



# Cassie

CENTRALIZED DATA REPOSITORY

## Web access to a fully searchable data repository

Ana Škrlin, Hospira Zagreb d.o.o.  
Velimir Gašparović, Ivan Klarić, Calyx d.o.o.

XIII. Annual INFOTEHNA Pharmaceutical Conference  
Wörthersee/Klagenfurt, Austria

calyx

# Agenda

- About us
- About Cassie
  - What Cassie is / isn't
  - Features
- Case Study: Hospira Zagreb
- The future of Cassie
- Q&A

# About us – Calyx d.o.o.

- Founded in 2007
- Providing IT solutions to pharma companies
- Clients include Pliva Hrvatska d.o.o., Hospira Zagreb d.o.o., Infotehna LLC
- A team of experienced software architects, developers, and testers working in pharma environment for the past five years
- A passion for the openness of our systems and for providing a great end-user experience
- Focused on providing IT solutions for business problems



# About Cassie - what Cassie is

- **Centralized Laboratory data repository**
  - **Collects data from all the laboratory workstations (and general purpose PCs)**
  - **Analyzes the data and makes it easily searchable**
  - **Archives the data to removable media for offsite storage (for security purposes, the data is still available on the server)**

# About Cassie - what Cassie isn't

- Cassie isn't a Chromatography Data System
- Cassie is used in some environments where CDS is already present
  - To handle the workstations which the CDS cannot handle due to compatibility issues
  - Can also be used to pick up data from the CDS – for additional redundancy and to make the data searchable through the Web GUI

# About Cassie - Features

## Backup features

- Data is automatically collected when the workstations are not in use
- Two types of backups
  - Full backups
  - Incremental backups
- The data is stored in a compressed format to optimize storage usage



# About Cassie - Features

## Web GUI

- Workstations view
- Data collection status
- Download the collected data
- Configuration
- Search
  - By workstation, file name, date
- Integrated with your IT infrastructure (Single Sign On)
- Two types of users
  - Administrators
  - Regular users



# Cassie Status Monitor

**Cassie Server Status**

**General Server Information**

- The servers PID is 1482, on host , version 3.2.0, started at 5/10 13:25.
- This status was generated at 5/10 13:30.
- The configuration was last loaded at 5/10 13:25.
- PCs will be next queued at 5/10 14:00.
- Other info:
  - 0 pending backup requests from last scheduled wakeup,
  - 0 pending user backup requests,
  - 0 pending command requests,
  - Pool is 0.14GB comprising 80 files and 172 directories (as of 5/10 01:00),
  - Pool hashing gives 0 repeated files with longest chain 0,
  - Nightly cleanup removed 0 files of size 0.00GB (around 5/10 01:00),
  - Pool file system was recently at 13% (5/10 13:28), today's max is 14% (5/10 01:00) and yesterday's max was 14%.

**Currently Running Jobs**

Host	Type	User	Start Time	Command	PID	Xfer PID
------	------	------	------------	---------	-----	----------

**Failures that need attention**

Host	Type	User	Last Try	Details	Error Time	Last error (other than no ping)
dellboy	full		5/10 13:00	XferLOG, Errors	5/10 13:00	No files dumped for share CS

- General Server stats
- Overview of currently running tasks
- Failures that need attention

# Workstations View

**Workstations Summary**

- This status was generated at 5/12 09:02.
- Pool file system was recently at 13% (5/12 09:00), today's max is 13% (5/12 01:00) and yesterday's max was 13%.

**Hosts with good Backups**

There are 1 hosts that have been backed up, for a total of:

- 3 full backups of total size 0.75GB (prior to pooling and compression),
- 6 incr backups of total size 0.00GB (prior to pooling and compression).

Host	User	#full	Full Age (days)	Full Size (GB)	Speed (MB/s)	#Incr	Incr Age (days)	Last Backup (days)	State	#Xfer errs	Last attempt
hal	dummy	3	1.7	0.25	22.59	6	0.5	0.5	idle	0	idle

**Hosts with no Backups**

There are 1 hosts with no backups.

