Wendelin Exanalytics

2020 Big Data with MariaDB

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

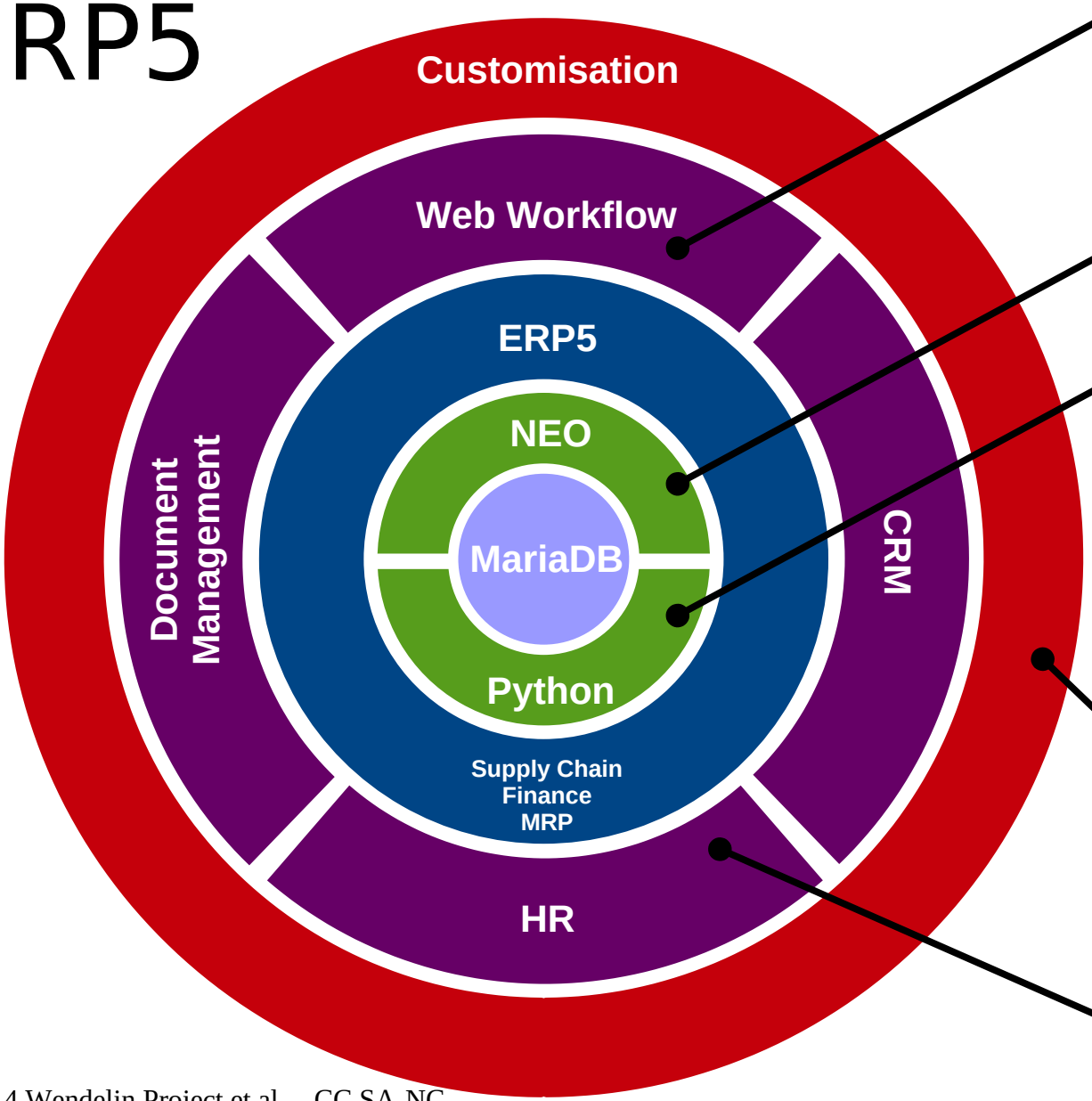
www.wendelin.io

Agenda

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

WENDELIN

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

WENDELIN

Terra-SAR X Satellite

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

WENDELIN

SANEF Group

MariaDB ERP5

Online sales and customer relation for ETC Tolling

Implemented in 4 months

Sanef

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

WENDELIN

Open Source ERP/CRM for S&P 100

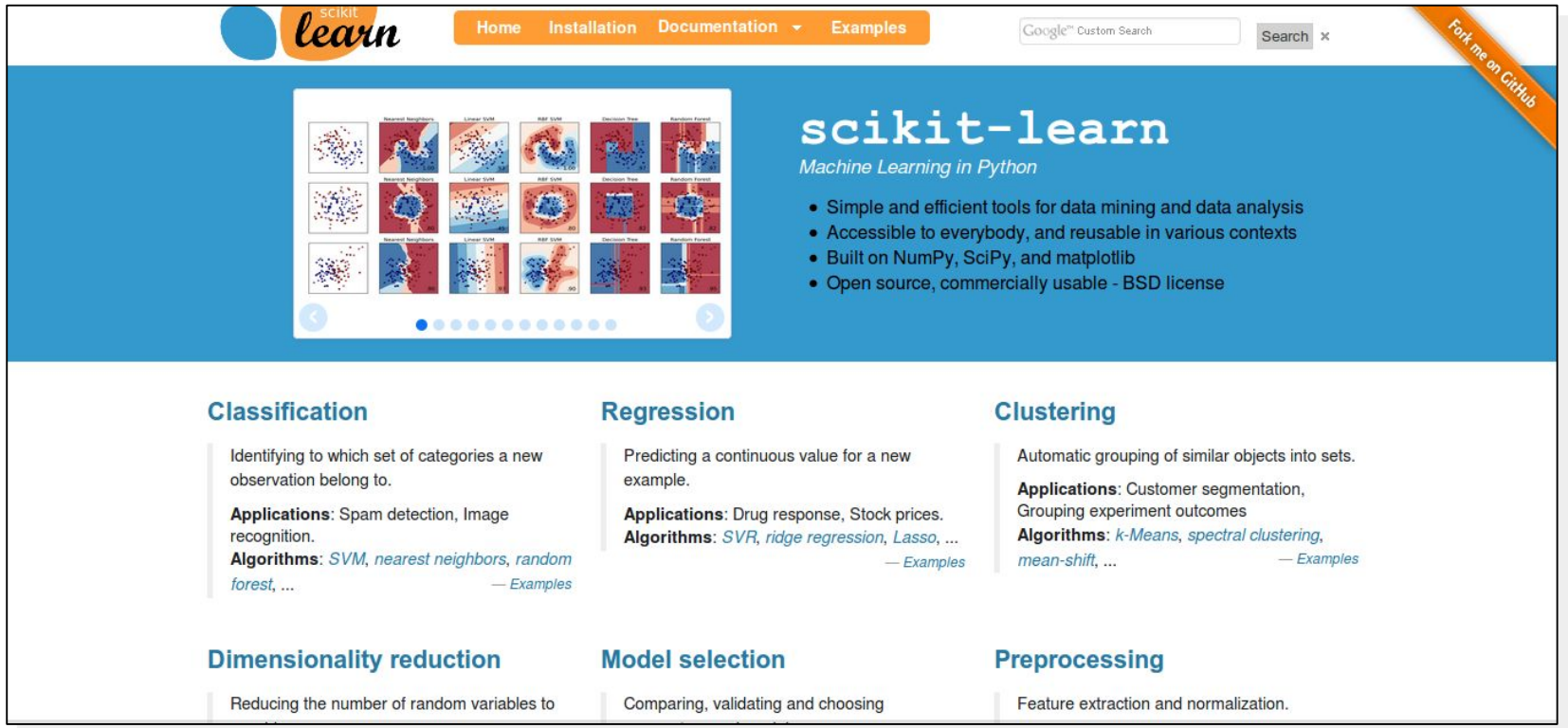
WENDELIN

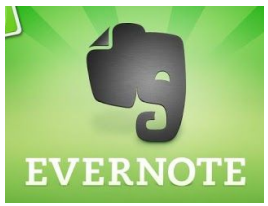
Agenda

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

WENDELIN

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]

