



Security in Data Science:

How are data scientists giving you a hard time?
What can you do to keep them happy?



We create powerful tools to push the boundaries
of analytics and predictive modeling



timi.eu

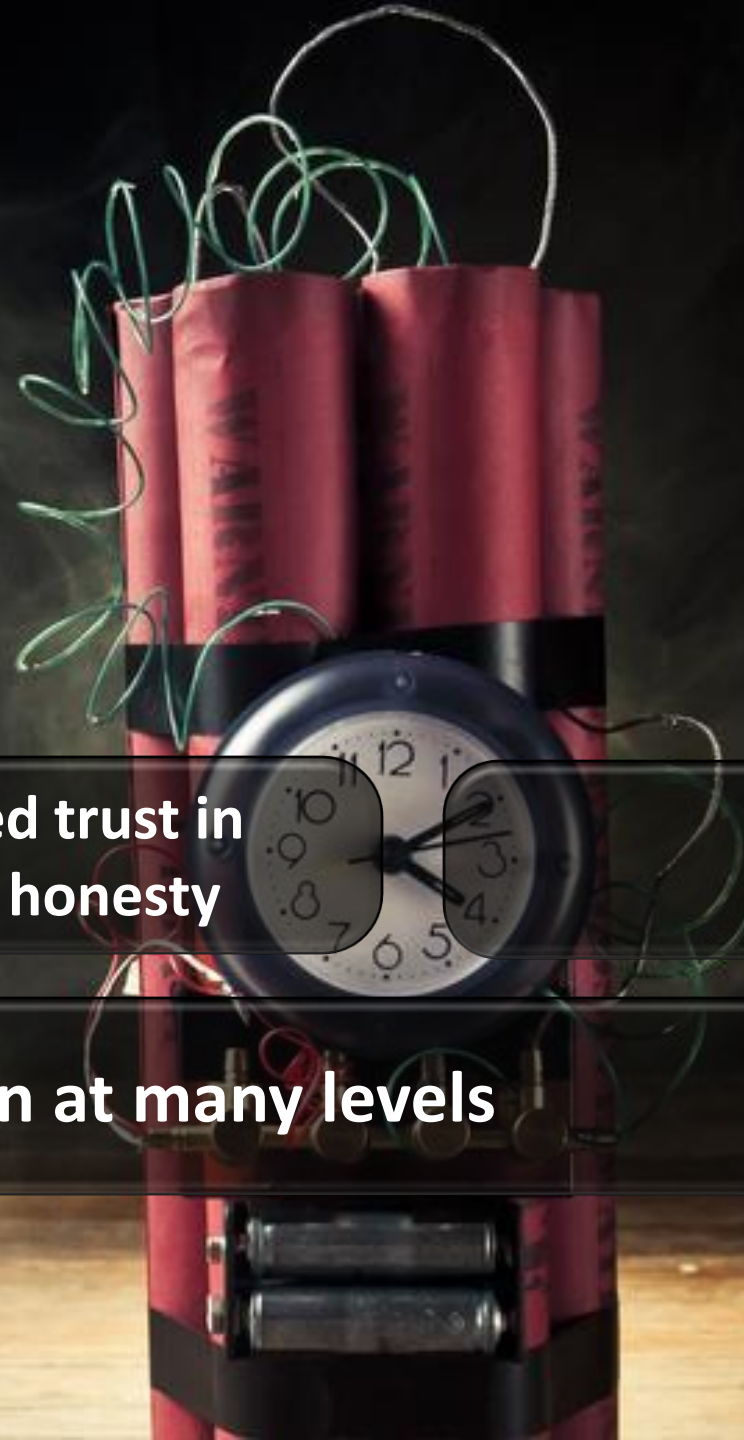
Data Science = Security Time Bomb !

Smelly Cheese

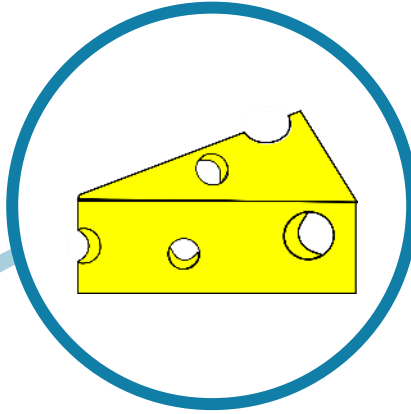
**Unchecked trust in
people's honesty**

GDPR

Lack of Encryption at many levels



Cyber criminals have your Data Scientists in Target



A wedge of Swiss cheese, likely Emmentaler, is shown against a dark background. The cheese is yellow and covered in numerous holes of various sizes. The text "Case 1" is overlaid on the cheese.

Case 1

Holes everywhere

Open source “Data” Frameworks: Some are more secure than others... Example 1:



**57 open
ports !**

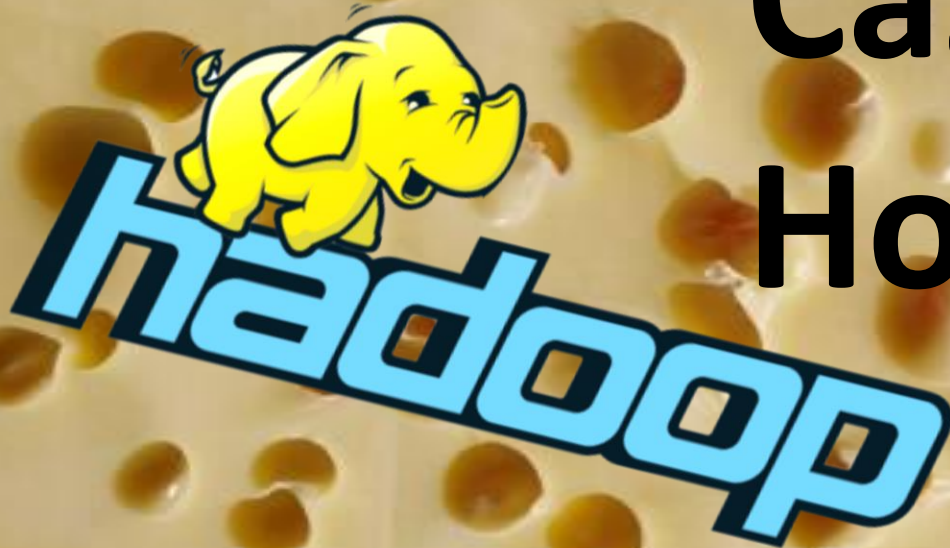
**⇒ 57 ways
to destroy
Everything**

TOOL	Port
HDFS WebUI for NameNode	9870/9871
HDFS Metadata service (NameNode)	hdfs://hdp-master:19000
HDFS Data Node	9864/9865 (https) and 9866 and 9867 (ipc)
HDFS Secondary NameNode	9868/9869
HDFS JournalNode	8485 (ipc) and 8480/8481 (http/https)
HDFS Aliasmap Server	50200
HDFS Namenode	Cloudera/HDP: 8020
HDFS Datanode (2)	50010 and 50020
HttpFS	14000
WebHDFS	Cloudera/HDP: 50070
YARN Resourcemanager	Cloudera: 8032 / HDP: 8050
YARN JobTracker	Cloudera/HDP: 8021
Hive Metastore (optional)	9083
YARN Resourcemanager Scheduler	8030
YARN collector-service.address	8048
YARN localizer.address	8040 (ipc)
YARN Resourcemanager Admin	8033

