

# CERN's Business Computing Accelerated by Pentaho

Where the infinitely large  
Meets the Infinitely small

Jan Janke  
Deputy Group Leader

CERN Administrative Information Systems Group



# CERN

## World's Leading Particle Physics Research Laboratory

- **Founded in 1954**
- **~2300 Staff**
- **~1600 other personnel**
- **12000 visiting scientists**
- **Budget of ~1 billion US\$**



# A Worldwide Community



Member States  
7115 researchers

Collaborators  
992 researchers

Associate  
774 researchers

Observers  
2558 researchers

Others  
817 researchers

# Birthplace of the Web

Frequently Asked Questions on WWW

## FREQUENTLY ASKED QUESTIONS ON W3

An FAQ list is really a cop-out from managed information. You should be able to find everything you want to know by browsing from the WWW project page, as everything should be arranged in a logical way. Here though are things which maybe didn't fit into the structure, with pointers to the answers which maybe did. Its an experiment, started May 92. The questioners are anonymous.

I am just starting: how do I find out more?[1]

How does www keep track of the available servers?[2]

How does W3 compare with WAIS and Gopher[3] ?

How do I create my own server[4] ?

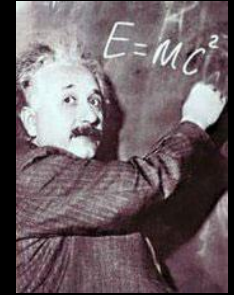
1-10, Up, <RETURN> for more, Quit, or Help: ■



# CERN's Mission

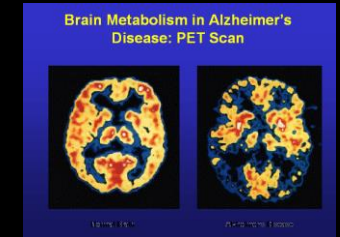
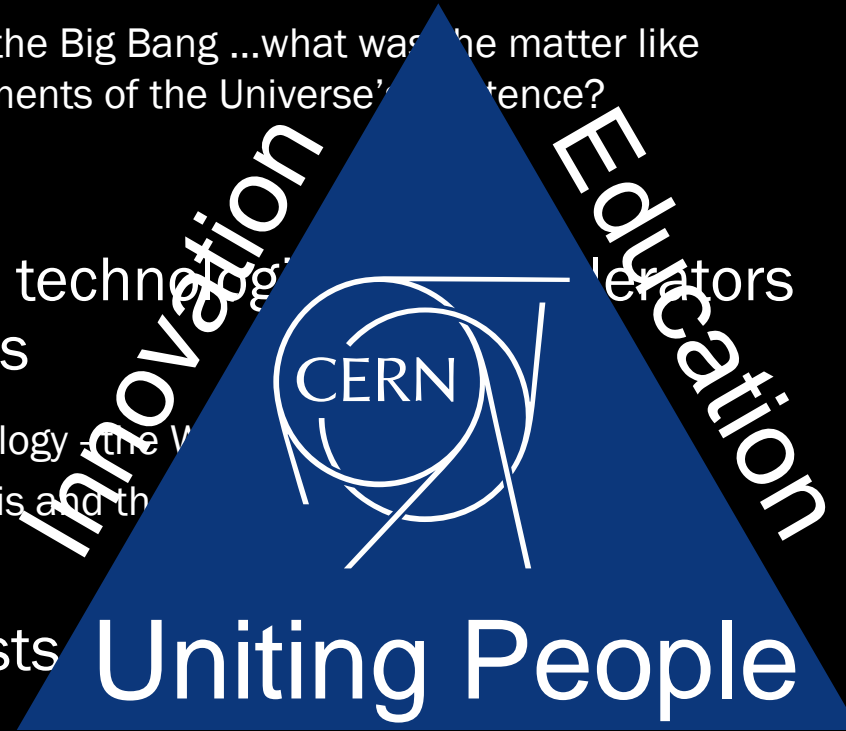
- **Advance** the frontiers of knowledge

E.g. the secrets of the Big Bang ...what was the matter like within the first moments of the Universe's existence?