Host	User	#full	Full Age (days)	Full Size (GB)	Speed (MB/s)	#Incr	Incr Age/days	Last Backup (days)	State	#Xfer errs	Last attempt
delboy		0		0.00		0			idle		backup failed (net connect: Connection refused)

- List of all workstations with latest backup status for each workstation

# Workstation Summary

**Workstation hal Backup Summary**

- This PC is used by **dummy**.
- Last email sent to **dummy** was at 5/10 14:25, subject **™**.
- Last status is state "idle" (idle) as of 5/10 14:00.
- Pings to hal have succeeded 25 consecutive times.
- Because hal has been on the network at least 7 consecutive times, it will not be backed up from 7:00 to 19:30 on Mon, Tue, Wed, Thu, Fri.

**User Actions**

**Backup Summary**

Click on the backup number to browse and restore backup files.

Backup#	Type	Filed	Level	Start Date	Duration/mins	Age/days	Server Backup Path
10	full	yes	0	4/29 20:00	0.2	10.8	/data/backupper/pc/hal/10
14	incr	no	1	5/3 20:00	0.1	6.8	/data/backupper/pc/hal/14
15	incr	no	1	5/4 20:00	0.1	5.8	/data/backupper/pc/hal/15
16	incr	no	1	5/5 20:00	0.1	4.8	/data/backupper/pc/hal/16
17	full	yes	0	5/6 20:00	0.2	3.8	/data/backupper/pc/hal/17
18	incr	no	1	5/7 20:00	0.1	2.8	/data/backupper/pc/hal/18
19	incr	no	1	5/8 20:00	0.1	1.8	/data/backupper/pc/hal/19
20	incr	no	1	5/9 20:00	0.1	0.8	/data/backupper/pc/hal/20

**Xfer Error Summary**

Backup#	Type	View	#Xfer errs	#bad files	#bad share	#tar errs
10	full	XferLOG, Errors	0	0	0	0
14	incr	XferLOG, Errors	0	0	0	0
15	incr	XferLOG, Errors	0	0	0	0
16	incr	XferLOG, Errors	0	0	0	0
17	full	XferLOG, Errors	0	0	0	0
18	incr	XferLOG, Errors	0	0	0	0
19	incr	XferLOG, Errors	0	0	0	0
20	incr	XferLOG, Errors	0	0	0	0

**File Size/Count Reuse Summary**

Existing files are those already in the pool; new files are those added to the pool. Empty files and SMB errors aren't counted in the reuse and new counts.

Backup#	Type	Totals			Existing Files		New Files	
		#files	Size/MB	MB/sec	#files	Size/MB	#files	Size/MB
10	full	86	248.5	24.85	86	248.5	0	0.0
14	incr	0	0.0	0.00	0	0.0	0	0.0
15	incr	0	0.0	0.00	0	0.0	0	0.0

- Backup control (start full/incremental)
- A list of all backups for a workstation
- File transfer error report (if any) for each retrieval
- Transfer statistics for each backup

# Browse Workstation Snapshots

The screenshot displays the Cassie backup management interface. The main window is titled "Backup browse for hal". It shows a list of files and folders from a workstation snapshot. The interface includes a sidebar with navigation options like "hal Home", "Workstations", "Server", and "Administration". The main content area shows a table of files with columns for Name, Type, Mode, #, Size, and Date modified. A "Restore selected files" button is visible at the bottom of the table.

Backup browse for hal

- You are browsing backup #20, which started around 5/9 20:00 (0.8 days ago).
- This display is merged with backup #19, #18, #17.
- Select the backup you wish to view: #20 - (5/9 20:00)
- show only this increment
- Enter directory: /
- Click on a directory below to navigate into that directory.
- Click on a file below to restore that file.
- You can view the backup history of the current directory.

