

AWG 2014

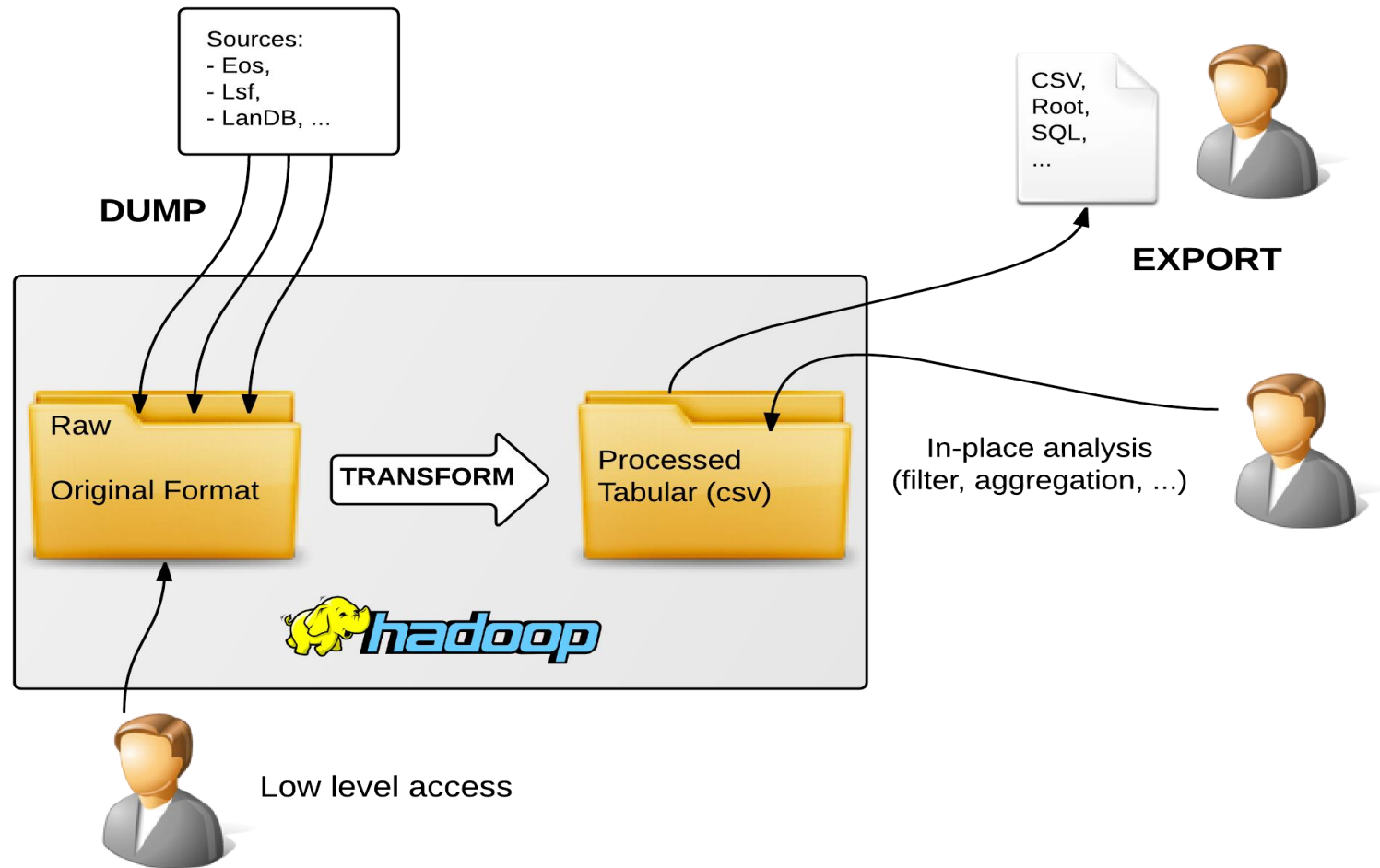
Data Model Description

Christian Nieke – IT-DSS

Disclaimer

- General ideas to start up the discussion
- Nothing is set in stone!

Architecture



Open questions

- Data dumping

- Time interval (daily, monthly...)
- Uniform format?

- Format of processed files

- CSV for now (simple)
- Technology review going on (Parquet, Avro...)

- Advanced Export interfaces

- Open research question

Data Description – WHAT do we need?

- Absolute Minimum:
 - List of data sets (EOS, LSF, ...)
 - Contact address for each data set!
 - List of attributes in **processed** data
 - Attribute names
 - Understandable description
 - Data type & unit

Data Description – WHAT do we need?

- Short Term:

- Links** between data sets

- EOS::