- **Develop** new technologies, accelerators and detectors

Information technology - the Web  
Medicine - diagnosis and therapy



- **Train** scientists



- **Unite** people from different countries and cultures



# Next Scientific Challenge:

Understand the very first moments of our Universe after the Big Bang



# World-class technology...

# The Coldest Temperatures



# The LHC



SUISSE  
FRANCE

LHCb

ATLAS

CERN Meyrin

CERN Prévessin

SPS - 7 km

PS - 0.28 km

CMS

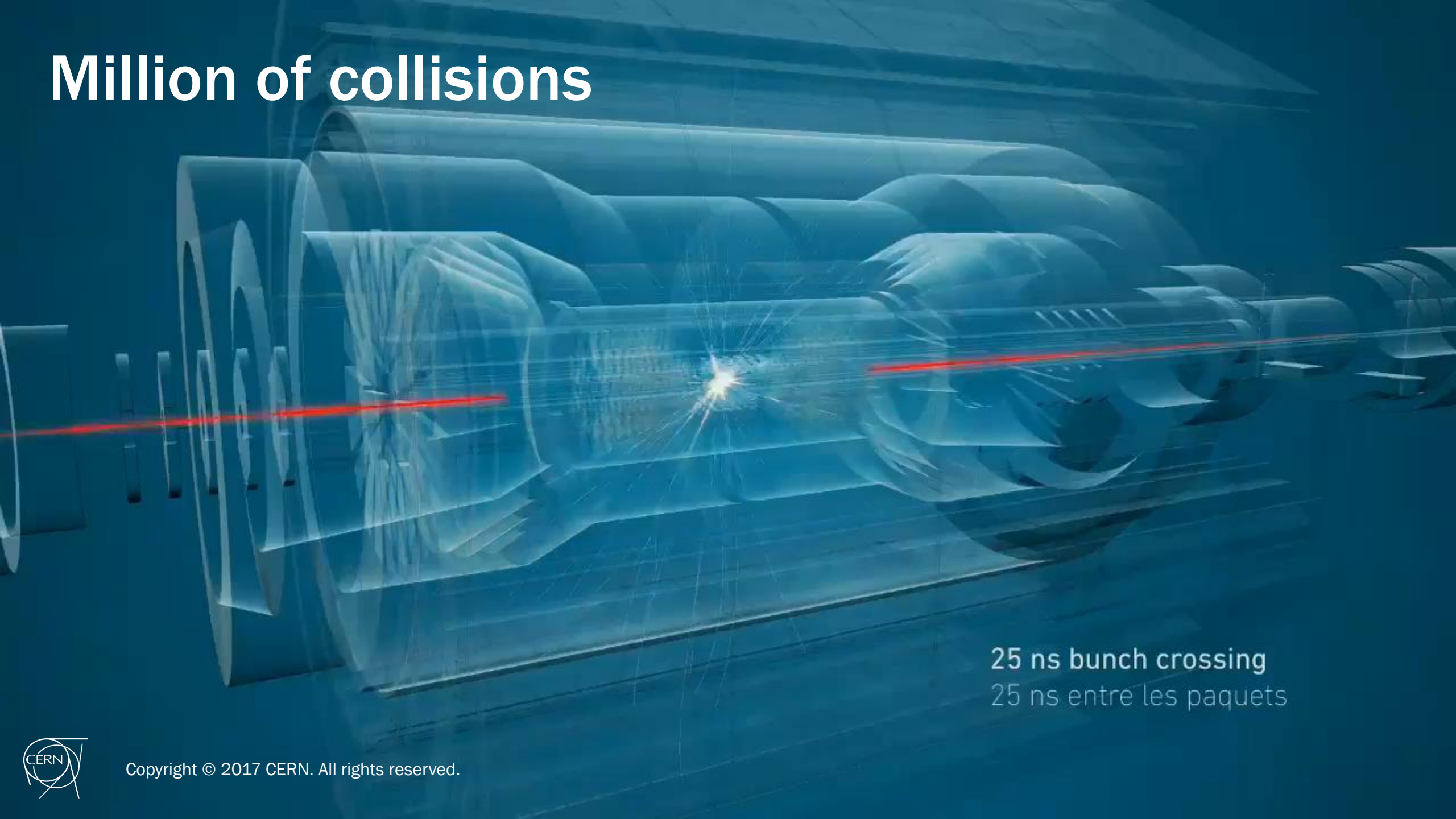
ALICE

LHC - 27 km





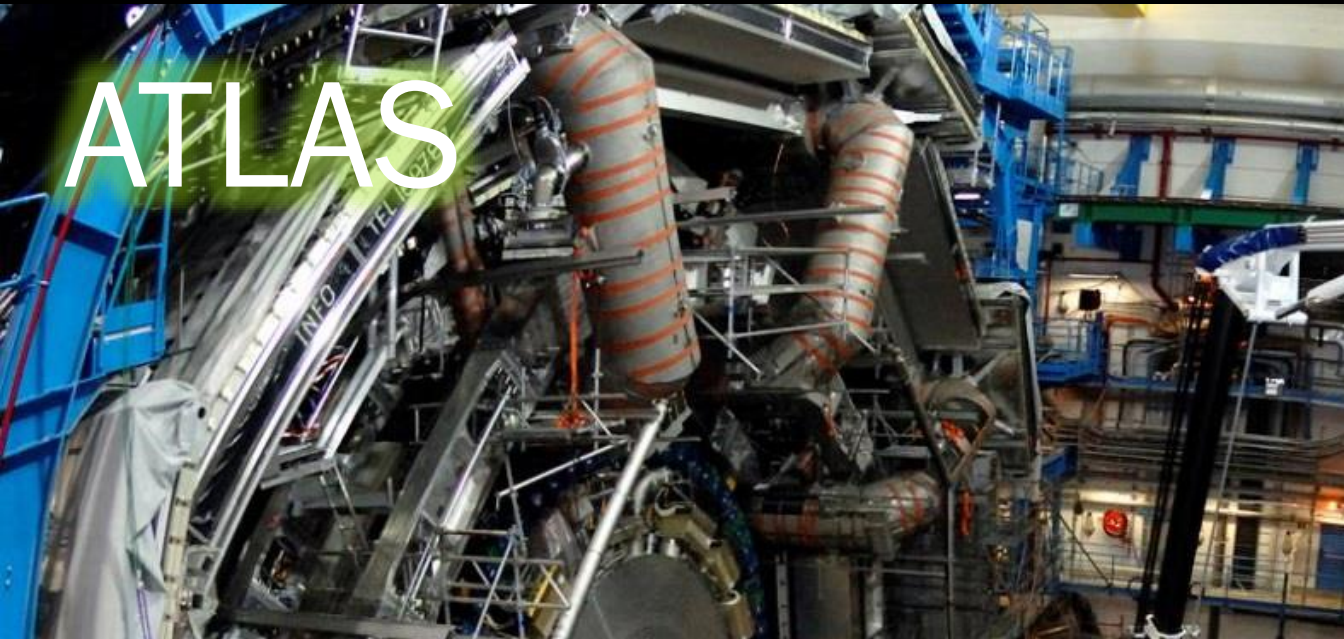
# Million of collisions



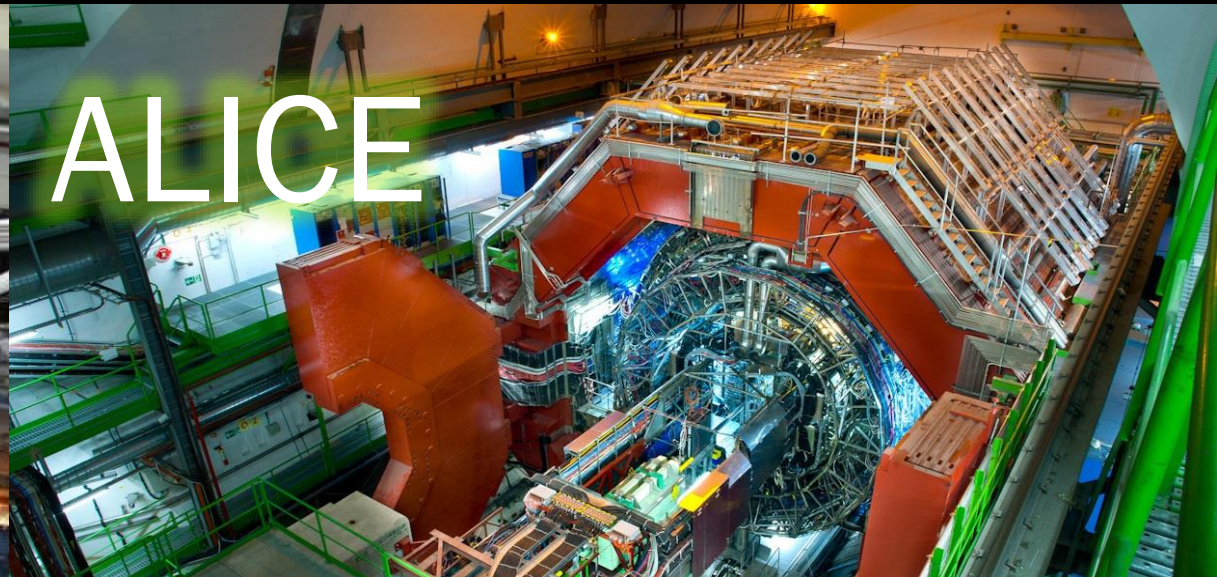
25 ns bunch crossing  
25 ns entre les paquets