Contents of asa

Name	Type	Mode	#	Size	Date modified
15	dir	01777	20	0	2011-04-21 16:20:58
25	dir	01777	20	0	2011-04-21 16:20:21
africa.jpg	file	0777	17	1920878	2011-03-02 18:09:33
ASA prurcnik za administratore.doc	file	0777	17	3988992	2011-04-05 13:48:52
ASA Test Case Procedures	dir	01777	20	0	2011-04-21 16:20:25
asa-150x50.png	file	0777	17	3932	2011-02-17 14:03:53
debian	dir	01777	20	0	2011-04-21 16:20:39
IBMIM_win32.exe	file	0777	17	97612664	2010-11-19 10:48:57
npp.5.8.6.Installer.exe	file	0777	17	4607330	2011-01-27 10:34:38
pathfinder.kc	file	0777	17	1168	2011-03-23 12:20:44
rsync	dir	01777	20	0	2011-04-21 16:20:41
rsync6.secrets	file	0777	17	18	2011-04-20 15:21:19
SOP001139.doc	file	0777	17	92160	2011-04-06 13:46:00
Test_Folder	dir	01777	20	0	2011-04-21 16:20:45
VMXP	dir	01777	20	0	2011-04-21 16:20:46

Select all

- Complete file/folder structure of a backup
- “System snapshot” view
- “New files” view
- Option to download files from the backup

# Search Repository

The screenshot displays the Cassie Search Repository interface. The top navigation bar is green with the Cassie logo and the word 'Search'. Below this, the search criteria are shown: 'Filename: exe' on share 'hal asa', showing 'all' files. The search results are displayed in a table with columns for Share, Type and Name, #, Size, Date, and Media. The results show 8 files found, with the first one being /Test\_Folder/putty.exe.

	Share	Type and Name	#	Size	Date	Media
1	halcasa	/Test_Folder/putty.exe	1	475136	2011-01-19 15:08:00	
2	halcasa	/rsync/rsync.exe	1	245248	2011-01-19 15:54:42	
3	halcasa	/Test_Folder/rsync/rsync.exe	1	245248	2011-01-19 15:54:42	
4	halcasa	/rsync/cygrunsv.exe	1	68096	2011-01-19 16:29:18	
5	halcasa	/Test_Folder/rsync/cygrunsv.exe	1	68096	2011-01-19 16:29:18	
6	halcasa	/debian/npp.5.8.6.Installer.exe	1	4607330	2011-01-27 10:34:38	
7	halcasa	/Test_Folder/npp.5.8.6.Installer.exe	1	4607330	2011-01-27 10:34:38	
8	halcasa	/npp.5.8.6.Installer.exe	1	4607330	2011-01-27 10:34:38	

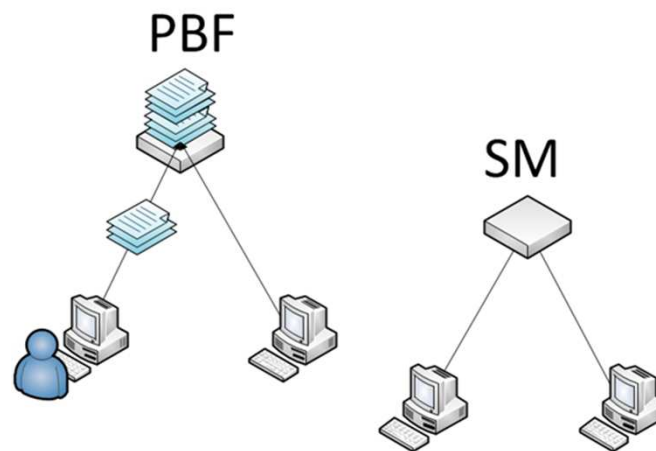
- Search across all folders for all backups for all workstations
- Filtering by date, workstation, folder, date