TOOL	Port
YARN Resourcemanager WebApp	8088
YARN Nodemanager WebApp	8042/8044 (http/https)
YARN sharedcache	8788
YARN sharedcache uploader server	8046
YARN sharedcache client server	8045
YARN amrm proxy	8049
YARN Timeline Service	10200 (ipc) or 8188/8190 (http/https)
YARN webapp	8089/8091 (http/https)
HDFS WebUI for NameNode	9870/9871
HDFS Metadata service (NameNode)	hdfs://hdp-master:19000
HDFS Data Node	9864/9865 (https) and 9866 and 9867 (ipc)
HDFS Secondary NameNode	9868/9869
HDFS JournalNode	8485 (ipc) and 8480/8481 (http/https)
HDFS Aliasmap Server	50200
MapReduce Job History	10020
MapReduce Job History UI	19888/19890 (http/https)
MapReduce History server admin	10033 (ipc)
HiveServer2 (optional)	TCP connection: 10000 / HTTP connection: 10001

Case 1

Holes Everywhere



databricks

Holes everywhere

7



=> Hadoop is extremely unsecure.

(e.g.: at a customer in LatAm: in 6 months, 2 times the whole Hadoop-based analytical infrastructure is lost. For the TIMi-based infrastructure: nothing is lost)

The Hadoop Solution:

- Don't let data scientists access raw data (they will not be very efficient!)
- Limit data access to the smallest team as possible
- Limit nature of data (1 cluster per project)
- Firewall everywhere

Avoid Hadoop like the Dark Plague: Literally use anything else!
Let's assumed you have this under control. How will we ruin your day?



7

Data Science frameworks: The recurrent problem

Passwords
are visible
“in plain sight”

```
import mysql.connector;
import pandas as pd;
conn = mysql.connector.connect(
    host='localhost', database='BookDB',
    user='root', password='root_pass');
cursor = conn.cursor();
cursor.execute('SELECT * FROM books');
df=cursor.fetchall();
df.to_csv('export.csv',sep=',',
    index=False);
cursor.close();
conn.close();
```



**Security is as robust
as our most dishonest
employee**



**He has all
the passwords!**

General Data Protection Regulation



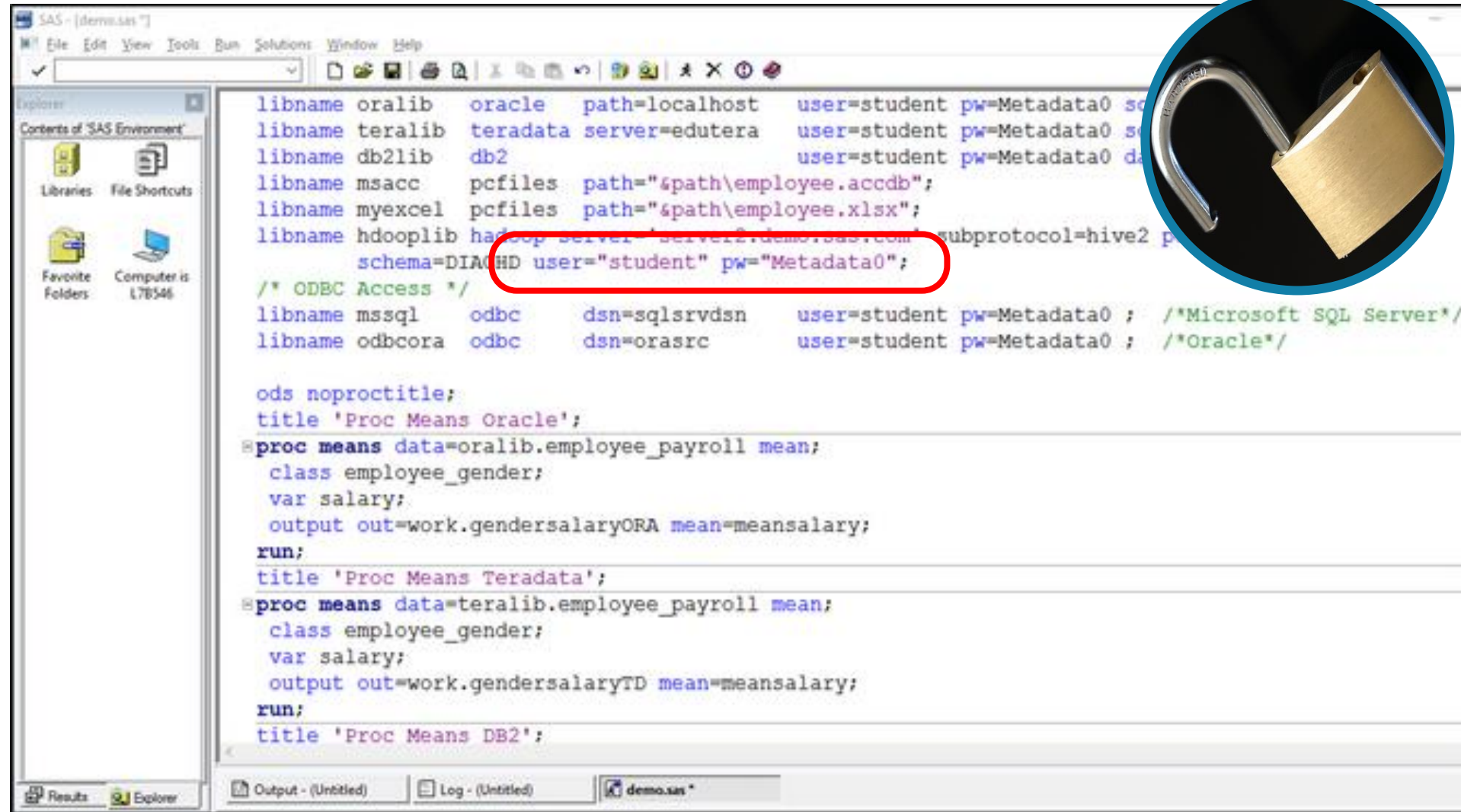
#	Fine	Non-US Company	Year
6	€35.3m	H&M	2020
7	€27.8m	TIM	2020
8	€22m	British Airways	2020
9	€20m	Clearview AI Inc.	2022
12	€16.7m	Wind Tre	2020
13	€14.5m	Deutsche Wohnen	2019
14	€12.25m	Vodafone Italia	2020
15	€11.5m	Eni Gas e Luce	2020
16	€10.4m	Notebooksbilliger.de	2021
18	€9.5m	Austrian Post	2021
19	€9m	Clearview AI Inc.	2022
20	€8.15m	Vodafone España	2021

Max. FINE: 20 million euros, or 4 % of total global turnover, whichever is higher.