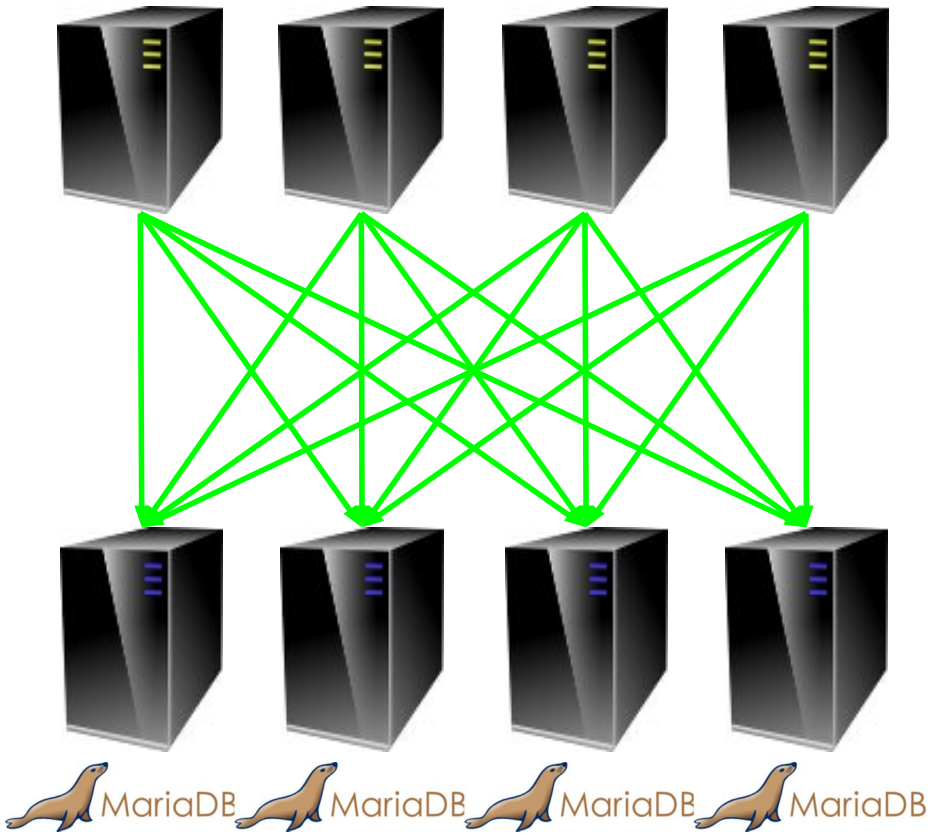
United States	11
France	10
Soviet Union (3) / Russia (6)	9
United Kingdom	6
Japan	3
Belgium	2
West Germany (1) / Germany (0)	1
Australia	1
British Hong Kong	1
Finland	1
Israel	1
Italy	1
Norway	1
New Zealand	1
Sweden	1
Vietnam	1

