Wendelin Exanalytics

2020 Big Data with MariaDB

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

www.wendelin.io

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

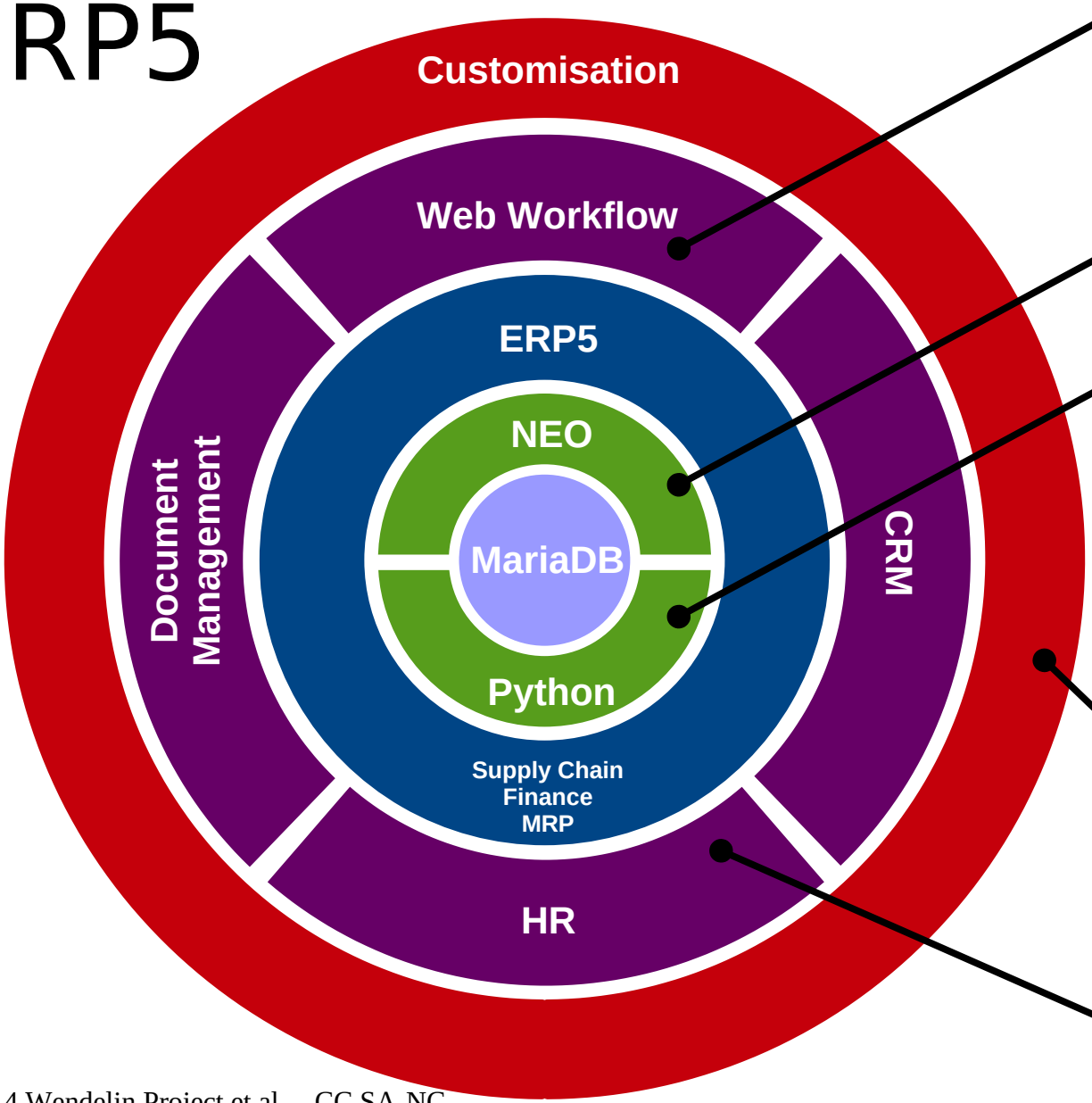
Agenda

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

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

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

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

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

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

WENDELIN

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

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

Open Source ERP/CRM for S&P 100

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

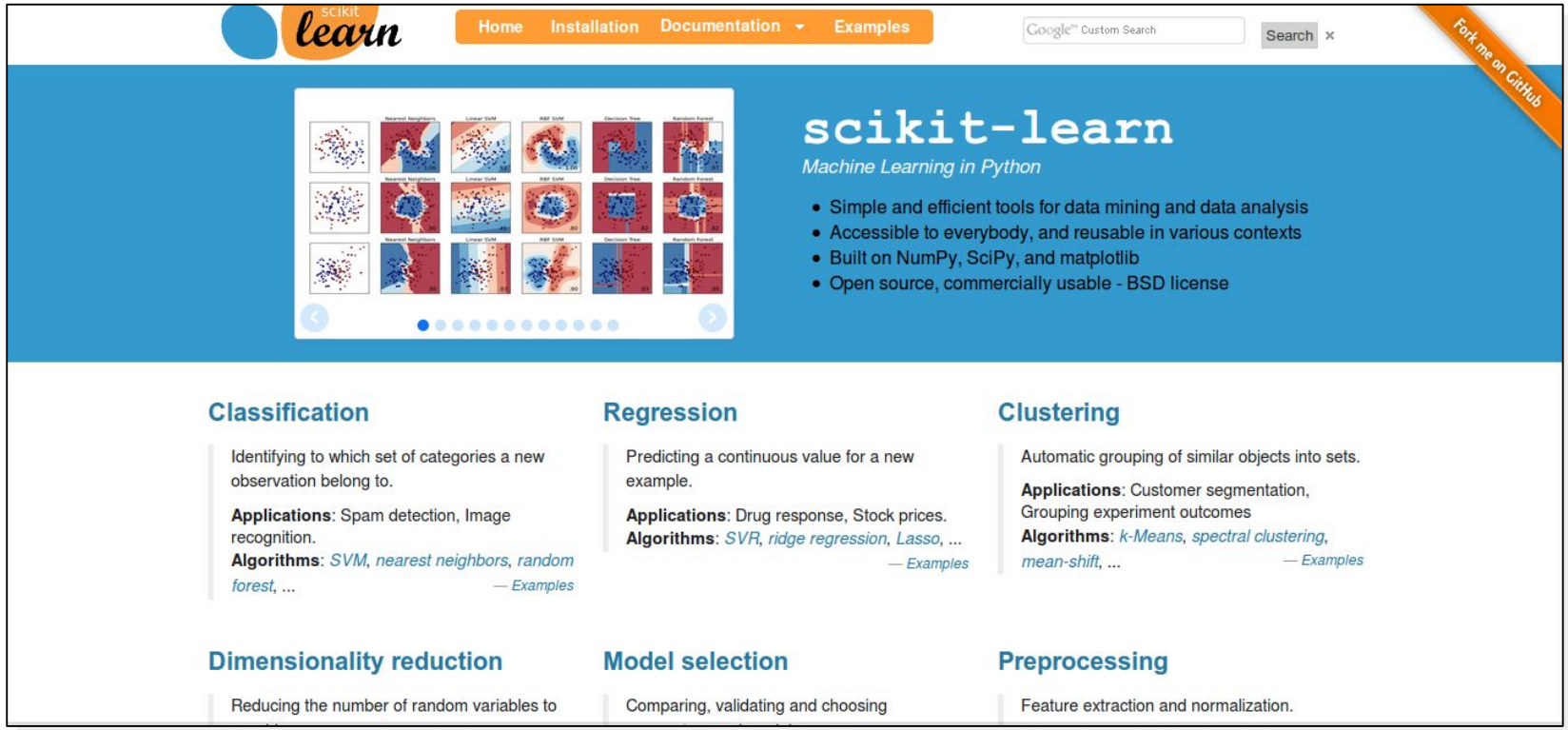
Agenda

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

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

WENDELIN

Take the Best Analytics scikit-learn.org



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

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

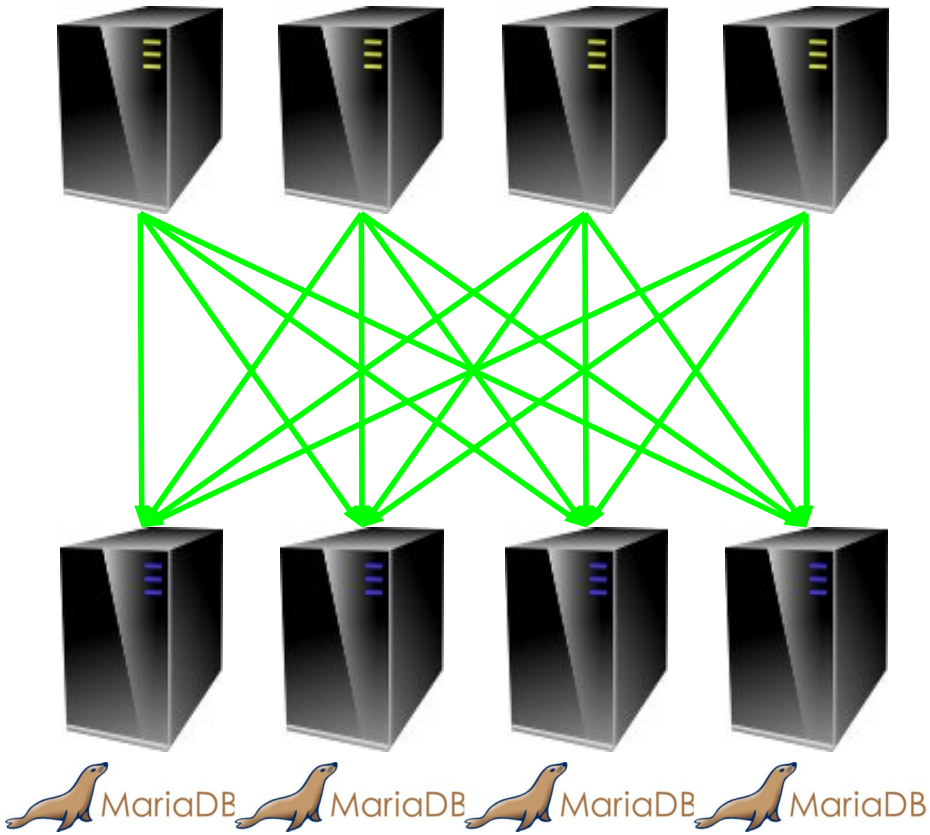
Wendelin Werner

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

WENDELIN

Add Distributed Storage neoppod.org

NEO



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

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

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