CASE 2: Passwords are visible “in plain sight” in Commercial Solutions

11

Example:

A screenshot of the SAS IDE interface. The main window displays SAS code. A red circle highlights the line: `libname hdooplib hadoop server="server.demo.sas.com" subprotocol=hive2 pw="Metadata0";`. The code includes various library definitions for Oracle, Teradata, DB2, and Hadoop, followed by PROC MEANS statements for analyzing payroll data. The interface includes a menu bar, a toolbar, and a sidebar with "Libraries" and "File Shortcuts".

```
libname orolib oracle path=localhost user=student pw=Metadata0;
libname teralib teradata server=edutera user=student pw=Metadata0;
libname db2lib db2 user=student pw=Metadata0;
libname msacc pcfiles path="%path%\employee.accd";
libname myexcel pcfiles path="%path%\employee.xlsx";
libname hdooplib hadoop server="server.demo.sas.com" subprotocol=hive2 pw="Metadata0";
/* ODBC Access */
libname mssql odbc dsn=sqlsrdsn user=student pw=Metadata0; /*Microsoft SQL Server*/
libname odbcora odbc dsn=orasrc user=student pw=Metadata0; /*Oracle*/

ods noproctitle;
title 'Proc Means Oracle';
proc means data=oralib.employee_payroll mean;
  class employee_gender;
  var salary;
  output out=work.gendersalaryORA mean=meansalary;
run;
title 'Proc Means Teradata';
proc means data=teralib.employee_payroll mean;
  class employee_gender;
  var salary;
  output out=work.gendersalaryTD mean=meansalary;
run;
title 'Proc Means DB2';
```



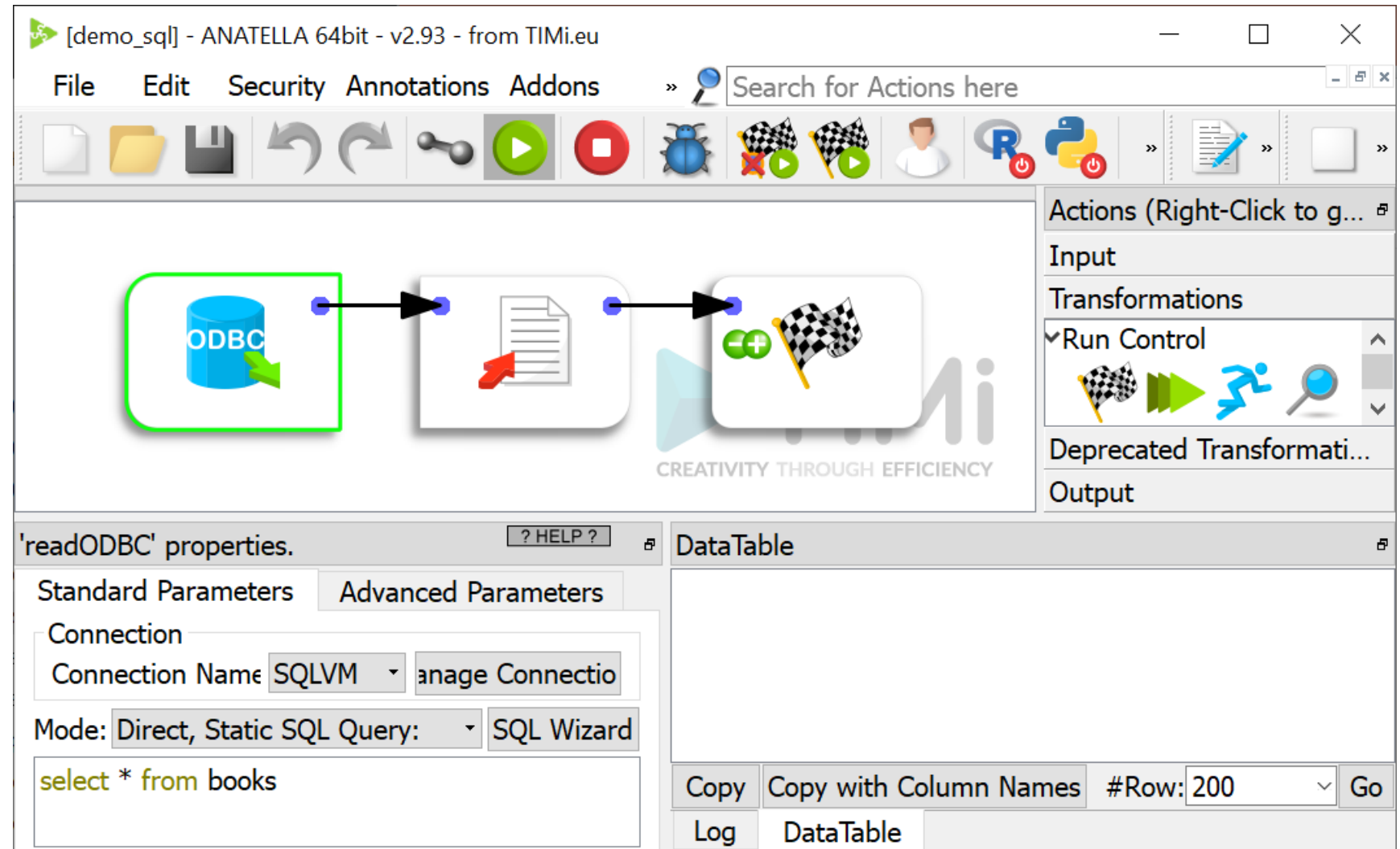
11



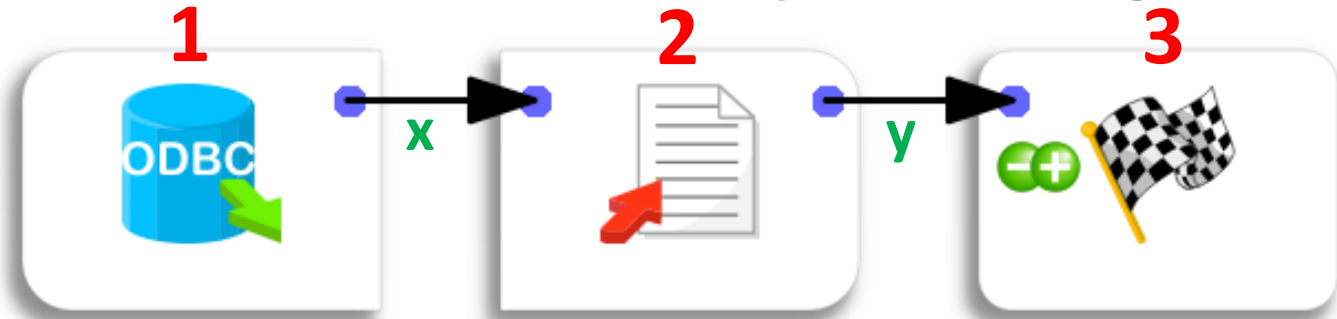
Be the hero data
security needs

CASE 2: Passwords are visible “in plain sight”

We DO
have
solutions!