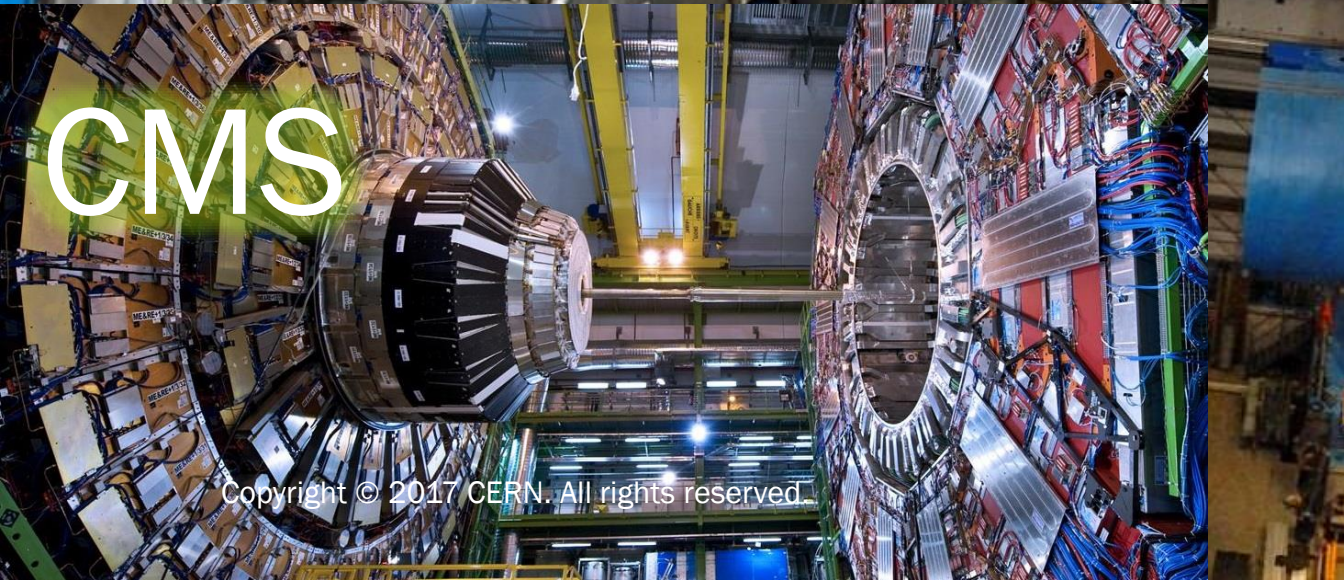
# Particle Detectors



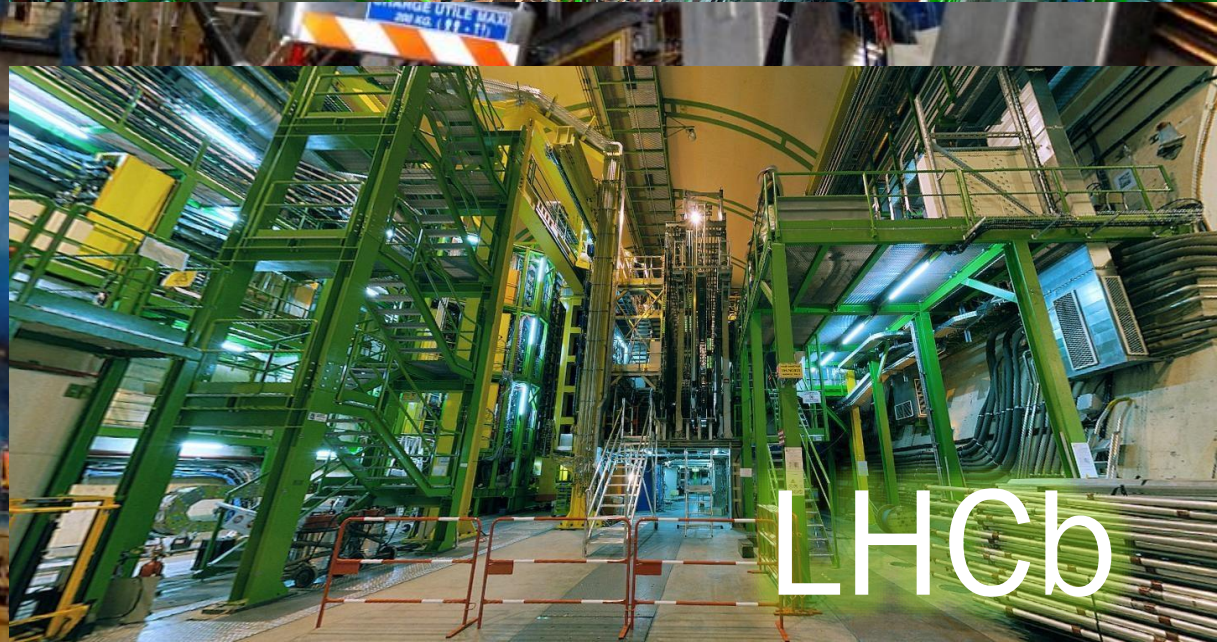
ATLAS



ALICE



CMS



LHCb

# Computer Centre (Tier 0)

Raw recording rate  $\sim 10 \frac{1}{2}$  GByte/sec

91,000 processing cores

30 PetaBytes of data storage on disk

70 PetaBytes of data storage on tape

$\sim 25$  PetaBytes/year





Grid computing...

