Wendelin Big Data

Industrial Monitoring Platform

2014-04-03 – Paris

Who are we?

- **Jean-Paul Smets**
- **Nexedi CEO**
- **Author of ERP5**
- **jp@nexedi.com**

- **Ivan Tyagov**
- **Senior Developer**
- **Wendelin project lead**
- **ivan@nexedi.con**

Who is missing?

- **Kirill Smelkov**
- **Senior Developer**
- **wendelin.core**

- **Sebastien Robin**
- **Project Director**
- **Author of POC**

Agenda

- **Where do we come from**
- **Wendelin Architecture**
- **Detailed Example**
- **Future Roadmap**

Where do we come from?

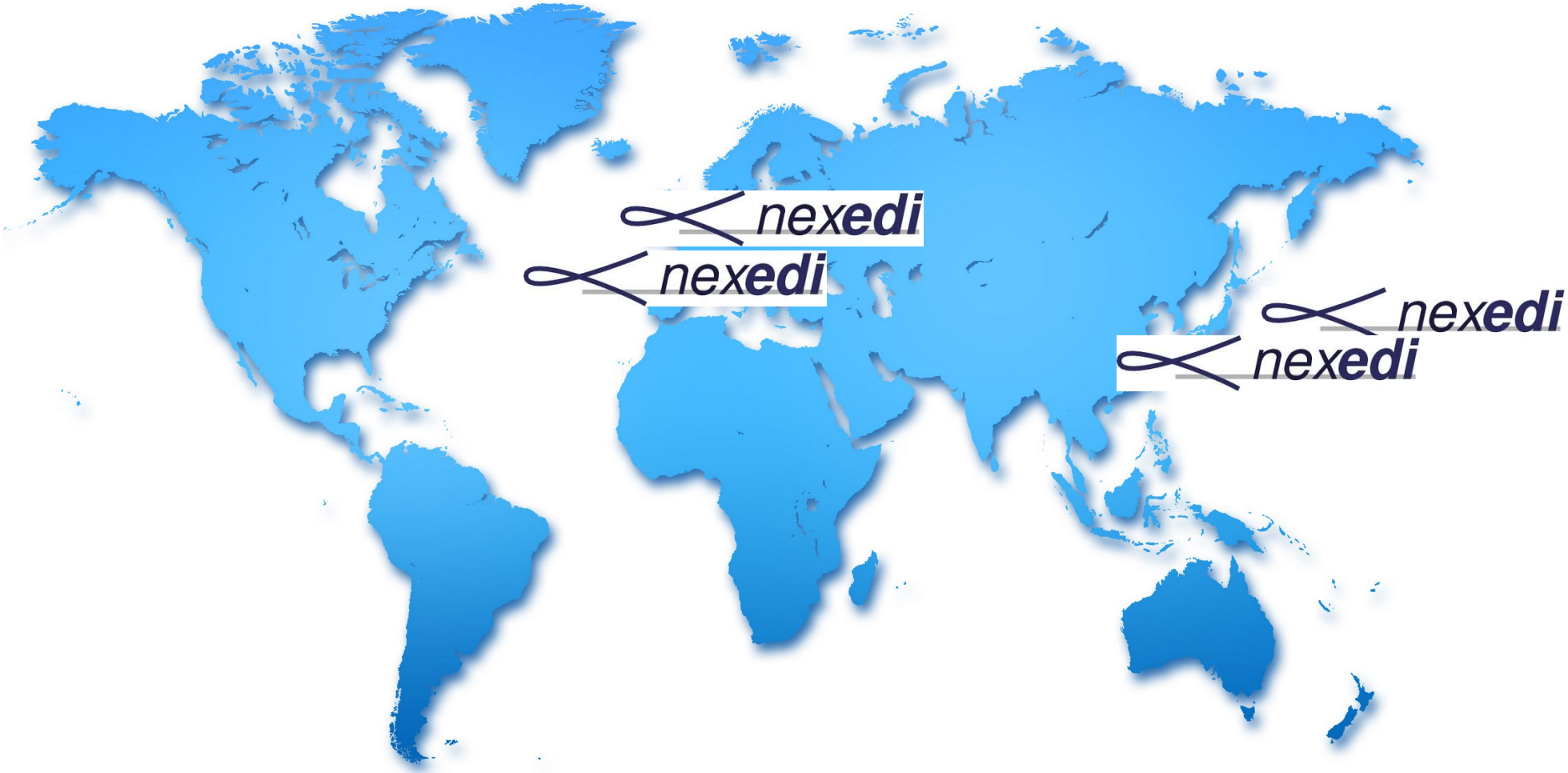
WENDELIN

nexedi

Nexedi

- **Possibly Largest OSS Publisher in Europe**
  - **ERP5: ERP, CRM, ECM, e-business framework**
  - **SlapOS: distributed mesh cloud operation system**
  - **NEO: distributed transactional NoSQL database**
  - **Wendelin: out-of-core big data based on NumPy**
  - **re6st: resilient IPv6 mesh overlay network**
  - **RenderJS: javascript component system**
  - **JIO: javascript virtual database and virtual filesystem**
  - **cloudooo: multimedia conversion server**
  - **Web Runner: web based Platform-as-a-Service (PaaS) and IDE**
  - **OfficeJS: web office suite based on RenderJS and JIO**

nexedi



nexedi

Aide et Action
L'Education change le monde

WEINPARIS.COM
BY ZAWEE

Liberté • Égalité • Fraternité
RÉPUBLIQUE FRANÇAISE
MINISTÈRE DE L'ÉCONOMIE ET DES FINANCES

BCEAO

HANGZHOU DIANZI UNIVERSITY

MITSUBISHI MOTORS

SOUTH AFRICA CAPAGO The Schengen link

AIRBUS DEFENCE & SPACE

Sanef

Kyorin

nexeai

Application Convergence

ERP5 + scikit learn

?

nexedi

Case 1: Wind Turbines

- **Collect logs**
- **Collect records** ERP5
- **Predict failure** scikit learn
- **Plan maintenance** ERP5
- **Reduce downtime**

  → **add X% profits**

nexedi

Case 2: Cars

- **Collect logs**
- **Collect records** ERP5
- **Predict failure** scikit learn
- **Plan maintenance** ERP5
- **Reduce downtime**

  → **increase loyalty**

nexedi

Case 3: Solar Energy

- **Collect logs**
- **Collect records** ERP5
- **Predict degradation** scikit learn
- **Plan maintenance** ERP5
- **Increase efficiency**

