Database Discovery: Identifying Hidden Risks and Sensitive Data
Table of Contents

Foreword .................................................................................................................. 3

Legacy Approaches to Database Discovery ............................................................ 4
  Environment Controls ......................................................................................... 4
  Intrusive Network Scanning ............................................................................... 4
  Intrusive Scanning Limitations ............................................................................ 5

New Non-Intrusive Approach to Database Discovery ............................................. 5
  DBN-6300 Database Discovery ............................................................................. 6

Database Software License Management .............................................................. 8
  DBN-6300 Database Discovery Reporting .......................................................... 8

Database Continuous Monitoring .......................................................................... 9

Conclusion ............................................................................................................. 9
Foreword

Knowing where your most valuable data assets are located is an essential first step in preventing data loss. Most security practices begin with a discovery phase to identify the organizations personal data, financial data, and intellectual property data. This will ensure that the implement security practices will protect these specific assets. As an information asset residing at the core of every critical business process, databases are the principle location for high-value organizational data. Information security standards and privacy regulations, such as the EU’s General Data Privacy Regulation (GDPR) and Payment Card Industry (PCI) explicitly specify the requirement for a comprehensive data discovery phase.

Keeping track of databases, and the associated data they host, turns out to be a significant challenge. Most organizations lack the necessary tools to track the location of personal and sensitive data repositories or even how they’re being used. The issue is far larger than a single or even a periodic initiative to create a personal data inventory. Given the dynamic nature of organizational data, the applications that process it, where its stored, and when it should be destroyed is constantly in flux and needs to be tracked on a continuous basis. Further, it is necessary to focus beyond just knowing where the databases are; it is also necessary to map and understand who/what is using those databases, and how. Without this complete, continuously updated understanding, it is not possible to protect the most critical information assets.
Legacy Approaches to Database Discovery

There are several legacy approaches used to populate the Configuration Management Database (CMDB) – we however won’t address the “word of mouth” approach used by some organizations as its shortcomings are obvious.

Environment Controls

The first of these legacy approaches is built on environmental controls. That is, every installation of a server and software is strictly controlled, and as, such no new databases can be installed without a set of authorizations. With this approach the CMDB of database installations is generally well understood and properly documented. The issue is that over time such controls inevitably break down and fail to address a special circumstance or a short-term urgency. In addition, this legacy approach only controls the installation of new database servers. Once the database server is handed over to a Database Administrator (DBA), the actual databases, stored data, and the access allowed to these data is outside the purview of security policy. Even when there are processes in place to control existing databases, short of periodically logging into each database server, with administrative permissions, and examining all databases and users configured, there would be no checks and balances to ensure the strict security policies are actually being followed. Finally, there’s no indication of where the database accesses are originating, and short of scanning the audit logs, the organization is missing critical information that could indicate a breach.

Intrusive Network Scanning

The second and far more common legacy approach for database discovery is intrusive network scanning. This is usually accomplished with a tool such as NMAP (Network Mapper). NMAP scans the network to identify the servers that respond to specific TCP ports. This technique, while better than a “word of mouth” approach or using environmental controls only, still has significant limitations. For example, scanning an entire IP address space with up to 65,535 TCP ports per IP address is a monumental task that will take a long period of time. It also generally requires either bypassing firewall rules intended to separate the network segments or installing scanning applications on each network segment. This form of intrusive scanning may also set off security alerts throughout the organization because it can appear to some security devices as if the organization is under an attack.

To improve efficiency, the number of ports scanned can be limited to a smaller subset. This is often done even, though it may not fully address the firewall
issue since even traditional database ports are often blocked by firewall rules. Worse yet is the assumption that databases operate only on their traditional, well-known TCP ports. That may not actually be the case. Both sanctioned and unsanctioned databases will often run on atypical TCP ports. Some organizations use TCP port numbers to identify a class of databases to simplify firewall rules. Other organizations use atypical TCP ports to separate production databases from development databases. Still other organizations may change from well-known TCP ports in an attempt to thwart hackers (who also often use NMAP tools).

**Intrusive Scanning Limitations**
Identifying database servers through intrusive scanning has serious limitations. Once the database server – located on a specific TCP port and IP address - is identified, the next task is to figure out what databases are hosted there. This means gaining access to the database. In some situations this will be straightforward because the server administrator will be known and they can be contacted. Assuming the administrator has the proper credentials it is possible to start probing the server to see what databases and users are configured. In most cases tracking done by the owner, gaining access to the server and probing it to acquire the necessary information is a daunting task.

Both the legacy database discovery methods described above (sometimes in combination) are used by many organizations to create a picture of the environment. The cost/complexity of these approaches are extensive in even a moderately sized environment and often database discovery is performed so infrequently that there is rarely ever a point in time where an accurate real-time picture of the database environment exists. In some organizations the attempt to do this “continuously” is often likened to painting a large bridge – by the time you finish painting the bridge, the bridge needs to be painted again. As a result there is no point in time where the organization has an accurate picture of the entire database environment.

**New Non-Intrusive Approach to Database Discovery**
The modern approach to database discovery is based on non-intrusive network monitoring with deep protocol analysis. DB Networks has pioneered such a database discovery approach with the DBN-6300. Database traffic is decoded directly from a copy of network traffic. With this method the DBN-6300 not only discovers databases but also identifies the applications and users accessing those databases and what they are doing. This approach goes well beyond simply identifying database servers. It’s able to discover the underlying databases as described by the database service name/id used to access them.
Therefore far greater insight into the database environment is now possible.

