inferapp

IT Discovery Service

Maintenance of
Software Discovery Library

www.inferapp.com

Contents

Contents

Introduction

Introduction

We are a company that focuses solely on maintaining the library for third-party use by IT asset management (ITAM) businesses. We recognize the full spectrum of challenges inherent in satisfying ITAM customers, and we believe our specialization in raw data analysis is the best way to overcome them.
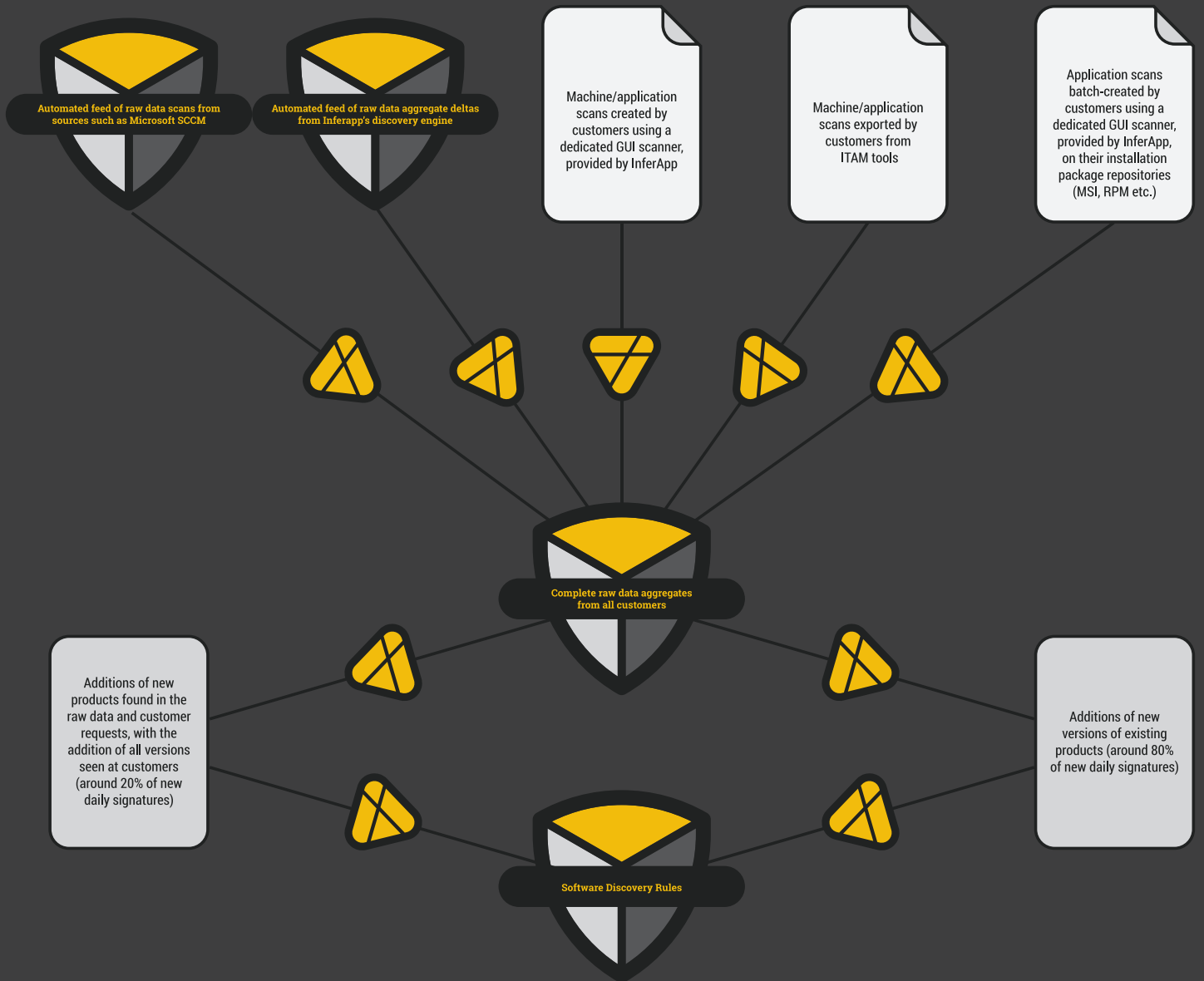
Inferapp continually collects its customers' raw data for analysis and publishes updates to the library every single business day. We do not simply normalize raw data. Instead, our software discovery rules are the result of careful analysis by specialized IT discovery professionals who proactively meet the needs of our customers.

We are able to add thousands of new quality rules every week. However, the total number of rules is irrelevant to accurate software discovery if the rules refer to old versions that are no longer in use, which, in dynamic IT environments of today, often occurs within days. What matters to accurate software inventory is how often and how fast the rules are being updated, rather than their total number.