And Multicloud Deployment slapos.org

Update software profile - helloworld (workspace/my_repository/software)

Profiles | Workspace | Run software | Software log | SR result | Run instance | Instance log | Inspect instance | Destroy instance

Edit software profile [More]:

```
[buildout]

extends =
# "slapos" stack describes basic things needed for 99.9% of SlapOS Software
# Releases
  ../../stack/slapos.cfg
# Extend here component profiles, like openssl, apache, mariadb, curl...
# Or/and extend a stack (lamp, tomcat) that does most of the work for you
# In this example we only need the dash binary to run a simple "hello world"
# shell script.
  ../../component/dash/buildout.cfg

parts =
# Call installation of slapos.cookbook egg defined in stack/slapos.cfg (neede
# in 99,9% of Slapos Software Releases)
  slapos-cookbook
# Call creation of instance.cfg file that will be called for deployment of
# instance
  template

# Download instance.cfg.in (buildout profile used to deployment of instance),
# replace all ${foo:bar} parameters by real values, and change $${foo:bar} to
# ${foo:bar}
[template]
recipe = slapos.recipe.template
url = ${:_profile_base_location_}/instance.cfg.in
output = ${buildout:directory}/instance.cfg
# MD5 checksum can be skipped for development (easier to develop), but must be filled for production
md5sum = 1fc461c00e86485bee77a942f39e3c43
mode = 0644
```