The  
largest  
computing  
grid



# Our Business Challenges



Copyright © 2017 CERN. All rights reserved.



# Our Business Challenges

Large Numbers of Highly Demanding Users



# Our Business Challenges

A Unique Organization



Copyright © 2017 CERN. All rights reserved.



# What do we do with Pentaho?

How do we tackle our business challenges?





# Personnel List

Report Options

Filters applied: Status Personnel: MP  At CERN?: Y  Organic Unit:   [Remove all filters](#)

Show subtotals by: Show no subtotals

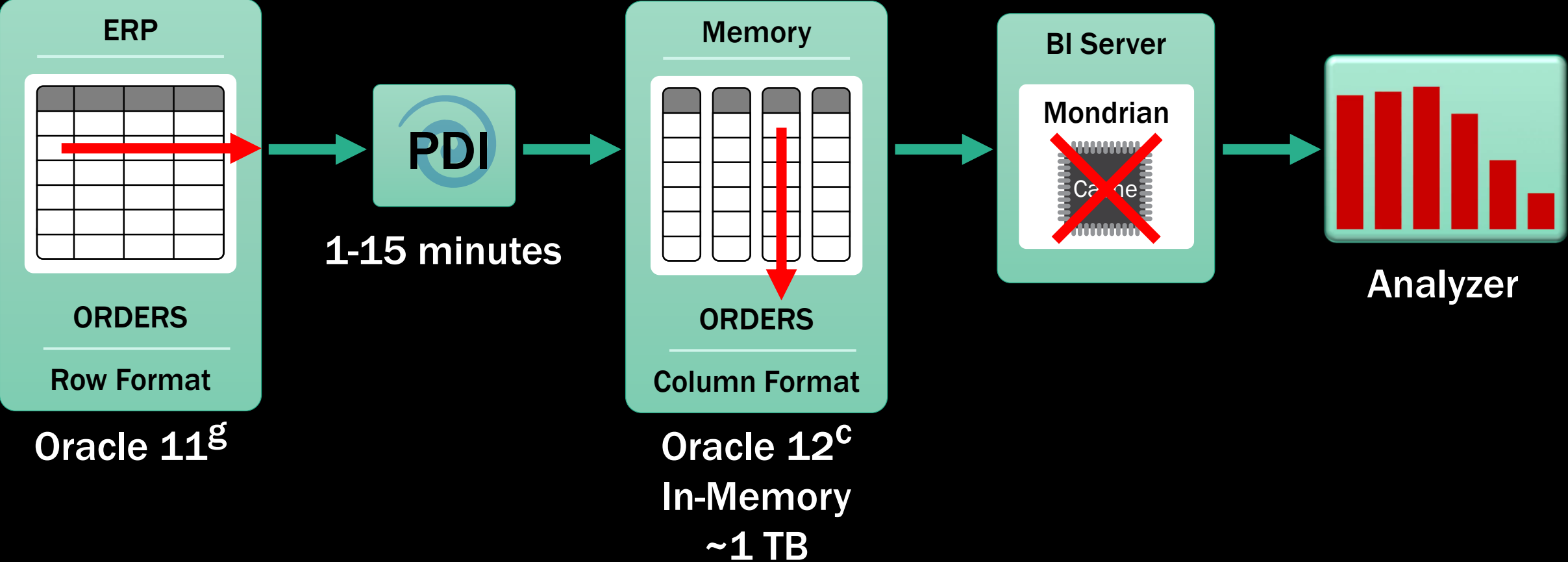
Person	Date of birth	Age	Primary Nationality	Contract Status	Organic Unit
Select...	Click to define a date range	0 → 2000	Select values...	Select values...	1 selected
[Name]		34	Italian	STAF Staff Member	
[Name]		34	Spanish	FELL Fellow	
[Name]		24	Portuguese	PJAS Project Associate	
[Name]		25	Spanish	PJAS Project Associate	
[Name]		28	Polish	FELL Fellow	
[Name]		38	German	STAF Staff Member	
[Name]		22	Greek	TECH Technical Student	
[Name]		30	Italian	FELL Fellow	
[Name]		46	Slovakian	STAF Staff Member	
[Name]		44	Swiss	STAF Staff Member	
[Name]		34	Dutch	STAF Staff Member	
[Name]		33	German	STAF Staff Member	
[Name]		27	Polish	FELL Fellow	
[Name]		51	Dutch	STAF Staff Member	

Show 25 table rows of 14 table rows 14 database results found

Jump to page  « 1 »



