# IMS Log Data Analysis and Visualization

Vaclav Koudelka

CA Technologies

November 2018

Session HG




# Disclaimer

- Certain information in this presentation may outline CA's general product direction.This presentation shall not serve to (i) affect the rights and/or obligations of CA or its licensees under any existing or future license agreement or services agreement relating to any CA software product; or (ii) amend any product documentation or specifications for any CA software product. This presentation is based oncurrent information and resource allocations as of May1, 2018and is **subject to change or withdrawal by CA at any time without notice**.**The development, release and timing of any features or functionality described in this presentation remain at CA's sole discretion**.
- Notwithstanding anything in this presentation to the contrary, upon the general availability of any future CA productrelease referenced in this presentation, CA may make such release available to new licensees in the form of a regularly scheduled major product release.Such release may be made available to licensees of the product who are active subscribers to CA maintenance and support, on a when and if-available basis.The information in this presentation is not deemed to be incorporated into any contract.
- 
- **THIS PRESENTATION IS FOR YOUR INFORMATIONAL PURPOSES ONLY**. CA assumes no responsibility for the accuracy or completeness of the information. TO THE EXTENT PERMITTED BY APPLICABLE LAW, CA PROVIDES THIS DOCUMENT "AS IS" WITHOUT WARRANTY OF ANY KIND, INCLUDING, WITHOUT LIMITATION, ANY IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, OR NONINFRINGEMENT.**In n**o event will CA be liable for any loss or damage, direct or indirect, in connection with this presentation, including, without limitation, lost profits, lost investment, business interruption, goodwill, or lost data, even if CA is expressly advised in advance of the possibility of such damages.

# About this presentation

- Practical introduction to IMS analytics
- Show one possible way how to approach IMS data analytics
- Demo samples included
- **NOT** ultimate guide to IMS data analytics

# Data science definition

*Despite the excitement around “data science”, “big data”, and “analytics", the ambiguity of these terms has led to poor communication between data scientists and those who seek their help.*[*]

Data mining - Nontrivial retrieval of implicit, formerly unknown and potentially useful information from data.

[*][Harlan Harris, Sean Murphy, and Marck Vaisman. Analyzing the Analyzers: An Introspective Survey of Data Scientists and Their Work. O'Reilly Media, Inc., 2013.]

# Resources

## Github https://github.com/volov0/IMS-analytics

- ftpdown.py – preprocessing  script, downloading logs
- imslog.py – preprocessing, transforming binary to dataframes
- logspecs.py – support functions for scripts
- logdownloader.py – support functions for preprocessing
- appetizer.ipynb – rba plot over time
- deadlock.ipynb – deadlock report
- dataframe.ipynb – introducing log dataframe
- extensions.ipynb – Data set extension plot
- whatsup.ipynb – display what is running in my IMS

# Information available in IMS logs

- Application statistics
- Database activity
- DB updates statistics
- CPU consumption
- Message queue statistics
- Abend information
- Deadlock data
- Monitor statistics
- Buffer statistics
- OLR statistics
- And many more…

# Tools – all open source

ANACONDA.

