



How to Measure Data Quality and Protect Against Software Audits

By Christof Beaupoil, co-founder and managing partner, Aspera

About the Author



Christof Beaupoil has over 10 years experience in software asset and license management. He studied mechanical engineering and information technology at RWTH Aachen, one of the most prestigious universities in Germany, and is a certified ITIL Foundation and Licensing Specialist.

As the leader of projects at Aspera, Beaupoil is responsible for process consulting for large organizations, project management and coaching, and Aspera managed services. In this role, he's successfully led numerous license management projects for international corporations in the healthcare industry, chemical industry, telecommunications, automotive, logistics, insurance, banking, and financial sectors.

Software asset management (SAM) covers a complex cross section of IT and business and can pose an integration problem between IT, purchasing, and finance. For this reason the focal point of license management is transparency. And consistent transparency can only be realized through consistent data quality management.

In order to achieve this it's necessary to gradually expand the coverage of data collection while also keeping the existing data up to date; and in doing so the most diverse process steps must be followed.

Data Gathering Challenges

Over the last decade organizations in all industries have been consolidating data centers. As data centers are merged and centralized, many enterprises choose to outsource management of the operations. Unfortunately, responsibility for license management has never been part of these outsourcing agreements, which results in negligent software cost management, and possibly compliance and legal trouble.

Not surprisingly, the enterprises that favor outsourcing also release a good portion of their in-house IT professionals that were in charge of managing the different data centers.

Now data center operations have been outsourced, but not license management and all the people at the company that knew anything about the servers and the software on them have been fired. Now, ask a simple question: "How many servers are running Oracle Database and which services are using these?" Nobody knows the correct answer.

The IT environment in any large organization contains various platforms from Windows and Linux/Unix to terminal servers, virtual servers, and virtual clients. This means, on the one hand, you have to gather technical data across numerous platforms and, on the other hand, the more software your company uses the more *and different* data must be gathered to manage licenses.

Data Comes from Processes and Tools Part I

Further complicating SAM is the fact that although software vendors are gradually modifying their license metrics to accommodate virtual environments (e.g. the Cloud), the management tools have not yet adapted to the changes. Ten years ago the majority of software was licensed on a per-installation, per-device, or per-user price model. Any tool that could discover software installations and/or devices could theoretically provide the necessary data to manage licenses. This is definitely no longer the case and processes are required to close the gap between today's software license models and the tools gathering data.

		Environment						Outsourced	Cloud	
		Windows	Linux/Unix	Terminal Server	Virtual (Server)	Virtual (Client)				
Software License Metric	Device	ARP/File Scanner (e.g. SCCM)	SIG Scanner (e.g. TAD4D)	LS Report & Application Matrix	vCenter			License demand reports	Capacity reports	
	Named User	ADS / User reports (e.g. MotioPI)								
	Concurrent User/Device	License Server (e.g. FlexLM, Rational License Server, Citrix License Management Console)								
	CPU/ Cores/ Units	SCCM	iQSonar		Microsoft SQL Server: SP					
	CAL	(e.g. EMC 2010)								
	Virtual Instance				vCenter					
	Capacity		ILMT	SAP LAW						
	Configuration	Deep scanner (e.g. iQSonar)								

The various platforms and differing license metrics require numerous tools to gather all the required data for license management. These are just a few examples to the tools that can be used.

What is Quality Data?

Quality data is having all the information you need when you need it. The three basic criteria to determine quality are: completeness, consistency, and timeliness.

Completeness

Data completeness in license management is three-dimensional. It requires having (a) full organizational data (legal entities, cost centers, users); (b) tools to gather data on each platform, and (c) comprehensive device data. The data is not complete if you need device information on a Unix server, but only have tools to gather data on Windows servers. If you need to know the number of CPUs on a device, but the asset inventory report does not have this, then you need to find a data source that does record this information and combine it with the other data you have.

This first report is grossly inadequate for license management.

Sample Report #1

Host Name	Software Product	Version	Publisher
SERV001	ocfs2console	1.2.4	Oracle
SERV002	Oracle Data Provider for .NET Help	9.2.040	Oracle Corporation
SERV003	Oracle Data Provider for .NET Help	9.2.040	Oracle Corporation
SERV004	Oracle Database	11.1.0	Oracle

Although there are still holes in this second inventory report there is significantly more data. Moreover, the report can be completed by filling the missing fields with worst-case-scenario values.