  → **add X% to profits**

nexedi

Wendelin Architecture

Standard Hardware no router / no SAN

inspur 浪潮 lenovo 联想 CORETO

x 160

- 2 x 10 Gbps
- 2 x 6 core Xeon CPU
- 512 GB RAM
- 4 x 1 TB SSD
- 1 x M2090 GPU

+

x 32

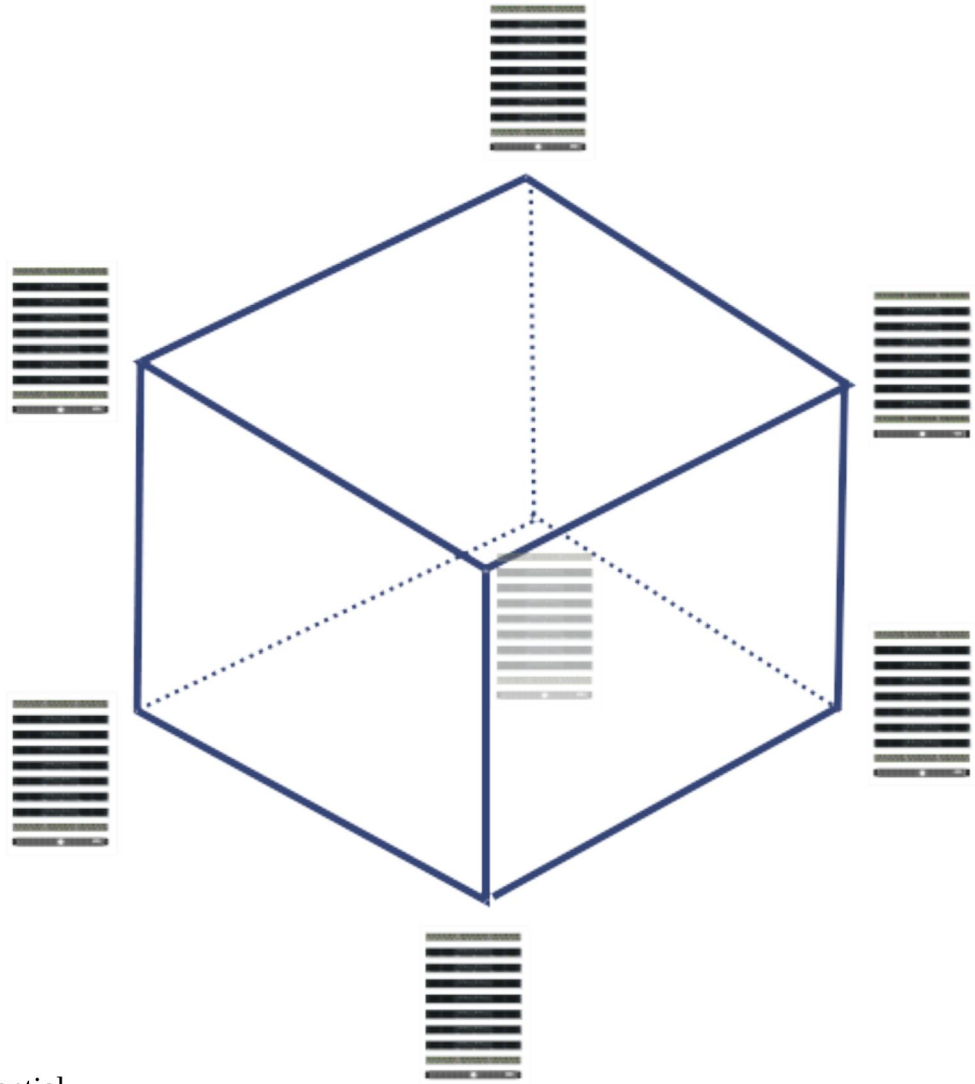
- 10 Gbps
- Unmanaged

+

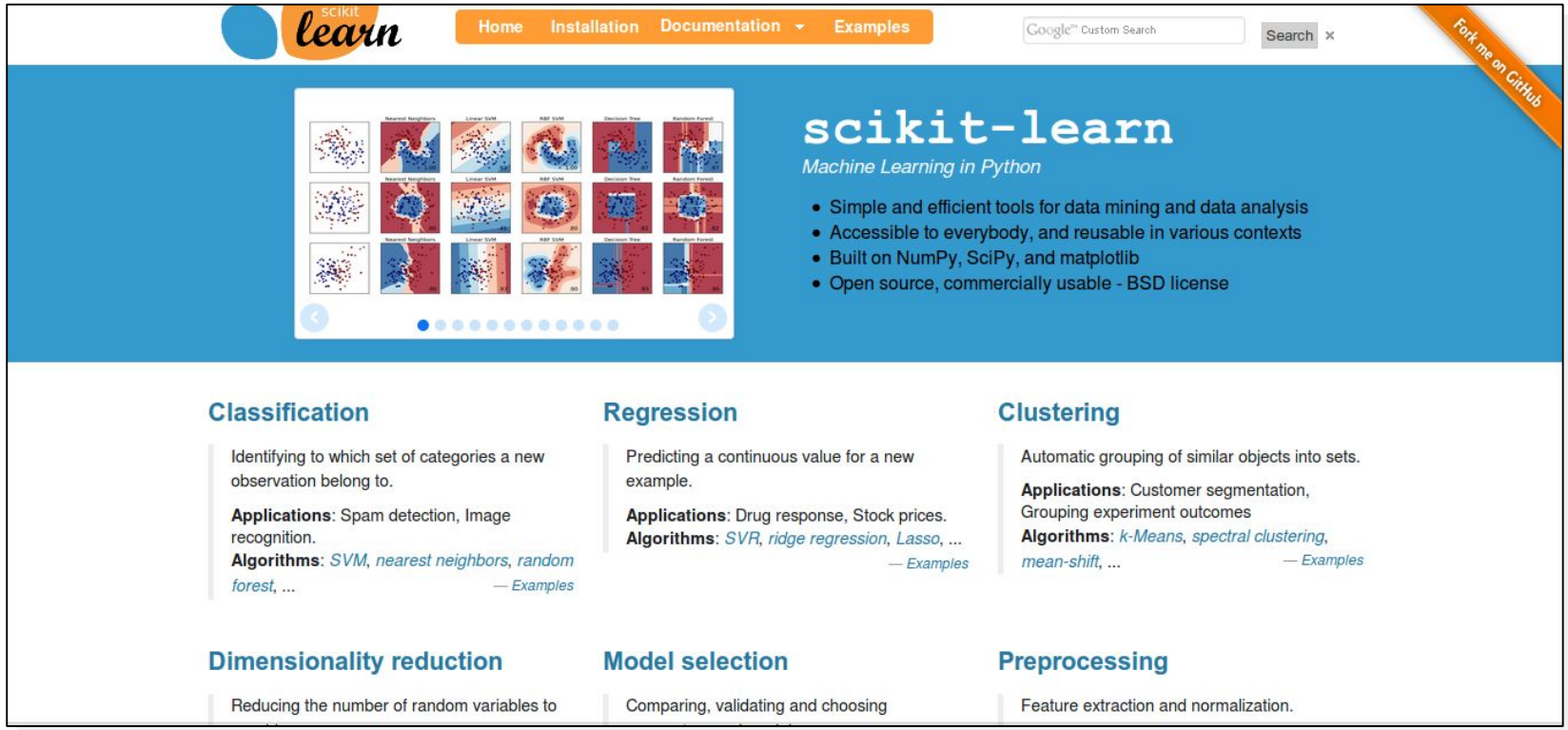
x 320

nexedi

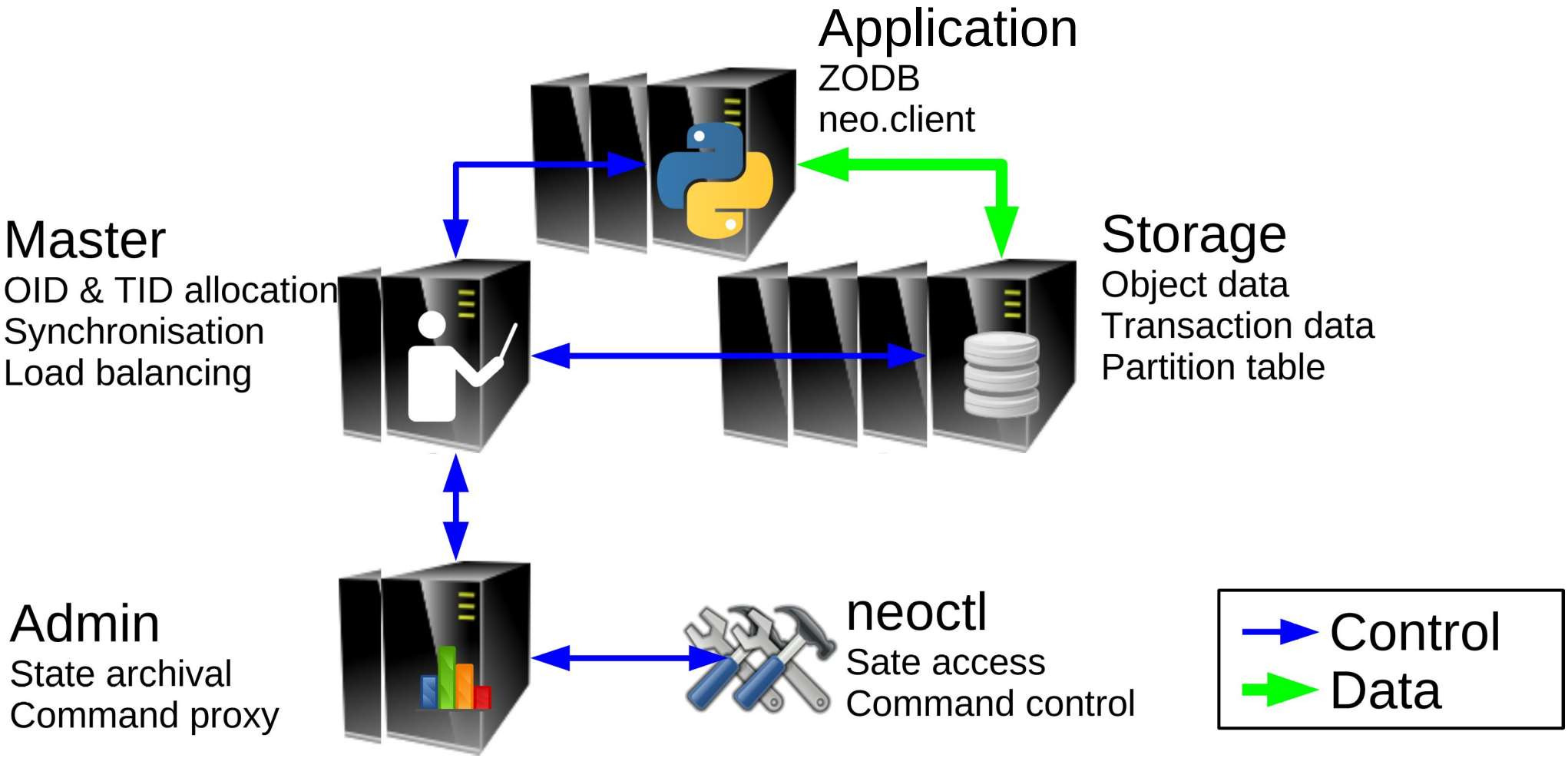
Wendelin Hypercube Datacenter



Take the Best Analytics scikit-learn.org



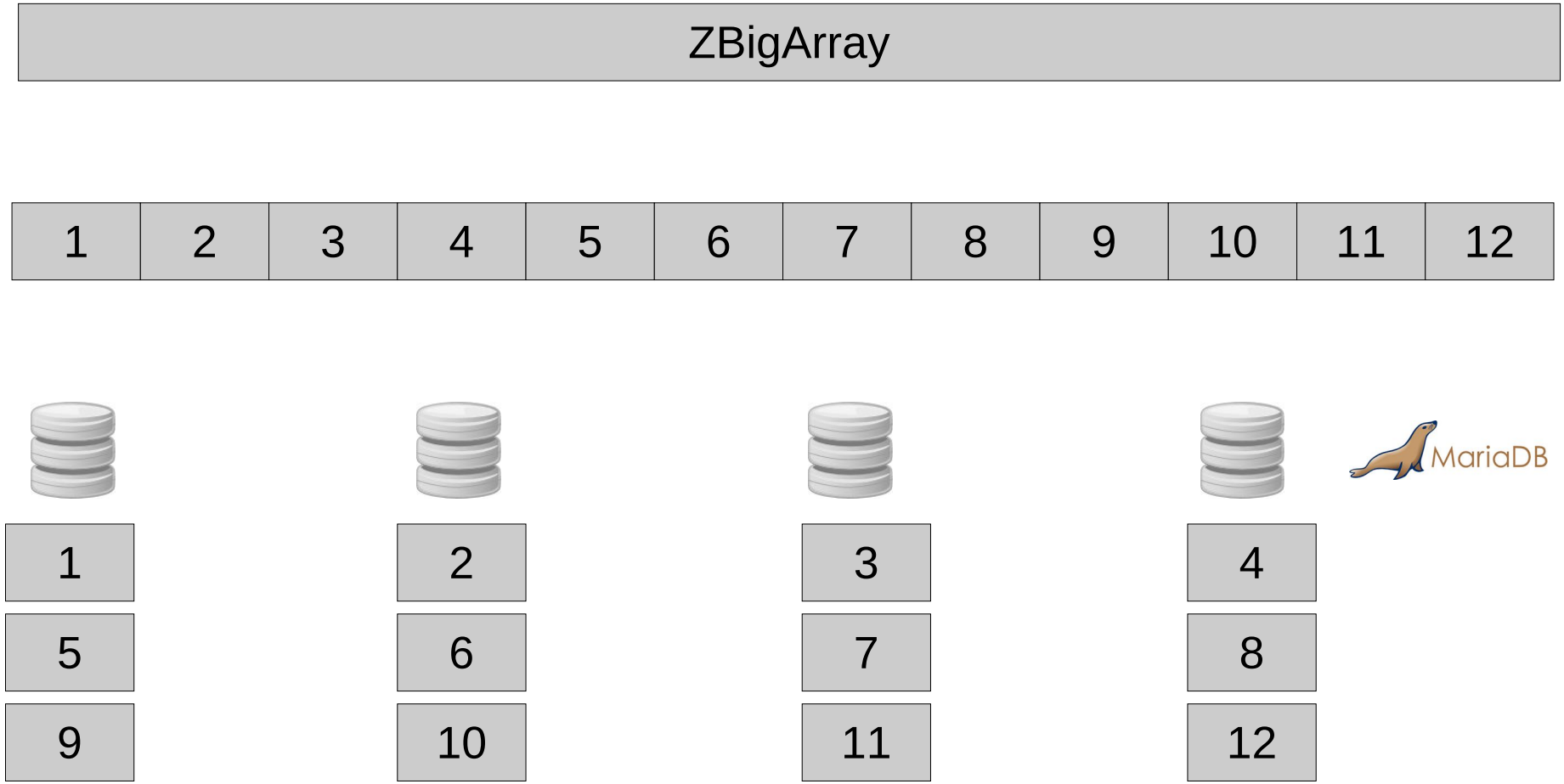
Add Distributed Storage neoppod.org



nexedi

“Magic” out-of-core for NumPy

PyData Paris 2015 – 16h45 Kirill Smelkov



nexedi

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

nexedi

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

nexedi

Wendelin Platform 100% open source

100% Python

Scikit Learn — Data Analytics

NEO — Distributed Storage

MariaDB

ERP5 — Elastic PaaS

SlapOS — Multicloud Deployment



Multi Data Center

nexedi

Wendelin Options 100% open source

100% Python

- OpenCV-Python
- NLTK
- Blaze
- Numba / Parakeet
- Pandas
- Scikit Learn
- NEO

Video Processing