CASE 2: Passwords are visible “in plain sight”



```
<?xml version="1.0" encoding="utf-8"?>
<ANATELLA version='2.93'>
  <GlobalParameters wDirLoc='1'>
    <ODBCConnections>
      <odbc name='SQLVM' link='SQLVM' login='root' ep='1' password='LAAAAGQDAADhAAAASFtQpxO08zUJI/Ufmtw58SQSS+zq6fBJghmR/3vCrQU='/>
    </ODBCConnections>
  </GlobalParameters>
  <ACTIONS>
    <readODBC module='DBConnectors' idx='1' x='0' y='0' odbcname='SQLVM'>
      <sql>
        select * from books
      </sql>
    </readODBC>
    <writeCSV idx='2' x='150' y='0' sep=',' filename='export.csv'/>
    <RunToFinishLine idx='3' x='300' y='0'/>
  </ACTIONS>
  <CONNECTORS>
    <Connection idxSrc='1' idxDest='2'/>
    <Connection idxSrc='2' idxDest='3'/>
  </CONNECTORS>
</ANATELLA>
```

1 (points to the `<readODBC>` action)

2 (points to the `<writeCSV>` action)

3 (points to the `<RunToFinishLine>` action)

x (points to the `<Connection idxSrc='1' idxDest='2'/>` connector)

y (points to the `<Connection idxSrc='2' idxDest='3'/>` connector)

Encrypted Database Password (points to the `password='LAAAAGQDAADhAAAASFtQpxO08zUJI/Ufmtw58SQSS+zq6fBJghmR/3vCrQU='` attribute)

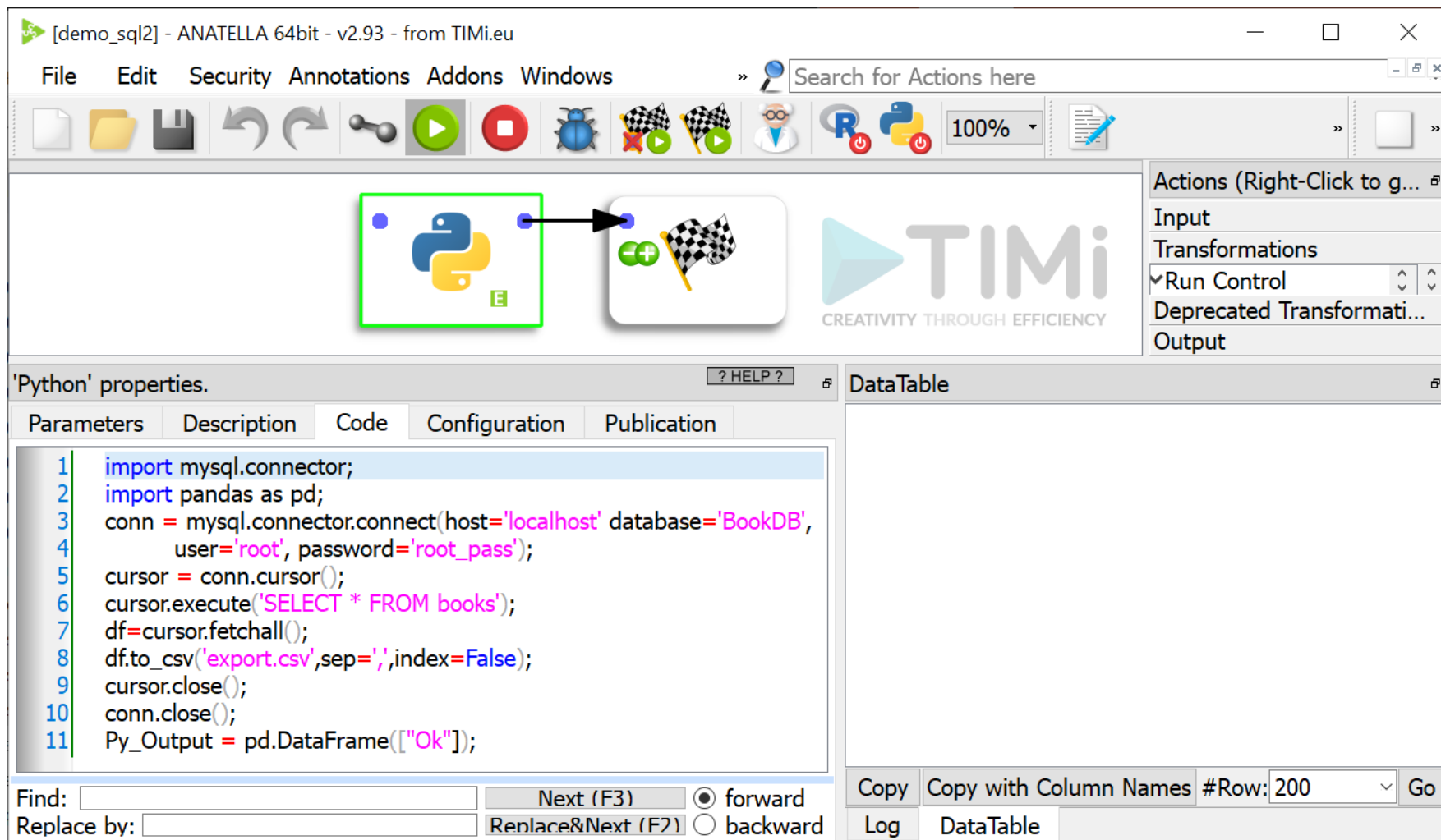
Open source
frameworks:

What can we
do with our
code?

```
import mysql.connector;
import pandas as pd;
conn = mysql.connector.connect(
    host='localhost', database='BookDB',
    user='root', password='root_pass');
cursor = conn.cursor();
cursor.execute('SELECT * FROM books');
df=cursor.fetchall();
df.to_csv('export.csv',sep=',',
    index=False);
cursor.close();
conn.close();
```



TIMi Framework 2: embedding python/R/JS code



The screenshot displays the TIMi Framework 2 interface. At the top, a menu bar includes File, Edit, Security, Annotations, Addons, and Windows. Below the menu is a toolbar with various icons, including a search bar labeled "Search for Actions here". The main workspace shows a Python icon (highlighted with a green box) connected to a TIMi icon (a play button with a checkered flag). The TIMi logo and tagline "CREATIVITY THROUGH EFFICIENCY" are visible in the background.

On the right side, there is a panel titled "Actions (Right-Click to g...)" with sections for Input, Transformations, Run Control, Deprecated Transformati..., and Output.

Below the workspace, there is a section titled "'Python' properties." with tabs for Parameters, Description, Code, Configuration, and Publication. The "Code" tab is active, showing the following Python code:

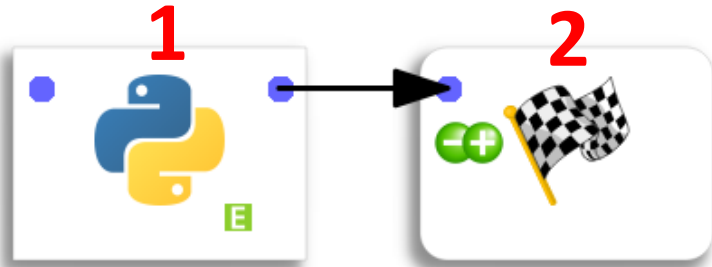
```

1 import mysql.connector;
2 import pandas as pd;
3 conn = mysql.connector.connect(host='localhost' database='BookDB',
4     user='root', password='root_pass');
5 cursor = conn.cursor();
6 cursor.execute('SELECT * FROM books');
7 df=cursor.fetchall();
8 df.to_csv('export.csv',sep=',',index=False);
9 cursor.close();
10 conn.close();
11 Py_Output = pd.DataFrame(["OK"]);
  
```

At the bottom, there is a search bar with "Find:" and "Replace by:" fields, and buttons for "Next (F3)", "Replace&Next (F2)", "forward", and "backward". On the right, there are buttons for "Copy", "Copy with Column Names", "#Row: 200", "Go", "Log", and "DataTable".