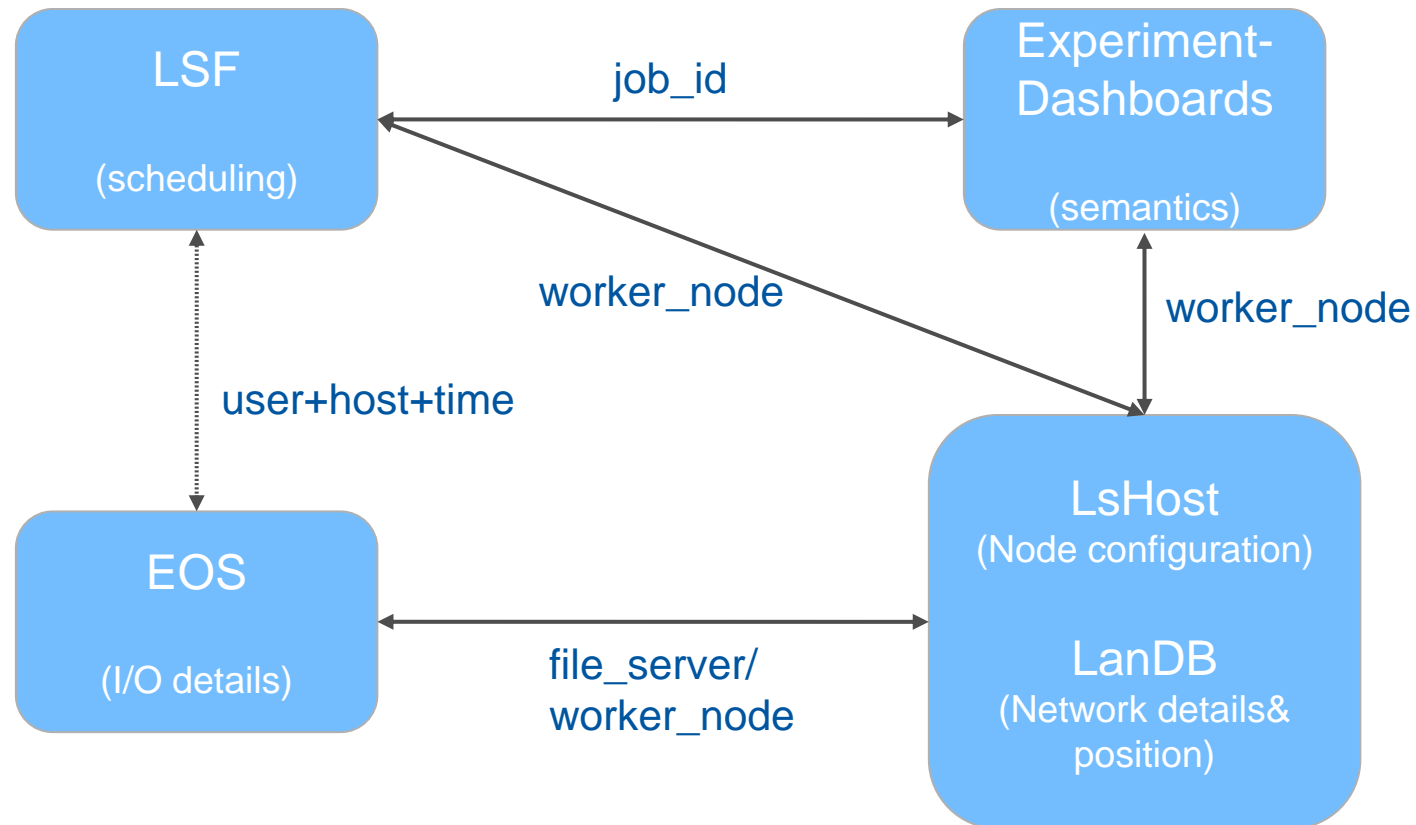
- List of attributes in **raw** data

- For discovery of features that should be extracted

- Names, optionally additional descriptions

Data Sources

- Graphical overview



Where to put it - TWiki Page?

EOS Logs

The logs for EOS transactions. Every entry corresponds to one file interaction, from open to close.

Contact

For AWG: christian.nieke@cernSPAMNOT.ch

Developer: andreas.joachim.peters@cernSPAMNOT.ch

Processed Parameters

Name	Type	Link	Description	Purpose	Remarks
td	String	-	client trace identifier (<username>.<pid>@<client-host>)	This identifier can be used to aggregate sessions (of the same user, with the same job on the same machine)	
path	String	-	Full namespace path to the file	The folder structure in the path could be usefull to match the file to a project/user/programm etc.	
ruid	Int	-	mapped unix user id	Allows aggregation by user	
guid	Int	-	mapped unix group id	Allows aggregation by group	
host	String	LanDB : name	Name of the disk server serving the file	Allows aggregation by user host and links to LanDB	
...					

Raw Parameters

Name	Type	Link	Description	Remarks
log	Int	-	internal EOS log id	
lid	Int	-	EOS file layout id	

TWiki

•Pros:

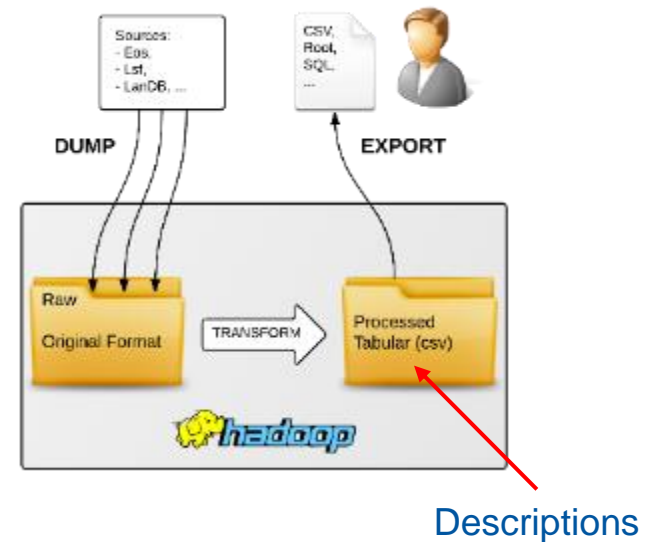
- Central
- Access for everybody
 - To manage ones own data
 - ... but also improve other peoples descriptions on demand
- Hyperlinks
- Fast & easy for the start**

•Cons:

- Requires manual changes

Later: Machine Readable Descriptions

- Simple description files
 - in the filesystem
 - description where the data is
 - can be used with parser



Discussion and demo...