Intel Russia / Willow / Itseez

Natural Language Tookit
U. Texas / Chalmers

Full out-of-core arrays
Continuum / DARPA

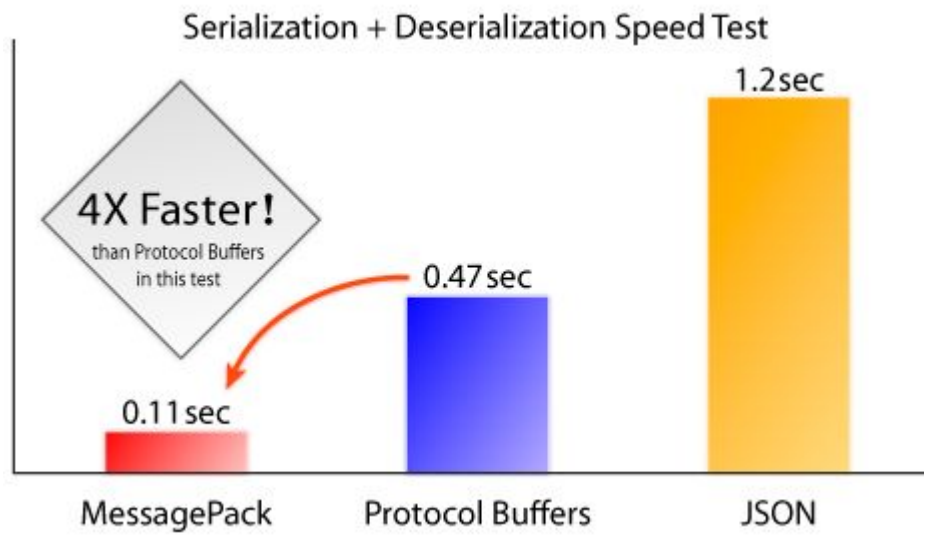
JIT compiler / type inference
Continuum / DARPA

Time sequence processing
DataPad / JP Morgan

nexedi

Data Ingestion: fluentd

- **Based on MsgPack middleware**
- **Created by TreasureData (BDaaS pioneers)**
- **Used by Amazon**
- **Numerous plugins**
- **Scalable and resilient**
- **Bandwidth saver**



Wendelin UI

Javascript

- **HTML5 Render** RenderJS
- **Data vizualisation**
- **Offline support** JIO

Python

- **Data access** REST API
- **Batch processing**

Wendelin Distinctive Advantages

- **Native out-of-core NumPy (scikit-learn, pydata)**
- **Native parallel processing**
- **Bare metal performance (GPU, FORTRAN)**
- **Transactions (ingestion, processing)**
- **NewSQL queries**
- **Built-in PaaS**
- **Lower deployment cost (10x less than...)**

nexedi

Detailed Example

WENDELIN

nexedi

Data Transportation **fluentd**

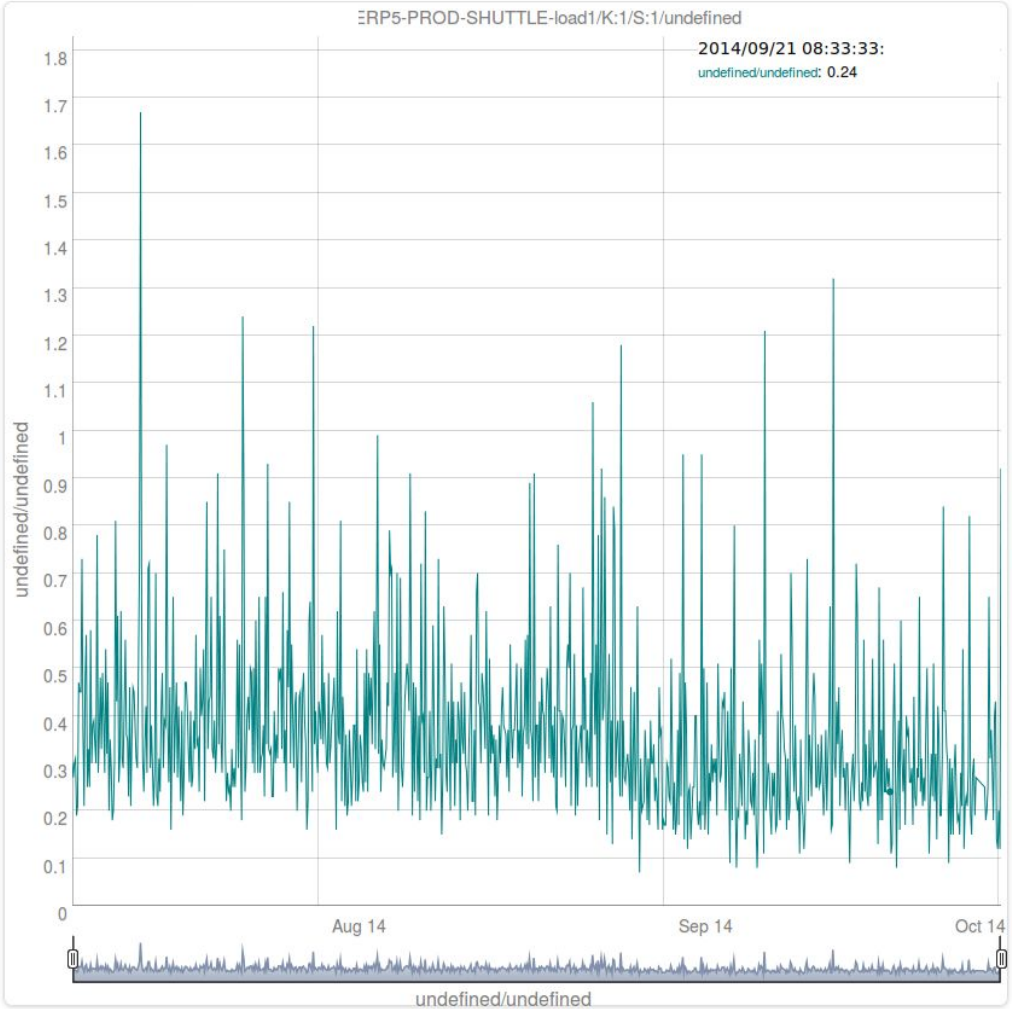
3 months benchmark
Frequent downtime (server, network)
Very poor networking (ADSL, 3G)