Wendelin Werner

WENDELIN

Add Distributed Storage neoppod.org

NEO



WENDELIN

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)
```

ERP5 PaaS

WENDELIN

And Multicloud Deployment slapos.org

MMC Rus

Aide et Action
L'Education change le monde

sanef

AIRBUS DEFENCE & SPACE

WENDELIN

Wendelin Exanalytics Core 100% open source

100% Python

Scikit Learn — Data Analytics

NEO — Distributed Storage

MariaDB

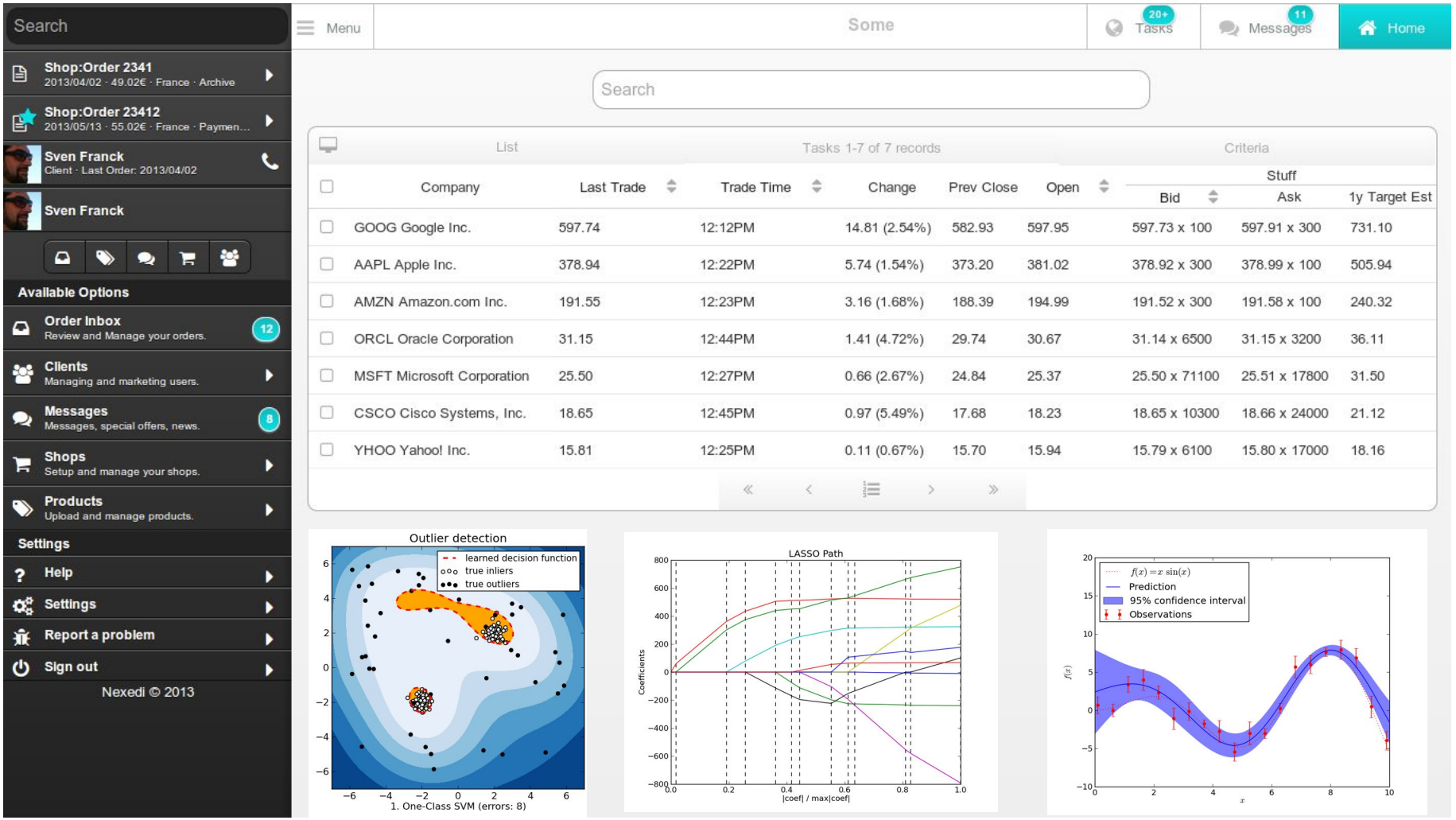
ERP5 — Elastic PaaS

SlapOS — Multicloud Deployment

Multi Data Center

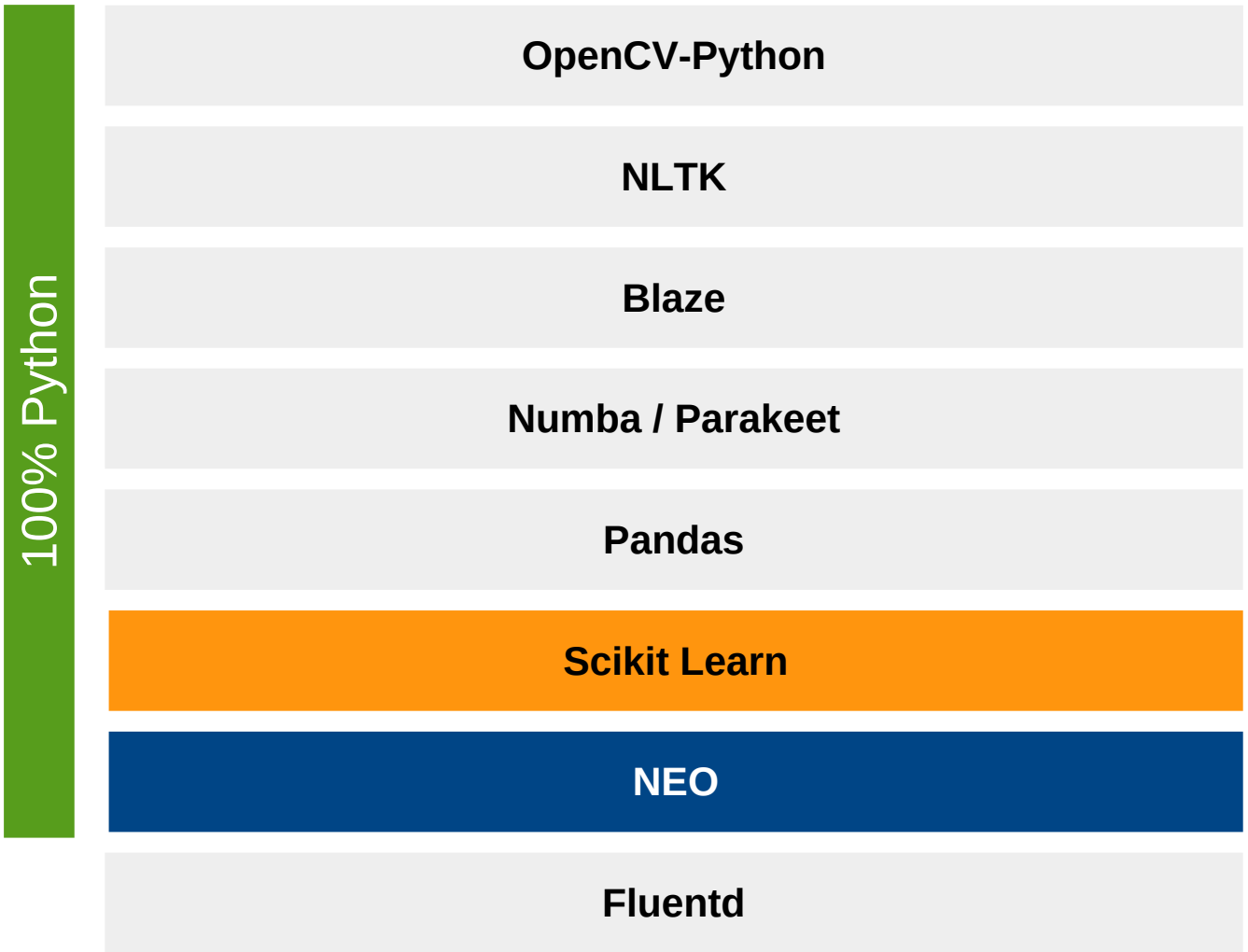
WENDELIN

Wendelin User Interface renderjs.org



WENDELIN

