



Wendelin Exanalytics

2020 Big Data with MariaDB

2014-04-03 – Santa Clara updated 2014-06-16



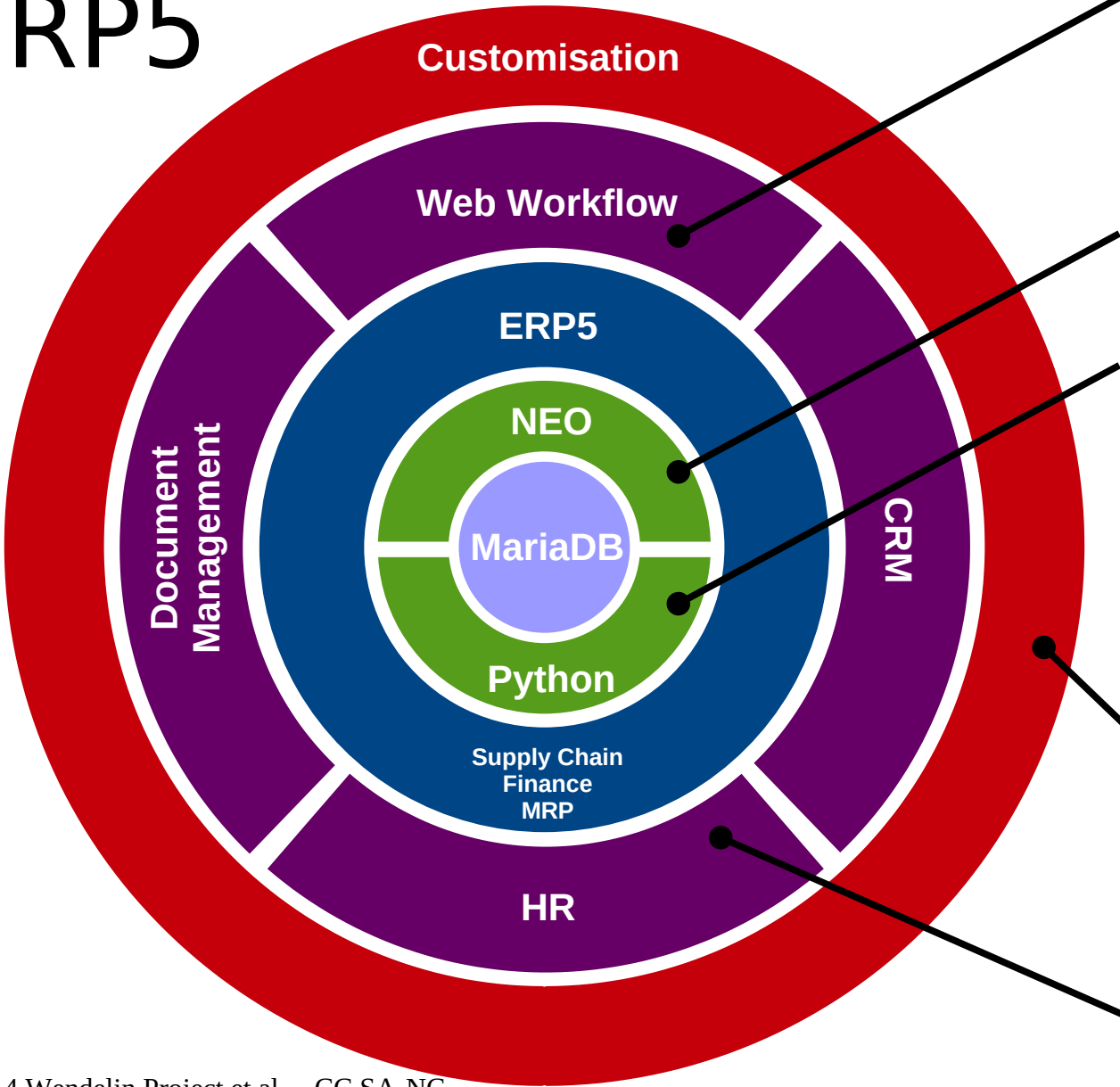
Agenda



Our background: ERP5

- **Our future: Wendelin Exanalytics**
- **Our challenge: out-of-core**

ERP5



Online contribution for 3rd parties
To-do lists
Notifications

Fine Grain Security
Full Traceability
Scalability

Flexibility
Rapid prototyping
Zope TTW on steroids

Banking
Aerospace
Health
Chemical
Government
NGO

Cloud Computing
Consulting
Mechanical

Careers and assignments
Payroll
Projects

Terra-SAR X Satellite



MariaDB



Management of sales and production of images



Compliant with ESA
standard (ECSS)



AIRBUS
DEFENCE & SPACE

Accessible to Airbus
partners and distributors
Interfaces with DLR
(Germany Space Agency)

« With ERP5, our partners all over the world can access our infrastructure and order online with complete security » Ralf Duering



SANEF Group



Online sales and customer relation for ETC Tolling



Implemented in 4 months



120.000 new customers / year
51.000 invoice/hour
7.000.000 contacts / year
250 users

« Web has become our primary sales channel. » Frédéric Charlier



Open Source ERP/CRM for S&P 100

Agenda

- **Our background: ERP5**



Our future: Wendelin Exanalytics

- **Our challenges with MariaDB**

Take the Best Analytics scikit-learn.org

scikit-learn
Machine Learning in Python

- Simple and efficient tools for data mining and data analysis
- Accessible to everybody, and reusable in various contexts
- Built on NumPy, SciPy, and matplotlib
- Open source, commercially usable - BSD license

Classification
Identifying to which set of categories a new observation belong to.
Applications: Spam detection, Image recognition.
Algorithms: SVM, nearest neighbors, random forest, ... — Examples

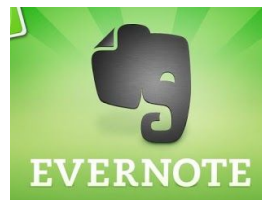
Regression
Predicting a continuous value for a new example.
Applications: Drug response, Stock prices.
Algorithms: SVR, ridge regression, Lasso, ... — Examples

Clustering
Automatic grouping of similar objects into sets.
Applications: Customer segmentation, Grouping experiment outcomes
Algorithms: k-Means, spectral clustering, mean-shift, ... — Examples

Dimensionality reduction
Reducing the number of random variables to

Model selection
Comparing, validating and choosing

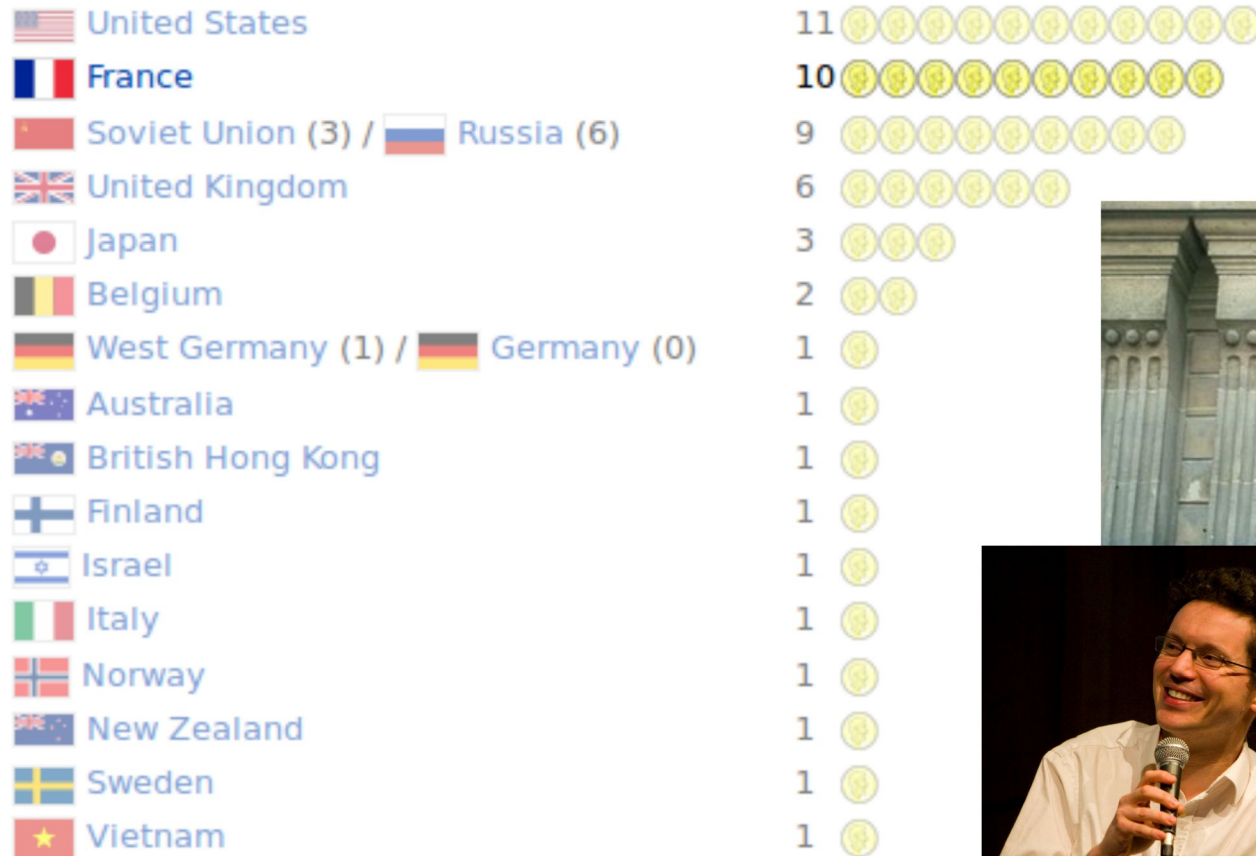
Preprocessing
Feature extraction and normalization.



