Agenda

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

Customisation

Web Workflow

Document Management

CRM

Python

Supply Chain
Finance
MRP

ERP5

NEO

MariaDB

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
Terra-SAR X Satellite

Management of sales and production of images

« 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

“Web has become our primary sales channel.” Frédéric Charlier

120,000 new customers / year
51,000 invoice/hour
7,000,000 contacts / year
250 users
Open Source ERP/CRM for S&P 100
Agenda

- Our background: ERP5
- Our future: Wendelin Exanalytics
- Our challenges with MariaDB
Take the Best Analytics  
sokit-learn.org
# Made by Great Mathematicians

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

## Number of Fields Medallists by country

<table>
<thead>
<tr>
<th>Country</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>United States</td>
<td>11</td>
</tr>
<tr>
<td>France</td>
<td>10</td>
</tr>
<tr>
<td>Soviet Union (3) / Russia (6)</td>
<td>9</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>6</td>
</tr>
<tr>
<td>Japan</td>
<td>3</td>
</tr>
<tr>
<td>Belgium</td>
<td>2</td>
</tr>
<tr>
<td>West Germany (1) / Germany (0)</td>
<td>1</td>
</tr>
<tr>
<td>Australia</td>
<td>1</td>
</tr>
<tr>
<td>British Hong Kong</td>
<td>1</td>
</tr>
<tr>
<td>Finland</td>
<td>1</td>
</tr>
<tr>
<td>Israel</td>
<td>1</td>
</tr>
<tr>
<td>Italy</td>
<td>1</td>
</tr>
<tr>
<td>Norway</td>
<td>1</td>
</tr>
<tr>
<td>New Zealand</td>
<td>1</td>
</tr>
<tr>
<td>Sweden</td>
<td>1</td>
</tr>
<tr>
<td>Vietnam</td>
<td>1</td>
</tr>
</tbody>
</table>

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

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

© 2014 Wendelin Project et al. – CC SA-NC
Wendelin Exanalytics Core 100% open source

100% Python

- Scikit Learn
- NEO
- ERP5
- SlapOS

Data Analytics
- Distributed Storage
- Elastic PaaS
- Multicloud Deployment
- Multi Data Center

© 2014 Wendelin Project et al. – CC SA-NC
Wendelin User Interface renderjs.org
<table>
<thead>
<tr>
<th>100% Python</th>
<th>100% open source</th>
</tr>
</thead>
<tbody>
<tr>
<td>NLTK</td>
<td>Natural Language Tookit U. Texas / Chalmers</td>
</tr>
<tr>
<td>Blaze</td>
<td>Full out-of-core arrays Continuum / DARPA</td>
</tr>
<tr>
<td>Numba / Parakeet</td>
<td>JIT compiler / type inference Continuum / DARPA</td>
</tr>
<tr>
<td>Pandas</td>
<td>Time sequence processing DataPad / JP Morgan</td>
</tr>
<tr>
<td>Scikit Learn</td>
<td>Reatime log collection Treasure Data / Amazon</td>
</tr>
<tr>
<td>NEO</td>
<td>Fluentd</td>
</tr>
</tbody>
</table>

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

Big Data System User

100%

Big Data System Supplier

100% open source

1 - 10%

Nexedi (WendelinCo)

proprietary

hardware

Scikit Learn

Extension 1

Extension 2

© 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') 1 Exabyte

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

Best out-of-core topology depends on the algorithm and array geometry
neo.ndarray out-of-core data

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

Master
OID & TID allocation
Synchronisation
Load balancing

Admin
State archival
Command proxy

Application
ZODB
neo.client

Storage
Object data
Transaction data
Partition table

neoctl
Sate access
Command control

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

Master
- OID & TID allocation
- Synchronisation
- Load balancing

Admin
- State archival
- Command proxy

Application
- ZODB
- neo.client

Storage
- Object data
- Transaction data
- Partition table

Control

Data
Object retrieval

Retrieve $x : \text{hash}(x._p\_oid)$

<table>
<thead>
<tr>
<th>Partition</th>
<th>Node</th>
<th>State</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>S1</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>S2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>S3</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Partition</th>
<th>Node</th>
<th>State</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>IP:PORT</td>
<td>?</td>
</tr>
<tr>
<td>S2</td>
<td>IP:PORT</td>
<td>Connected</td>
</tr>
<tr>
<td>S3</td>
<td>IP:PORT</td>
<td>?</td>
</tr>
</tbody>
</table>
Rodmap

- Q2 2014: neo.ndarray
- Q3 2014: developer release of Wendelin
- Q4 2014: neo.ndarray with simple optimizations
- Q1 2014: 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