# Case Study – Hospira Zagreb

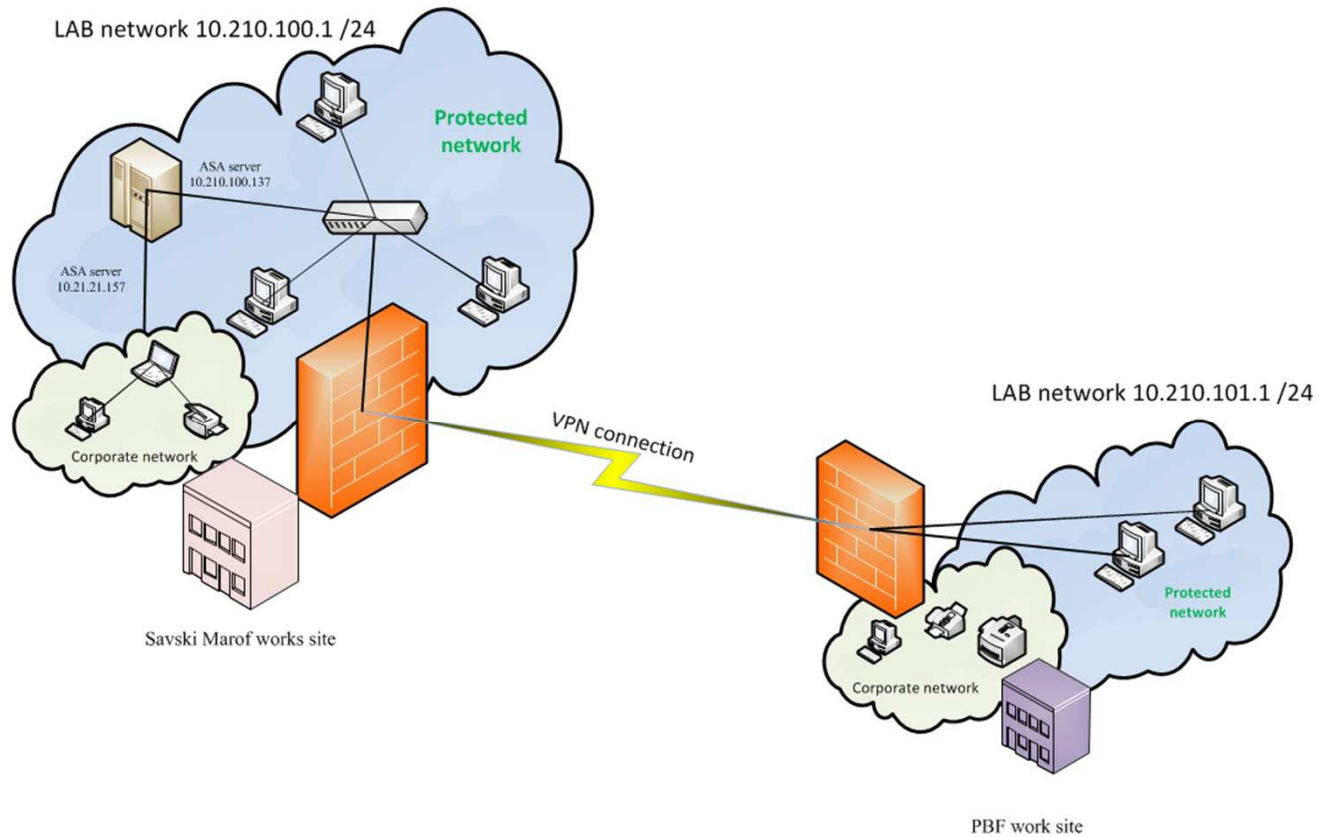
- Hospira Zagreb d.o.o., established in 2009. after spin off of biotechnology division from PLIVA Croatia Ltd.
- Main activity: Manufacturing and testing of biologics API and drug product
- Annex 11: “...Regular back-ups of all relevant data should be done...” (GMP requirement)
- Before Cassie: QC data was backed-up and archived manually

# Case Study – Hospira Zagreb

- Hospira Zagreb is located at two sites: Zagreb (PBF) and Savski Marof (SM), Croatia
- The quality control laboratory has 22 workstations (analytical instruments): 9 at PBF and 13 at SM
- Workstations include: HPLC, UV/VIS spectrofotometers, fluorimeter and circular dichroism