Made by Great Mathematicians

http://en.wikipedia.org/wiki/Fields_Medal

Number of Fields Medallists by country [\[edit\]](#)

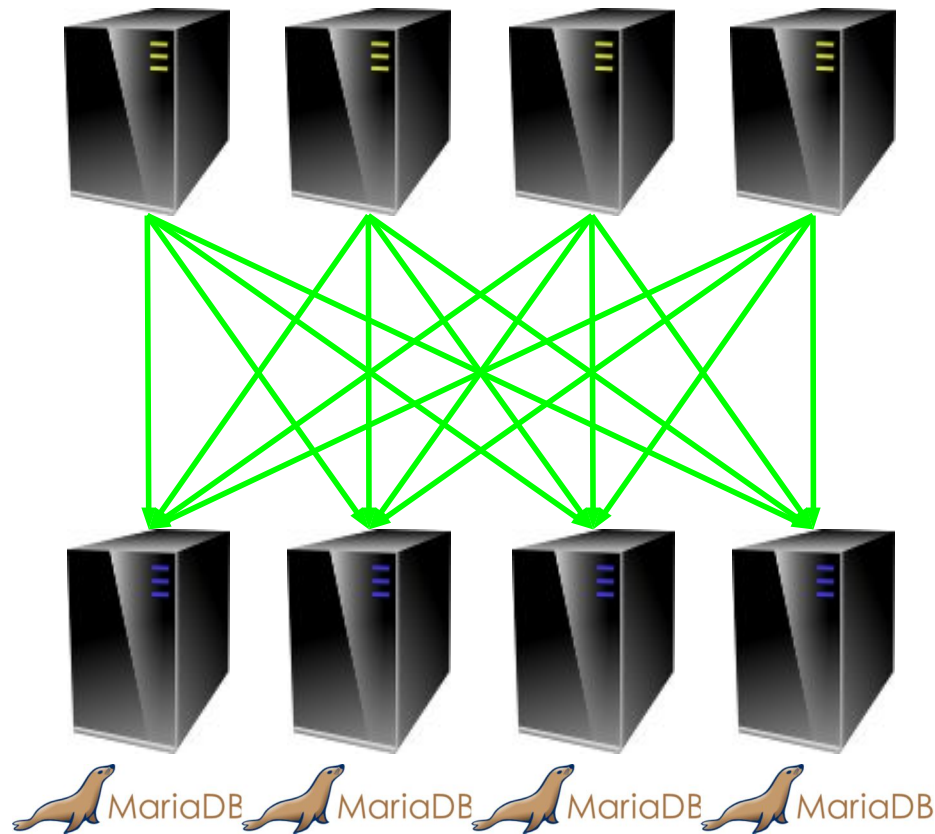


Wendelin Werner

Add Distributed Storage neopod.org



NEO



Add Elastic PaaS erp5.com

```
# Initialize data
data_size = 1000000
server_count = 1000
chunk_size = data_size / server_count
data = array(data_size)

# Process data in parallel on each server (Map Reduce, Batch, etc.)
for server in server_count:
    data.activate().process(server*chunk_size, chunk_size)
```



And Multicloud Deployment slapos.org



The screenshot shows the Slapos web interface for editing a software profile. The browser address bar indicates the URL is `workspace/my_repository/software`. The interface has a navigation menu with options: Profiles, Workspace, Run software, Software log, SR result, Run instance, Instance log, Inspect instance, and Destroy instance. The main content area is titled "Edit software profile [More]:" and contains a code editor with the following configuration:

```
0 [buildout]
1
2 extends =
3 # "slapos" stack describes basic things needed for 99.9% of SlapOS Software
4 # Releases
5 ../../stack/slapos.cfg
6 # Extend here component profiles, like openssl, apache, mariadb, curl...
7 # Or/and extend a stack (lamp, tomcat) that does most of the work for you
8 # In this example we only need the dash binary to run a simple "hello world"
9 # shell script.
10 ../../component/dash/buildout.cfg
11
12 parts =
13 # Call installation of slapos.cookbook egg defined in stack/slapos.cfg (needs
14 # in 99,9% of Slapos Software Releases)
15 slapos-cookbook
16 # Call creation of instance.cfg file that will be called for deployment of
17 # instance
18 template
19
20 # Download instance.cfg.in (buildout profile used to deployment of instance),
21 # replace all ${foo:bar} parameters by real values, and change ${foo:bar} to
22 # ${foo:bar}
23 [template]
24 recipe = slapos.recipe.template
25 url = ${:_profile_base_location_}/instance.cfg.in
26 output = ${buildout:directory}/instance.cfg
27 # MD5 checksum can be skipped for development (easier to develop), but must be filled for production
28 md5sum = 1fc461c00e86485bee77a942f39e3c43
29 mode = 0644
30
```

A "Save" button is located at the bottom left of the code editor area. To the right of the code editor, the Slapos logo (a green cloud shape) and the word "SLAPOS" are displayed.