Save

SLAPOS

MMC Rus

Aide et Action
L'Education change le monde

sanef

AIRBUS DEFENCE & SPACE

WENDELIN

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

Wendelin Exanalytics Core 100% open source

100% Python

Scikit Learn — Data Analytics

NEO — Distributed Storage

MariaDB

ERP5 — Elastic PaaS

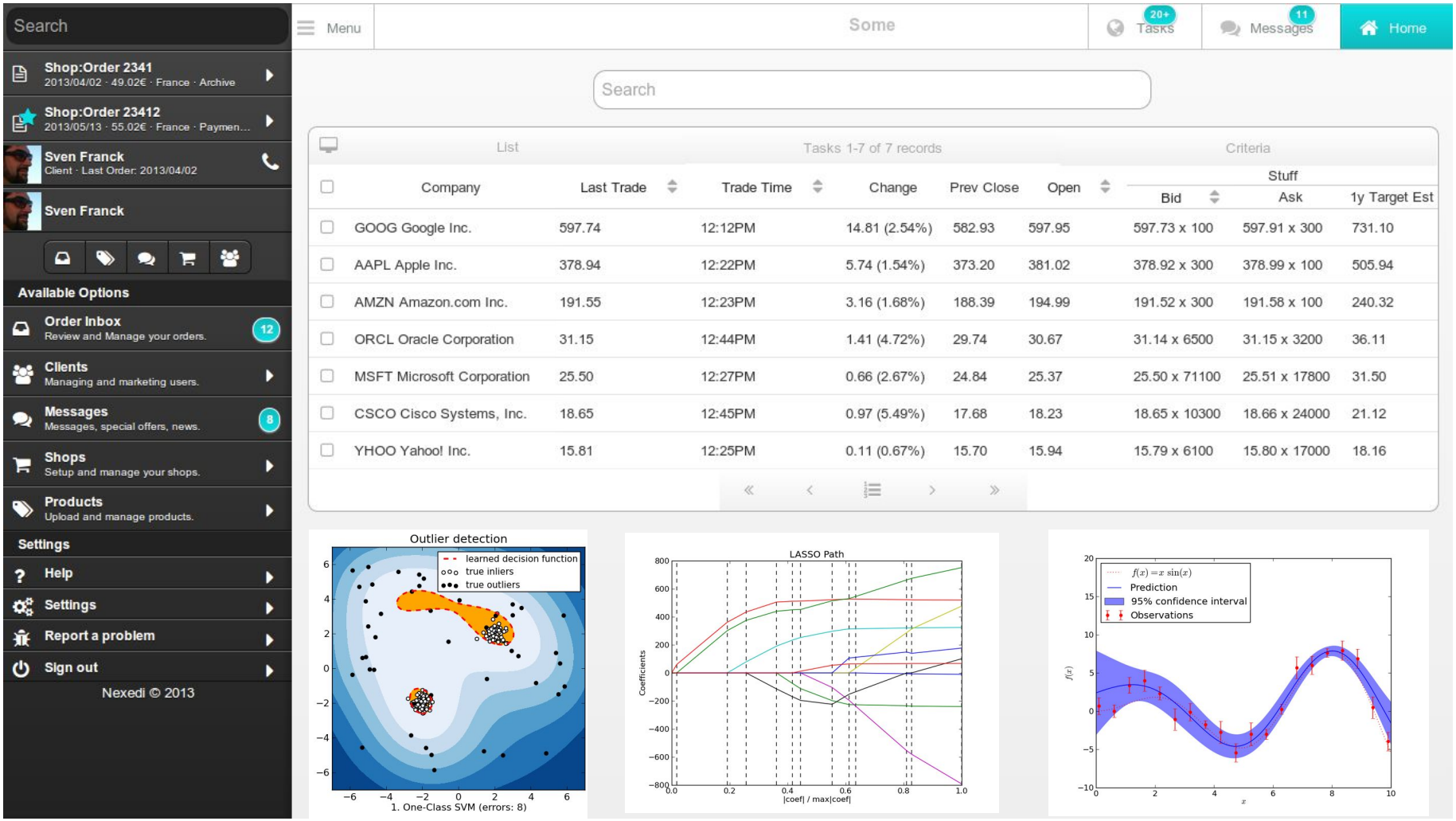
SlapOS — Multicloud Deployment

Multi Data Center

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

WENDELIN

Wendelin User Interface renderjs.org



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

WENDELIN

Wendelin Options 100% open source

100% Python

OpenCV-Python	**Video Processing** Intel Russia / Willow / Itseez
NLTK	**Natural Language Tookit** 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