Wendelin Options 100% open source



Video Processing
Intel Russia / Willow / Itseez

Natural Language Tookit
U. Texas / Chalmers

Full out-of-core arrays
Continuum / DARPA

JIT compiler / type inference
Continuum / DARPA

Time sequence processing
DataPad / JP Morgan

Reatime log collection
Treasure Data / Amazon

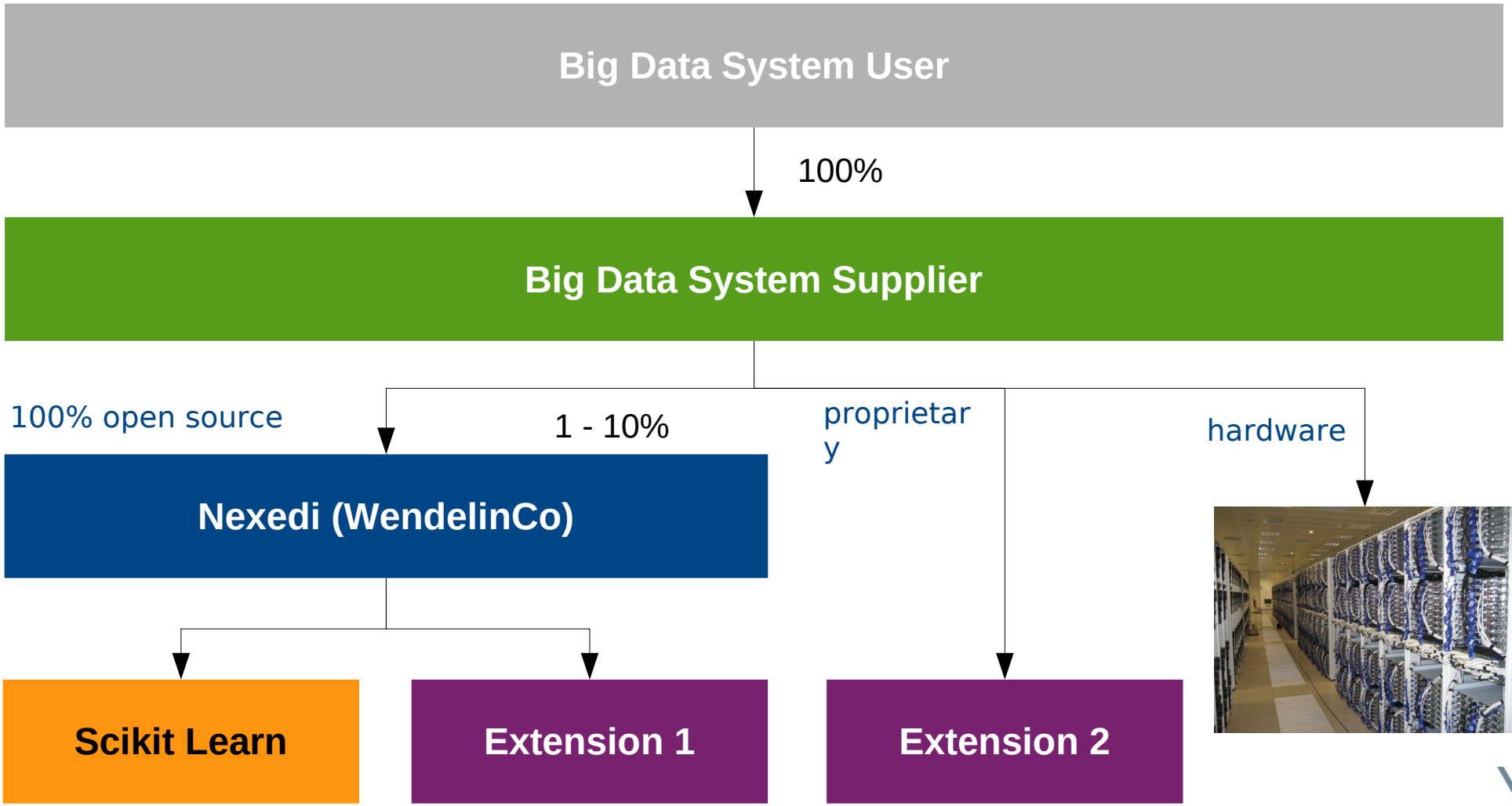
WENDELIN

Wendelin Applications

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

WENDELIN

Business Model: German Style No VC



WENDELIN

Agenda

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

WENDELIN

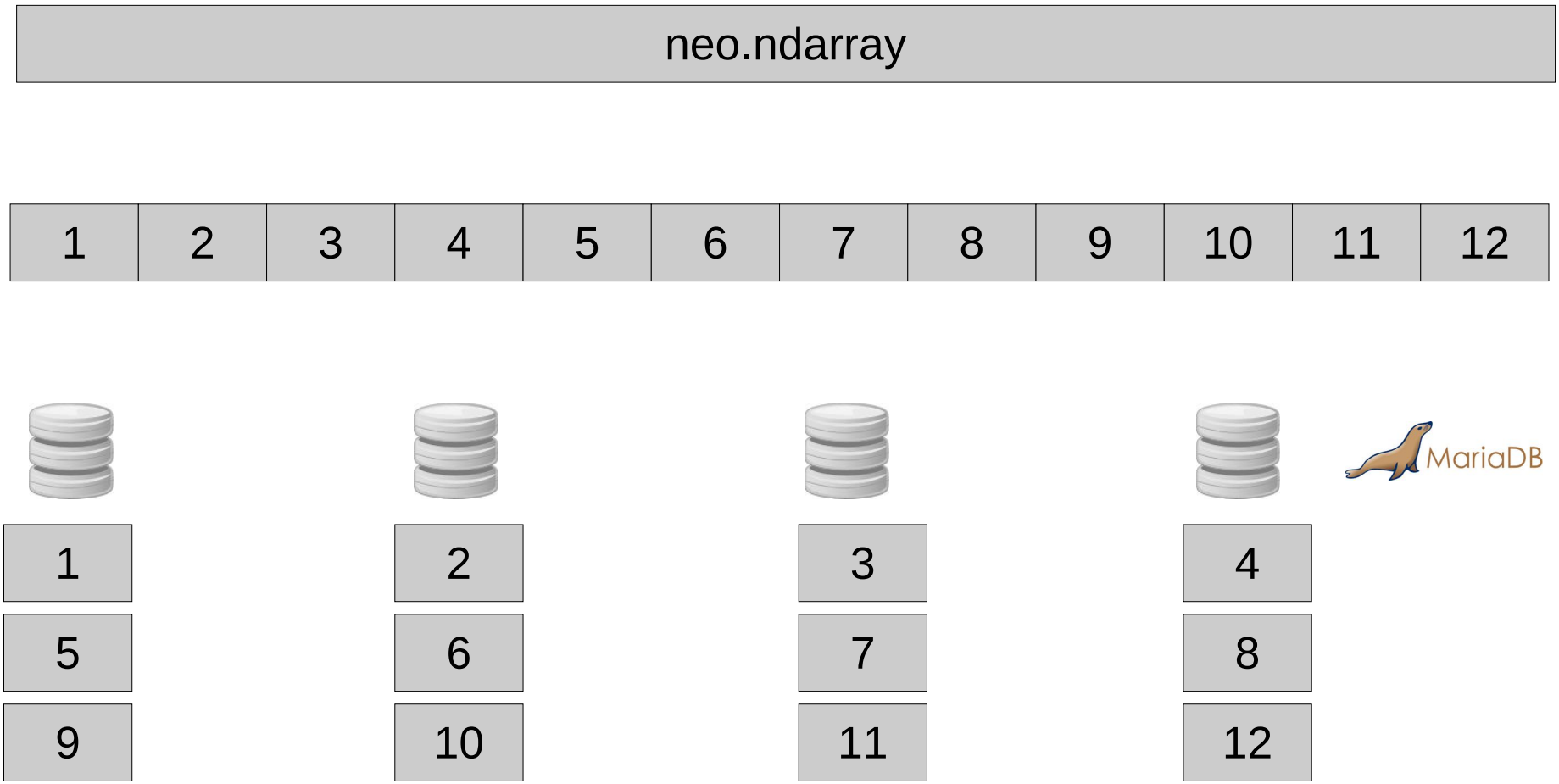
Out-of-core arrays