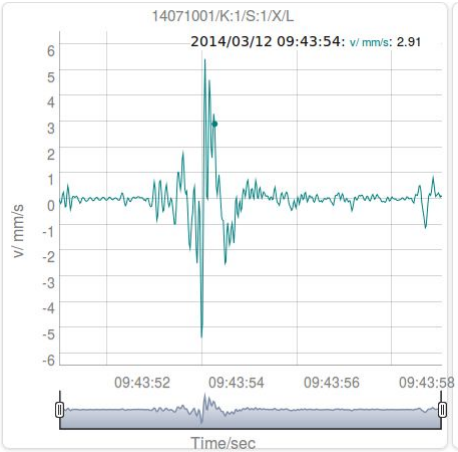
< 0.001% loss

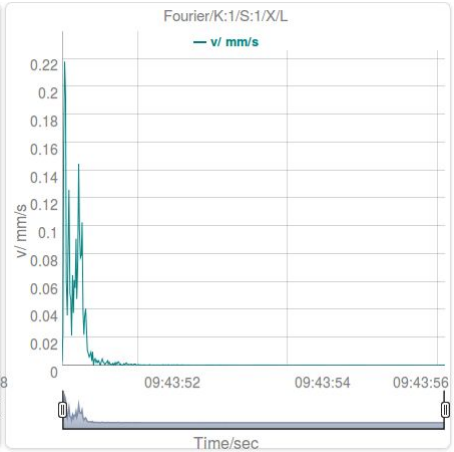
nexedi

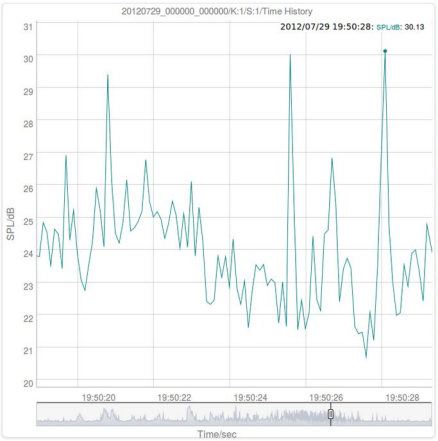
UI: HTML5 Components **RenderJS**

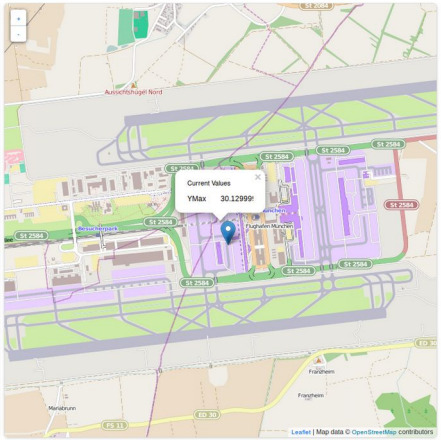
Extend UI Components **RenderJS**





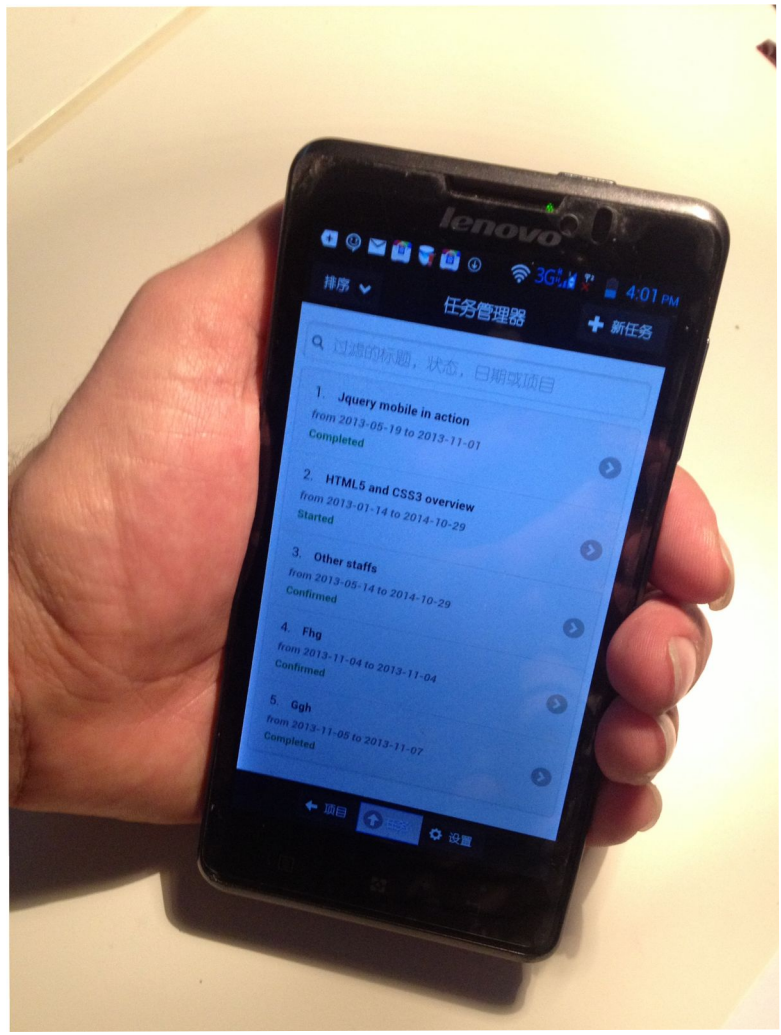




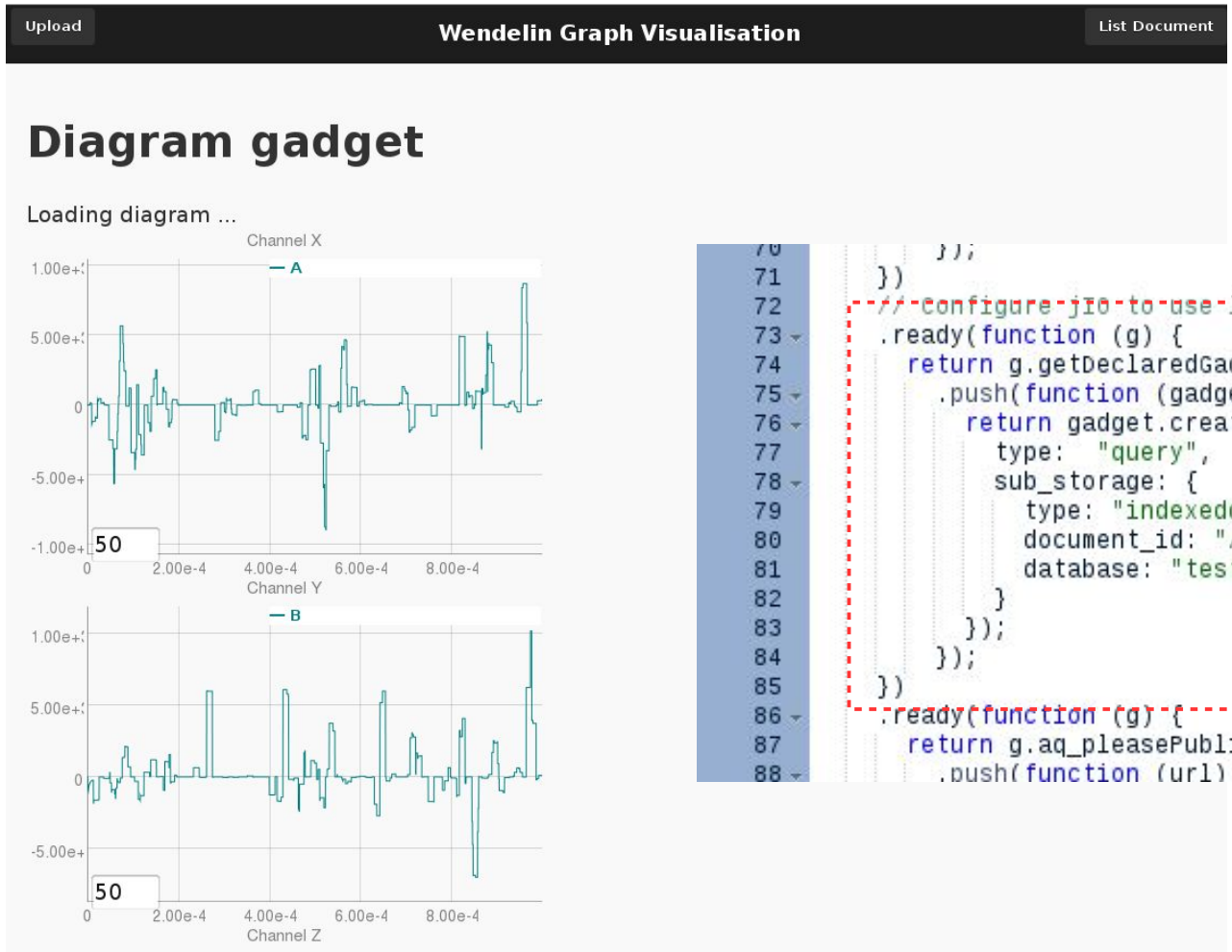


nexedi

UI : Responsive **RenderJS**



UI: Offline / Other Backends JIO




```
70        });
71    })
72    // configure jio to use localstorage
73    .ready(function (g) {
74      return g.getDeclaredGadget("JIO")
75        .push(function (gadget) {
76          return gadget.createJio({
77            type:  "query",
78            sub_storage: {
79              type: "indexeddb",
80              document_id: "/",
81              database: "test_ivan"
82            }
83          });
84        });
85    })
86    .ready(function (g) {
87      return g.aq_pleasePublishMyState({page: 'listbox'})
88        .push(function (url) {
```

Data Science in Javascript vs. Python ?

10.0.111.35:12002/erp5/web_page_module/woelfel_agorithm_js/WebPage_view?ignore_layout:int

```
/*global XMLHttpRequest, JSON, RSVP, window, UriTemplate, console,      lAlgorithm, webMonitoring, Blob, Math, Float32Array, FileReader, FFT
/*jslint nomen: true, indent: 2 */


/*
  ALGORITHM
  */

(function (window) {
  "use strict";
  var       lAlgorithm = {};

        lAlgorithm.doMax = function (array, params, output_buffer) {
    console.log("getMax reading chunk", array.length);
    console.log("getMax channel", params.channel);
    console.log("getMax algo status start pos", params.algorithm_status.start_position);
    console.log("getMax algo status stop pos", params.algorithm_status.stop_position);
    console.log("getMax, arrayBefore", array[0]);
    console.log("getMax, output_buffer", output_buffer);
    var i, max = -Infinity, array_length;
    array_length = array.length;
    for (i = 0; i < array_length; i = i + 1) {
      max = Math.max(max, array[i]);
    }
    for (i = 0; i < array_length; i = i + 1) {
      array[i] = max;
    }
    console.log("getMax, arrayAfter", array[0]);
    if (output_buffer.meta_data === undefined) {
      output_buffer.meta_data = params.meta_data;
    }
    output_buffer.write(array);

  };

        lAlgorithm.doEscalatedMax = function (array, params, output_buffer) {
    console.log("doEscalatedMax, array.length", array.length, array[0], array[1]);
    var i, max, previous_max_dict, array_length, algorithm_result;
    algorithm_result = params.algorithm_result || {};
    previous_max_dict = algorithm_result.previous_max_dict || {};
    algorithm_result.previous_max_dict = previous_max_dict;
    max = previous_max_dict[params.channel] || -Infinity;
    console.log("doEscalatedMax, previous_max", max, "for channel", params.channel);
    array_length = array.length;
    for (i = 0; i < array_length; i = i + 1) {
      max = Math.max(max, array[i]);
    }
```

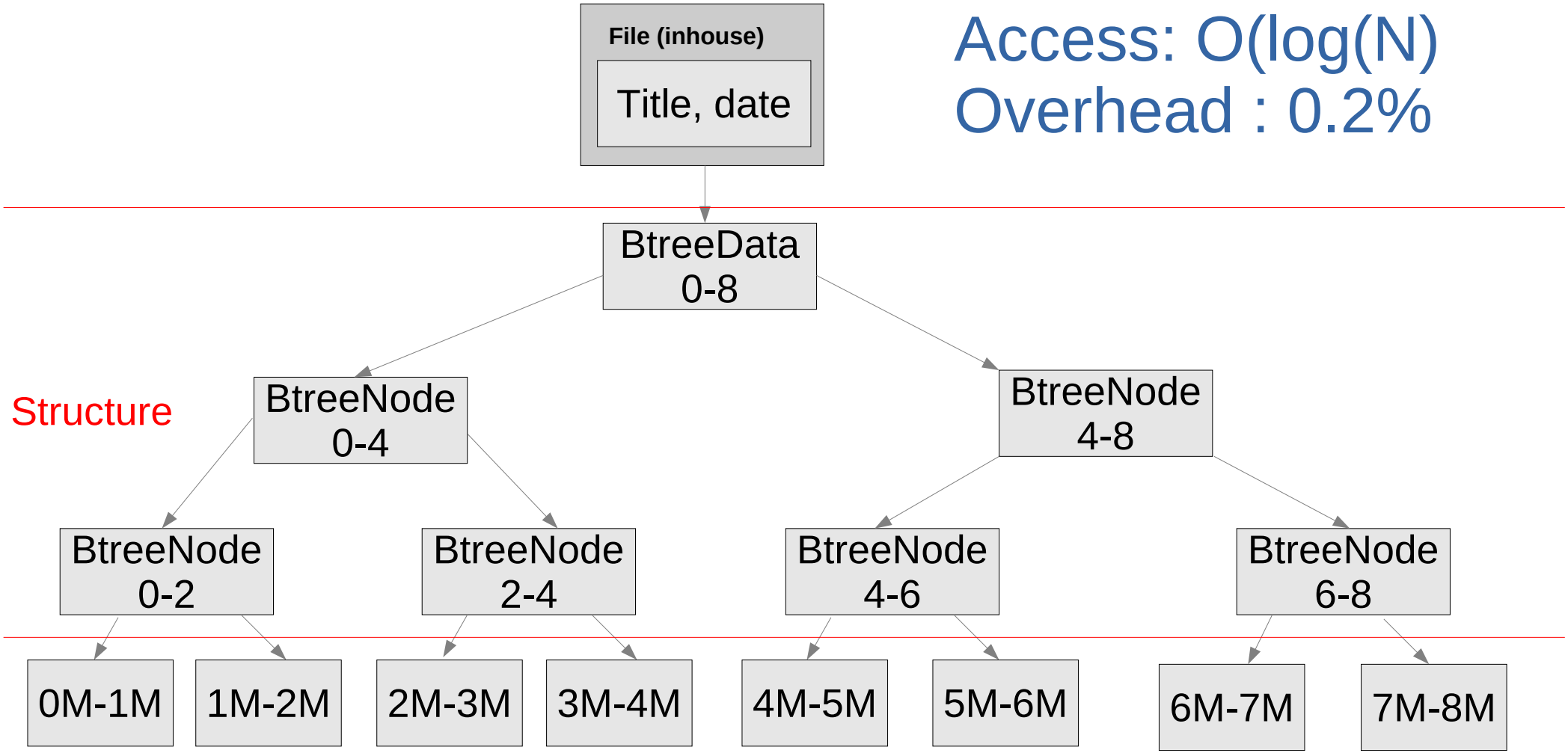
Data Sciences in Javascript ? phantomjs

- **Small data on client side** ☑
- **Small data on server side** ☑
- **Medium data (> 1 GB) in JS** ⛔
- **Out-of-core data in JS** ⛔
- **PyData compiled in JS** ⛔
- **PyData in NaCl / PNaCl** ☐