MMC Rus

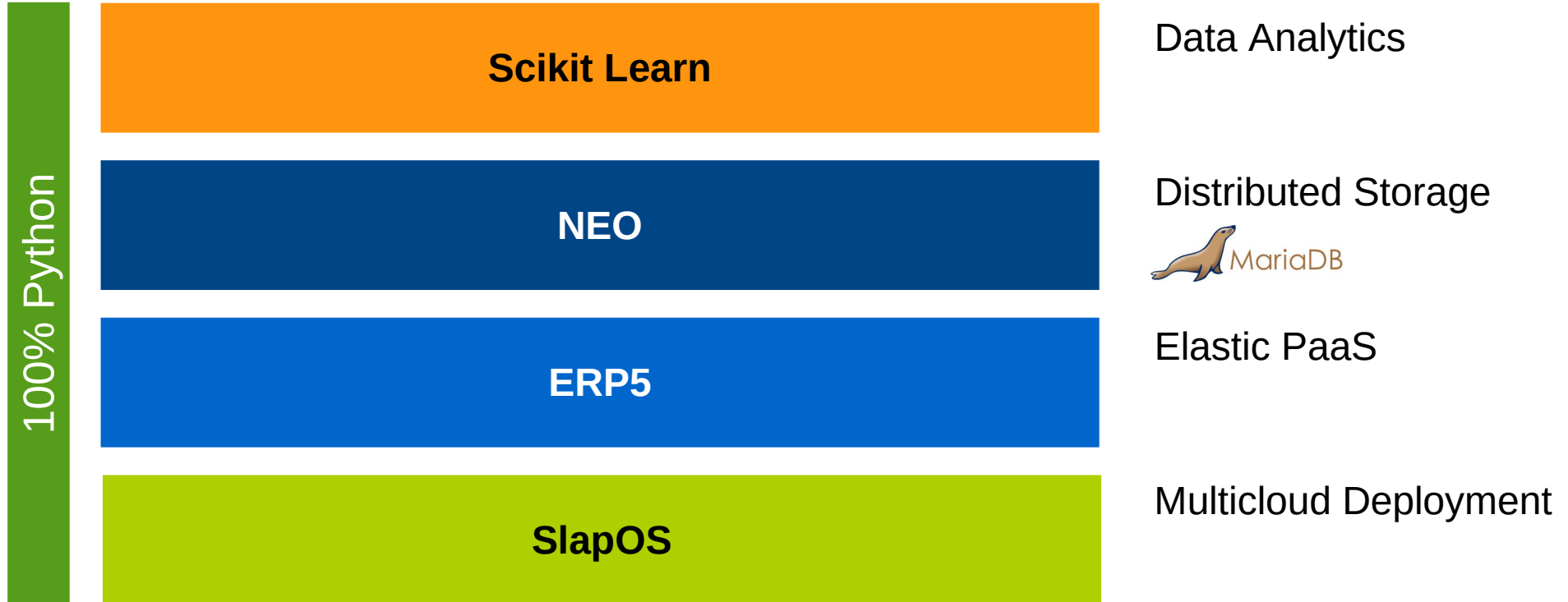


L'Education change le monde



Wendelin Exanalytics Core 100% open

source



Multi Data Center

Wendelin User Interface renderjs.org

Search

Menu

Some

20+ Tasks
11 Messages
Home

Shop:Order 2341
2013/04/02 · 49.02€ · France · Archive

Shop:Order 23412
2013/05/13 · 55.02€ · France · Paymen...

Sven Franck
Client · Last Order: 2013/04/02

Sven Franck

Available Options

Order Inbox
Review and Manage your orders. 12

Clients
Managing and marketing users.

Messages
Messages, special offers, news. 8

Shops
Setup and manage your shops.

Products
Upload and manage products.

Settings

Help

Settings

Report a problem

Sign out

Nexedi © 2013

List Tasks: 1-7 of 7 records Criteria

	Company	Last Trade	Trade Time	Change	Prev Close	Open	Stuff		1y Target Est
							Bid	Ask	
<input type="checkbox"/>	GOOG Google Inc.	597.74	12:12PM	14.81 (2.54%)	582.93	597.95	597.73 x 100	597.91 x 300	731.10
<input type="checkbox"/>	AAPL Apple Inc.	378.94	12:22PM	5.74 (1.54%)	373.20	381.02	378.92 x 300	378.99 x 100	505.94
<input type="checkbox"/>	AMZN Amazon.com Inc.	191.55	12:23PM	3.16 (1.68%)	188.39	194.99	191.52 x 300	191.58 x 100	240.32
<input type="checkbox"/>	ORCL Oracle Corporation	31.15	12:44PM	1.41 (4.72%)	29.74	30.67	31.14 x 6500	31.15 x 3200	36.11
<input type="checkbox"/>	MSFT Microsoft Corporation	25.50	12:27PM	0.66 (2.67%)	24.84	25.37	25.50 x 71100	25.51 x 17800	31.50
<input type="checkbox"/>	CSCO Cisco Systems, Inc.	18.65	12:45PM	0.97 (5.49%)	17.68	18.23	18.65 x 10300	18.66 x 24000	21.12
<input type="checkbox"/>	YHOO Yahoo! Inc.	15.81	12:25PM	0.11 (0.67%)	15.70	15.94	15.79 x 6100	15.80 x 17000	18.16

Outlier detection

LASSO Path

Wendelin Options

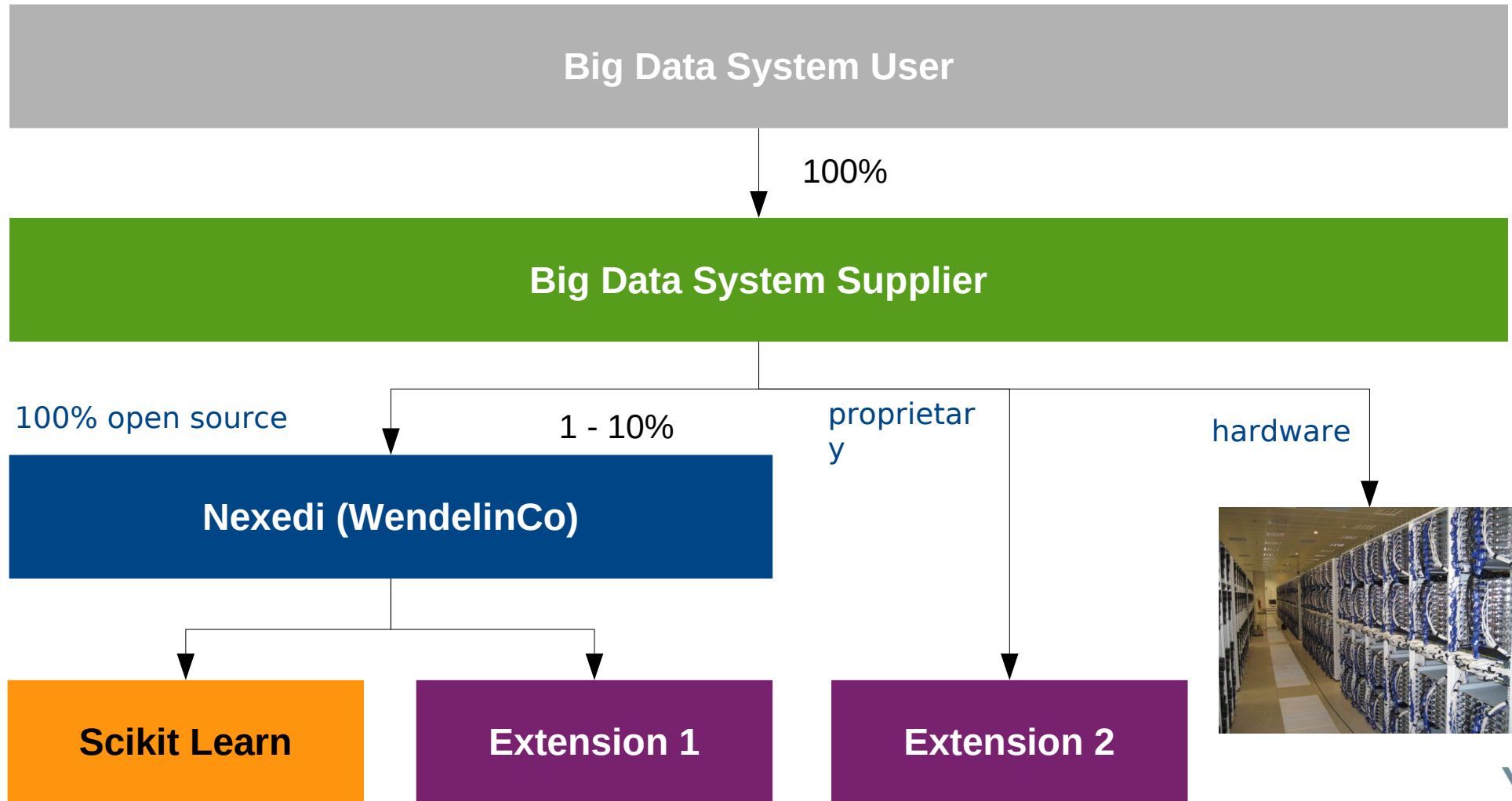
100% open source

100% Python	OpenCV-Python	Video Processing Intel Russia / Willow / Itseez
	NLTK	Natural Language Toolkit U. Texas / Chalmers
	Blaze	Full out-of-core arrays Continuum / DARPA
	Numba / Parakeet	JIT compiler / type inference Continuum / DARPA
	Pandas	Time sequence processing DataPad / JP Morgan
	Scikit Learn	
	NEO	
	Fluentd	Reatime log collection Treasure Data / Amazon

Wendelin Applications

- **Intrusion detection**
- **Fraud detection**
- **Business and economic prevision**
- **Marketing**
- **Media analysis**
- **Public security**
- **Brain Computer Interface**
- **Internet Of Things**

Business Model: German Style No VC



Agenda

- **Our background: ERP5**
- **Our future: Wendelin Exanalytics**



Our challenge: out-of-core

Out-of-core arrays

```
# Numpy
```

```
np.ndarray(shape=(2,2), dtype=float, order='F')
```

```
# Out-of-core data
```