```
# 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

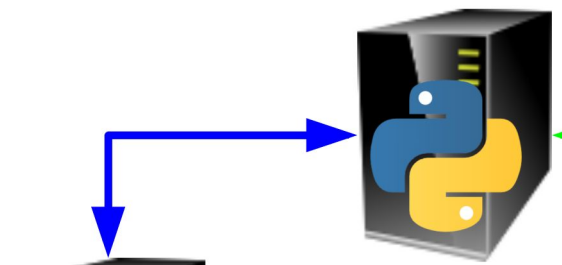
WENDELIN

neo.ndarray out-of-core data



WENDELIN

NEO Overview



Application
ZODB
neo.client

Master
OID & TID allocation
Synchronisation
Load balancing

Storage
Object data
Transaction data
Partition table

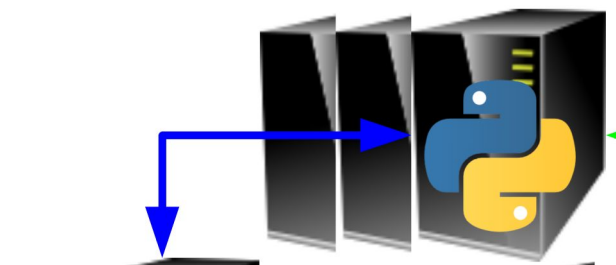
Admin
State archival
Command proxy

neoctl
Sate access
Command control

Control
Data

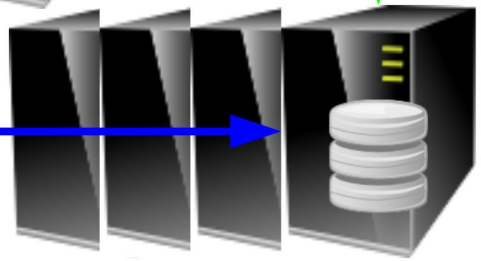
WENDELIN

NEO Overview



Application
ZODB
neo.client

Master
OID & TID allocation
Synchronisation
Load balancing



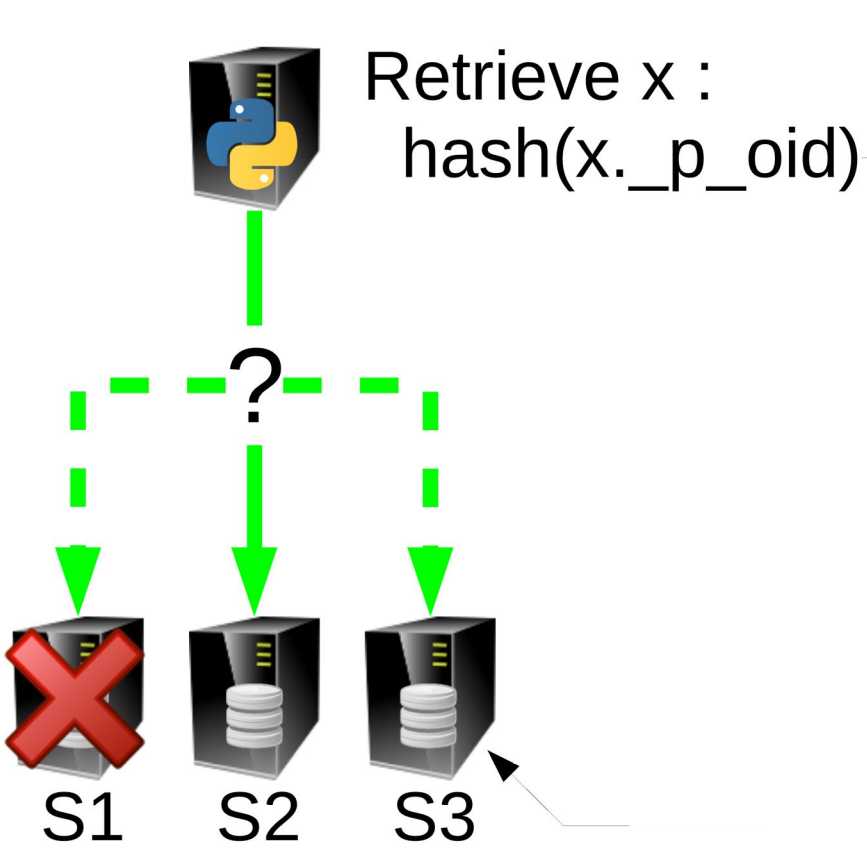
Storage
Object data
Transaction data
Partition table

Admin
State archival
Command proxy

neoctl
Sate access
Command control

WENDELIN

Object retrieval



Partition	Node	State
► 0	S1	
	S3	
1	S2	
	S3	
...	...	...

Variable

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

Variable

WENDELIN

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**

WENDELIN

Challenges

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

WENDELIN

Wendelin Exanalytics

2020 Big Data with MariaDB

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

www.wendelin.io

Wendelin Exanalytics

2020 Big Data with MariaDB

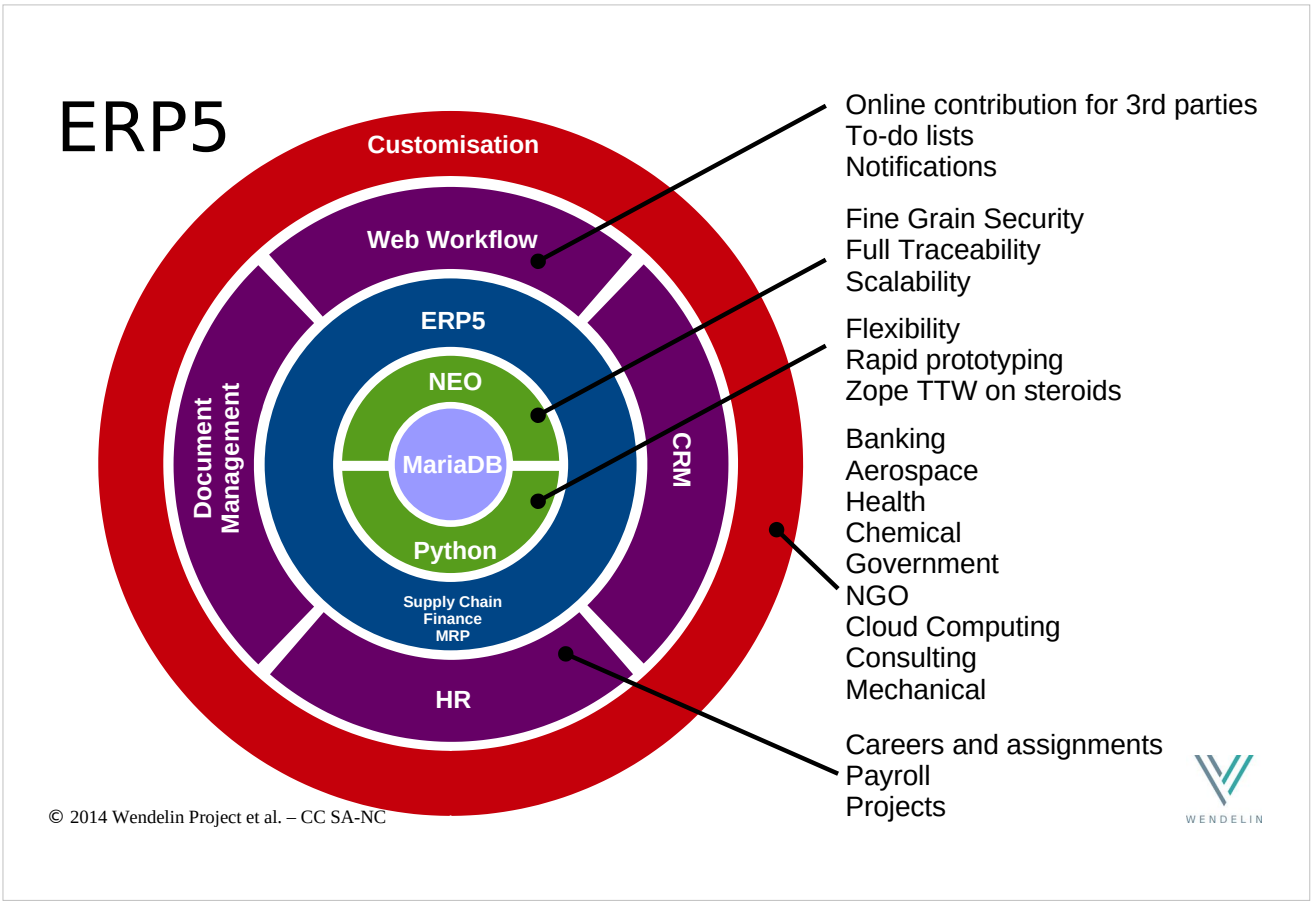
2014-04-03 – Santa Clara updated 2014-06-16