nexedi

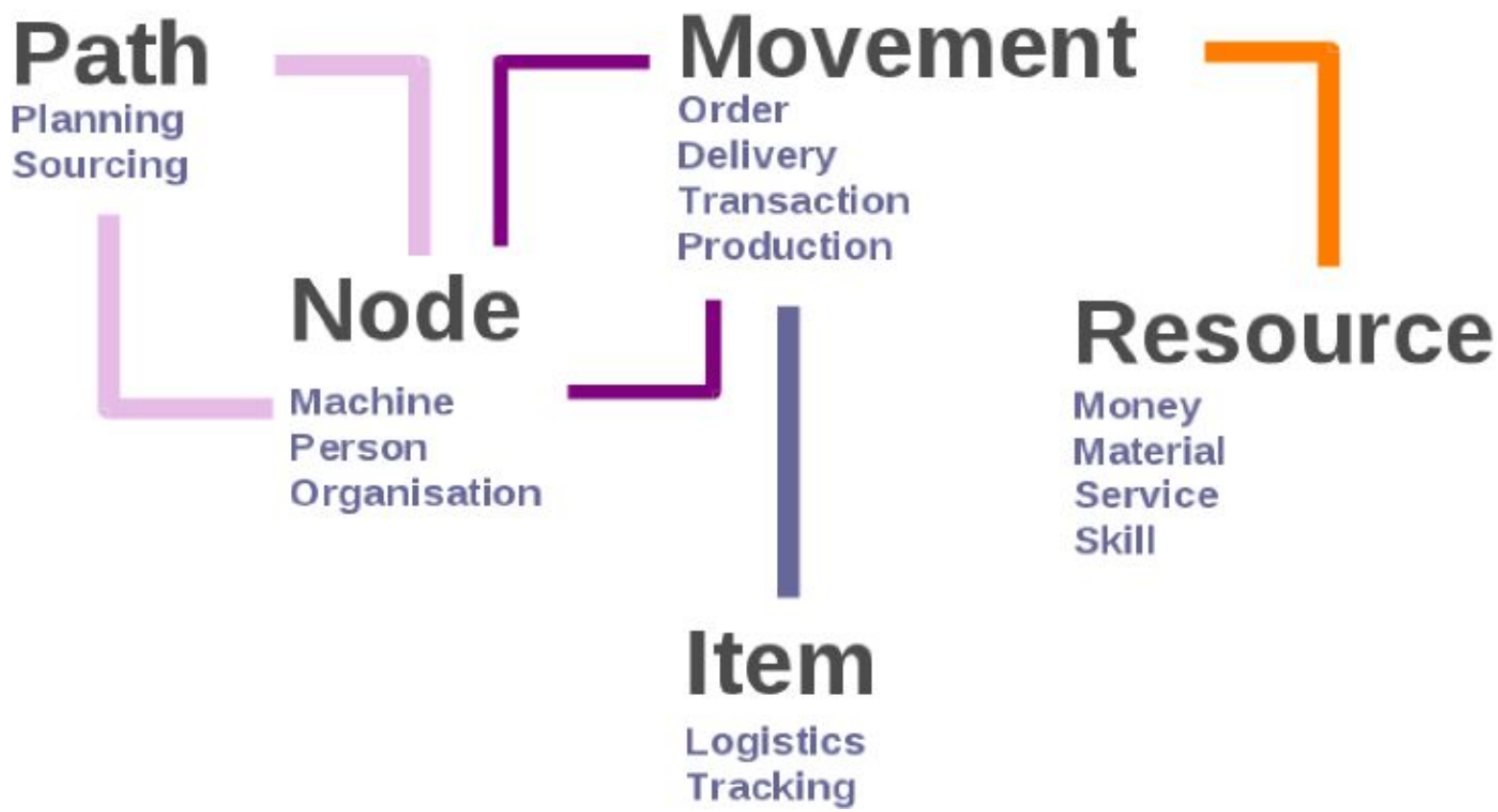
Storing large streams in NEO

Access: O(log(N)
Overhead : 0.2%



nexedi

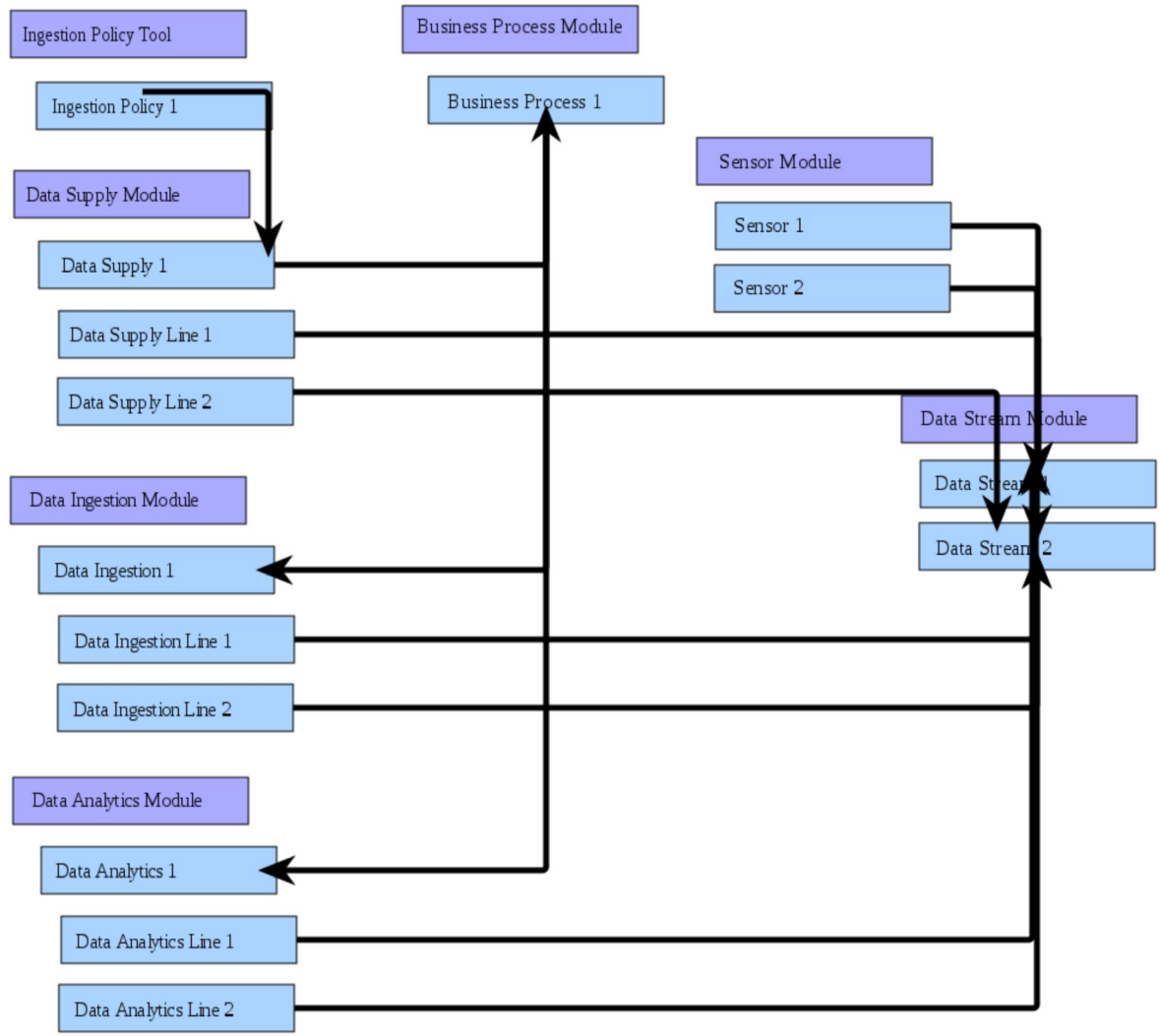
UBM Monitoring Model?



UBM Business Model

- **Movement – ingestion of data**
- **Resource – type of data (ex. memory log)**
- **Node – data source, data owner**
- **Path – data source registration**
- **Item – sensor, data itself, license, data set**

UBM Business Model



What UBM gets us for free

- **Accounting, billing and payment**
- **User registration and management**
- **Rule based security model**
- **Customer relationship management**
- **Web Content Management**

→ **save 12+ months and > 200 K€ on any Big Data project**

nexedi

Future Roadmap

WENDELIN

nexedi

Roadmap

www.wendelin.io

- **Mainly accelerate learning curve**
  - ☐ **Universal packaging**
  - ☐ **Ready to use examples**
  - ☐ **Act as a backend to ipython notebook**
  - ☐ **Port joblib to CMFActivity**
- **Yet, you can start using part of Wendelin now!**
  - ☑ **wendelin.core out-of-core for NumPy** PyData Paris 2015 – 16h45 Kirill Smelkov
  - ☑ **JIO abstract data access library**
  - ☑ **RenderJS components** http://learn.renderjs.org
  - ☑ **UI sample application** https://lab.nexedi.cn/Tyagov/wendelin/
  - ☑ **Open Source**

nexedi

R&D Partners

www.wendelin.io

- **Wendelin-IA (FSN)**
  - **Nexedi**
  - **Abilian**
  - **2nd Quadrant**
  - **Paris 13**
  - **IMT**
  - **INRIA / ENS**
  - **MMC Rus (Ru)**
  - **X Corp**

- **Windelin (Eurostars)**
  - **Nexedi (FR)**
  - **MariaDB (FI)**
  - **Y Corp (DE)**

Wendelin Big Data

Industrial Monitoring Platform

2014-04-03 – Paris

WENDELIN

Wendelin Big Data

Industrial Monitoring Platform

2014-04-03 – Paris

nexedi

Who are we?

- **Jean-Paul Smets**
- **Nexedi CEO**
- **Author of ERP5**
- **jp@nexedi.com**

- **Ivan Tyagov**
- **Senior Developer**
- **Wendelin project lead**
- **ivan@nexedi.con**

Who is missing?

- **Kirill Smelkov**
- **Senior Developer**
- **wendelin.core**

- **Sebastien Robin**
- **Project Director**
- **Author of POC**

Agenda

- **Where do we come from**
- **Wendelin Architecture**
- **Detailed Example**
- **Future Roadmap**

Where do we come from?