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

WENDELIN

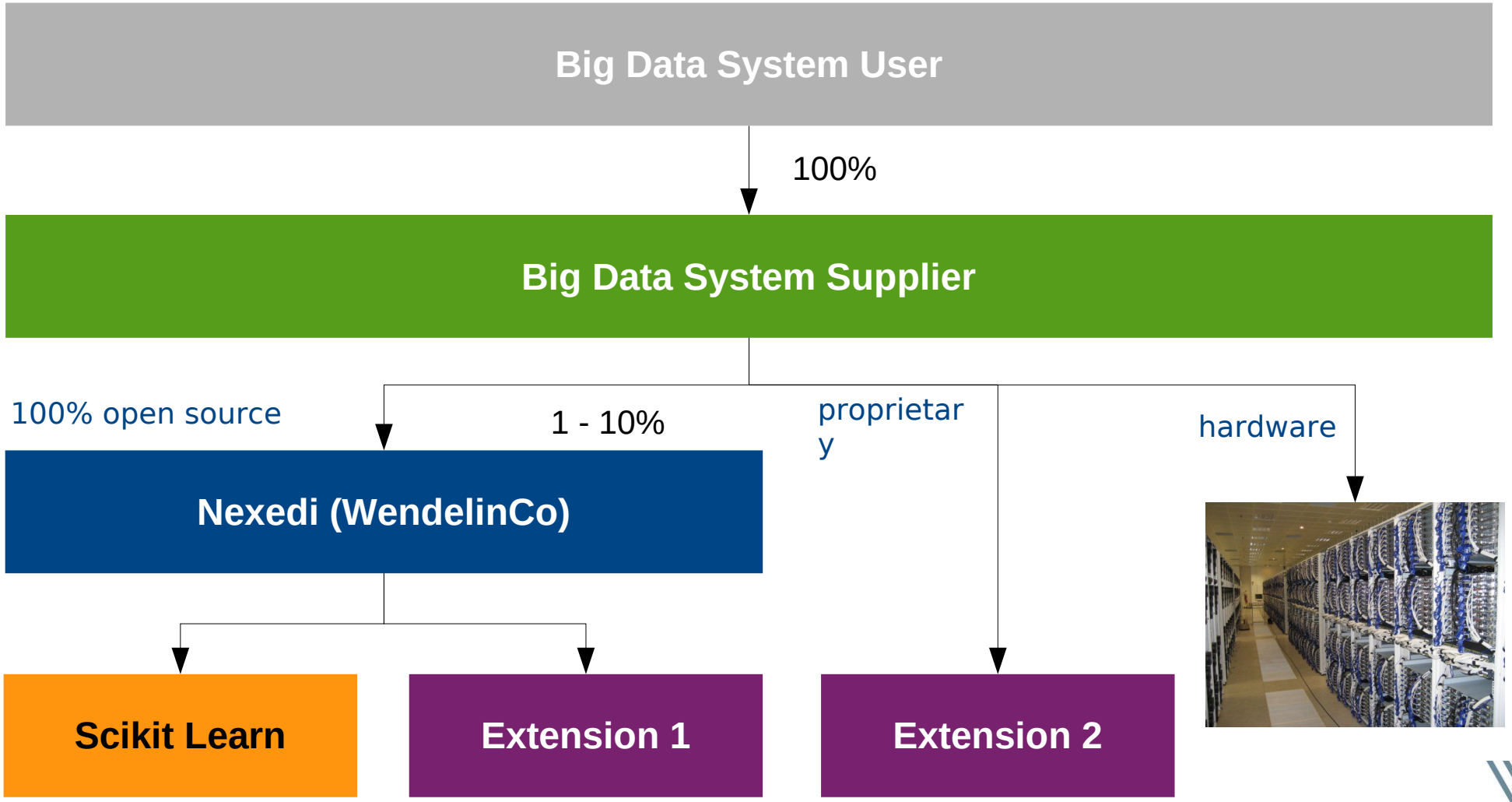
Wendelin Applications

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

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

WENDELIN

Business Model: German Style No VC



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

WENDELIN

Agenda

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

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

WENDELIN

Out-of-core arrays

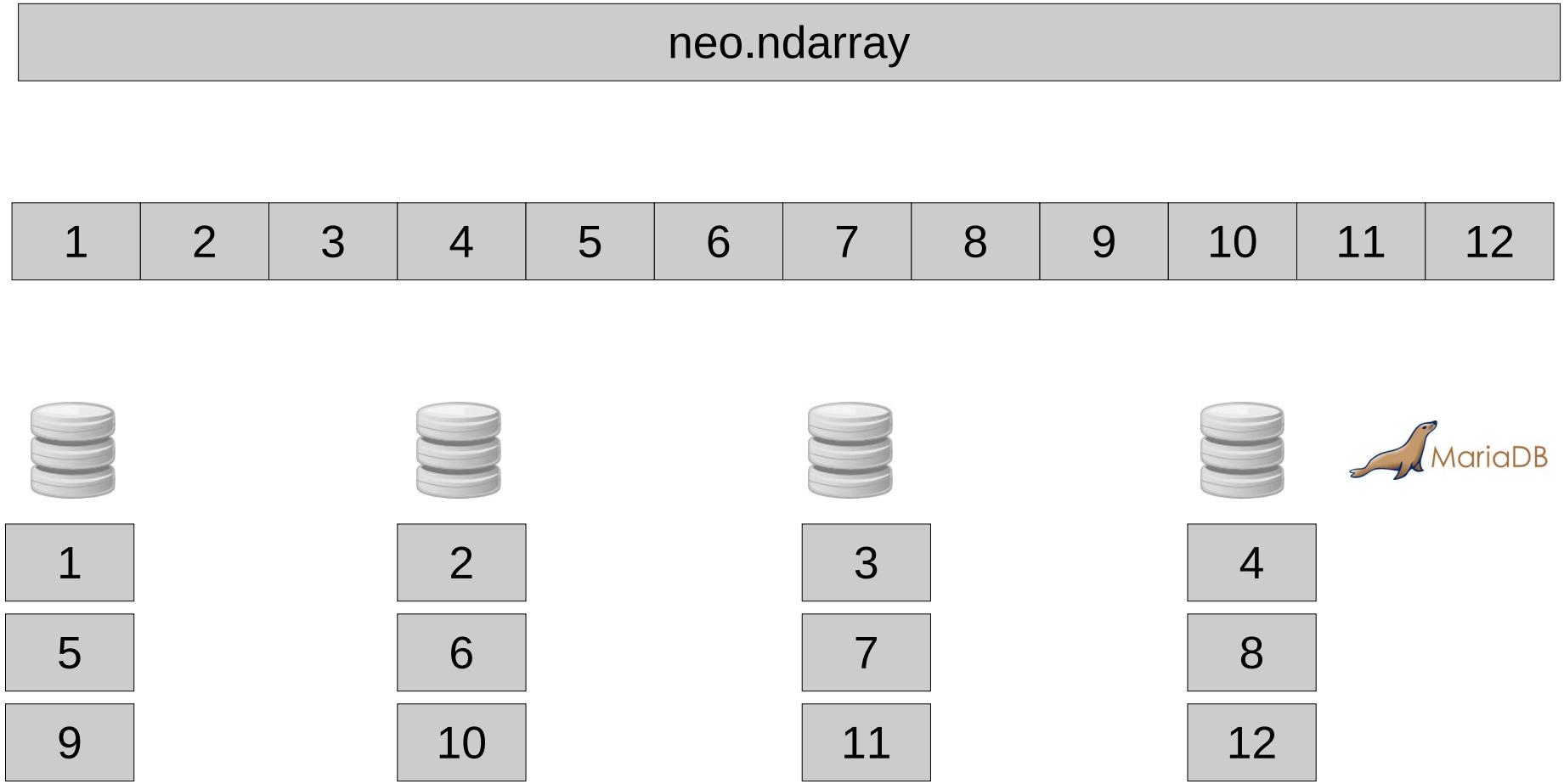
```
# Numpy
np.ndarray(shape=(2,2), dtype=float, order='F')

# 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

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

WENDELIN

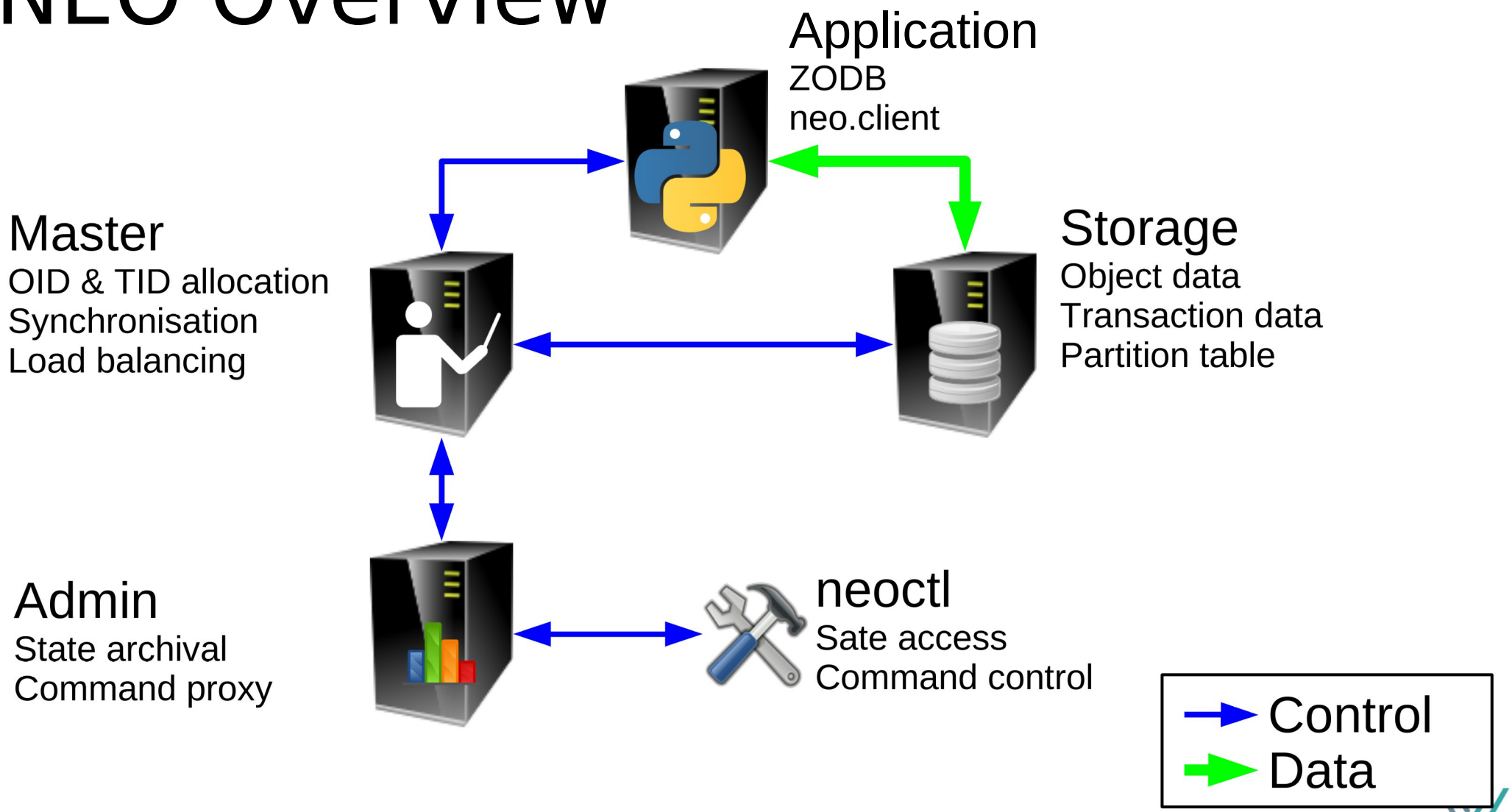
neo.ndarray out-of-core data



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

WENDELIN

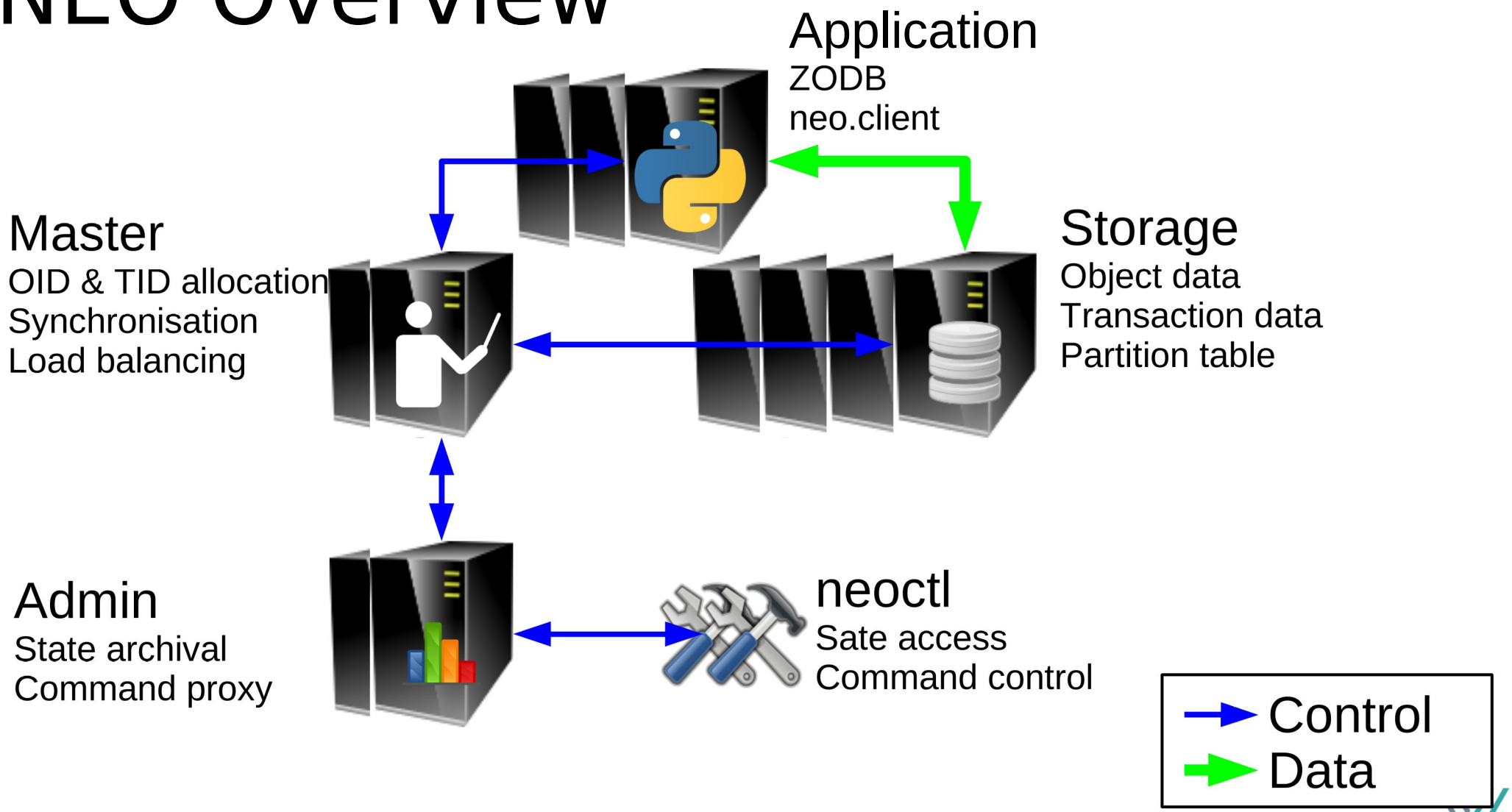
NEO Overview



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

WENDELIN

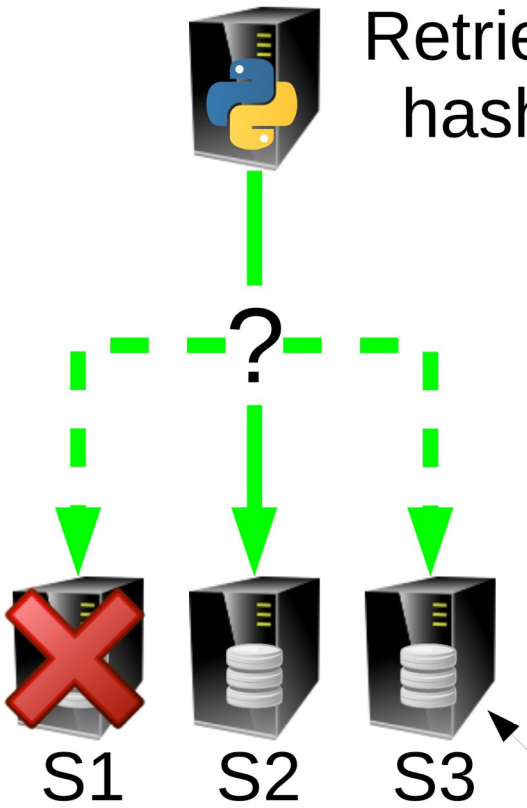
NEO Overview



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

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

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