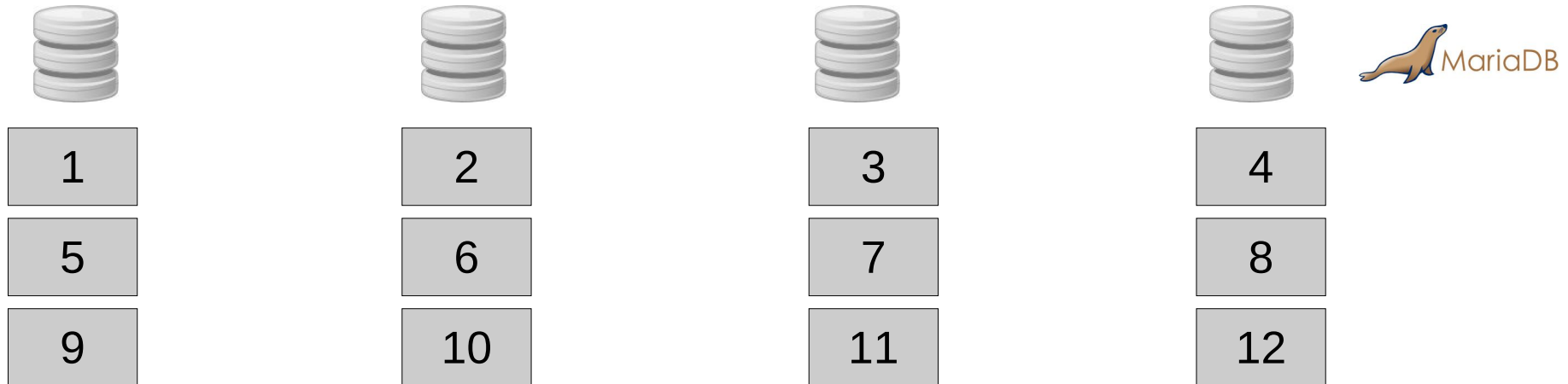
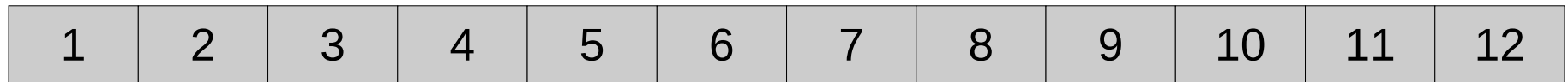
```
np.ndarray(shape=(1e18,2), dtype=float, order='F') 2 Exabyte
```

```
# Full out-of-core
```

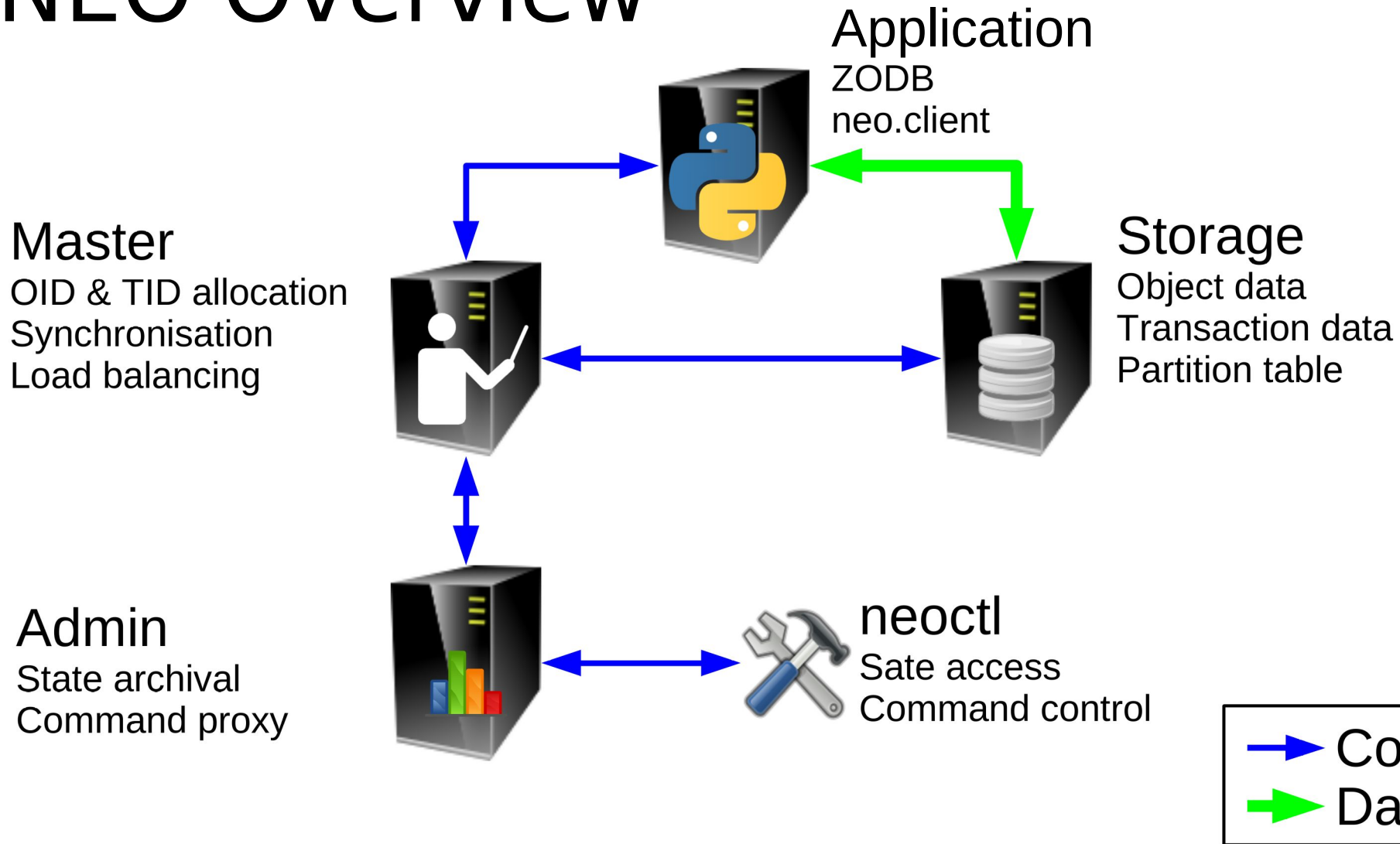
```
np.ndarray(shape=(1e9,2e9), dtype=float, order='F') 2 Exabyte
```