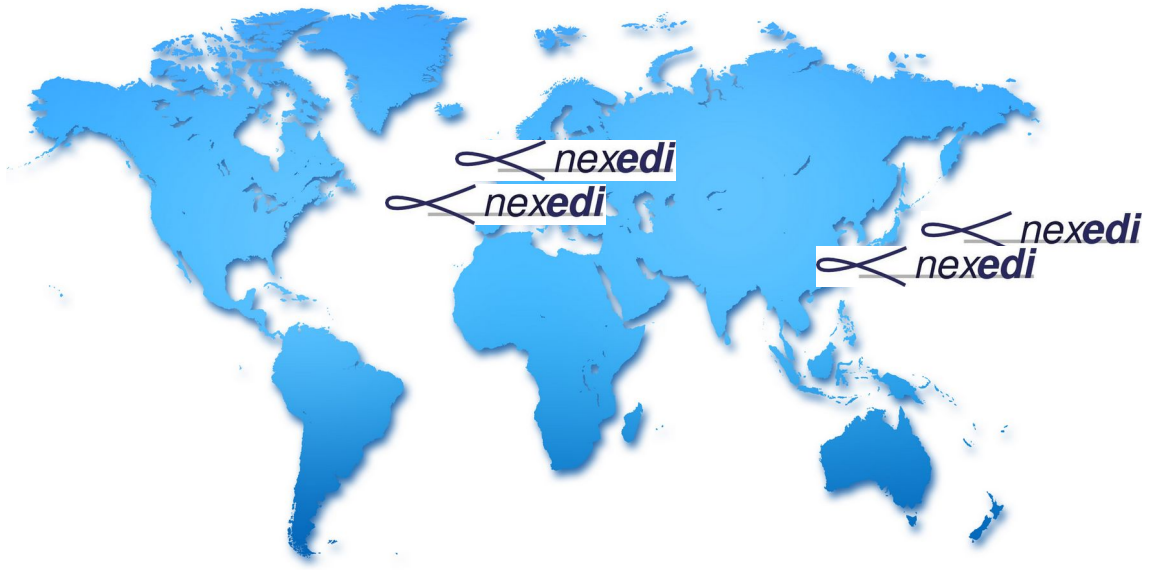
WENDELIN

nexedi

The solution that was deployed at the Lightning Protection Center complies is based on open source software – with full access to source code – and does not use software made by IBM, Oracle or EMC. It is thus a “No IOE” compliant solution, in line with directives published by Chinese governments for certain markets.

Nexedi

- **Possibly Largest OSS Publisher in Europe**
  - **ERP5: ERP, CRM, ECM, e-business framework**
  - **SlapOS: distributed mesh cloud operation system**
  - **NEO: distributed transactional NoSQL database**
  - **Wendelin: out-of-core big data based on NumPy**
  - **re6st: resilient IPv6 mesh overlay network**
  - **RenderJS: javascript component system**
  - **JIO: javascript virtual database and virtual filesystem**
  - **cloudooo: multimedia conversion server**
  - **Web Runner: web based Platform-as-a-Service (PaaS) and IDE**
  - **OfficeJS: web office suite based on RenderJS and JIO**



nexedi

Application Convergence

ERP5 + scikit learn

?

nexedi

Case 1: Wind Turbines

- **Collect logs**
- **Collect records** ERP5
- **Predict failure** scikit learn
- **Plan maintenance** ERP5
- **Reduce downtime**

  → **add X% profits**

nexedi

Case 2: Cars

- **Collect logs**
- **Collect records** ERP5
- **Predict failure** scikit learn
- **Plan maintenance** ERP5
- **Reduce downtime**

  → **increase loyalty**

nexedi

Case 3: Solar Energy

- **Collect logs**
- **Collect records** ERP5
- **Predict degradation** scikit learn
- **Plan maintenance** ERP5
- **Increase efficiency**

  → **add X% to profits**

nexedi

Wendelin Architecture

WENDELIN

nexedi

The solution that was deployed at the Lightning Protection Center complies is based on open source software – with full access to source code – and does not use software made by IBM, Oracle or EMC. It is thus a “No IOE” compliant solution, in line with directives published by Chinese governments for certain markets.