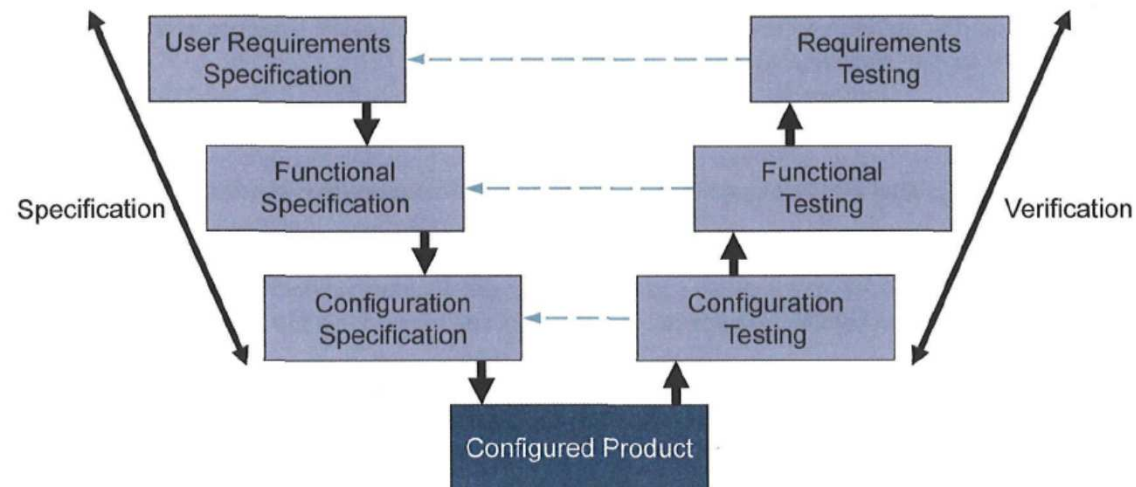
# Case Study – Hospira Zagreb



Hospira network diagram with Cassie

# Case Study – Hospira Zagreb

- Annex 11: Principle requirement: “...The application should be validated”
- Software products which are configured for a specific business process are typically classified as GAMP category 4
- V-validation model (according to GAMP 5)



# Case Study – Hospira Zagreb

- Operational Requirements
  - Data backup (full or incremental, backup restore, etc..)
  - Data archiving
  - Search options (by date period, by workstation and by filename)
  - Password requirements
  - Privileges (user and administrator authorities)
  - Access and visibility
  - Configuration (adding and removing workstations, adding shared folders, etc...)
  - Log files (“audit trail”)

# Case Study – Hospira Zagreb

- **Functional Requirements**
  - Automatic operation on Win 2000, Win XP and Win NT operating systems
  - Designed to support a minimum of at least 50 workstations (10 GB/month)
  - Data collection and security:
    - Data is fetched from each workstation and stored on Cassie server automatically
    - Only newly created and edited files are transferred to the backup server