www.wendelin.io

Agenda

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

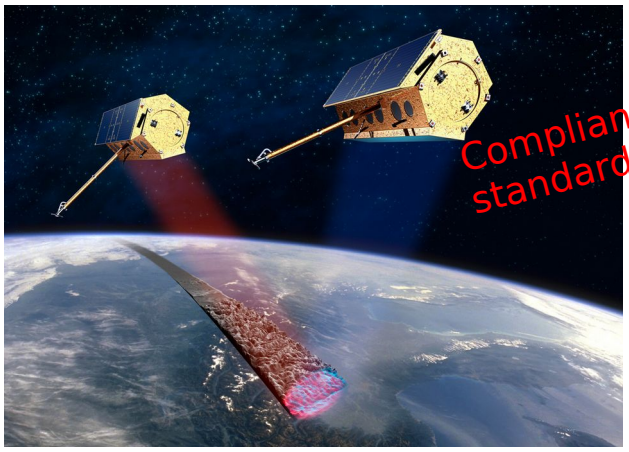
WENDELIN



Terra-SAR X Satellite

MariaDB ERP5

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

WENDELIN

SANEF Group

MariaDB ERP5

Online sales and customer relation for ETC Tolling

Implemented in 4 months

Sanef

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

WENDELIN

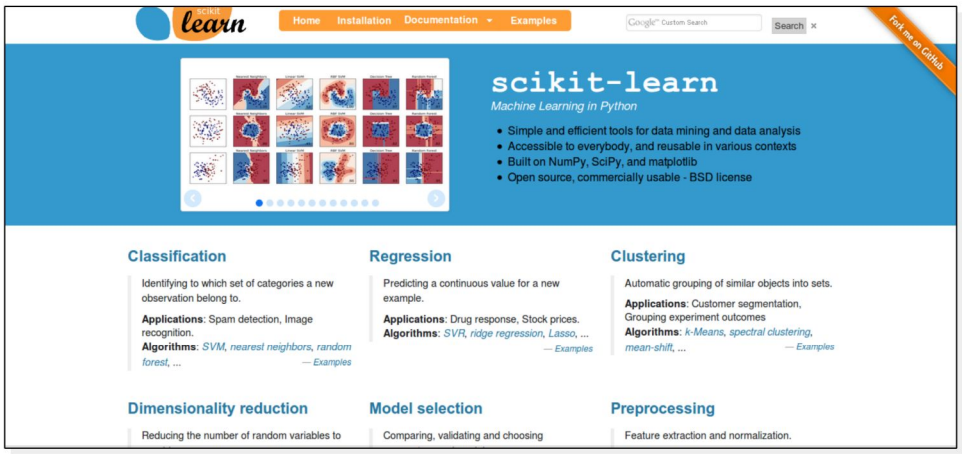
Open Source ERP/CRM for S&P 100

Agenda

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

WENDELIN

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]

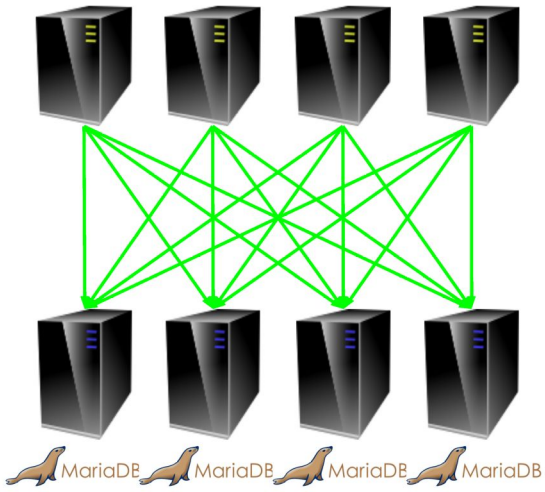
United States	11
France	10
Soviet Union (3) / Russia (6)	9
United Kingdom	6
Japan	3
Belgium	2
West Germany (1) / Germany (0)	1
Australia	1
British Hong Kong	1
Finland	1
Israel	1
Italy	1
Norway	1
New Zealand	1
Sweden	1
Vietnam	1