Best out-of-core topology depends on the algorithm and array geometry

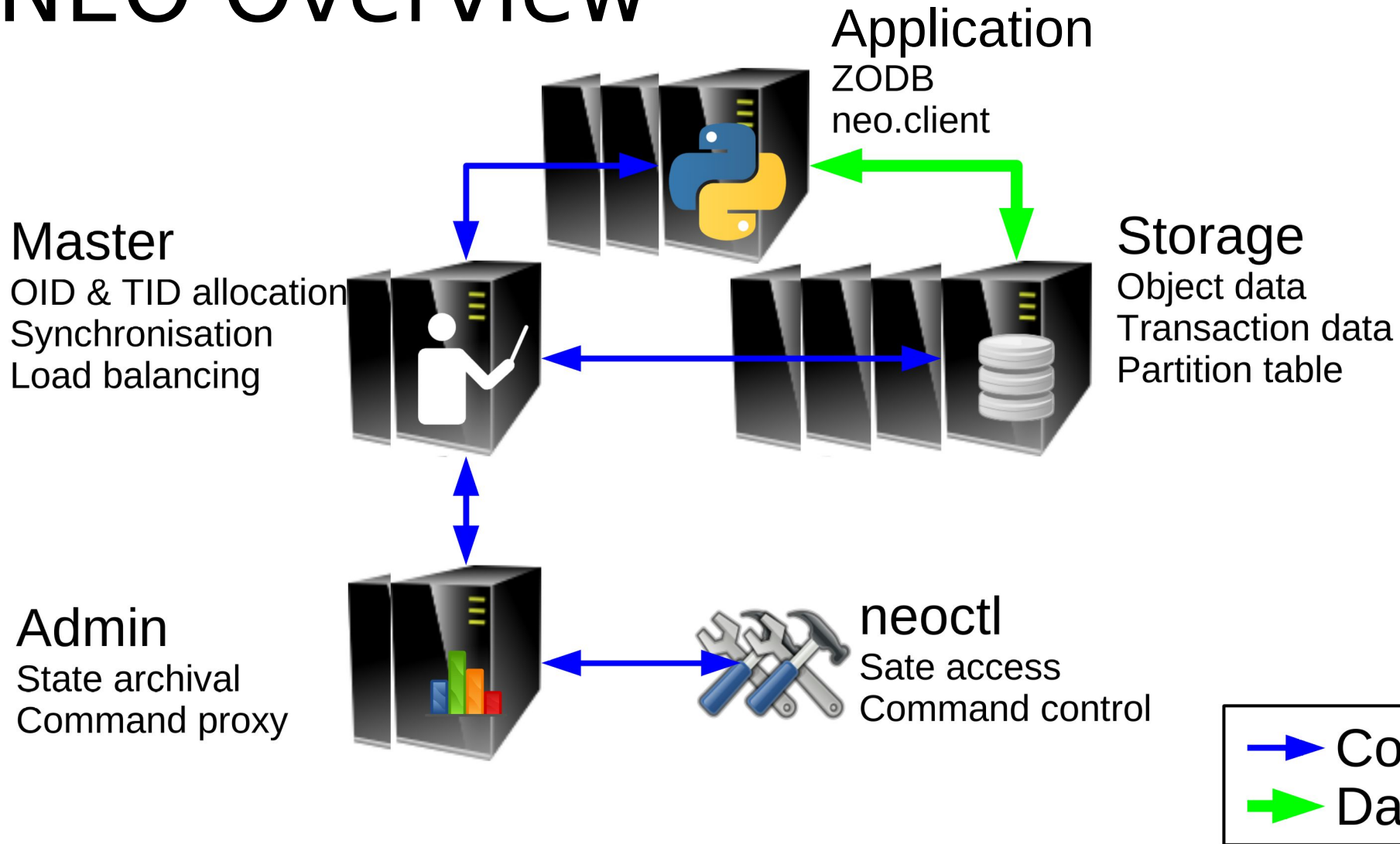
neo.ndarray out-of-core data



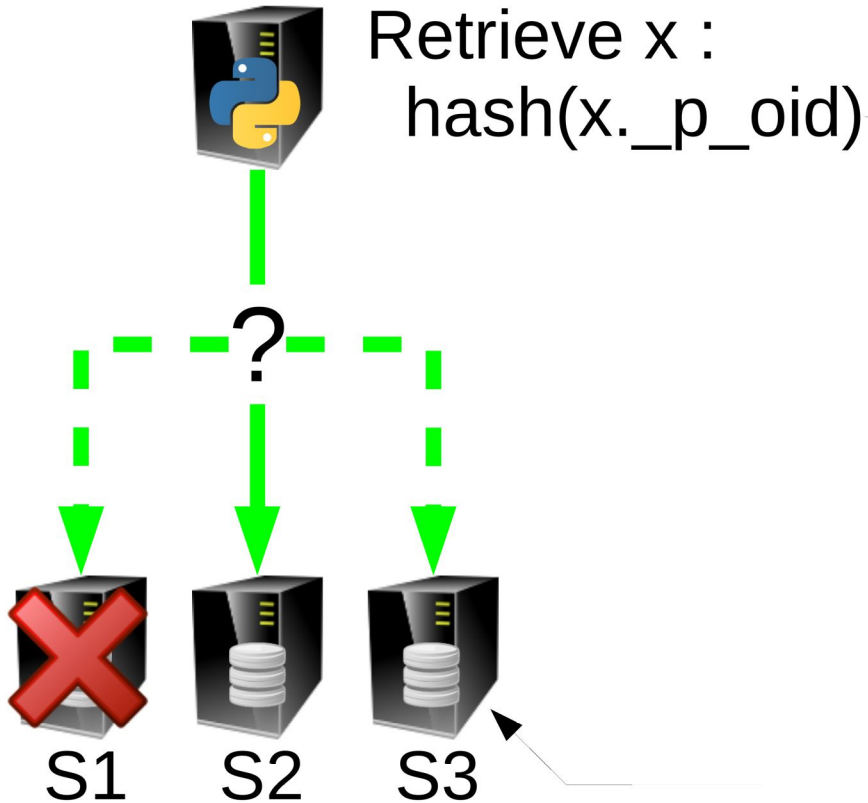
NEO Overview



NEO Overview



Object retrieval



Partition	Node	State
0	S1	Red
	S3	Green
1	S2	Green
	S3	Green
...

Variable

Parition	Node	State
S1	IP:PORT	?
S2	IP:PORT	Connected
S3	IP:PORT	?

Variable

Roadmap

- **Q3 2014: neo.ndarray**
- **Q3 2014: developer release of Wendelin**
- **Q4 2014: neo.ndarray with simple optimizations**
- **Q1 2015: mariadb embedded**
- **Q2 2015: coloured caches**
- **Q3 2015: coloured caches with C client cache**
- **Q4 2015: GO storage**

Challenges

- **Reduce latency → embedded mariadb ?**
- **Reduce SQL overhead → precompile queries ?**
- **Reduce copies → BLOB protocol ?**
- **Accelerate storage → C++ ? GO ?**
- **Optimize cache → colored caching**



Wendelin Exanalytics

2020 Big Data with MariaDB

2014-04-03 – Santa Clara updated 2014-06-16





Wendelin Exanalytics *2020 Big Data with MariaDB*

2014-04-03 – Santa Clara updated 2014-06-16



© 2014 Wendelin Project et al. – CC SA-NC

www.wendelin.io

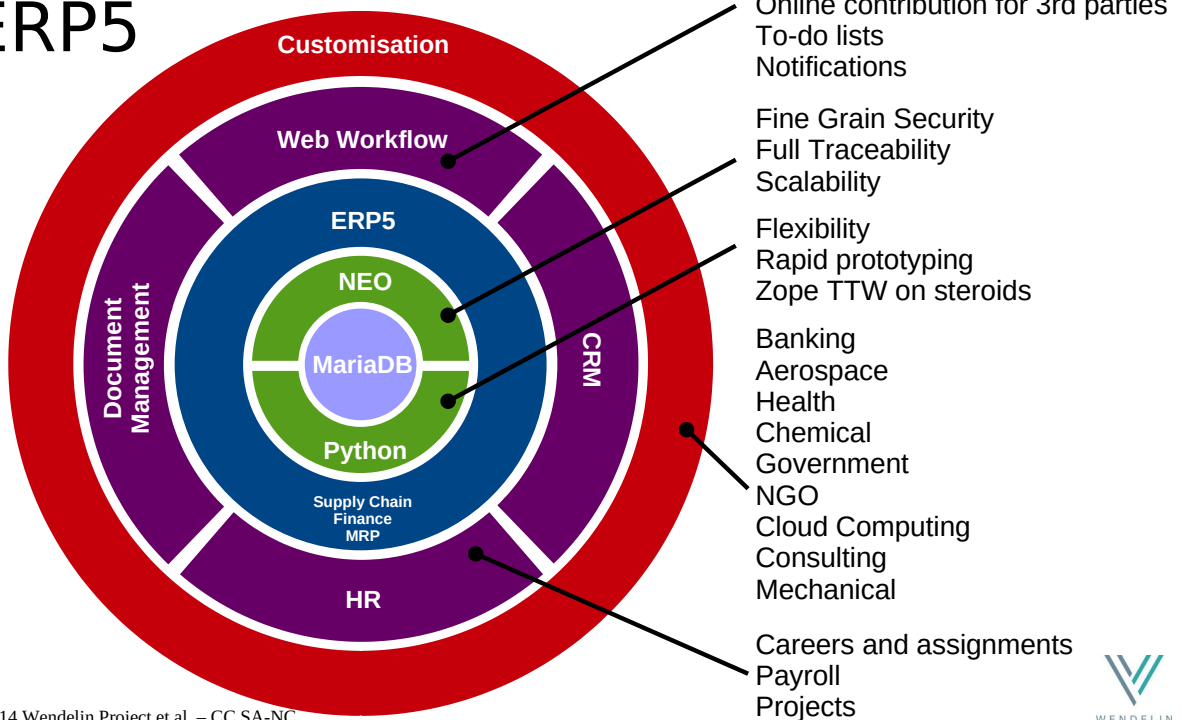


Agenda



- Our background: ERP5**
- **Our future: Wendelin Exanalytics**
- **Our challenge: out-of-core**