TIMi Framework 3

ABSTRACTION LAYER ABOVE CODE



“Normal User Mode” view:

'Python' properties. ? HELP ?

Parameters Description Code Configuration Publi

Script name: onTemplate + Add parameter - Remove

Description	Value
Database login	root
Database password	••••••
Database command(SQL)	SELECT * FROM books

```

<?xml version="1.0" encoding="utf-8"?>
<ANATELLA version='2.93'>
  <GlobalParameters wDirLoc='1'>
  </GlobalParameters>
  <ACTIONS>
    <Python idx='1' x='0' y='0' id='My_pythonTemplate' revision='0.01'>
      <Description></Description>
      <Parameters></Parameters>
      <Script>
import mysql.connector;
import pandas as pd;
conn = mysql.connector.connect(host='localhost' database='BookDB',
                               user='root', password='root_pass');
cursor = conn.cursor();
cursor.execute('SELECT * FROM books');
df=cursor.fetchall();
df.to_csv('export.csv',sep=',',index=False);
cursor.close();
conn.close();
Py_Output = pd.DataFrame(["Ok"]);
      </Script>
    </Python>
    <RunToFinishLine idx='2' x='150' y='0' />
  </ACTIONS>
  <CONNECTORS>
    <Connection idxSrc='1' idxDest='2' />
  </CONNECTORS>
</ANATELLA>
  
```

1

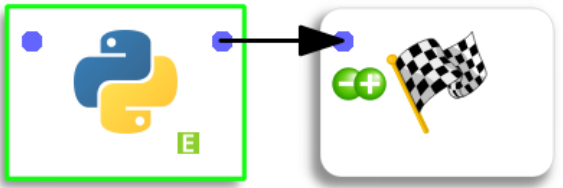
Visible Password for Database!!
THIS IS BAD !!

2

TIMi Framework 3: ABSTRACTION LAYER ABOVE CODE

18

“Expert Mode” view:



'Python' properties.

Parameters Description Code Configuration Publication

Script name: My_My_pythonTemplate

Description	Value	id in code	type	Parameter
Partition Type	No partition			
Database login	root	idLogin	string	
Database password	•••••	idPass	password	
Database command(SQL)	SELECT * FROM books	idSQLCommand	string	

'Python' properties.

Parameters Description Code Configuration Publication

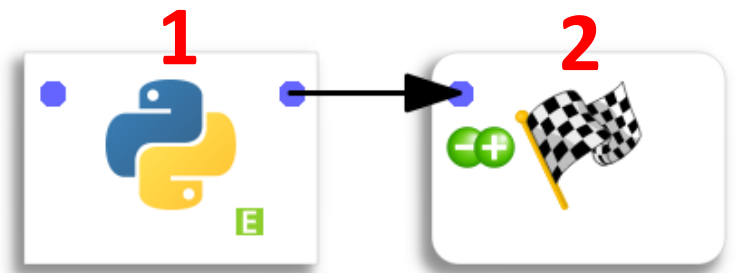
```
1 import mysql.connector;
2 import pandas as pd;
3 conn = mysql.connector.connect(host='localhost' database='BookDB',
4     user=idLogin, password=idPass);
5 cursor = conn.cursor();
6 cursor.execute(idSQLCommand);
7 df=cursor.fetchall();
8 df.to_csv('export.csv',sep=',',index=False);
9 cursor.close();
10 conn.close();
11 Py_Output = pd.DataFrame(["Ok"]);
```

Find: Next (F3) ☒ forward
Replace by: Replace&Next (F2) ☐ backward

18

TIMi Framework 3: ABSTRACTION LAYER ABOVE CODE

19



```
<?xml version="1.0" encoding="utf-8"?>
<ANATELLA version='2.93'>
  <GlobalParameters wDirLoc='1'></GlobalParameters>
  <ACTIONS>
    <Python idx='1' x='0' y='0' id='My_pythonTemplate' revision='0.01'>
      <Description></Description>
      <Parameters>
        <Parameter id='idLogin' text='Database login' type='string'>root</Parameter>
        <Parameter id='idPass' text='Database password' type='password'>HAAAAG4AGNuXOmTHQnA==</Parameter>
        <Parameter id='idSQLCommand' text='Database command(SQL)' type='string'>SELECT * FROM books</Parameter>
      </Parameters>
      <Script>
import mysql.connector;
import pandas as pd;
conn = mysql.connector.connect(host='localhost' database='BookDB',
    user=idLogin, password=idPass);
cursor = conn.cursor();
cursor.execute(idSQLCommand);
df=cursor.fetchall();
df.to_csv('export.csv',sep=',',index=False);
cursor.close();
conn.close();
Py_Output = pd.DataFrame(["Ok"]);
      </Script>
    </Python>
    <RunToFinishLine idx='2' x='150' y='0' />
  </ACTIONS>
  <CONNECTORS>
    <Connection idxSrc='1' idxDest='2' />
  </CONNECTORS>
</ANATELLA>
```

