



# Google Search Appliance – Database Crawler Feature Snippet



## Business Overview

The new Google Search Appliance Database Crawler provides access to the valuable information stored inside company databases and delivers comprehensive, unified search across structured and unstructured data. With the database crawler, you can provide users access to key corporate data housed in document management systems, enterprise applications, or custom database systems. Any content stored in standard enterprise relational databases can be easily searched by users.

## Acquiring Content

The Google Search Appliance Database Crawler uses the industry standard Structured Query Language (SQL) to define which information should be extracted from the database. The search administrator first defines the connection information including the database hostname and port, the database name, and access credentials (username and password.) to setup a connection to the desired database. The administrator then defines the Crawl Query, a SQL statement which retrieves the desired records from the database.

## Example:

For an “employee” database with these fields:  
employee\_id, first\_name, last\_name, email, dept

The crawl query might look like this:  
SELECT employee\_id, first\_name, last\_name, email, dept  
FROM employee

This will retrieve all of the employee records from the database that match the crawl query. The records are pulled from the database onto the search appliance, where they can be indexed. The incremental crawl query allows the administrator to define a SQL statement that targets insertions, updates, and deletions in the database to keep the searchable content fresh and up to date. Administrators can select exactly which fields are crawled and it is very straightforward to exclude fields from being crawled or served in result sets to users.

## Indexing Records

The crawl query, as defined above, generates a recordset and the resultant database records are brought onto the search appliance for indexing. The search appliance converts each database record into an HTML document by applying an XSLT stylesheet. The administrator can choose to use the default stylesheet, which generates a simple name / value paired table for each record, or upload a custom stylesheet to format the information as desired. Through the use of industry standard XSLT, the administrator can format the database record as best suited for search users, including generating meta tags for special and restricted searches.

---

## SPECIFICATIONS

---

### Supported Databases:

IBM DB2  
Microsoft SQLServer  
MySQL  
Oracle  
Sybase

---

### Google Search Appliance

Hardware: v4.0 or greater  
Software: v4.2 or greater

---

## ORDERING INFORMATION

---

[www.google.com/enterprise](http://www.google.com/enterprise)

Email [appliance1@google.com](mailto:appliance1@google.com)

---

## Serving Results

Now that the database content is indexed, it is available for users to search against. Because the Google Search Appliance maintains one unified index of all enterprise content, users are not forced to search a separate interface to retrieve results that live in a corporate database. Search administrators can provide search to a single dataset, or all enterprise content with the use of unlimited collections of the Google Search Appliance.

The search administrator can choose how results are displayed when a user performs a search on the Google Search Appliance and selects a database result. The administrator can choose to set a serve query, which will retrieve the most up-to-date record for a matching result, and apply the previously defined stylesheet to the record for display to the user.

### Example:

Serve query:

```
SELECT employee_id, first_name, last_name, email, dept
FROM employee
WHERE employee_id = ?
```

The Primary Key Field for this example is: employee\_id

Alternatively, the search administrator could make use of a pre-existing web-based version of the database record by defining the Serve URL Field. The serve URL Field instructs the appliance to use the value of one of the fields previously defined in the Crawl Query that contains the URL that should be displayed. For example, in a company directory, if an HTML page exists for each record, and the links are always in the same format (such as [http://corp.company.com/hr/Joe\\_Employee.html](http://corp.company.com/hr/Joe_Employee.html)), then the appliance displays that link when it serves results.

### Supported Databases:

The Google Search Appliance Database Crawler supports IBM DB2, Microsoft SQLServer, MySQL, Oracle, and Sybase. The Google Search Appliance uses native JDBC database drivers supplied by the database manufacturers. These database drivers and versions are supported:

- MySQL: JDBC version 3.0
- Oracle: Oracle JDBC Driver version - 10.1.0.2.0
- Sybase: jConnect (TM) for JDBC(TM)/5.5(Build 25137)/P/EBF10000/JDK12/Sun Jan 6 21:57:04 2002
- Microsoft SQLServer: JDBC version 2.2
- IBM DB2: JDBC version 8.1