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**

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

WENDELIN

Challenges

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

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

WENDELIN

Wendelin Exanalytics

2020 Big Data with MariaDB

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

www.wendelin.io

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

Wendelin Exanalytics

2020 Big Data with MariaDB

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

www.wendelin.io

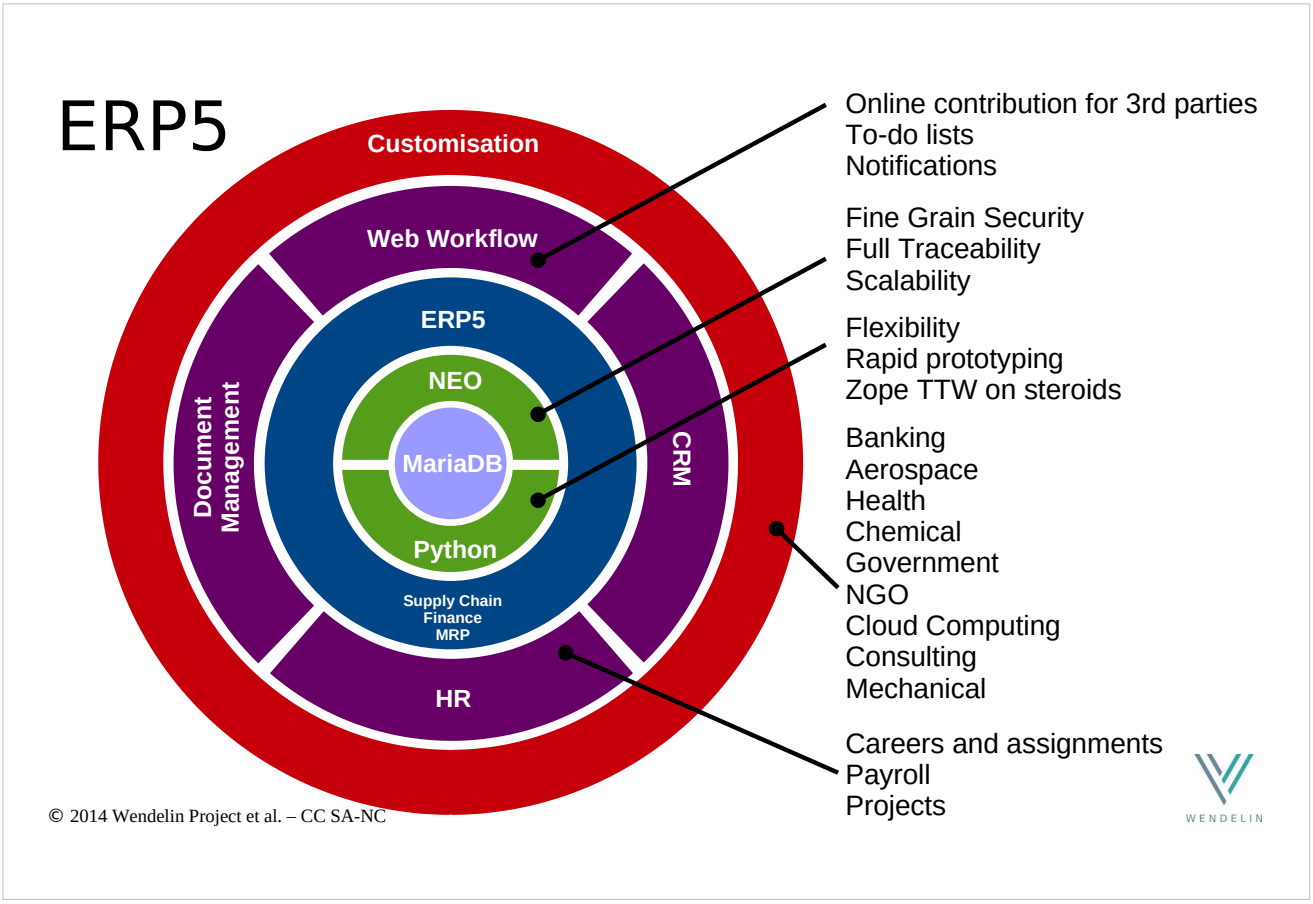
© 2014 Wendelin Project et al. – CC SA-NC

Agenda

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

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

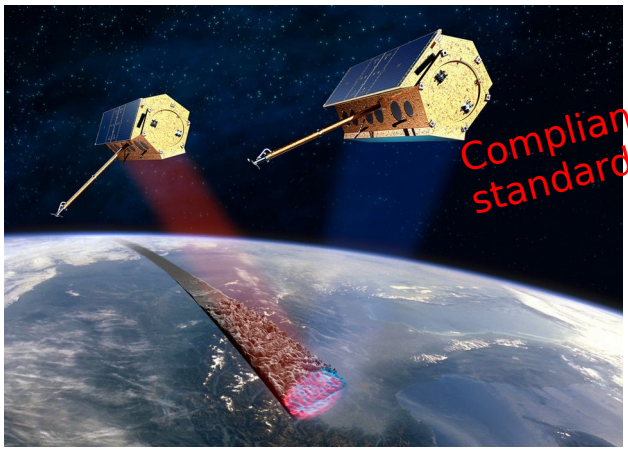
WENDELIN


ERP5
Customisation
Web Workflow
ERP5
NEO
MariaDB
Python
Supply Chain
Finance
MRP
Document Management
CRM
HR
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
© 2014 Wendelin Project et al. – CC SA-NC
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

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

WENDELIN

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

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

MariaDB ERP5