Professional maintenance of discovery rules requires a rigorous adherence to a set of best practices because, unfortunately, software detection has been arbitrary in its naming conventions, as opposed to raw data collection, which is now relatively standardized. There are also many alternative rule creation choices, and some of them will cause misdetections that could be very difficult to recognize later on.

Inferapp | March 2015

Raw Data
Analysis Process

Automated feed of raw data scans from sources such as Microsoft SCCM

Automated feed of raw data aggregate deltas from Inferapp's discovery engine

Machine/application scans created by customers using a dedicated GUI scanner, provided by InferApp

Machine/application scans exported by customers from ITAM tools

Application scans batch-created by customers using a dedicated GUI scanner, provided by InferApp, on their installation package repositories (MSI, RPM etc.)

Complete raw data aggregates from all customers

Additions of new products found in the raw data and customer requests, with the addition of all versions seen at customers (around 20% of new daily signatures)

Additions of new versions of existing products (around 80% of new daily signatures)

Software Discovery Rules

Raw Data Analysis Process

Inferapp receives raw data from various sources, the most important of which is the automated feed of scans from sources such as Microsoft SCCM. Our discovery engine is also capable of preparing raw deltas that can be automatically sent to Inferapp for analysis. Some customers may use GUI tools to get raw data for specific applications. For example, a batch scanner for installation package repositories (MSI, RPM etc.) largely automates the job, provided a customer maintains such a repository. Finally, some ITAM tools offer front end functionality that allows for export of a machine scan, which can then also be used by Inferapp to create reliable signatures.

All the raw data is aggregated into one database so that our IT discovery professionals can be continuously alerted to new products and versions. First, we make sure all new versions of existing products get added as they arrive daily in the raw data. This constitutes over 80% of daily additions to the library. Next, we add products requested by customers, along with all versions of these products our customers have. Finally, we proactively look for and add new products to the raw data that are not yet present in the library. This third step is crucial because we want to save our customers from the additional work of having to request new products in the first place. We fully realize that our guaranteed one business day turnaround still introduces an unwelcome delay to our library's users, so we do our best to minimize the need for such requests. Typically, requests are only needed for products that are especially difficult to gather exclusively from the raw data.

Challenges in the Creation of Discovery Rules

Challenges in the Creation of Discovery Rules

Signature publisher and product names in the library reflect current values for software products, so they may differ from what can be found in the raw data elements that have triggered particular software detection results, e.g., executables, addremoves or pkginsts. They may also differ because publisher and product names are placed in raw data elements by software developers for internal purposes, mostly to help in source control, and they often bear a quite loose relation to the licensable software applications actually installed. Various smaller executables of same product family that by themselves do not require a license will often have the resource's product name exactly the same as one of the main/parent/associated licensable products. It is completely at the application developer's discretion how this is configured, but it is usually done with no regard to software asset management.

For example, product name found in PowerPoint Viewer's executable (pptview.exe) states Microsoft Office even though it is not the same as a Microsoft Office license. Its file description does explain it is a viewer, and we know this suite well enough to know the viewer is not licensable. However, there are product families from IBM, Oracle, SAP and others, where such descriptions are missing or are too cryptic to be of any informational value for software asset management. Our job is made still more difficult by the lack of consistent software naming conventions. Publishers, products and version names vary with the data source, so rigorous adherence to one set of naming conventions is a must to avoid signature duplicates.

In any environment, there also exists a good deal of installer exes with file names that have version resources very similar to installed legitimate instances, the only difference being their file size or file version format. Similarly, exes may be legitimate but located under patch, setup, backup and similar file paths that clearly indicate the software has not actually been installed.

Preferably, a file based rule should only contain the core executable required for an application to run. An application, however, rarely consists of only a single executable. It usually contains many raw data elements such as files and registry keys. When other applications or other application editions/versions have similar files, including the core executable, as they often do, then the maintenance of such globally unique rules for each application requires extensive expertise. On top of that, applications may be configured differently during installation and usage, so the same exact versions may look quite dissimilar on different machines, even in the same environment. Rules are also at risk of becoming incorrect as soon as the application's new version, or an altogether different application, happens to include the same files.

Existing rules may therefore occasionally have to be changed to take all this into account and Inferapp has the capacity to make the necessary corrections so they are transparent to our library's users. It is our job to maintain discovery rules to meet the above challenges and produce accurate inventory results.

inferapp

We hope this paper has answered some of your initial questions regarding Inferapp's Software Discovery Library.

For more information, please contact us at **info@inferapp.com**