ERP5



© 2014 Wendelin Project et al. – CC SA-NC



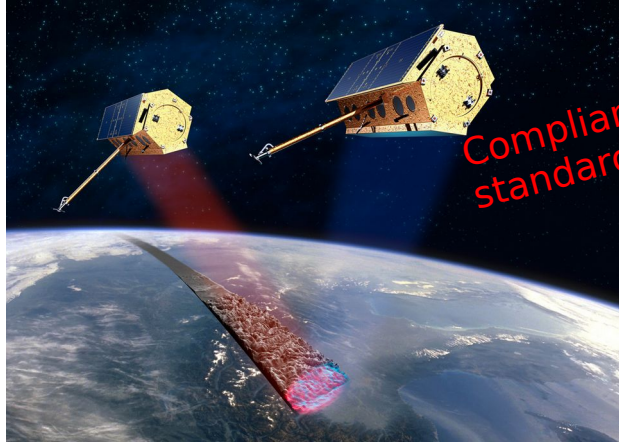
Terra-SAR X Satellite



MariaDB



Management of sales and production of images



Compliant with ESA
standard (ECSS)



AIRBUS
DEFENCE & SPACE

Accessible to Airbus
partners and distributors
Interfaces with DLR
(Germany Space Agency)

« With ERP5, our partners all over the world can access our infrastructure
and order online with complete security » Ralf Duering

© 2014 Wendelin Project et al. – CC SA-NC



SANEF Group



Online sales and customer relation for ETC Tolling



120.000 new customers / year
51.000 invoice/hour
7.000.000 contacts / year
250 users

« Web has become our primary sales channel. » Frédéric Charlier

© 2014 Wendelin Project et al. – CC SA-NC






Open Source ERP/CRM for S&P 100

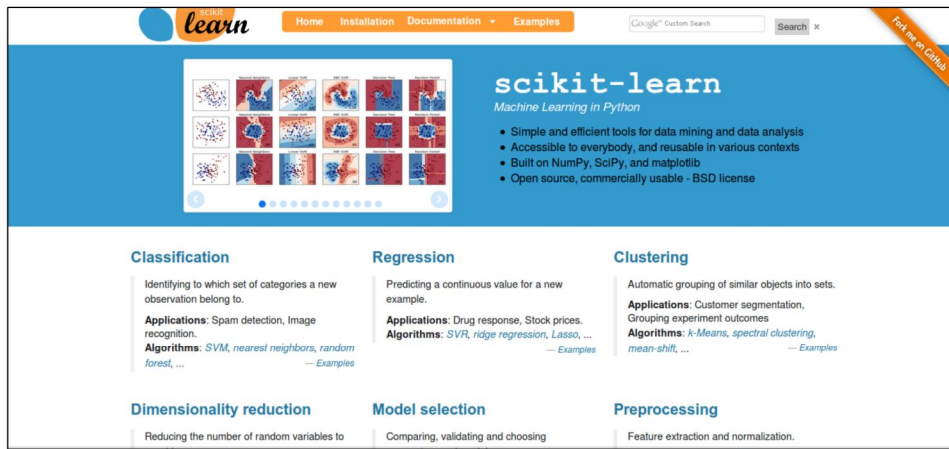
© 2014 Wendelin Project et al. – CC SA-NC



Agenda

- **Our background: ERP5**
-  **Our future: Wendelin Exanalytics**
- **Our challenges with MariaDB**

Take the Best Analytics scikit-learn.org



Made by Great Mathematicians

http://en.wikipedia.org/wiki/Fields_Medal

Number of Fields Medallists by country [\[edit\]](#)

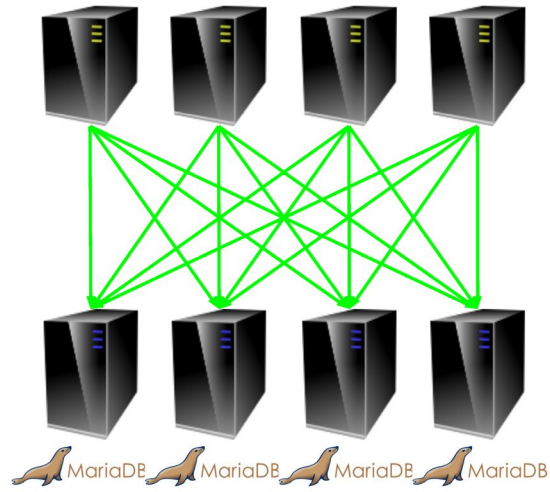


Wendelin Werner



WENDELIN

Add Distributed Storage neoppod.org



© 2014 Wendelin Project et al. – CC SA-NC



Add Elastic PaaS erp5.com

```
# Initialize data
data_size = 1000000
server_count = 1000
chunk_size = data_size / server_count
data = array(data_size)

# Process data in parallel on each server (Map Reduce, Batch, etc.)
for server in server_count:
    data.activate().process(server*chunk_size, chunk_size)
```



And Multicloud Deployment slapos.org

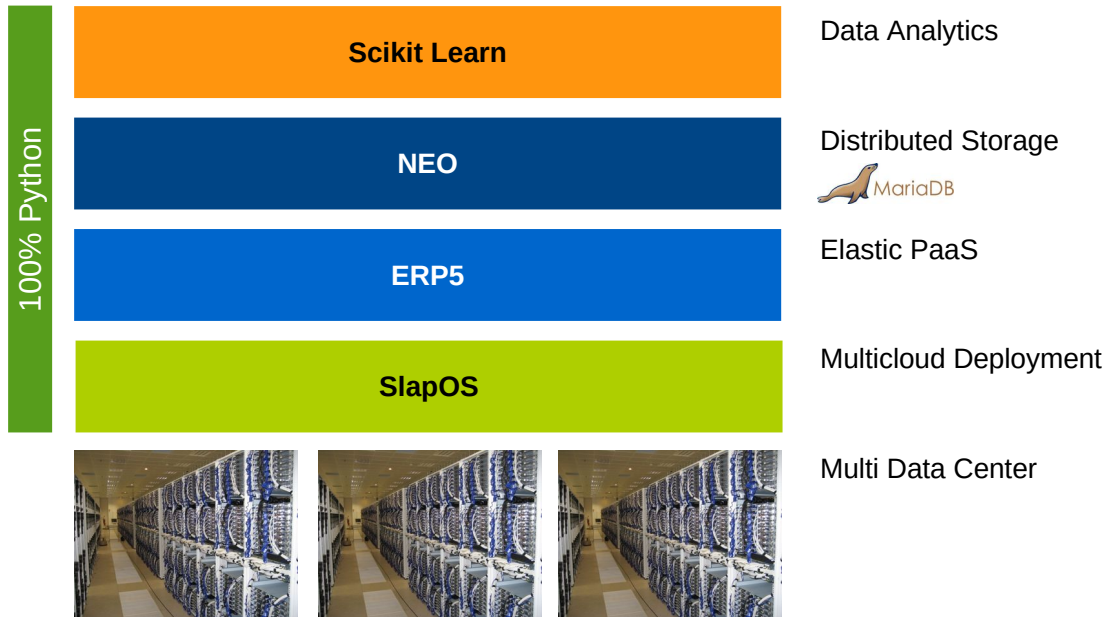
```

0 [buildout]
1
2 extends =
3 # "slapos" stack describes basic things needed for 99.9% of SlapOS Software
4 # Releases
5 ../../stack/slapos.cfg
6 # Extend here component profiles, like openssl, apache, mariadb, curl...
7 # Or/and extend a stack (lamp, tomcat) that does most of the work for you
8 # In this example we only need the dash binary to run a simple "hello world"
9 # shell script.
10 ../../component/dash/buildout.cfg
11
12 parts =
13 # Call installation of slapos.cookbook egg defined in stack/slapos.cfg (neede
14 # in 99.9% of Slapos Software Releases)
15 slapos-cookbook
16 # Call creation of instance.cfg file that will be called for deployment of
17 # instance
18 template
19
20 # Download instance.cfg.in (buildout profile used to deployment of instance).
21 # replace all ${foo:bar} parameters by real values, and change $$${foo:bar} to
22 # ${foo:bar}
23 [template]
24 recipe = slapos.recipe.template
25 url = ${:_profile_base_location_}/instance.cfg.in
26 output = ${buildout:directory}/instance.cfg
27 # MD5 checksum can be skipped for development (easier to develop), but must be filled for production
28 md5sum = 1fc461c00e86485bee77a942f39e3c43
29 mode = 0644
30
    
```



Wendelin Exanalytics Core 100% open

source



© 2014 Wendelin Project et al. – CC SA-NC



Wendelin User Interface renderjs.org

The screenshot displays the Wendelin User Interface with a dark sidebar on the left and a main content area. The sidebar includes sections for Search, Shop:Order, Clients, Messages, Shops, Products, and Settings. The main content area features a search bar, a table of stock market data, and three data visualization plots.