Open Source ERP/CRM
for S&P 100

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

WENDELIN

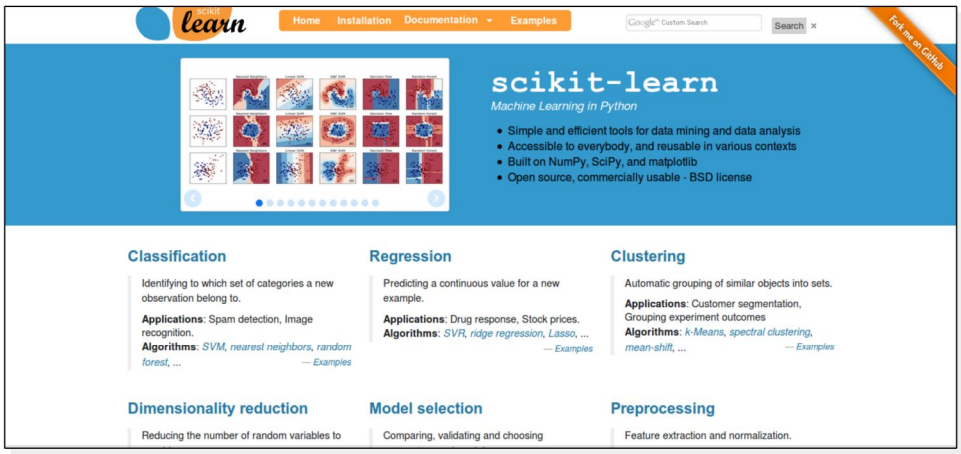
Agenda

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

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

WENDELIN

Take the Best Analytics scikit-learn.org



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

Made by Great Mathematicians

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

Number of Fields Medallists by country [edit]

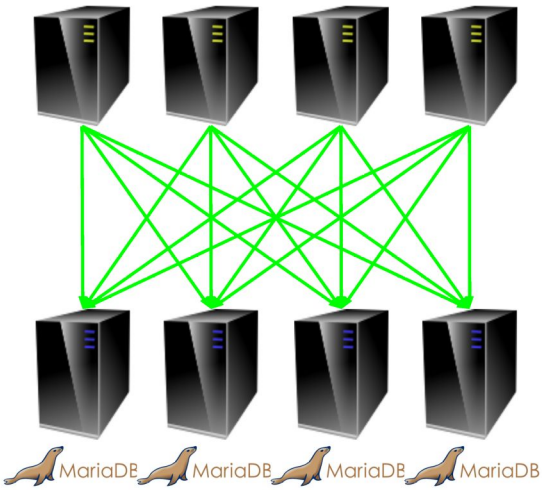
United States	11
France	10
Soviet Union (3) / Russia (6)	9
United Kingdom	6
Japan	3
Belgium	2
West Germany (1) / Germany (0)	1
Australia	1
British Hong Kong	1
Finland	1
Israel	1
Italy	1
Norway	1
New Zealand	1
Sweden	1
Vietnam	1

Wendelin Werner

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

Add Distributed Storage neoppod.org

NEO



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

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

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

WENDELIN

And Multicloud Deployment slapos.org

MMC Rus

Aide et Action
L'Education change le monde

sanef

AIRBUS DEFENCE & SPACE

WENDELIN

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