Standard Hardware no router / no SAN

inspur 浪潮 lenovo 联想 CORETO

x 160

- 2 x 10 Gbps
- 2 x 6 core Xeon CPU
- 512 GB RAM
- 4 x 1 TB SSD
- 1 x M2090 GPU

+

x 32

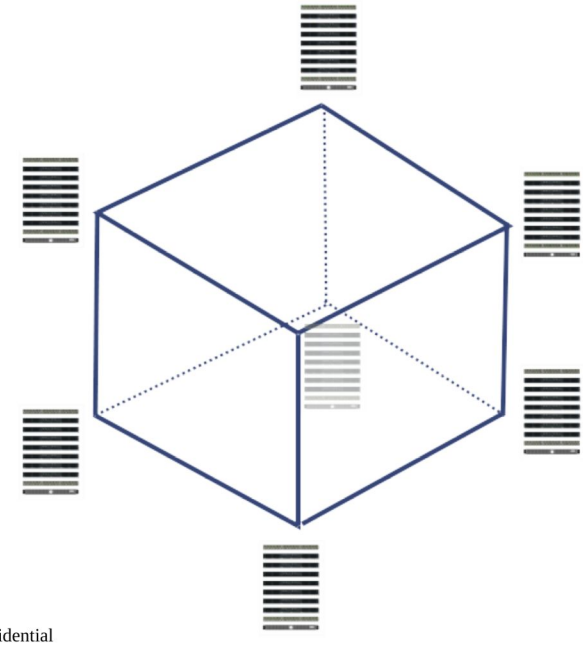
- 10 Gbps
- Unmanaged

+

x 320

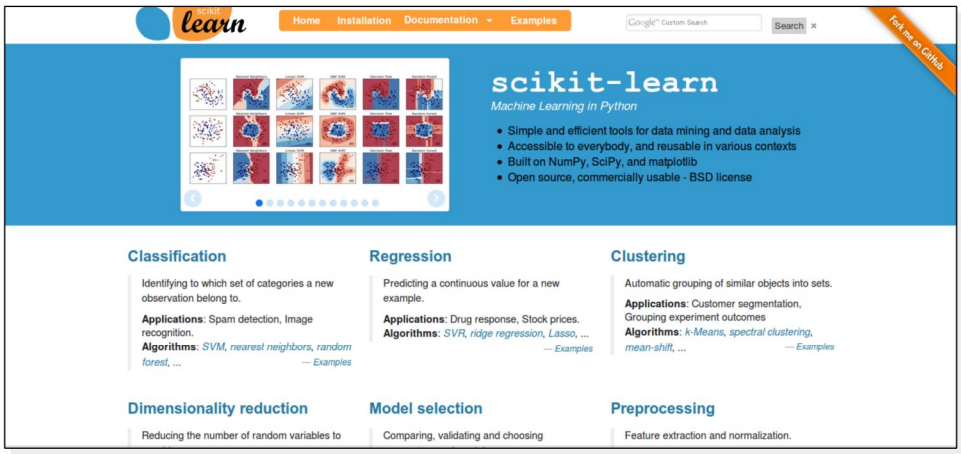
nexedi

Wendelin Hypercube Datacenter

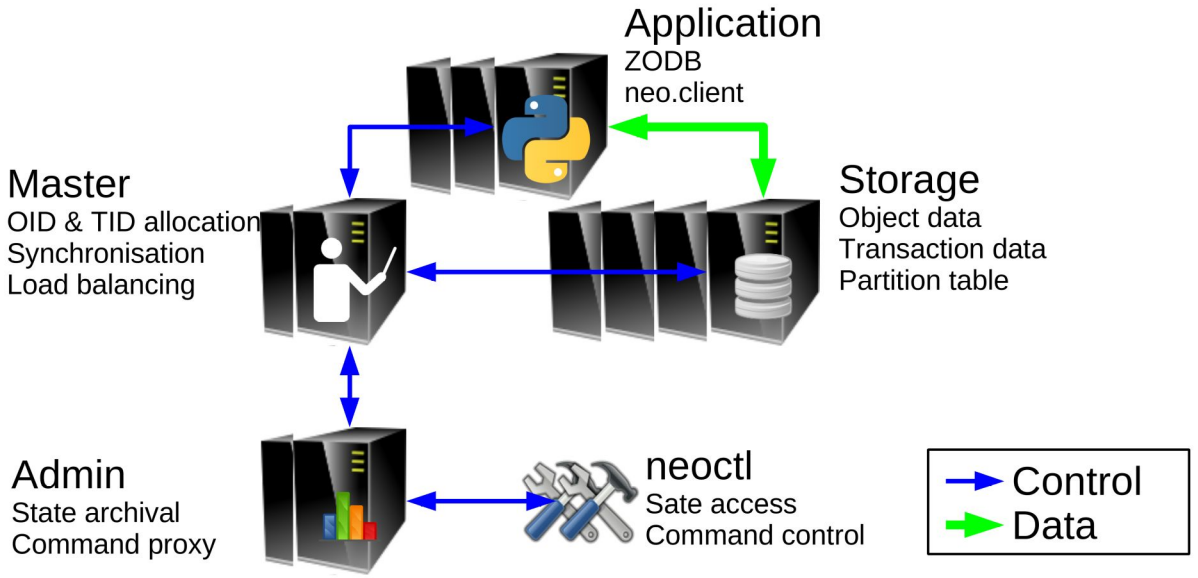


nexedi

Take the Best Analytics scikit-learn.org



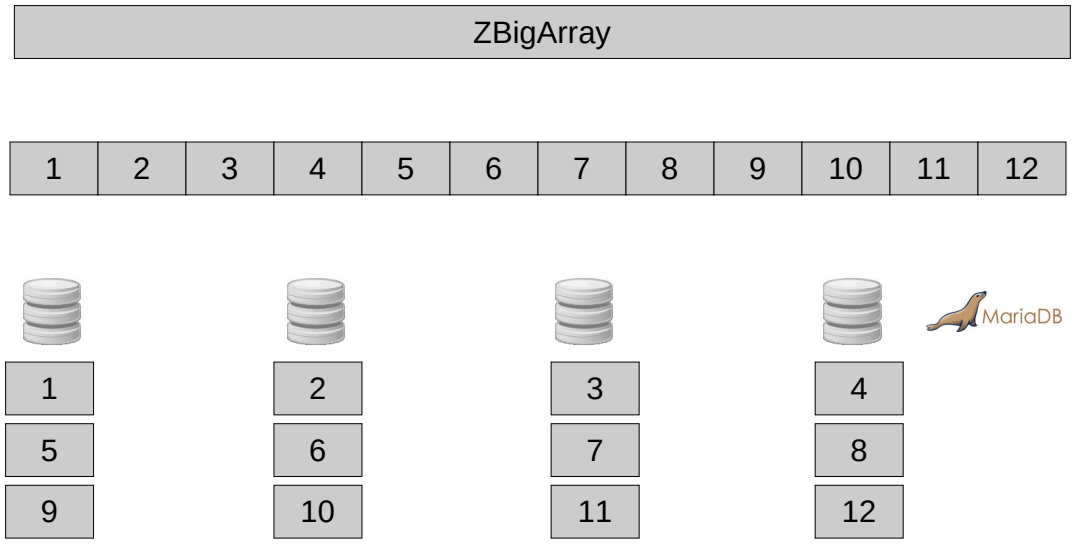
Add Distributed Storage neoppod.org



nexedi

“Magic” out-of-core for NumPy

PyData Paris 2015 – 16h45 Kirill Smelkov



nexedi

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

nexedi

And Multicloud Deployment slapos.org

MMC Rus

Aide et Action
L'Education change le monde

sanef

AIRBUS DEFENCE & SPACE

nexedi

Wendelin Platform 100% open source

100% Python

Scikit Learn — Data Analytics

NEO — Distributed Storage

MariaDB

ERP5 — Elastic PaaS

SlapOS — Multicloud Deployment

Multi Data Center

nexedi

Wendelin Options 100% open source

100% Python

- OpenCV-Python
- NLTK
- Blaze
- Numba / Parakeet
- Pandas
- Scikit Learn
- NEO

Video Processing
Intel Russia / Willow / Itseez

Natural Language Tookit
U. Texas / Chalmers

Full out-of-core arrays
Continuum / DARPA

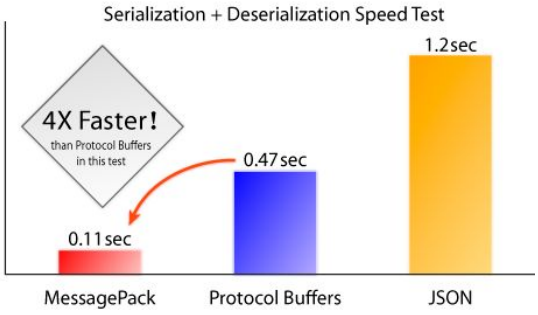
JIT compiler / type inference
Continuum / DARPA

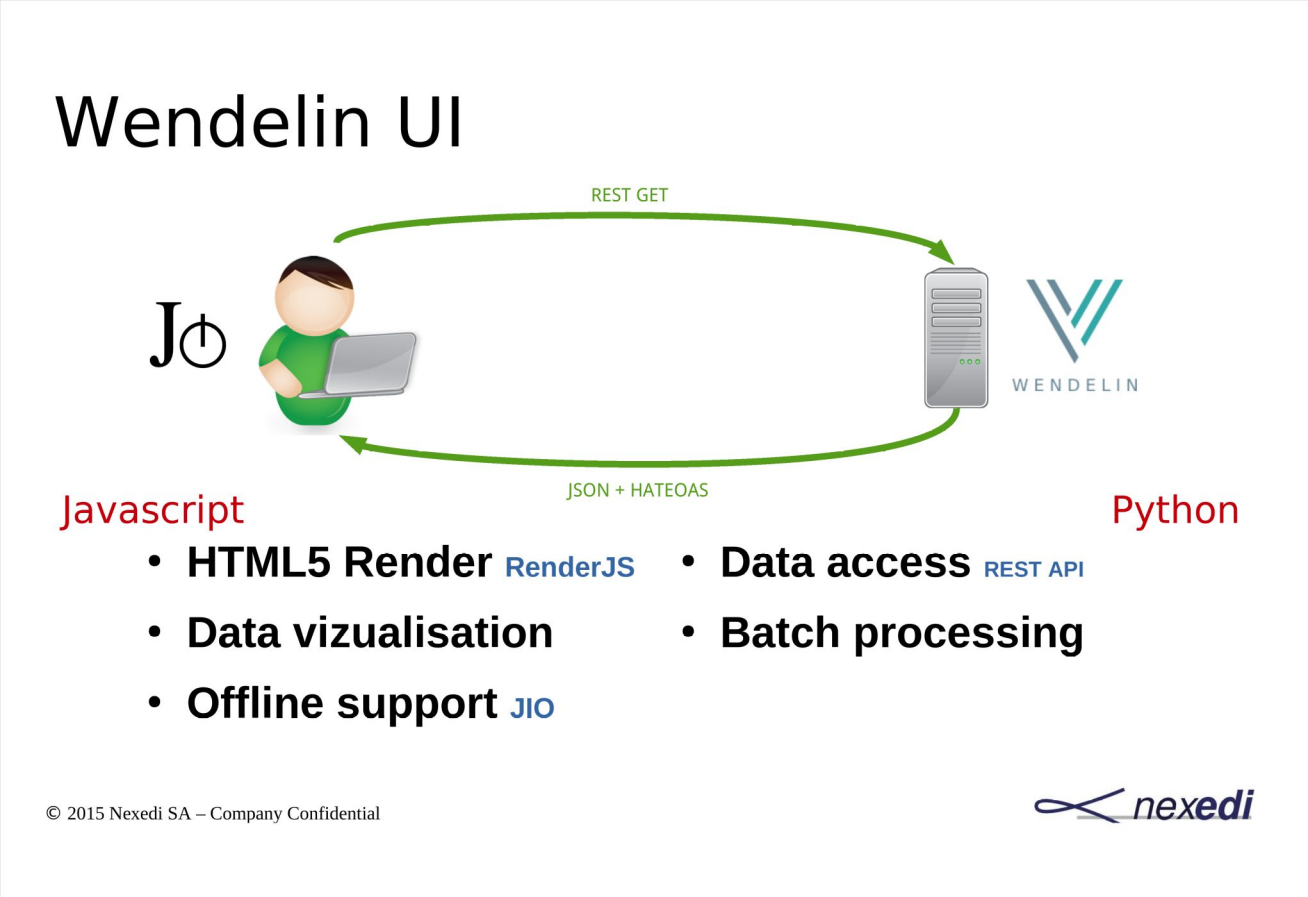
Time sequence processing
DataPad / JP Morgan

nexedi

Data Ingestion: fluentd

- **Based on MsgPack middleware**
- **Created by TreasureData (BDaaS pioneers)**
- **Used by Amazon**
- **Numerous plugins**
- **Scalable and resilient**
- **Bandwidth saver**





The experimental HTML5 UI of ERP5 uses a library called JIO to abstract the relation between the browser and the server.

The browser sends REST requests over HTTP.

The server returns JSON data over HTTP with a self-discoverable format, something called HATEOAS.

The user interface is implemented as a javascript application that runs on the browser side. HTML is generated on the browser side by Javascript code. Form data is prepared by Javascript code and sent as JSON to the server. Many features of the application are still available even offline.

The role of the server in this architecture is only to provide access to the data and to validated the data before updating records in the dabase, using global consistency rules.

Wendelin Distinctive Advantages

- **Native out-of-core NumPy (scikit-learn, pydata)**
- **Native parallel processing**
- **Bare metal performance (GPU, FORTRAN)**
- **Transactions (ingestion, processing)**
- **NewSQL queries**
- **Built-in PaaS**
- **Lower deployment cost (10x less than...)**

nexedi

Detailed Example

WENDELIN

nexedi

The solution that was deployed at the Lightning Protection Center complies is based on open source software – with full access to source code – and does not use software made by IBM, Oracle or EMC. It is thus a “No IOE” compliant solution, in line with directives published by Chinese governments for certain markets.

Data Transportation fluentd

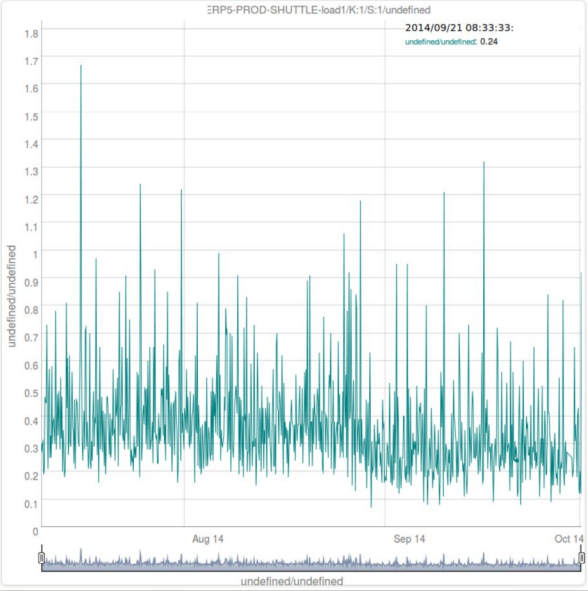
3 months benchmark
Frequent downtime (server, network)
Very poor networking (ADSL, 3G)