Wendelin Werner

WENDELIN

Add Distributed Storage neoppod.org

NEO



WENDELIN

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)
```

ERP5 PaaS

WENDELIN

And Multicloud Deployment slapos.org

MMC Rus

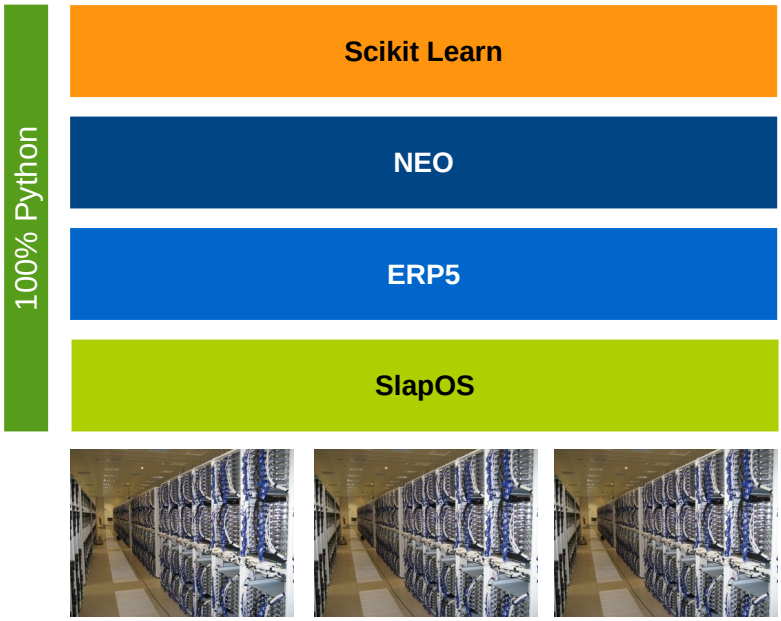
Aide et Action
L'Education change le monde

sanef

AIRBUS DEFENCE & SPACE

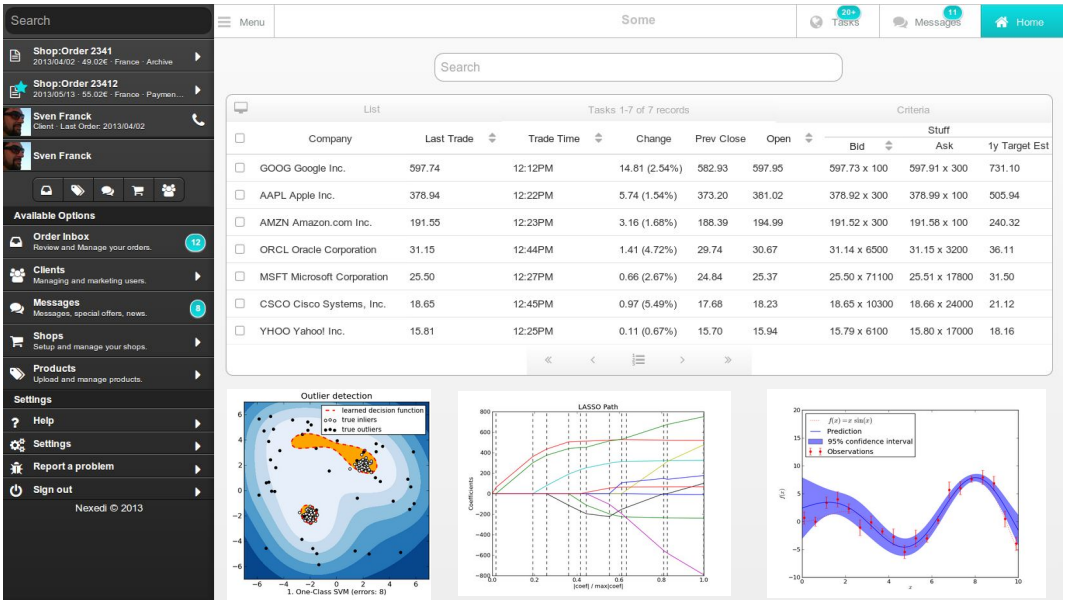
WENDELIN

Wendelin Exanalytics Core 100% open source



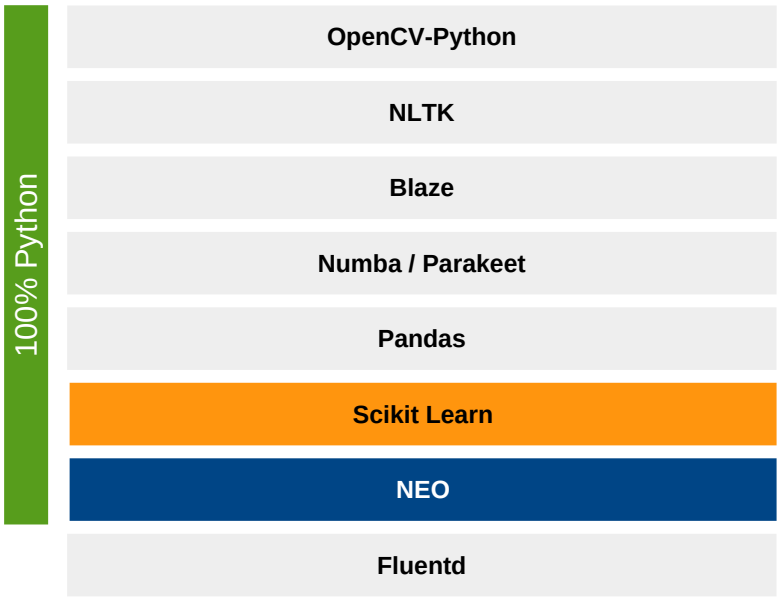
WENDELIN

Wendelin User Interface renderjs.org



WENDELIN

Wendelin Options 100% open source



Video Processing
Intel Russia / Willow / Itseez

Natural Language Tookit
U. Texas / Chalmers

Full out-of-core arrays
Continuum / DARPA

JIT compiler / type inference
Continuum / DARPA

Time sequence processing
DataPad / JP Morgan

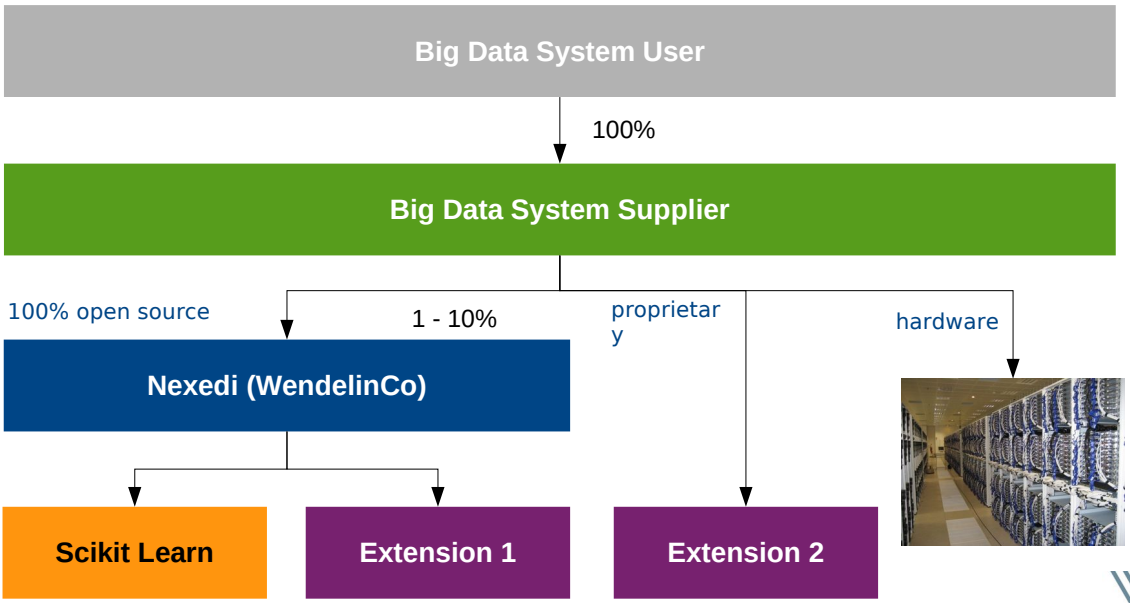
Reatime log collection
Treasure Data / Amazon

WENDELIN

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



WENDELIN

Agenda

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

WENDELIN

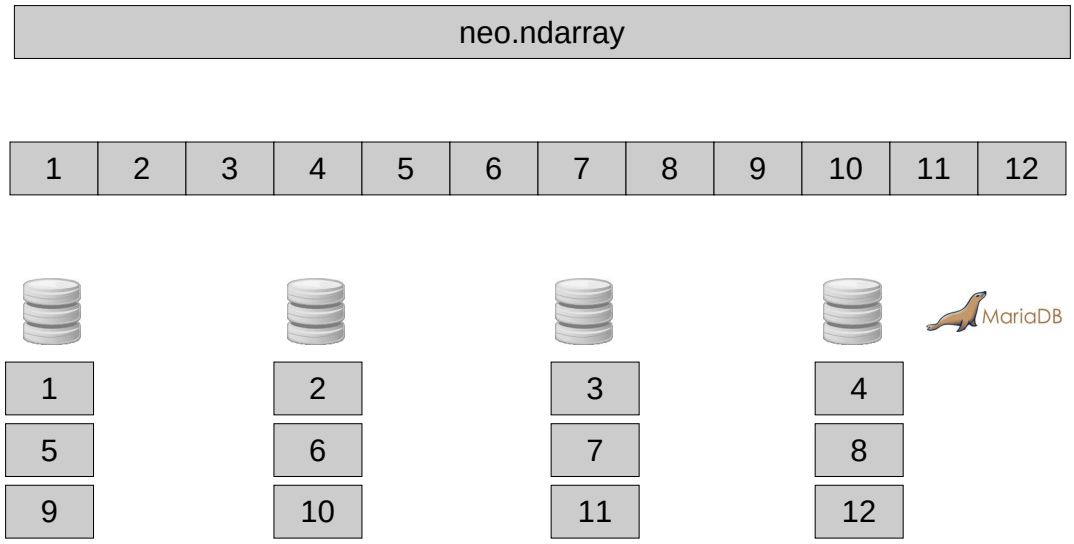
Out-of-core arrays