Wendelin Exanalytics Core 100% open source

100% Python

Scikit Learn — Data Analytics

NEO — Distributed Storage

MariaDB

ERP5 — Elastic PaaS

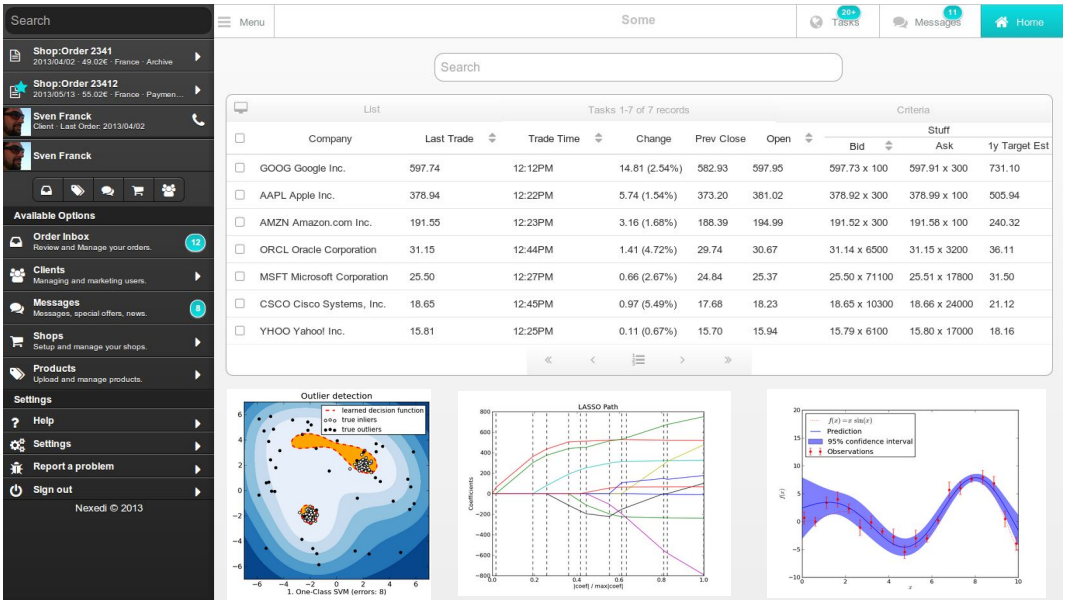
SlapOS — Multicloud Deployment

Multi Data Center

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

WENDELIN

Wendelin User Interface renderjs.org



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

WENDELIN

Wendelin Options 100% open source

100% Python

OpenCV-Python	**Video Processing** Intel Russia / Willow / Itseez
NLTK	**Natural Language Tookit** 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

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

WENDELIN

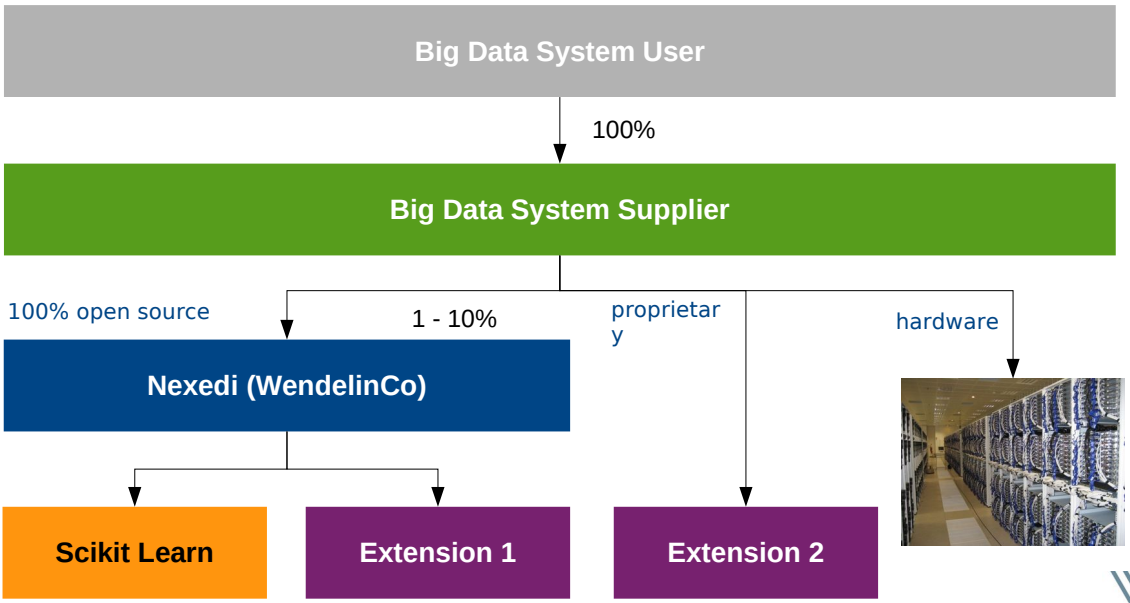
Wendelin Applications

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

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

WENDELIN

Business Model: German Style No VC



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

WENDELIN

Agenda

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

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

WENDELIN

Out-of-core arrays

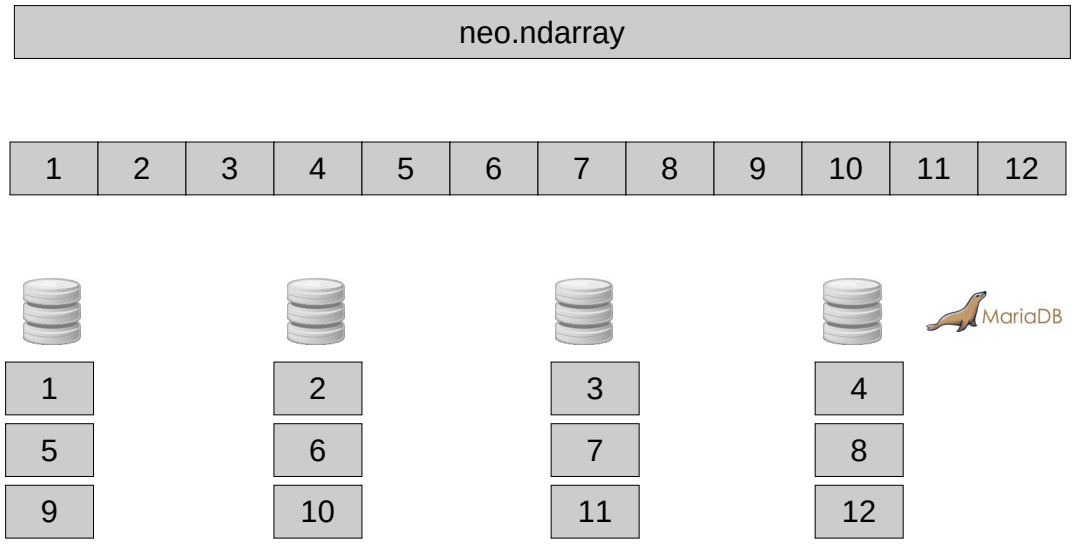
```
# Numpy
np.ndarray(shape=(2,2), dtype=float, order='F')