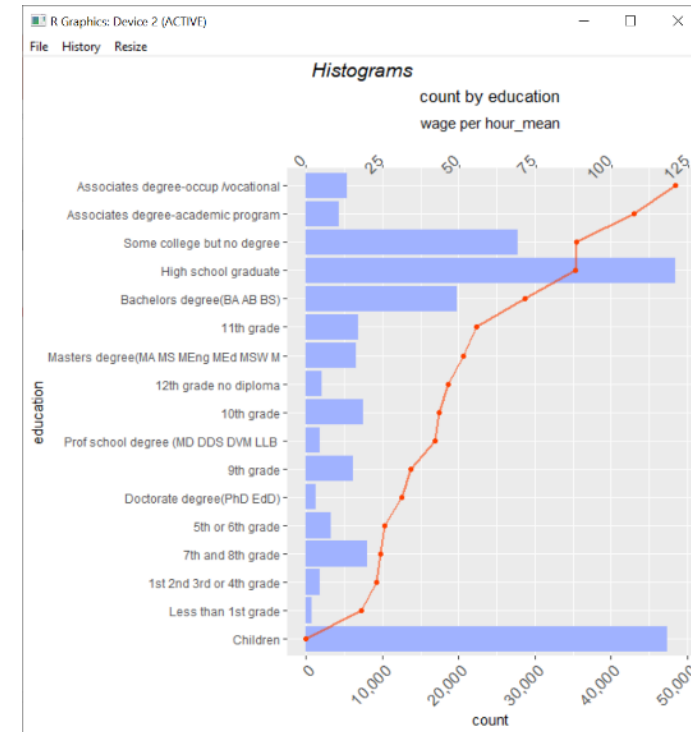
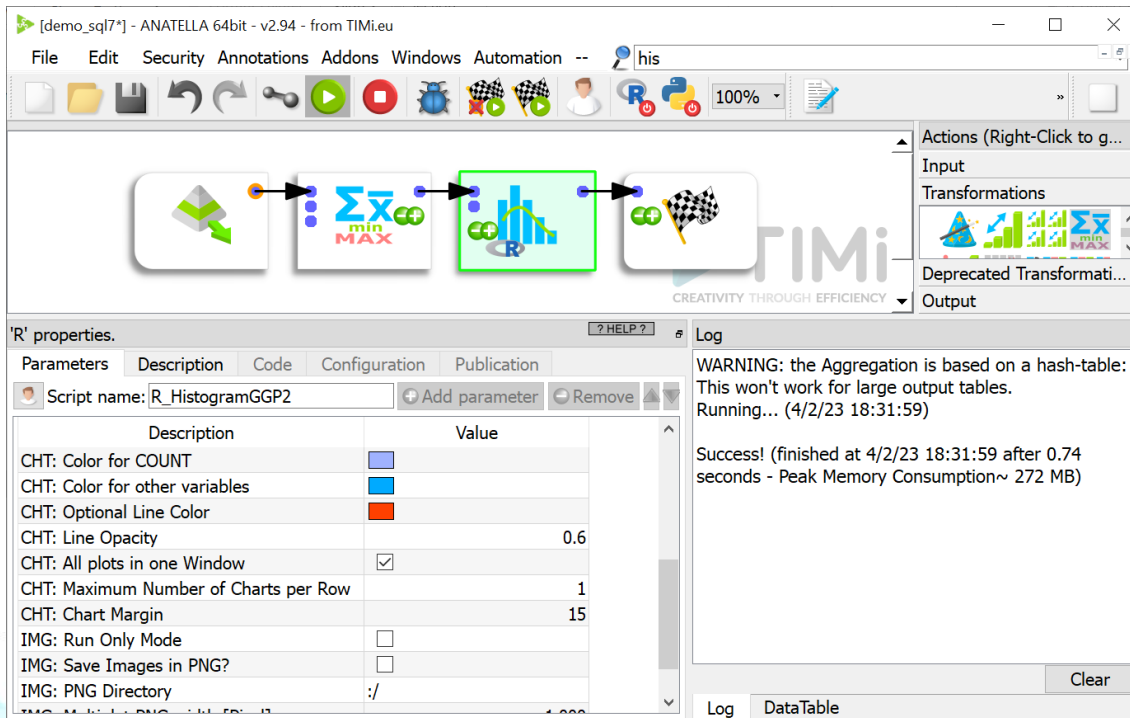
Password for Database is encrypted !!

Password for Database is not visible anymore !!

19

TIMi Framework 4: Slight digression: ABSTRACTION LAYER

Use the **FULL POWER** of R/Python/JS without Seeing/Typing one line of code !
Execute/Share/Parametrize/Update boxes based on (encrypted) R/Python/JS code.



TIMi Framework 5: Encryption: ...of the box parameters

21

“Normal user” Mode view:

The screenshot shows the TIMi Framework 5 interface. The main workspace displays a workflow with a Python node (highlighted with a green box) and a TIMi node. The 'Python' properties panel is open, showing a table of parameters. The 'Database password' field is circled in red, and a red arrow points to it. The 'Database login' field is set to 'root', and the 'Database command(SQL)' field is set to 'SELECT * FROM books'.

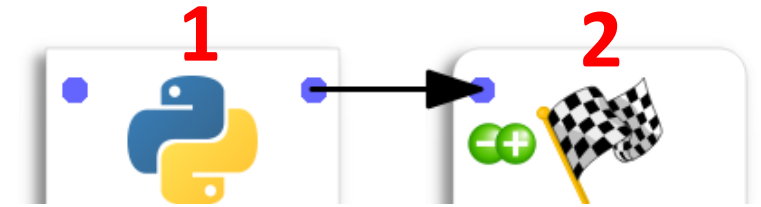
Parameters

Description	Value
Database login	root
Database password
Database command(SQL)	SELECT * FROM books

21

TIMi Framework 4: Encryption: ...of python/R/JS Code

22



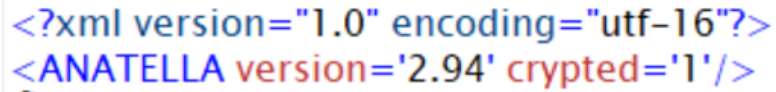
```
<?xml version="1.0" encoding="utf-8"?>
<ANATELLA version='2.93'>
  <GlobalParameters wDirLoc='1'></GlobalParameters>
  <ACTIONS>
    <Python idx='1' x='0' y='0' id='My_pythonTemplate' crypted='1' revision='0.01'>
      <Description></Description>
      <Parameters>
        <Parameter id='idLogin' text='Database login' type='string'>root</Parameter>
        <Parameter id='idPass' text='Database password' type='password'>HAAAAAgCAACax/VfSrL9gQ==</Parameter>
        <Parameter id='idSQLCommand' text='Database command(SQL)' type='string'>SELECT * FROM books</Parameter>
      </Parameters>
      <Script>
nAAAAAgCAAIAgAAusyFWhYsSRsg9cb/giumZOdcZF1PrsewLRNXu8JwxSLq1GZtzt2mszlwabs6HRG1U5+trERZYMmbpQYUw2Wop1RM
kyklk9u8JVcqEb84TuzXLskjd2y78pG3+U3TzjPCNmBhNKNRuUe3m70EEsV6AGOfQncIPWNCMZbfSUmKiYnuF2ZvaDYkvhVNx4pG+mm
      </Script>
    </Python>
    <RunToFinishLine idx='2' x='150' y='0' />
  </ACTIONS>
  <CONNECTORS>
    <Connection idxSrc='1' idxDest='2' />
  </CONNECTORS>
</ANATELLA>
```

Database Password still encrypted

Code is encrypted !!
CODE CANNOT BE TEMPERED WITH HERE !!

1

22

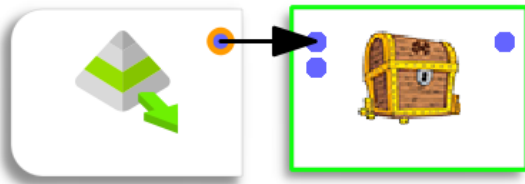
[illegible]

BONUS:

1. Give/Remove to the users the right to EDIT/VIEW the graph (to protect your IP!)
2. All graph can still be EXECUTED from the command-line (or from the Scheduler)
3. Expiration dates on graphs.

TIMi Framework 4: Encryption: ...of the data

Bidirectional Data Encryption (DES, 3DS):
Military grade:



'Encrypt' properties. ? HELP ?