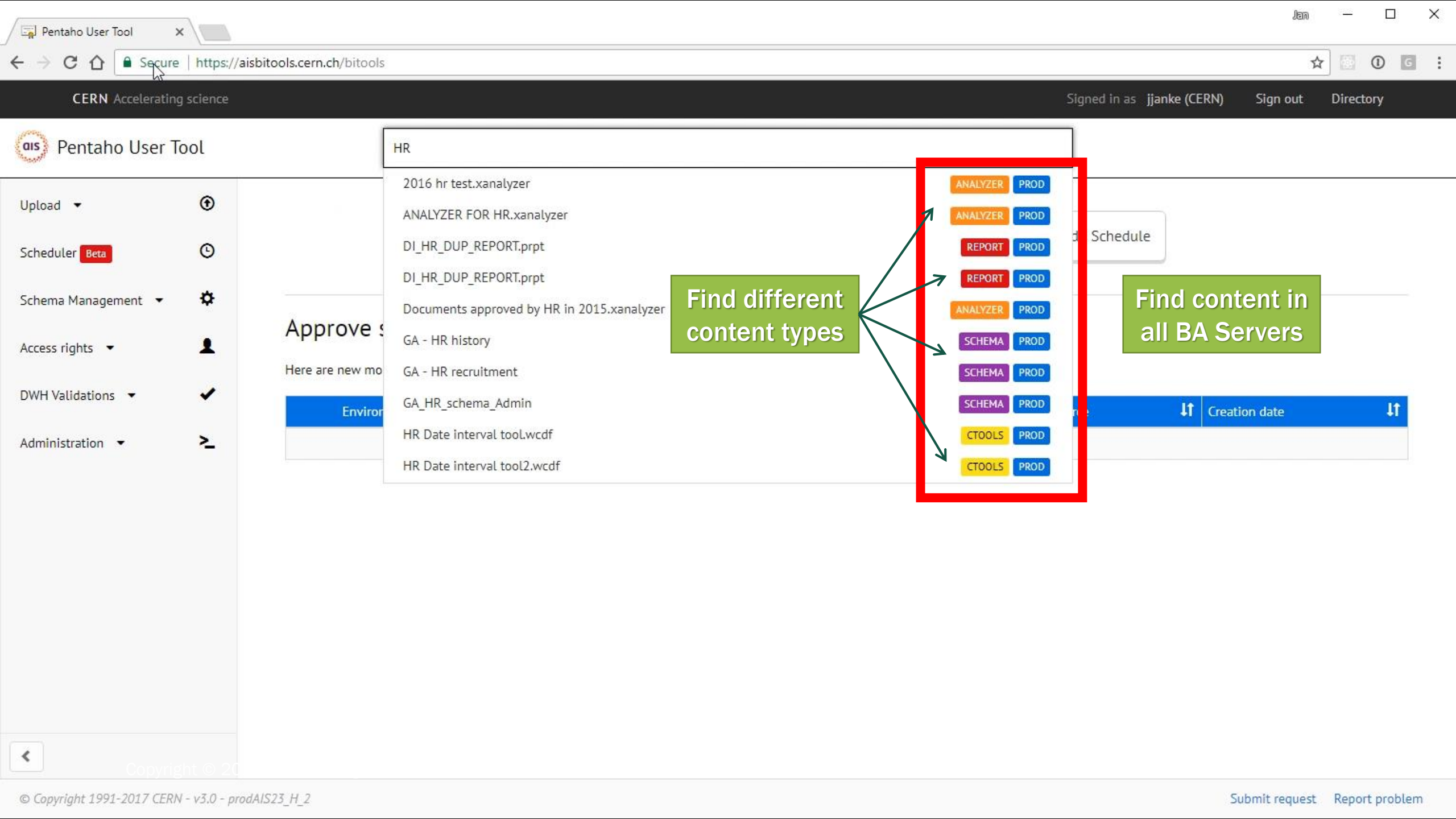
# Performance



# How do we manage Pentaho?

Content management, multi tenancy, scheduling





Find different content types

Find content in all BA Servers

HR

- 2016 hr test.xanalyzer
- ANALYZER FOR HR.xanalyzer
- DI\_HR\_DUP\_REPORT.prpt
- DI\_HR\_DUP\_REPORT.prpt
- Documents approved by HR in 2015.xanalyzer
- GA - HR history
- GA - HR recruitment
- GA\_HR\_schema\_Admin
- HR Date interval tool.wcdf
- HR Date interval tool2.wcdf

- ANALYZER PROD
- ANALYZER PROD
- REPORT PROD
- REPORT PROD
- ANALYZER PROD
- SCHEMA PROD
- SCHEMA PROD
- SCHEMA PROD
- CTOOLS PROD
- CTOOLS PROD