Company	Last Trade	Trade Time	Change	Prev Close	Open	Bid	Ask	1y Target Est
GOOG Google Inc.	597.74	12:12PM	14.81 (2.54%)	582.93	597.95	597.73 x 100	597.91 x 300	731.10
AAPL Apple Inc.	378.94	12:22PM	5.74 (1.54%)	373.20	381.02	378.92 x 300	378.99 x 100	505.94
AMZN Amazon.com Inc.	191.55	12:23PM	3.16 (1.68%)	188.39	194.99	191.52 x 300	191.58 x 100	240.32
ORCL Oracle Corporation	31.15	12:44PM	1.41 (4.72%)	29.74	30.67	31.14 x 6500	31.15 x 3200	36.11
MSFT Microsoft Corporation	25.50	12:27PM	0.66 (2.67%)	24.84	25.37	25.50 x 71100	25.51 x 17800	31.50
CSCO Cisco Systems, Inc.	18.65	12:45PM	0.97 (5.49%)	17.68	18.23	18.65 x 10300	18.66 x 24000	21.12
YHOO Yahoo! Inc.	15.81	12:25PM	0.11 (0.67%)	15.70	15.94	15.79 x 6100	15.80 x 17000	18.16

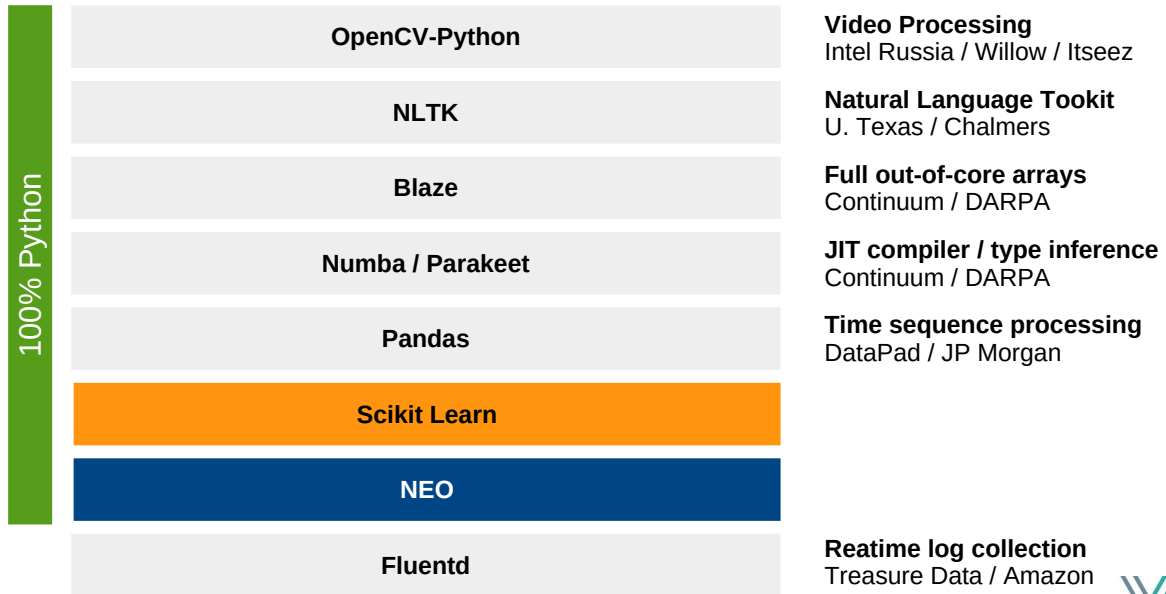
The visualization plots include:

- Outlier detection:** A scatter plot showing data points with a decision boundary and outliers.
- LASSO Path:** A plot showing the path of coefficients for different variables as the regularization parameter λ varies.
- Prediction plot:** A plot showing the predicted function $f(x)$ and its 95% confidence interval against observations.

© 2014 Wendelin Project et al. – CC SA-NC



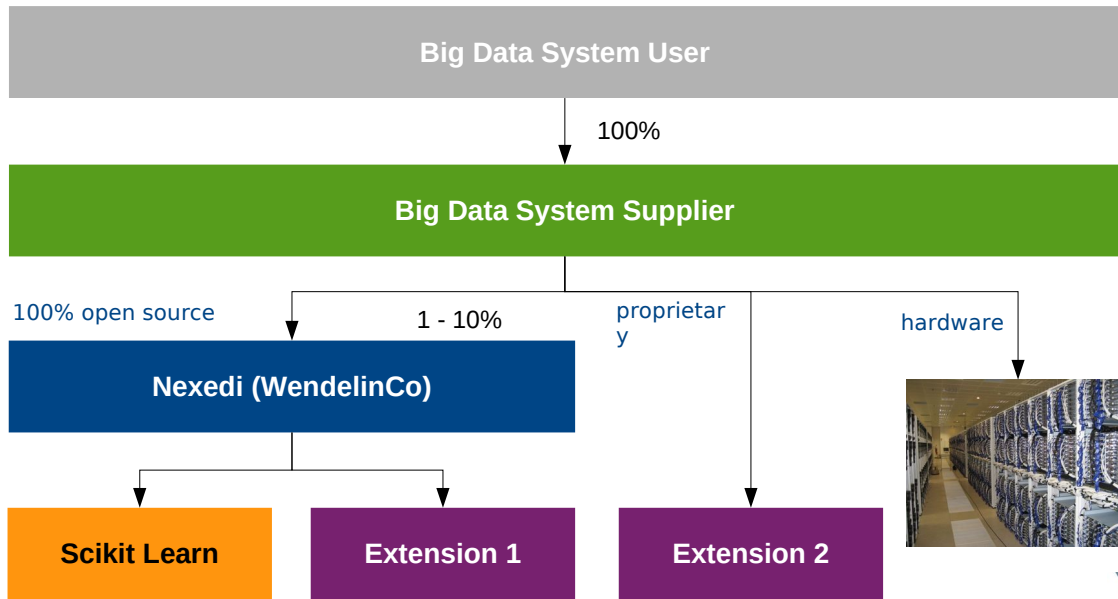
Wendelin Options 100% open source



Wendelin Applications

- **Intrusion detection**
- **Fraud detection**
- **Business and economic prevision**
- **Marketing**
- **Media analysis**
- **Public security**
- **Brain Computer Interface**
- **Internet Of Things**