General Parameters Extra Parameters

☐ Encrypt ☒ Decrypt

Columns to process:

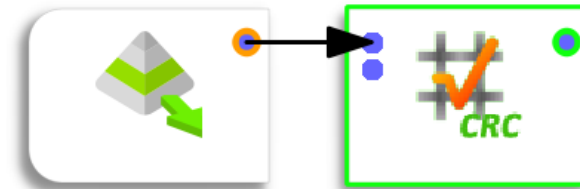
Encryption Key

Encryption Key is stored: ☐ in a file: ☒ in a string:

Key:

Decryption Key can be also stored
in a **deported** storage

Unidirectional Data Encryption (MD5, SHA256, ...):



'ActionCRC' properties. ? HELP ?

Standard Parameters Advanced Parameters

Mode: ☒ Fixed: ☐ Dynamic:

Column name with result:

Compute Checksum on:

(Requires you to keep a “translation table”)

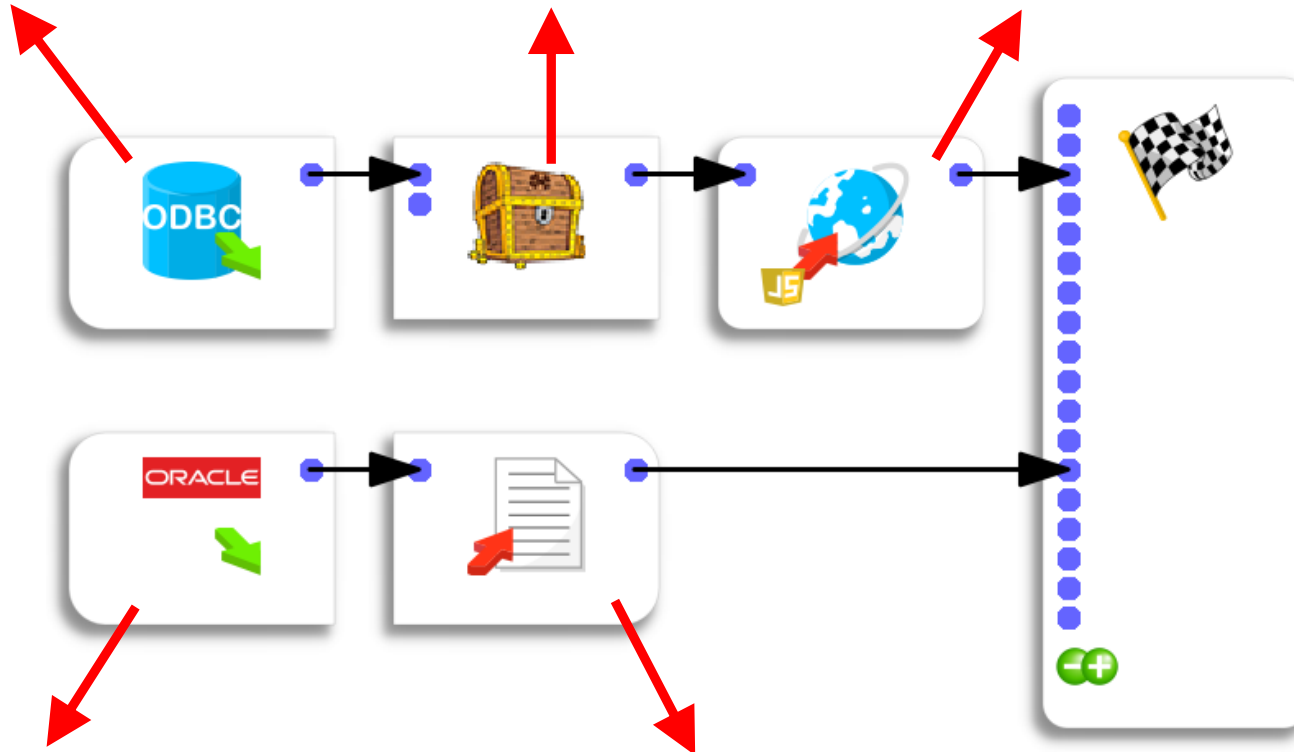
TIMi Framework 5: Deported Password Storage

25

Get Database Password from
Meta-Data Repository

Decrypt
Password

Set Password as the
TRANSIENT Global Parameter "X"



Use password saved into the
TRANSIENT Global Parameter "X"
to extract data from Database

Save extracted
data into CSV file

25



One more level

How do we protect critical processes?

TIMi Framework 6: Protecting Execution against thieves

27

On an UN-authorized machine:

The screenshot displays the TIMi Framework 6 interface. The top menu bar includes File, Edit, Security, Annotations, Addons, and Windows. A search bar for actions is present. The main workspace shows a Python script icon connected to a TIMi logo. The 'Python' properties panel is open, showing the following code:

```
1 if hwid!='4WCWLQBuTAaO_':  
2     quit('FORBIDDEN MACHINE');  
3  
4 print("Hello World!")  
5  
6 Py_Output = pd.DataFrame(["Ok"]);
```

The Log panel on the right shows the execution status: "Running... (1/2/23 12:35:57)". An error message is highlighted in a red circle:

```
ERROR:  
FORBIDDEN MACHINE  
Exception ignored in: <__main__.StdoutCatcher object  
at 0x000000001263E4E0>
```

The bottom of the interface includes a Find and Replace section with options for Next (F3), forward, backward, and Replace&Next (F7).

27

TIMi Framework 6: Protecting Execution against thieves

28

It's easy to prevent thieves to execute your critical processes !

In Python:

```
if hwid!='4WCWLQBuTAaO' :  
    quit('FORBIDDEN MACHINE');
```

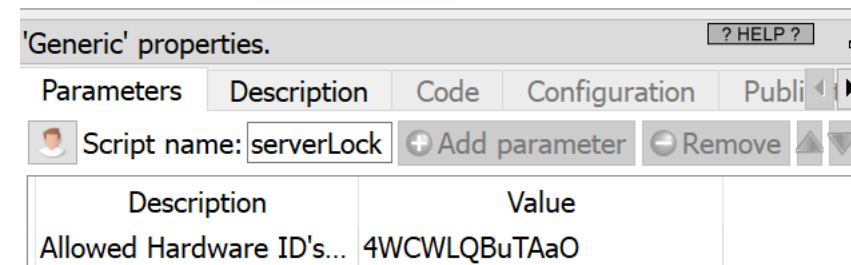
In JavaScript:

```
if (hardwareID!="4WCWLQBuTAaO")  
    throw "ERROR: FORBIDDEN MACHINE";
```

In R:

```
if (hwid!="4WCWLQBuTAaO")  
{  
    stop('FORBIDDEN MACHINE');  
}
```

As an Anatella box:



28

TIMi Framework : Cyber Security Summary

Five Encryption Levels:

1. All Passwords for native Anatella boxes (ODBC,OleDB,Cloud,etc.) are encrypted.
2. The Passwords for custom R/Python/JS boxes are encrypted.
3. The code of R/Python/JS boxes are encrypted.
4. Whole Anatella graphs are encrypted.
5. Data is encrypted

What can we do more than encryption?