pandas

$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$

matplotlib

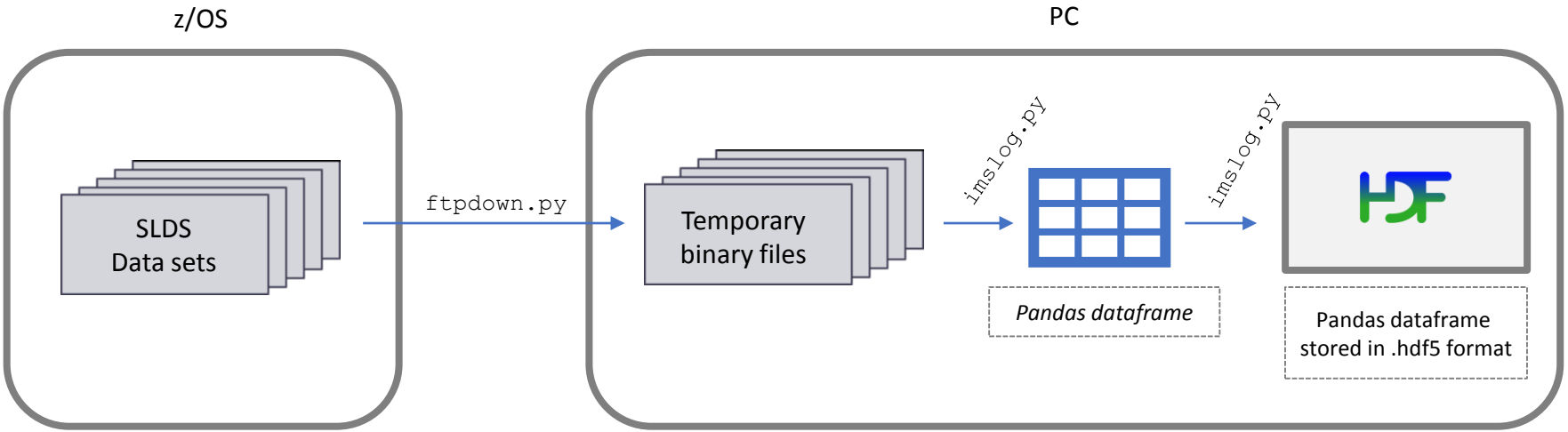
jupyter

HDF

python™

APACHE Spark™

# Preprocessing Data Flow



# Dataframe concept

- Fundamental data structure provided by Pandas (and Spark)
- Contains row for each record and a column for each variable
- Provides methods for accessing, modifying and exporting data

| | type | datetime | psb | extime | ccode | job | step | dlicnt | vsamrio | vsamwio | osamrio | osamwio | totio | iotime | lktime |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **951** | 07 | 2018-08-06 11:54:25.997874 | HPOIO | 9093425 | 0 | HPOD158R | STPRRC00 | 0 | 0 | 0 | 0 | 0 | 0 | 12 | 12 |
| **978** | 07 | 2018-08-06 11:55:06.219089 | HPOIO | 8568734 | 0 | HPOD158R | STPRRC00 | 0 | 0 | 0 | 0 | 0 | 0 | 12 | 12 |
| **950** | 07 | 2018-08-06 11:59:30.614678 | HPOIO | 8898327 | 0 | HPO7300X | STPRRC00 | 0 | 0 | 0 | 0 | 0 | 0 | 12 | 12 |
| **987** | 07 | 2018-08-06 12:06:16.777518 | HPOIO | 9304078 | 0 | HPO7300R | STPRRC00 | 0 | 0 | 0 | 0 | 0 | 0 | 12 | 12 |
| **14649** | 07 | 2018-08-06 12:06:26.448871 | PSBHPO04 | 1632971573 | 0 | HPO7300R | STPRRC00 | 2200 | 187 | 733 | 0 | 0 | 920 | 268775980 | 12 |
| **34448** | 07 | 2018-08-06 12:07:02.198581 | HPODOPT1 | 15522441876 | 0 | HPO7300R | $$@ | 104161 | 11217 | 0 | 0 | 0 | 11217 | 2298755228 | 5341708 |
| **69792** | 07 | 2018-08-06 12:07:21.535785 | PSBHPO04 | 9460267441 | 0 | HPO7300S | STPRRC00 | 31910 | 1557 | 5213 | 0 | 0 | 6770 | 2016884332 | 1324 |

# Deadlock report – DFSERA10 alternative

DFSERA10 control statements:

```
CONTROL  CNTL
OPTION   PRINT OFFSET=5,VALUE=67FF,FLDLEN=2,COND=E,EXITR=DFSERA30
END
```

Deadlock report produced by DFSERA10/DFSERA30:

```
RESOURCE DMB-NAME LOCK-LEN LOCK-NAME
01 OF 02 DVAS7310    08    0018F801C28201D7
KEY IS ROOT KEY OF DATA BASE RECORD ASSOCIATED WITH LOCK
KEY=(7033000163)
       IMS-NAME TRAN/JOB PSB-NAME DBD/PCB# PST#  RGN  CALL LOCK  LOCKFUNC STATE
WAITER IMSX     HPO7300S PSBHPO04 INDEXPCB 00002 BMP  ISRT GBIDP 22400318  06-P
BLCKER IMSX     HPO7300R PSBHPO04 -------- 00001 BMP  ---- ----- --------  04-P
.....................................................................................
RESOURCE DMB-NAME LOCK-LEN LOCK-NAME     - WAITER FOR THIS RESOURCE IS VICTIM
02 OF 02 DVAS7310    08    0018F801C28201D7
KEY IS ROOT KEY OF DATA BASE RECORD ASSOCIATED WITH LOCK
KEY=(0000101000)
       IMS-NAME TRAN/JOB PSB-NAME DBD/PCB# PST#  RGN  CALL LOCK  LOCKFUNC STATE
WAITER IMSX     HPO7300R PSBHPO04 INDEXPCB 00001 BMP  ISRT GBIDP 22400318  06-P
BLCKER IMSX     HPO7300S PSBHPO04 -------- 00002 BMP  ---- ----- --------  06-P
```

# Apache Spark vs Pandas

## Spark

- Cluster computing framework
- Scalability
- Similar API to Pandas

## Pandas

- In-memory single server
- Dataframe concept
- Better performance in small scale
- More flexible


**Vaclav Koudelka**
Principal Software Engineer
Vaclav.Koudelka@ca.com

https://github.com/volov0/IMS-analytics

https://linkedin.com/in/václav-koudelka-17736350


# Thank You.