Business Model: German Style No VC



© 2014 Wendelin Project et al. – CC SA-NC



Agenda

- **Our background: ERP5**
- **Our future: Wendelin Exanalytics**



Our challenge: out-of-core

Out-of-core arrays

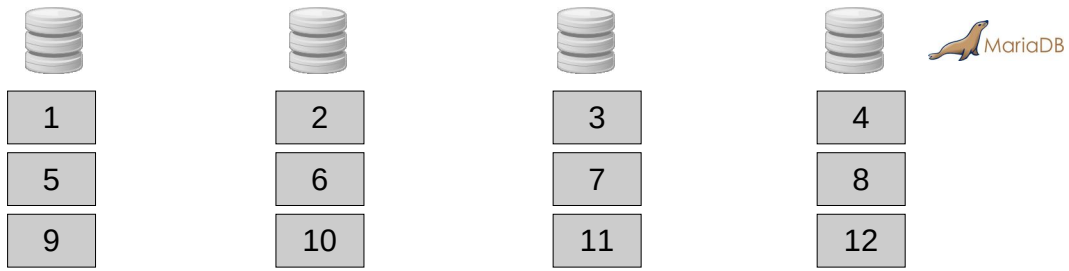
```
# Numpy
np.ndarray(shape=(2,2), dtype=float, order='F')

# Out-of-core data
np.ndarray(shape=(1e18,2), dtype=float, order='F')      2 Exabyte

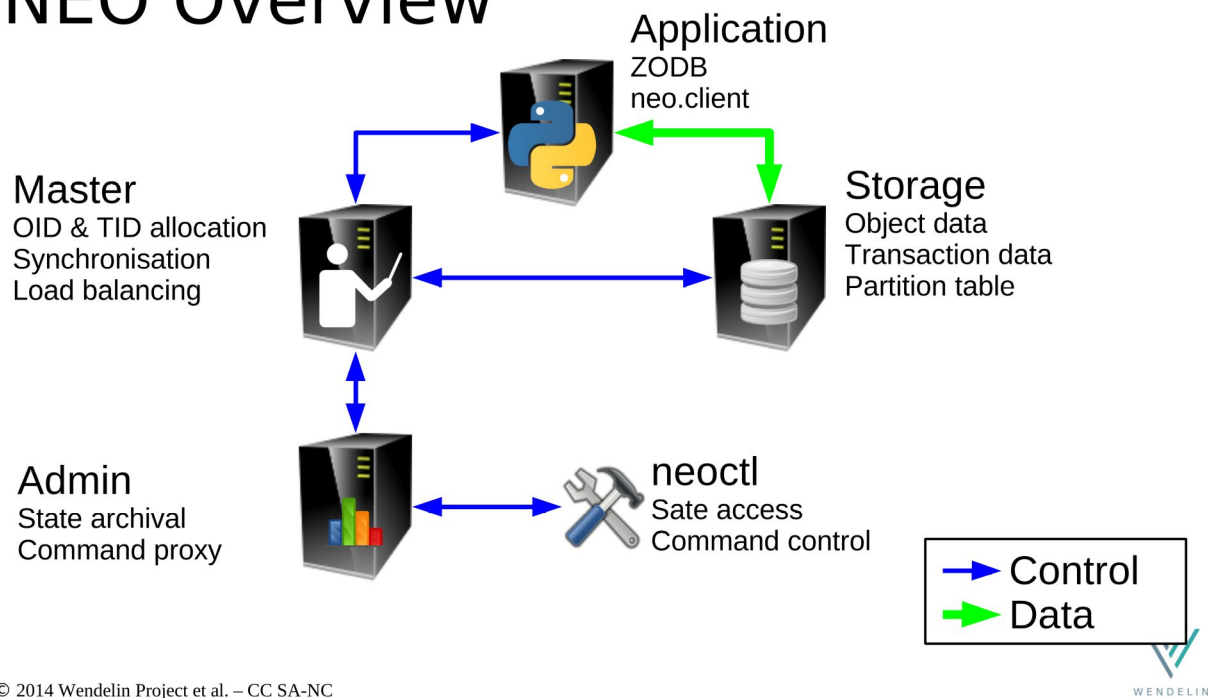
# Full out-of-core
np.ndarray(shape=(1e9,2e9), dtype=float, order='F')      2 Exabyte
```

Best out-of-core topology depends on the algorithm and array geometry

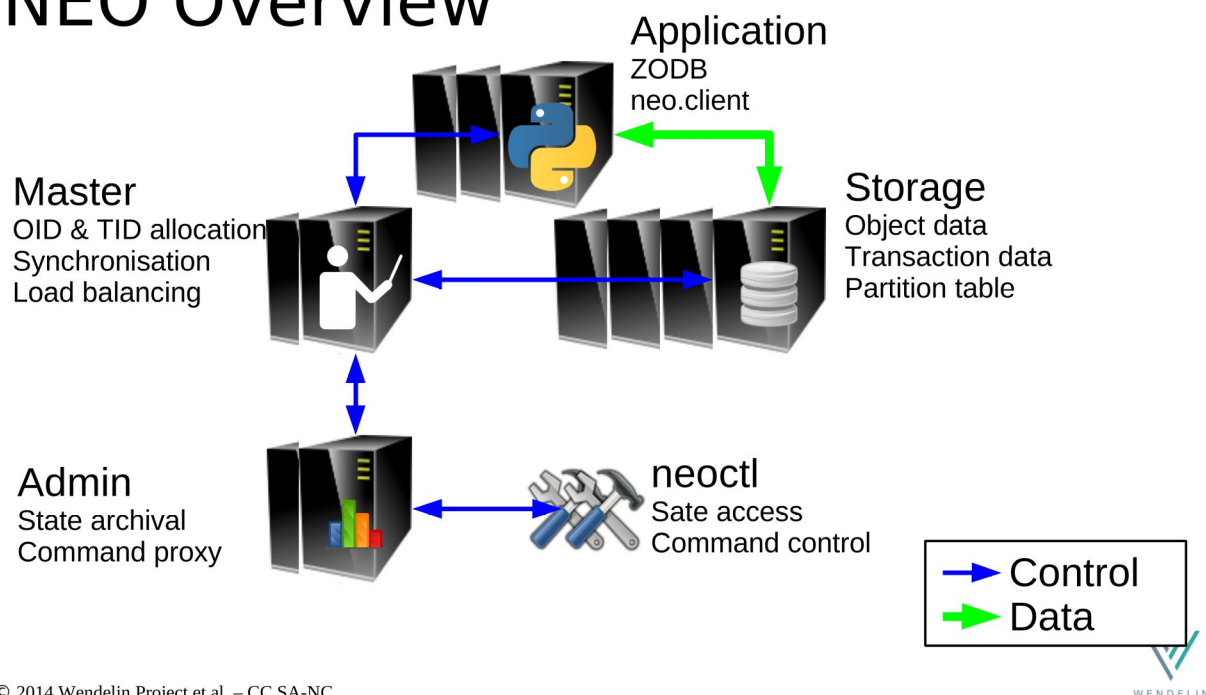
neo.ndarray out-of-core data



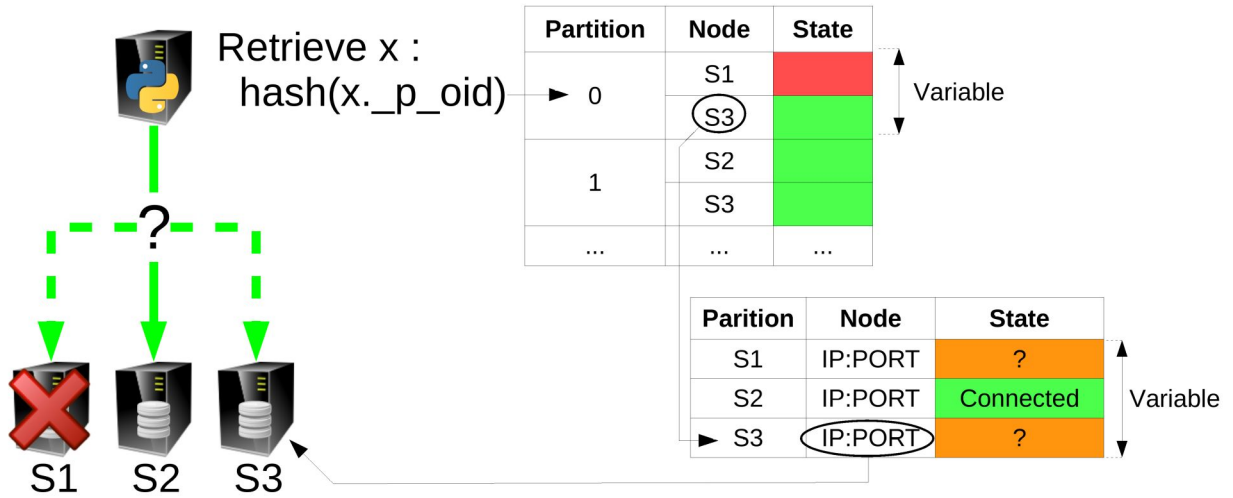
NEO Overview



NEO Overview



Object retrieval



Roadmap

- **Q3 2014: neo.ndarray**
- **Q3 2014: developer release of Wendelin**
- **Q4 2014: neo.ndarray with simple optimizations**
- **Q1 2015: mariadb embedded**
- **Q2 2015: coloured caches**
- **Q3 2015: coloured caches with C client cache**
- **Q4 2015: GO storage**

Challenges

- **Reduce latency** → **embedded mariadb ?**
- **Reduce SQL overhead** → **precompile queries ?**
- **Reduce copies** → **BLOB protocol ?**
- **Accelerate storage** → **C++ ? GO ?**
- **Optimize cache** → **colored caching**



Wendelin Exanalytics *2020 Big Data with MariaDB*

2014-04-03 – Santa Clara updated 2014-06-16



© 2014 Wendelin Project et al. – CC SA-NC

www.wendelin.io