1. All to **execute/share/parametrize/update** boxes based on (encrypted) R/Python/JS code.
=> **ABSTRACTION LAYER FOR Analytical Culture**
2. Deported Password storage / Deported Decryption Key
3. Prevent to steal whole Anatella graphs.



TESTIMONIALS

***"TIMi/Anatella is something that you cannot live without.
With a good NVMe SSD and TIMi you can do it all."***

Artur Grzebowski, Data Scientist at Volvo Cars

"The perfect "cognition amplifier" for data scientists."

Rabie Nait-Abdallah, University Professor of Data Science in LatAm.

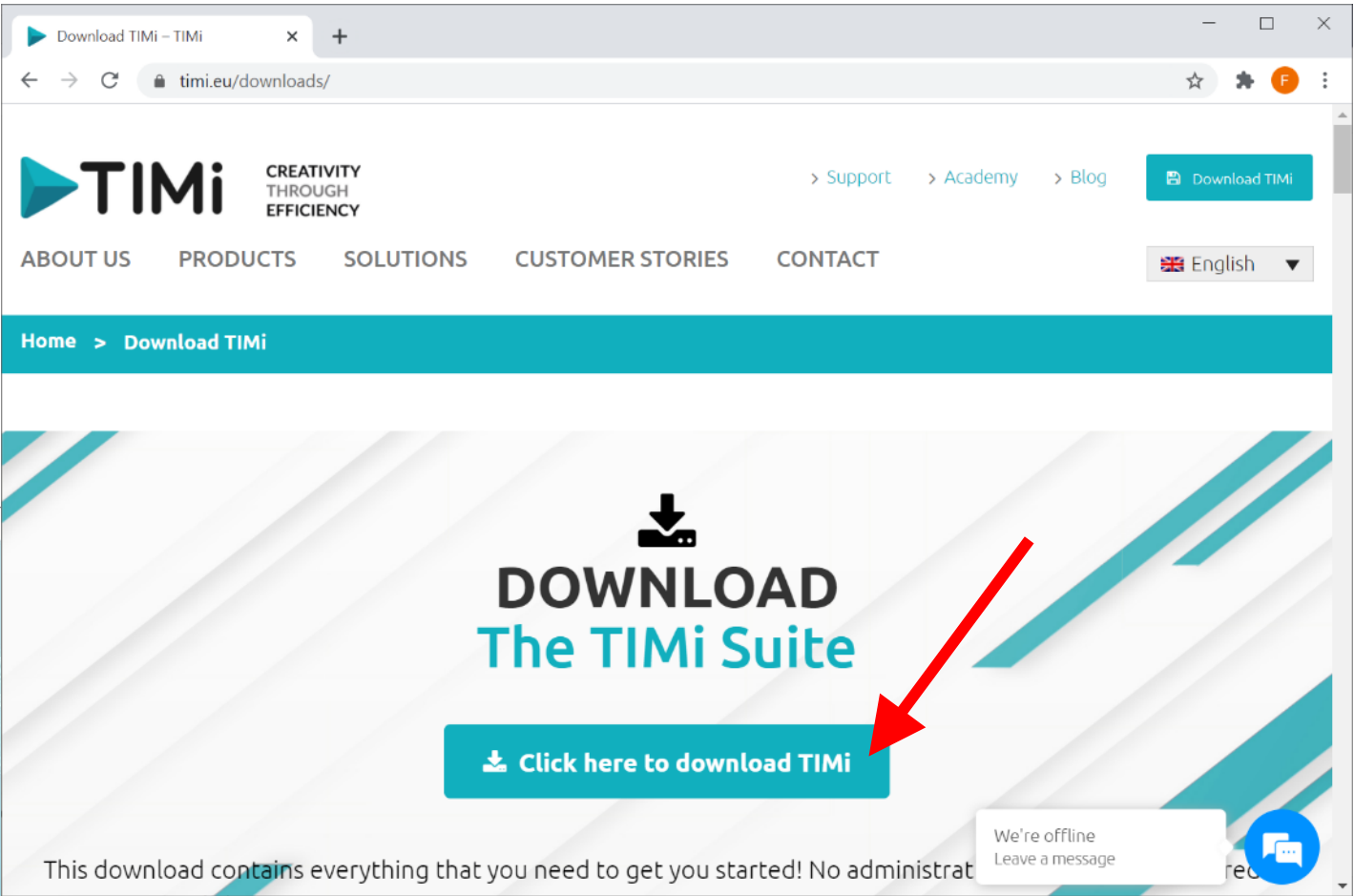
***"You can easily automate in a few mouse clicks
the most complex data processes."***

Alejandro García Cabrera, Head of Metering, Enel



Download & Install TIMi (Trial)
Open in a browser:

<https://timi.eu/downloads/>



Thanks for your Attention

For more information, please consult our website:

<http://timi.eu>

Mission: Enable everyone to extract knowledge and value from their data.
To boldly go where no data scientist has gone before.

To reach our goal: we created a new analytical platform: TIMi.

With TIMi, we changed forever the way you do analytics!



About TIMi: A few Key Events

Since 2007, TIMi grew from a niche Auto-ML solution into a full Data Science Suite

First commercial version of TIMi Modeler, adapted to the needs of key TELECOM clients. Bouygue Telecom is our first client. **AUTOML is BORN**

2007



2015

Geographic expansion in LatAm with the opening of TIMi LatAm in Bogotá. Scotiabank is our first LatAm client (Peru)

2017



2018 - today

Strengthen Academic Ties
Cloud Integration
Continuous Improvements



2010

Launch of Anatella Beta, allowing for the first time Big Data Processing on a laptop



2016



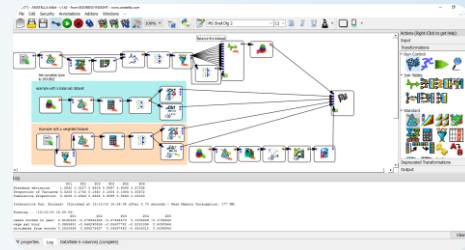
Connect

To anything



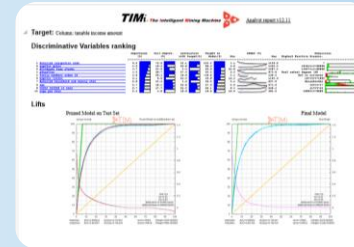
Transform

With ease and Speed



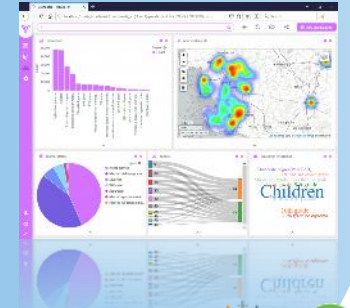
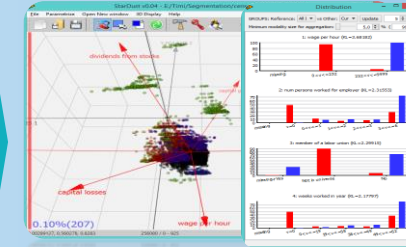
Model

More Speed & Precision



Integrate

Visualize & Share results



is the fastest solution for predictive analytics and data Management available today. Process Billions of record at minimum infrastructure cost.