebi, con el fin de dar a conocer lo que se puede llegar a obtener, así como Demos Online en la web de Stratebi:



Data Ingestion
Manipulation
Integration



Enterprise and
Ad Hoc Reporting



Data Discovery
Visualization



Predictive
Analytics