Schedule

Approve s

Here are new mo

Environ

Creation date

# SCHEMA GA - HR history

Match from: NAME CONTENT

How this schema was found

- ais Pen
- Upload
- Scheduler
- Schema Ma
- Access right
- DWH Validat
- Administrat

Schema Reports Properties Access Rights

```

<Schema name="GA - HR history" description="Personnel History Cube as of a given date" me
  <Annotations>
    <Annotation name="Created">
      <![CDATA[Snapshot cube based on the aisdwh_stats.PERSON_CONTRACT_SNAPSHOT view]]>
    </Annotation>
  </Annotations>
  <Dimension type="TimeDimension" visible="true" highCardinality="false" name="dKeyDate"

```

Decentralised administration of schemas!  
 Select who can manage which schemas – no need to be too generous w. full admin rights.

78 Dashboards matching HR

2 3 Next

# Multi-Tenancy: Decentralise Administration

CERN Accelerating science

Signed in as jjanke (CERN) Sign out

ais Pentaho User Tool

Search for Pentaho content

Accessible to domain administrators

Upload

- Upload Mondrian Schema
- Upload Mondrian Security

Scheduler **Beta**

Schema Management

Access rights

DWH Validations

### Upload Mondrian Schema

Upload to any available BA server

Select file

Choose environment

Choose owner

Choose datasource

Upload

GA - HR history - JJ Test.mondrian.xml

GA - HR history Copy - JJ Test

PROD

CREATE

aisbi-ais-admins

AIS\_DWH

Confirm up

New creations require approval by global Pentaho admin team.

Schema Management

- Approve schemas 0

Cancel Upload



# Uncover Dependencies

HR| SCHEMA GA - HR history Operations ^

Match from: NAME CONTENT

Schema Reports Properties Access Rights

Display 10 records per page Search 10 years

Name
<input type="checkbox"/> 02_F_Evolution in the number of Staff Members, Fellows and MPAs over the last 10 years (Students & Trainees and Apprentices excluded).xanalyzer
<input type="checkbox"/> 03_F_Evolution in the Percentage of Female Members of the Personnel over the last 10 years.xanalyzer
<input type="checkbox"/> 08_F_Staff Members by Professional Category over the past 10 years.xanalyzer
<i>Path</i> /home/CERN/TPR/Personnel Statistics/08_F_Staff Members by Professional Category over the past 10 years.xanalyzer
<input type="checkbox"/> 08_T_Staff Members by Professional Category over the past 10 years.xanalyzer
<input type="checkbox"/> 12_F_Evolution in the Average Number of Years of Service of Staff Members over the last 10 year.xanalyzer



# Bulk Update Features

The screenshot displays the Pentaho user interface. On the left, a navigation sidebar includes 'Upload', 'Scheduler Beta', and 'Schema Management'. Under 'Schema Management', there are two items: 'Approve schemas' with a '0' indicator, and a red-bordered box containing 'Schema names bulk update' and 'Data sources bulk update'. The main content area is titled 'Bulk update of the Data sources' and features a progress bar with two steps: 'Choose environment' (with a 'PROD' button) and 'Select the Data source replace' (with an 'aisbip\_dwh' button). Below the progress bar, two green callout boxes provide details: 'Update schema names and all dependent reports' and 'Exchange a data source for another (new) one'.

Upload

Scheduler **Beta**

Schema Management

- Approve schemas 0
- Schema names bulk update
- Data sources bulk update

## Bulk update of the Data sources

Choose environment

PROD

Select the Data source replace

aisbip\_dwh

Update schema names and all dependent reports

Exchange a data source for another (new) one



# Lifecycle

## Select Data Source

Specify which data source you want to analyze. Each data source consists of a schema and a cube in the schema. This will determine which fields you will have available. ?

Data Sources:

- GA - Fidelio Data Mart: Direction Forecast
- GA - Fidelio Data Mart: Forecast
- GA - Fidelio Data Mart: Maintenance
- GA - Fidelio Data Mart: Reservations
- GA - HR history: Personnel**
- GA - HR recruitment: Recruitment
- GA - Infor EAM Data Mart: Booked Hours Cube
- GA - Infor EAM Data Mart: Equipment Count Cube
- GA - Infor EAM Data Mart: Stock Count Cube

Show schemas under development

Hide schemas under development

- UD - Fidelio Data Mart: Forecast
- UD - Fidelio Data Mart: Maintenance
- UD - Fidelio Data Mart: Reservations
- UD - HR Frontlines Internal Activities: Tickets
- UD - HR history: Personnel**
- UD - Intangible Assets Data Mart for Reports: Intangible Cube
- UD - JMT Data Mart: Costs JMT
- UD - JMT Data Mart: Exchange rate JMT