# 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

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

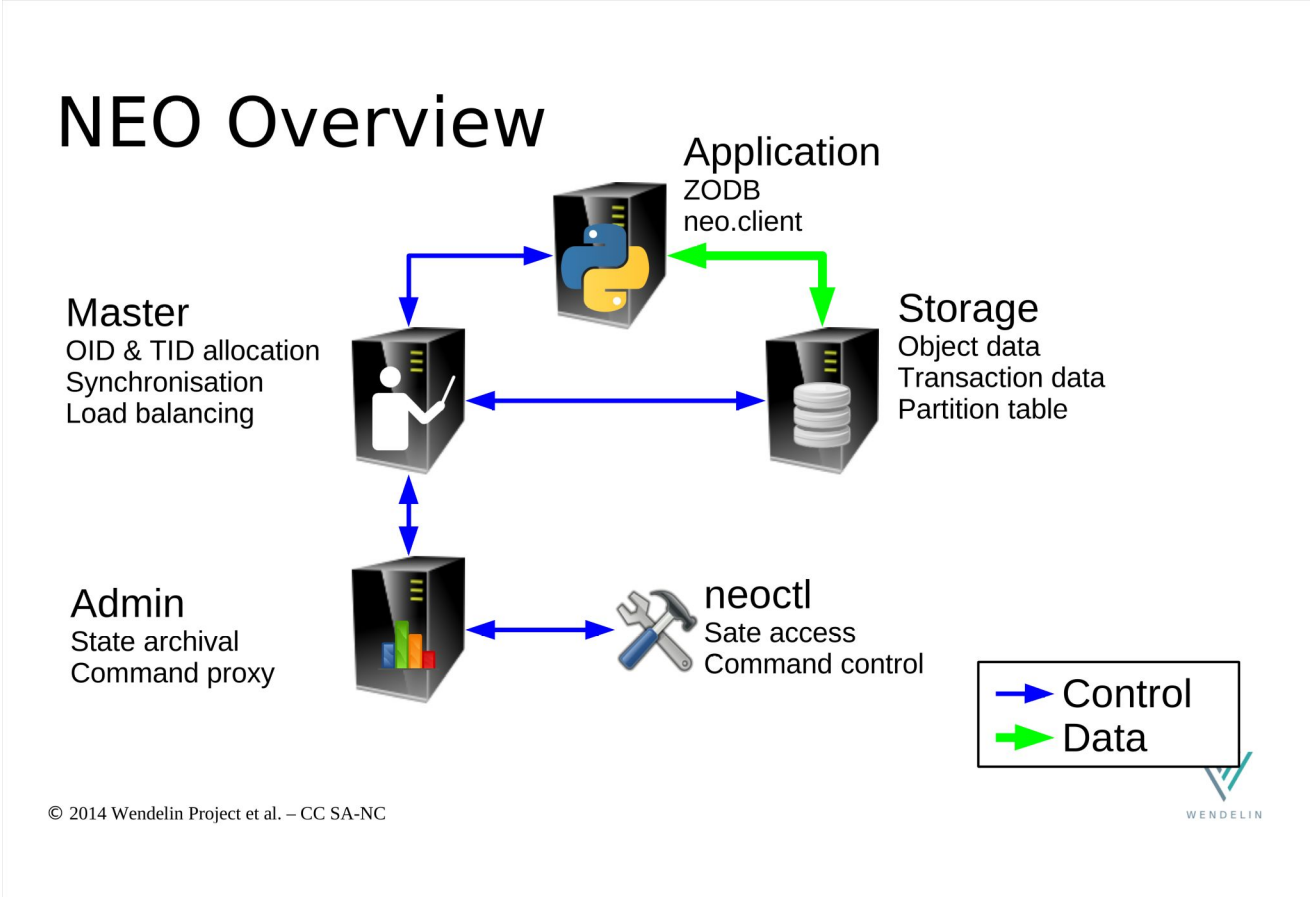
WENDELIN

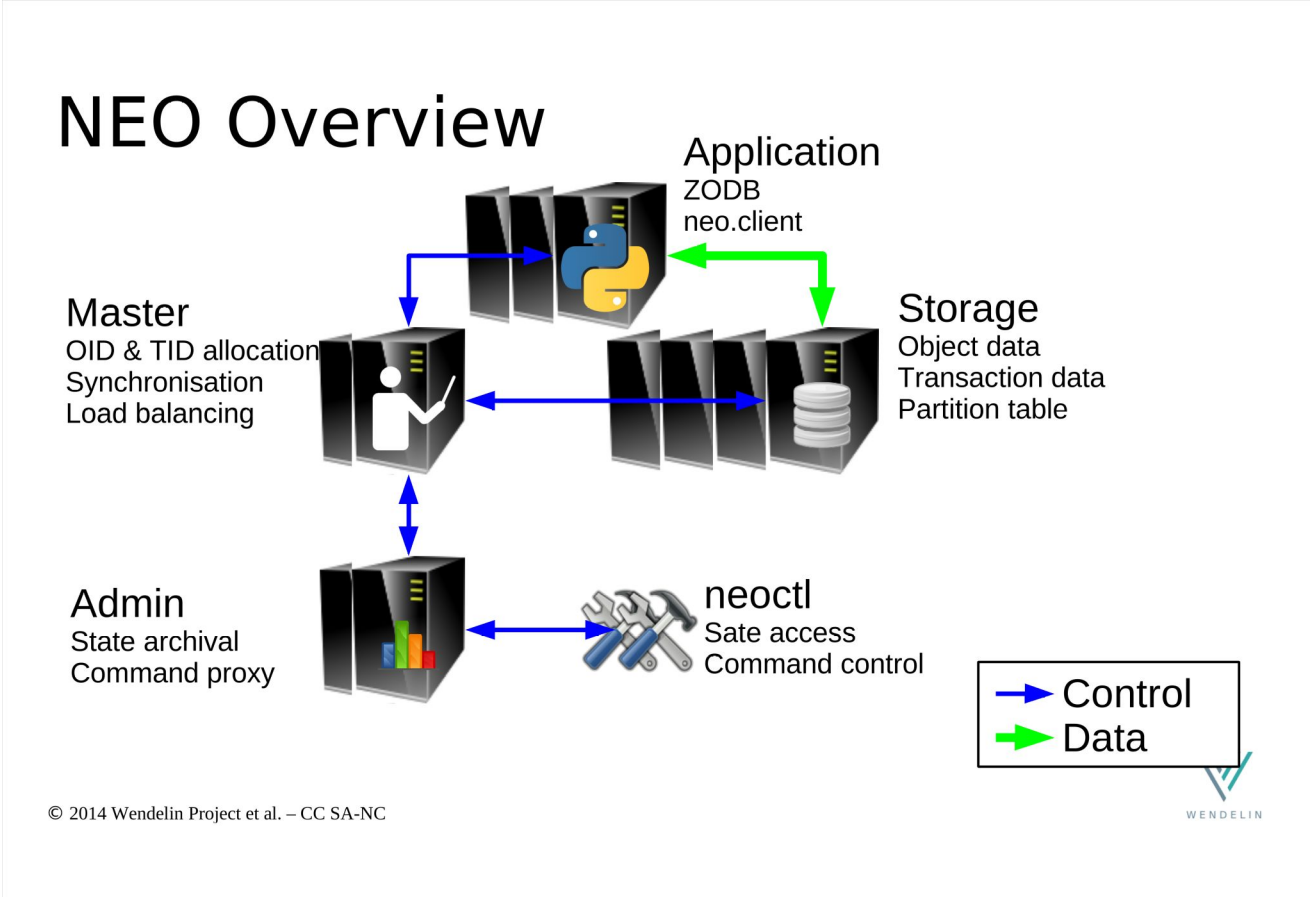
neo.ndarray out-of-core data



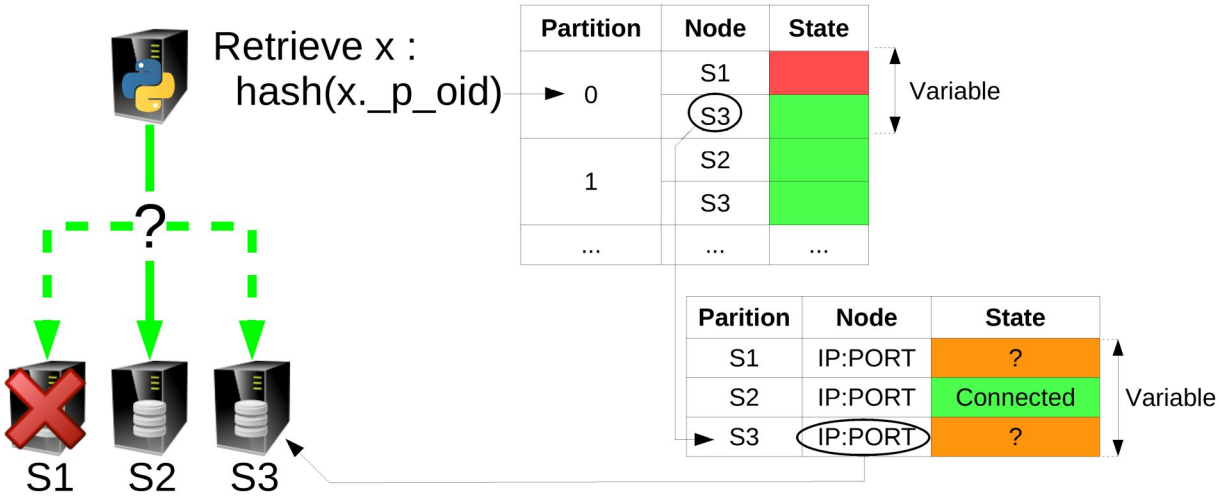
© 2014 Wendelin Project et al. – CC SA-NC

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

Object retrieval



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

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**

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

WENDELIN

Challenges

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

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

WENDELIN

WENDELIN

Wendelin Exanalytics

2020 Big Data with MariaDB

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

MariaDB ERP5

www.wendelin.io

WENDELIN

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