Sample Report #2

Logical Host	Physical Host	Cluster	Edition	Build Number	Partitioning Method	Processor Type	CPUs	Cores	Options
LOG5	PHYS1	CLUST1	Enterprise	10.2.0.3.0	Soft Partitioning	Intel(R) Xeon(TM) CPU 3.40GHz	4	8	RAC
LOG6	PHYS2	CLUST1	Enterprise	10.2.0.3.0	Physical Host	Intel(R) Xeon(TM) CPU 3.40GHz			RAC
LOG7	PHYS3	CLUST1	Enterprise	10.2.0.3.0	Physical Host		4	8	RAC
LOG8	PHYS4	CLUST1	Enterprise	10.2.0.3.0	Physical Host	Intel(R) Xeon(TM) CPU 3.40GHz	4	8	RAC
LOG9	PHYS5	CLUST1		10.2.0.4.0	Physical Host	Intel(R) Xeon(R) CPU E7440 @ 2.40GHz	4	16	RAC
LOG10	PHYS6	CLUST1	Enterprise	10.2.0.4.0	Physical Host	Intel(R) Xeon(R) CPU E7440 @ 2.40GHz	4	16	RAC
LOG11	PHYS7	CLUST1	Enterprise	10.2.0.4.0	Physical Host	Intel(R) Xeon(R) CPU E7440 @ 2.40GHz	4	16	RAC

Consistency

Consistency draws on the reliability of the data. If device "SRV004" is recorded in your CMDB, then you should also find "SRV004" in your discovery data. If this is not the case, your data is not consistent. Similarly, if "SRV004" is marked in the CMDB as having 4 cores, then the scan data should provide the same information. Expected results vs. delivered results are a key factor to determine data consistency. If you expect 3,000 servers to be inventoried and only 1500 show up in the data, then you have a consistency problem.

Timeliness

Quality data is up-to-date and on time. Having 80% of the data for all software in the organization all time is useless when you need 100% of the data for just one product. For example, your company needs to negotiate new purchase conditions with a software publisher. In order to complete this task you need information about *all* software and licenses in the organization from this publisher. Without this data your company is entering the negotiations blind. In other words: If the data is not current and available at the time you need it, then the quality is terrible.

Measuring SAM Data Quality

The key performance indicator (KPI) Expected vs. Delivered Certainty Rate is a proven method to measure data quality and progress.

This KPI is calculated based data completeness and consistency. In the example below, the graph shows the data quality certainty rate for Oracle Database software in two legal entities. Completeness is determined by the data containing all relevant values for managing Oracle Database license metrics:

- 1. Completeness of devices:** Legal Entity 1 should have 980 devices. However, as marked under “Delivered,” only complete data on 400 devices is provided, meaning only 41% of device data is available.
- 2. Completeness of software:** Of the 400 devices with complete data, only 200 of them (50%) have complete software data, or rather all relevant software information (product, version, edition, etc.) necessary to manage Oracle Database licenses.
- 3. Signed off:** Of the 200 devices with complete device *and* software data, the person responsible for the devices could validate that 100 of them have consistent data.
- 4. Certainty:** Combining these factors the final data quality certainty rate for Legal Entity 1 at the end of week one is very low: 10%. Legal Entity 2 has a data quality certainty rate of 60% at the end of week one; the organization overall is 35%.

The target for data quality certainty is determined by the risk avoidance factor commonly stated in the software purchase agreement as part of the audit clause. The greater part of vendors requires a 95% risk avoidance rate, and therefore, in our example, our target certainty rate is 95%.

Sample KPI Expected vs. Delivered Certainty Rate

Product:			Week 1		Week 2		TARGET
	Expected	Delivered	Expected vs. Delivered (%)	Expected	Delivered	Expected vs. Delivered (%)	
Oracle Database	Expected	Delivered	Expected vs. Delivered (%)	Expected	Delivered	Expected vs. Delivered (%)	TARGET
Legal Entity 1							
Completeness of devices	980	400	41%	980	900	92%	
Completeness of software	400	200	50%	900	850	94%	
Signed off	200	100	50%	850	850	100%	
Legal Entity 2			10%			87%	95%
Completeness of devices	400	300	75%	400	390	98%	
Completeness of software	300	243	81%	390	390	100%	
Signed off	243	240	99%	390	390	100%	
			60%			98%	95%
Overall %			35%			92%	95%

A very important aspect of defending your company during an audit is not only to show compliance, but to prove that the underlying data is of high quality so the compliance results can be trusted. In this case, if the organization were to be audited by Oracle, it would not only have a difficult time to show compliance, but assuring Oracle that the data is reliable would be nearly impossible.

Data Comes from Processes and Tools Part II

Effective SAM relies on the information provided, and is therefore driven by data management. The majority of organizations already have the necessary tools in place to act as data sources for SAM. For this reason, your SAM solution provider should work with you to improve the return on investment for these tools, by adapting and improving existing processes to account for the limitations and to enable higher quality data.

The process mistake most organizations make is to first purchase the software licenses, install the software, and then try to gather the necessary data. This process needs to be turned up-side-down: Once a need for new software is established but before the licenses are purchased IT needs to determine how it can manage the software. If an organization does not have a reliable method to track devices, then it should definitely not purchase licenses with a device or installation-based price model. The same software could be licensed through a different price model, or the organization could discuss different metric options with the vendor. What SAM comes down to is: Don't buy it unless you can "count" it!

Software License Procurement Process to Ensure High Data Quality