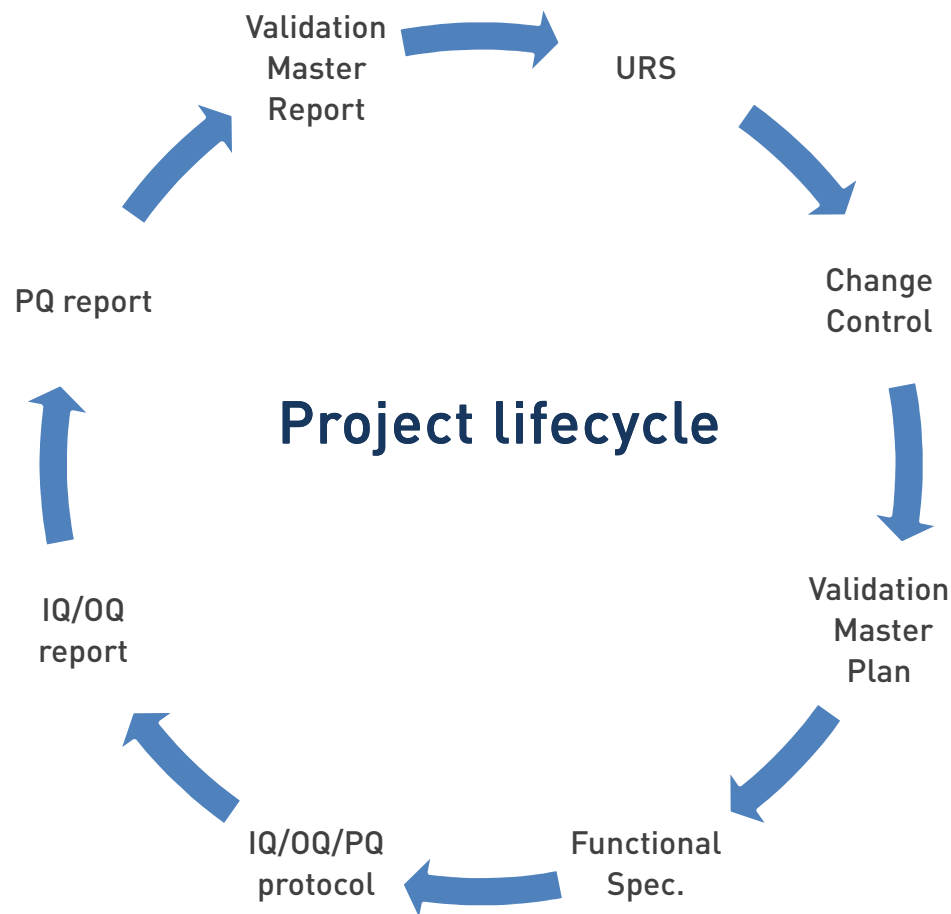
# Case Study – Hospira Zagreb

- **Engineering Constraints**
  - Workstations being backed up must support Microsoft Windows
  - Software constraints:
    - The solution must be able to backup data from Microsoft Windows
  - Hardware constraints:
    - The server needs to be designed to support up to 50 workstations with the capacity of up to 5 years worth of data
    - The server needs to be able to transfer all new daily data in the given timeframe (midnight to 6 AM)

# Case Study – Hospira Zagreb

- **Quality Requirements**
  - OQ / PQ performed for each operational requirement stated in URS
  - Each test step was described and the expected result was stated
  - Tests were approved by QA before testing started
  - Documented evidence collected to support the test result (screen shots)
  - Before go live, training of users and administrators was performed

# Case Study – Hospira Zagreb



# Case Study – Hospira Zagreb

- Fully GMP compliant product
- Developed in line with Annex 11 requirements
- Suitable for use in the pharmaceutical industry both in R&D and QC GMP environments
- Acknowledgements:
  - Marko Vuletić, Hospira Zagreb d.o.o.
  - Mihajlo Ceraj-Cerić, Infotehna d.o.o.
  - Neven Kristijan, Calyx d.o.o.

# The future of Cassie



## ENHANCED INTEGRATION WITH LAB WORKSTATIONS

- reading and interpreting collected measurement data
  - Project Name, Experimenter, Experiment Method, Date/time added to snapshot data should yield better search ability
  - Will cover all major vendors & devices



## FURTHER INTEGRATION WITH EXISTING IT INFRASTRUCTURE

- Integration with your IT technical support system
  - Automatic trouble detection and ticket issuing to your IT administrators
- Integrated with the laboratory network domain controller
  - Laboratory networks often aren't connected to corporate networks and lack centralized user management - Cassie will fix that
- Providing weekly newsletter report for administrators and key persons
  - What projects are being works on, how many measurements were made, per users' usage statistics and more...

# The future of Cassie



## CASSIE ELN

- ELN = Electronic Laboratory notebook
- Cassie is already handling experiment data, why not also handle the Laboratory Notebook process and integrate it with experiment data
  - A lot of companies are still doing their laboratory diaries/notebooks manually
  - Laboratory Notebook data correlated with raw experiment data is a powerful dataset which will give laboratory administrators great insight into what is really going on with their projects
- Cassie ELN will
  - Ease the process of writing reports – a lot of data will be prefilled
  - Automate the integration with Cassie backups
  - Provide great laboratory insights to the administrators
    - Who did what, when, where and what data came out of the experiment

# Questions?



**Cassie**  
CENTRALIZED DATA REPOSITORY

calyx