```
# 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
```

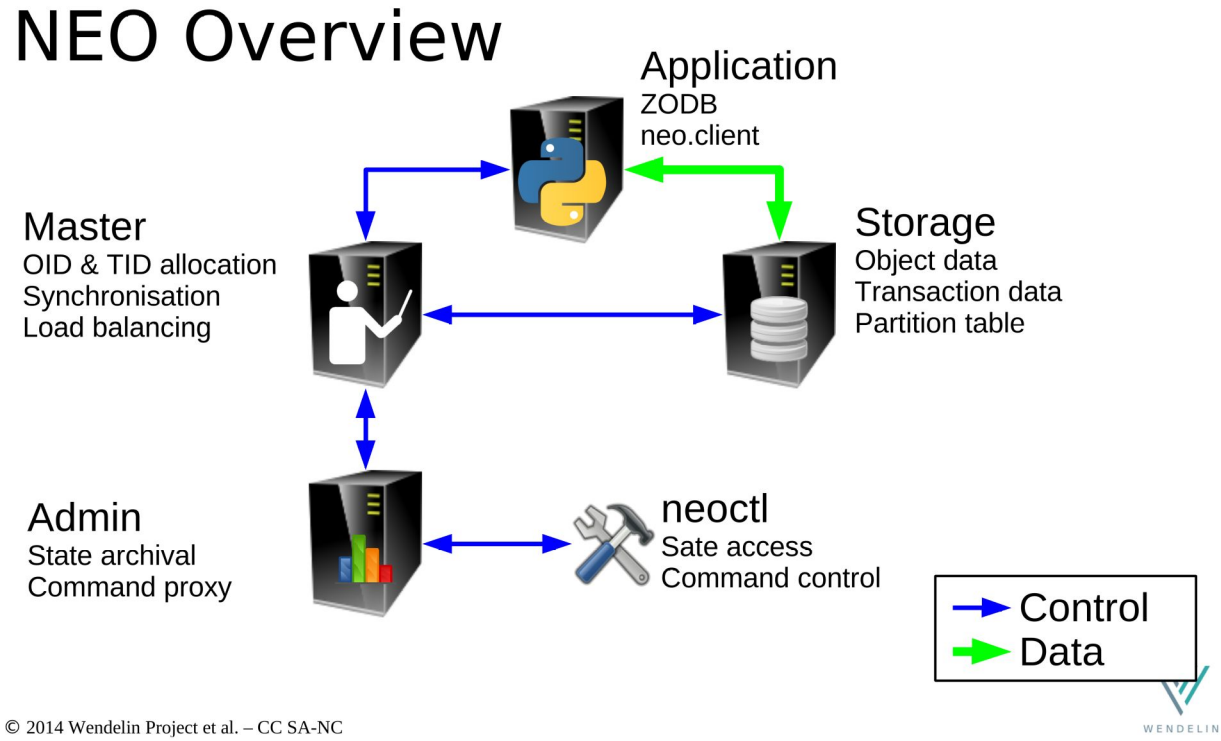
Best out-of-core topology depends on the algorithm and array geometry

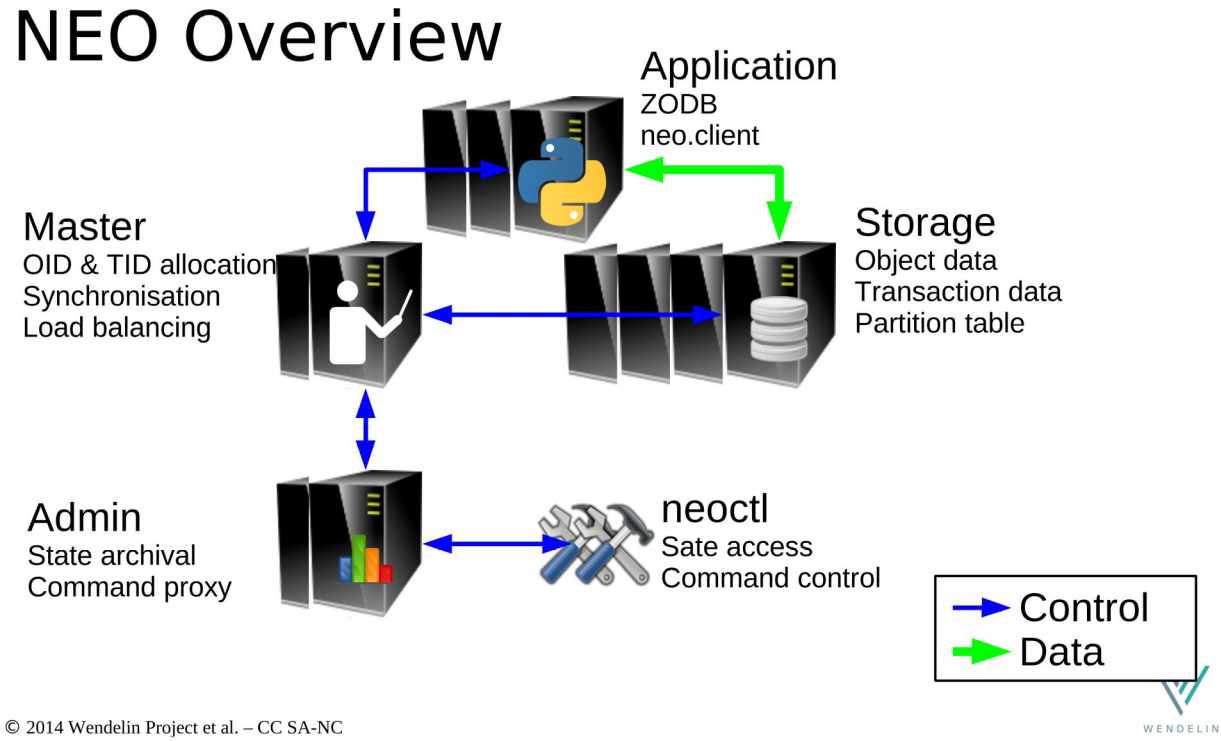
WENDELIN

neo.ndarray out-of-core data

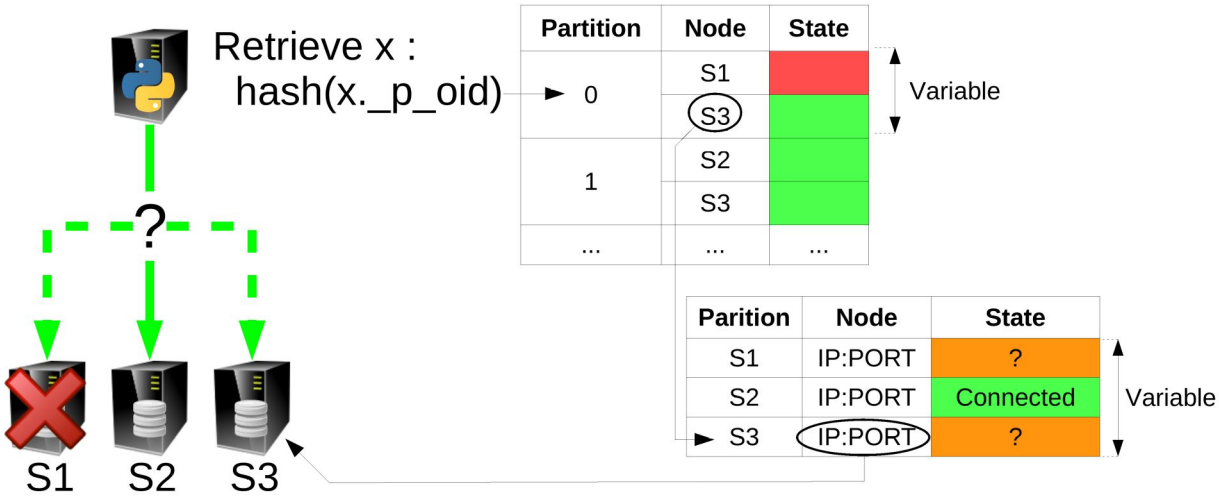


WENDELIN





Object retrieval



WENDELIN

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

www.wendelin.io