**DBN-6300 Database Discovery**

When the DBN-6300, with up to six network traffic capture interfaces, is connected to a primary network traffic concentration point (figure 1) it can immediately begin to gather and analyze database traffic.

The DBN-6300 does not rely on the TCP port numbers to identify database traffic. Rather it decodes all the packets to determine which connections are carrying database traffic regardless of the TCP port it’s assigned to. The DBN-6300 then extracts the service name, login credentials, SQL statements and RPCs being used to communicate with the database. The information is recorded for each database and it’s associated with the connecting client or application. This provides a comprehensive picture of the interactions between the database and those individuals and applications accessing its data.

Once deployed the DBN-6300 quickly builds a complete picture of the database environment automatically. This includes a list of all databases, the clients that access the databases, the SQL dialect used to interact with them, and the volume of access/activity those clients execute. This information is available interactively within the product’s user interface. A visual depiction of database to client connectivity is shown in figure 2. This map, which represents the high
level view of the database/client connectivity can be drilled into to reveal specific details on any particular interaction. This information is available in tabular form as shown in figure 3. Further, this information is available in downloadable form for use in external analytics, Personal Data Inventory, or for CMDB reconciliation.

![Figure 2](image)

**Figure 2**

![Figure 3](image)

**Figure 3**

Using this detailed information provides a number of extremely valuable benefits. The ability to identify all databases quickly ensures that only those
databases with a legitimate purpose and whose operation adheres to business/security policies are operational. Discovered databases operating outside of this realm can then be easily identified, and the organization can begin to address them accordingly. Using the discovered database IP address it can be immediately determined whether the database resides in a security domain consistent with the sensitivity of the data it hosts. The same holds true for the discovered clients and applications accessing the databases. This, capability combined with complexity (based on the diversity of SQL statements) and execution rates, further aids in the understanding and prioritization of potential threats. While the DNB-6300 goes way beyond these capabilities with its Application Breach Detection, Insider Threat Detection, and Compromised Credential Detection capabilities, database discovery provides a substantial benefit to any organization’s security posture.

Using these capabilities greatly simplifies a range of both security and operations tasks. Ensuring information assets within a PCI environment, for example, are properly segmented from other areas of the organization is a straightforward task. Identifying lateral breaches from compromised servers and accesses to secure database assets from low security network segments becomes very obvious to security personnel.

Database Software License Management
Another benefit of database discovery is that it ensures the organization’s CMDB data is current. This enables better management of database assets and database software licenses – which can have substantial budgetary impact. There’s often an organizational disconnect between the individuals paying the annual database software licensing expenses and those tasked with managing database assets. In many cases there’s an opportunity for IT expense reduction. Using traffic characterization, underutilized databases servers can be identified and targeted for consolidation, further improving cost efficiency (this not only addresses the database licensing issue, but rack space, power, air conditioning, maintenance, etc.). For any organization about to embark on a datacenter consolidation project, database discovery in a critical first step.

DBN-6300 Database Discovery Reporting
There are a variety of methods to interactively investigate the information from the DBN-6300 database discovery capability. Database discovery comes with an analytics capability that analyzes the data collected and identifies a number of key security risks that may be present. See figure 4. When necessary the data can also be easily exported.
Reports from the DBN-6300 provide both a summary as well as detailed analysis of the key security risks with the corresponding risk level along with the supporting data. These risks can be further explored using the interactive interface, which allows users to drill all the way down to the specific SQL statement from any client or server, providing an unprecedented level of insight and full understanding of the risks within the organizations database environment.

**Database Continuous Monitoring**

The DBN-6300 continuously monitors the network to identify changes to the environment on a real-time basis. For example, newly created database servers, database services, or clients each result in an event being generated indicating the discovery. Any appearance of a rogue database results in an immediate alert that is sent to the SOC. Further, a client or application accessing a database for the first time will also generate an alert. Through these mechanisms an up to date view of the state of the database environment is readily available. Additionally, it is possible to configure views like a network map to indicate recent changes to the connectivity, thereby creating a Network Operations Center display that can be used to actively track and analyze changes in the database network.

**Conclusion**

Regulations, such as the GDPR, require organizations know where the personal data they process is stored. Also, given the importance of the proprietary and
financial data contained in today’s enterprise databases it is critical to have a complete and up to date understanding of the database infrastructure. Legacy tools available to create and maintain a view of both the databases and applications / users that connect to them are resource intensive, highly invasive, and of limited effectiveness. Legacy tools result in an out of date and incomplete view of the organizations databases. Legacy tools offer little to no insight into who is using the databases nor in what way they are being used. DB Networks’ DBN-6300 has simplified the process of building a comprehensive and continuous inventory of databases, the clients that connect to them, and the activities those clients perform. This reduces the efforts associated with managing the trust boundaries around sensitive data. The ability to non-intrusively build a complete picture of the database environment and continuously monitor it for changes vastly improves the effectiveness of every security operation. Further, the characterization of each database’s activity provides substantial opportunities to reduce database software licensing expenses and server complexity by identifying underutilized databases that could be consolidated.

Learn more
To find out more about how to gain deep visibility into your database networks and improve your database security, contact DB Networks at +1-800-598-0450 or email info@dbnetworks.com.

DB Networks®
15015 Avenue of Science
Suite 150
San Diego, CA 92128
Tel: +1-800-598-0450
www.dbnetworks.com