By convention, name starts w. "UD" (under development)

Publish to BA server

Create and test reports

Prepare new data source

Approval for release

ust ally

available)



# Scheduler Enhancements

Schedules Add Schedule

Schedule Name	Repeats	Reports	Output	Last Run	Next Run	Created by	Status	Available actions
My Test Schedule	Single Execution: 22.08.2017 11:33	<b>Schedule several reports together</b> /home/CERN/CMMS/Alarm Systems/GAS/Trim reports/01 - Liste Des Installations Gaz.prpt /home/CERN/TPR/Personnel Statistics/01_F_Members of the PersonnelXanalyzer	JAN JANKE (FAP-AIS-GI)	22.08.2017 11:33	-	JJANKE	expired	

Manage Schedules

Schedule Name	Repeats	Source File	Output Location	Last Run	Next Run	Created By	Status
Items under contract per pl...	Every Wednesday at 1:00:00 PM	/home/cgarino/ Items under contract per planner	/home/cgarino/Scheduled execu...	Aug 16, 2017 1:00:0...	Aug 23, 2017 1:00:0...	cgarino	Normal
PIE_NB OF SALES AND TUR...	The first Monday of every mo...	/home/cgarino/ PIE_NB OF SALES AND TURNOVER P...	/home/cgarino/Scheduled execu...	Aug 7, 2017 12:00:0...	Sep 4, 2017 12:00:0...	cgarino	Normal
Pickings for the last 30 days	CRON: 0 45 8 ? * 1-5 *	/home/cgarino/ Pickings for the last 30 days	/home/cgarino/Scheduled execu...	Aug 16, 2017 8:45:0...	Aug 17, 2017 8:45:0...	cgarino	Normal
Pickings per week per ware...	Every Monday	/home/cgarino/ Pickings per week per warehouse	/home/cgarino/Scheduled execu...	Aug 14, 2017 10:00:0...	Aug 21, 2017 10:00:0...	cgarino	Normal

**Suitable for simpler most frequent needs**

# Scheduler Enhancements

Check previous executions.  
Download results from past runs.

Execution	Results
04.08.2017 16:55	<a href="#">Download</a>
08.08.2017 13:58	<a href="#">Download</a>
04.08.2017 16:43	<a href="#">Download</a>

Parameter Values

Standard

SQL Based

Tables *	<input type="checkbox"/> COR_D_PERSON_INFO	
	<input type="checkbox"/> COR_D_PERSON_ASSIGNMENT	
	<input type="checkbox"/> COR_F_PERSON_BRIDGE	
Aliases	COR_D_PERSON_INFO	C
	COR_D_PERSON_ASSIGNMENT	B
	COR_F_PERSON_BRIDGE	A

Conditions de la requête SQL

```
a.assignment_id = b.dwh_bus_id  
and a.info_id = c.dwh_bus_id  
and a.dwh_bus_start <= sysdate  
and a.dwh_bus_end > sysdate  
and a.dwh_current = 'Y'  
and b.dwh_current = 'Y'  
and c.dwh_current = 'Y'  
and b.organic_unit_name = 'FAP-AIS-GI'
```

ATT - Attestation d'emploi

Base report parameters on SQL.

	<input checked="" type="checkbox"/> SQL Based	
CERN ID*	<input checked="" type="checkbox"/>	C.CERN_ID
Attestation for this Date (dd/mm/yyyy)*	<input type="checkbox"/>	13/10/2017
Options	<input type="checkbox"/>	Pension fund History Hide Stamp Conducting Research
Language*	<input type="checkbox"/>	Pref. Language
Notes	<input type="checkbox"/>	placeholder
<input type="button" value="Enregistrer le calendrier"/>		<input type="button" value="Annuler"/>



# Version Control Integration



git

## Post-Receive WebHooks

Bitbucket will send a POST request to the URLs provided after code has been pushed to your Git repository. The body of the POST request contains information about the repository where the change originated, a list of commits, and the user that made the push.

You can find the details in our documentation.

URL \*

TPR - Monitoring and

## Source

master

test\_gitintegration / home / CERN / TPR / **\_Private** /

..

report\_v7.prpt

publish report saved from prd 7.1

testreportgit2.prpt

Test upload

- > GS-AIS
- > HR
- > INFOR
- > JMT
- > ServiceNow
- > SITE SERVICES
- > SUPPLY\_CHAIN
- ▼ TPR
  - > \_Dashboards
  - \_Private**

- Enable change management
- Keep a clean approved “production” environment



# Thank You!

- We are eager to share our experiences
- What are you doing with and around Pentaho?
- Visit CERN
  - Virtually at <https://home.cern>
  - In Person, book a (free) visit: <https://visit.cern/>

# Questions?