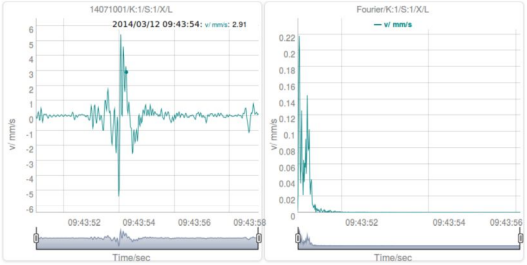
< 0.001% loss

nexedi

UI: HTML5 Components **RenderJS**

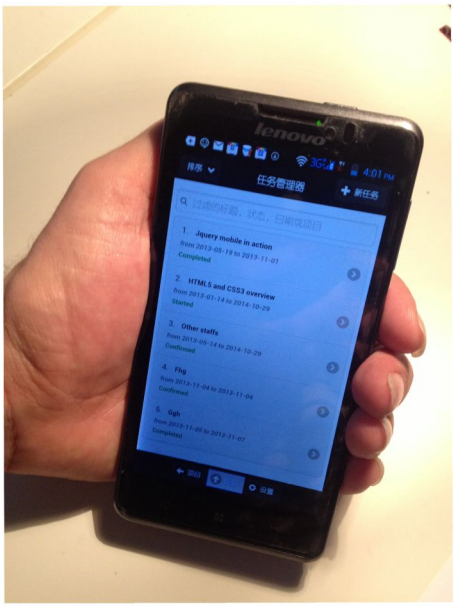
Extend UI Components **RenderJS**



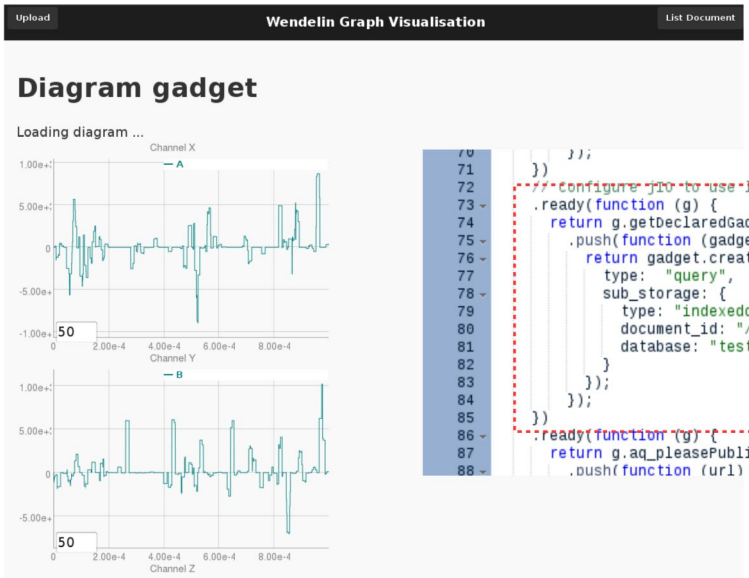


nexedi

UI : Responsive RenderJS

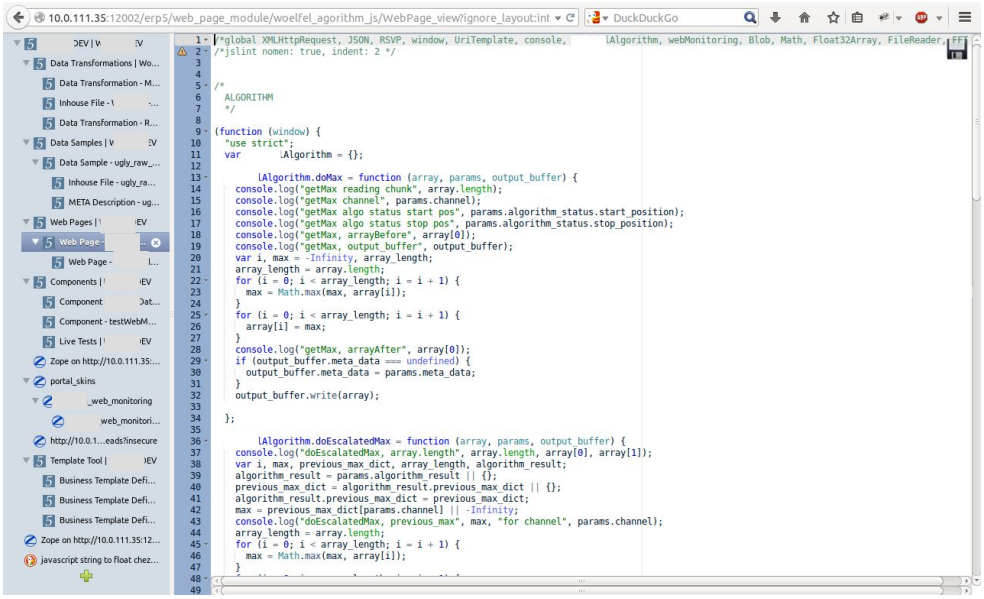


UI: Offline / Other Backends JIO




```
  })
  // configure jio to use localstorage
  .ready(function (g) {
    return g.getDeclaredGadget("JIO")
      .push(function (gadget) {
        return gadget.createJio({
          type:  "query",
          sub_storage: {
            type: "indexeddb",
            document_id: "/",
            database: "test_ivan"
          }
        });
      });
  })
  .ready(function (g) {
    return g.aq_pleasePublishMyState({page: 'listbox'})
      .push(function (url) {
```

Data Science in Javascript vs. Python ?



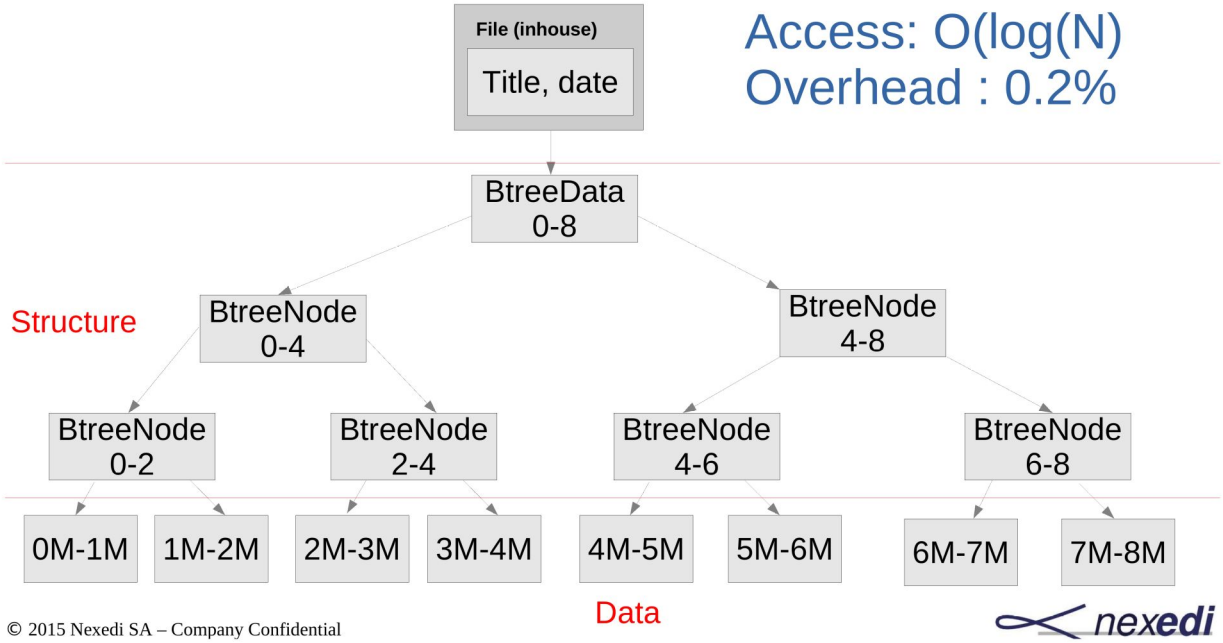
nexedi

Data Sciences in Javascript ? phantomjs

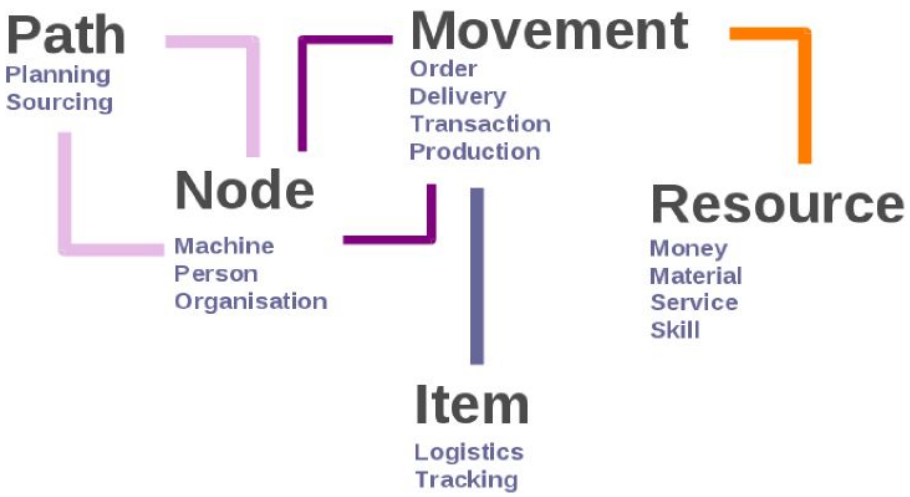
- Small data on client side ✅
- Small data on server side ✅
- Medium data (> 1 GB) in JS ⛔
- Out-of-core data in JS ⛔
- PyData compiled in JS ⛔
- PyData in NaCl / PNaCl ☐

nexedi

Storing large streams in NEO



UBM Monitoring Model?

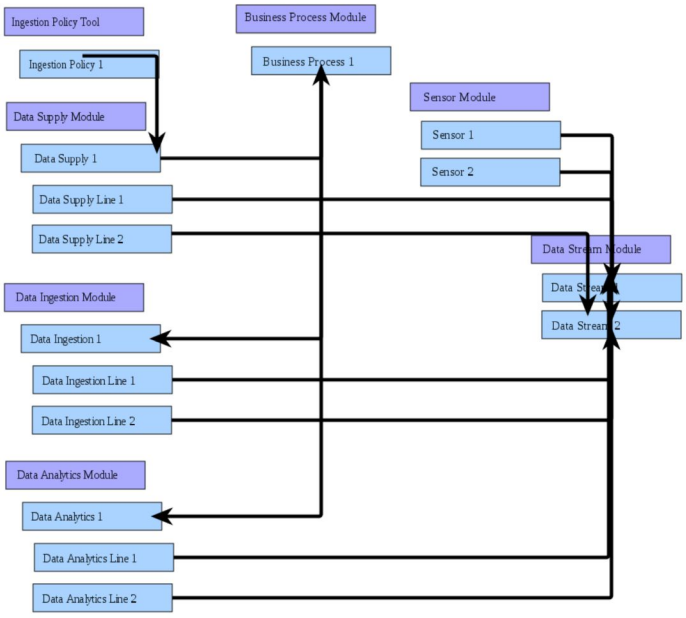


nexedi

UBM Business Model

- **Movement – ingestion of data**
- **Resource – type of data (ex. memory log)**
- **Node – data source, data owner**
- **Path – data source registration**
- **Item – sensor, data itself, license, data set**

UBM Business Model



What UBM gets us for free

- **Accounting, billing and payment**
- **User registration and management**
- **Rule based security model**
- **Customer relationship management**
- **Web Content Management**

→ **save 12+ months and > 200 K€ on any Big Data project**

nexedi

Future Roadmap

WENDELIN

nexedi

The solution that was deployed at the Lightning Protection Center complies is based on open source software – with full access to source code – and does not use software made by IBM, Oracle or EMC. It is thus a “No IOE” compliant solution, in line with directives published by Chinese governments for certain markets.

Roadmap

www.wendelin.io

- **Mainly accelerate learning curve**
  - ☐ **Universal packaging**
  - ☐ **Ready to use examples**
  - ☐ **Act as a backend to ipython notebook**
  - ☐ **Port joblib to CMFActivity**
- **Yet, you can start using part of Wendelin now!**
  - ☑ **wendelin.core out-of-core for NumPy** PyData Paris 2015 – 16h45 Kirill Smelkov
  - ☑ **JIO abstract data access library**
  - ☑ **RenderJS components** http://learn.renderjs.org
  - ☑ **UI sample application** https://lab.nexedi.cn/Tyagov/wendelin/
  - ☑ **Open Source**

nexedi

R&D Partners

www.wendelin.io

- **Wendelin-IA (FSN)**
  - **Nexedi**
  - **Abilian**
  - **2nd Quadrant**
  - **Paris 13**
  - **IMT**
  - **INRIA / ENS**
  - **MMC Rus (Ru)**
  - **X Corp**
- **Windelin (Eurostars)**
  - **Nexedi (FR)**
  - **MariaDB (FI)**
  - **Y Corp (DE)**

WENDELIN

Wendelin Big Data

Industrial Monitoring Platform

2014-04-03 – Paris

nexedi