Getting Started With

Hands-on exercises included

IBM Data Studio for DB2

Ideal for application developers and administrators

by:
Debra Eaton
Vitor Rodrigues
Manoj K. Sardana
Michael Schenker
Kathryn Zeidenstein
Raul F. Chong

DB2 ON CAMPUS BOOK SERIES

First Edition (December 2009)

Second printing (September 2010)

© Copyright IBM Corporation 2009, 2010. All rights reserved.

IBM Canada
8200 Warden Avenue
Markham, ON
L6G 1C7
Canada

The licensed program described in this document and all licensed material available for it are provided by IBM under terms of the IBM Customer Agreement, IBM International Program License Agreement or any equivalent agreement between us.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

All statements regarding IBM's future direction or intent are subject to change or withdrawal without notice, and represent goals and objectives only.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs. The sample programs are provided "AS IS", without warranty of any kind. IBM shall not be liable for any damages arising out of your use of the sample programs.

References in this publication to IBM products or services do not imply that IBM intends to make them available in all countries in which IBM operates.

If you are viewing this information softcopy, the photographs and color illustrations may not appear.

About the authors

Debra Eaton is a software information technology specialist with IBM's Central North Technical Sales Team in Chicago. For 20 years she has worked with Fortune 500 companies on a variety of application development projects. Her specialty is database applications and their Integrated Development Environments. Debra has trained a variety of IBM DB2 customers in the area of Application Development, spoken about DB2 application development at International DB2 Users Group (IDUG) and the IBM DB2 Technical conference, and authored several redbooks, white papers and developerWorks tutorials. She can be reached via e-mail at deaton@us.ibm.com.

Vitor Rodrigues is a Software Engineer at the IBM Silicon Valley Lab working on Optim™ Development Studio and Data Studio products. Previously held positions include Technical Enablement and Quality Assurance roles in the Data Studio and DB2 pureXML organizations. Prior to joining IBM, Vitor graduated in Computer and Software Engineering from University of Minho, Portugal. He is an IBM Certified Solution Developer for XML and Related Technologies and an IBM Certified Database Administrator - DB2 9 DBA for Linux, UNIX and Windows. Vitor has co-authored several articles and tutorials for developerWorks.

Manoj K. Sardana is a staff software engineer working with IBM India software labs. He holds a bachelor’s degree in Computer Science from NITK Surathkal, India. He has worked on various projects within the DB2 team and at the time of this publication is working with the pureQuery development team. He has previously worked on developing the sample application for the new features of DB2 and on functional verification testing for DB2. Manoj is an IBM certified application developer and advance database administrator for DB2 V9. He is also an IBM certified solution developer for XML and related technologies. Manoj likes technical writing and teaching and has presented at various conferences and published articles. In his free time he likes to play with kids and listen to music.

Michael Schenker is a software engineer at IBM's Silicon Valley Laboratory in San Jose, Calif. He joined IBM in 2002 and works in the IBM Data Server Tooling area. His subject of expertise is the Web service enablement of IBM's data servers. He holds a master's degree in computer sciences from the University of Applied Sciences in Leipzig, Germany.

Kathryn Zeidenstein is a member of the Data Studio and Optim Solutions technical enablement team and has responsibility for community building and communications with the technical community. She has many years of experience with IBM starting out as an Information Developer for DB2 for z/OS®, managing the SQL Standards team, managing editor for the Information Management zone on developerWorks® and as product manager and marketing manager for several Information Management products. She has authored or co-authored numerous articles on developerWorks and in other publications. She holds a master’s degree in Professional Writing from Illinois State University.

Raul F. Chong is the DB2 on Campus program manager based at the IBM Toronto Laboratory, and a DB2 technical evangelist. His main responsibility is to grow the DB2 community around the world, helping members interact with one another, and contributing

Contributors

The following people edited, reviewed, provided content, and contributed significantly to this book.

Contributor	Company/University	Position/Occupation	Contribution
Agatha Colangelo	YCDSB: Adult & Continuing Education	Instructor. DB2 on Campus Community President	Editing and reviewing.
Tina Chen	IBM, Silicon Valley Laboratory	Data Studio Solution Architect	Sample database, reviewing, guidance.
Clifford Chu	IBM, Silicon Valley Laboratory	Lead developer, routine tooling	Review and guidance.
Joseph Fontana	IBM Silicon Valley Laboratory and Northern Illinois University	Intern, Optim Solutions Technical Enablement.	Testing and review.
Philip Gunning	Gunning Technology Solutions, LLC	Principal Consultant	Review.
Holly Hayes	IBM, Silicon Valley Laboratory	Integrated Data Management solutions evangelist	Review and guidance
Jayashree Ramachandran	IBM, India Laboratory, Bangalore	Software Engineer, Optim Database Administrator	Review and contributions to Chapter 2.
Marcos Ramirez	IBM Silicon Valley Laboratory and San Jose State University	Intern, Optim Solutions Technical Enablement.	Testing and review.
Thomas Sharp	IBM, Silicon Valley Laboratory	Architect, Routine tooling	Technical and editorial review.

9. What is the name of the Eclipse view used to browse of the projects that hold SQL scripts, Data Web Services artifacts, and stored procedures?
   A. Thin Client
   B. Data Source Explorer
   C. Data Project Explorer
   D. Outline
   E. None of the above
10. In which Eclipse view do results of SQL operations appear?
   A. Data Source Explorer
   B. Properties
   C. Data Project Explorer
   D. Editor
   E. None of the above

Exercise 4: Try adding a user and see how many privileges you can give. Browse through all the possible privileges.

Exercise 5: Create a table where the value for a specific column will always be generated by the DB2 based on an expression defined by you.

2.10 Summary

In this chapter you have learned about instances, how to create a database, get connected to it, and create a connection profile. You have also learned how to create different object once you are connected to the database. At the end you have learned how to add new users and give them the privileges to access the different database objects.

2.11 Review questions

1. What are the two options for layout of the Data Source Explorer? Briefly describe each one.
2. How can you generate an Entity-Relationship diagram in Data Studio?
3. Related database objects are grouped together in a _____.
4. Why are connection profiles useful?
5. When creating a new user in Data Studio, which tab in the object editor enables you to specify which objects that person has access to?
6. When connecting to a database, which of these is a mandatory parameter?
   A. Port number
   B. Hostname/IP Address
   C. User name/password
   D. All of the above
   E. None of the above
7. Which editor can be used to alter table properties?
   A. SQL Editor
   B. Data Object Editor
   C. Routine Editor
   D. Database table editor
   E. All of the above
8. Which of these objects is not associated with a schema?

A. Table

B. View

C. Index

D. Table space

E. Column
Hint: Browse through the database tree and see what is not included under the Schema folder

9. While creating a table, when is an index automatically created?

A. When you define the primary key

B. When you define a NOT NULL constraint for the column

C. When the column value is defined as auto-generated

D. No index gets created automatically

E. All of the above

10. You can create a view using:

A. A full select on a table

B. By selecting a list of columns from the table

C. By joining mutliple tables

D. All of the above

E. None of the above

2. What are the file formats supported for exporting data?
3. Which file format allows you to create the table while exporting?
4. Name the two different logging methods supported in DB2 data serveres.
5. What two operations are combined to compose the Recover operation?
6. A system-managed tablespace can have containers of the following type:
   A. Directory
   B. File
   C. Raw Device
   D. All of the above
   E. None of the above
7. A buffer pool is used mainly
   A. To store the intermediate results of queries
   B. To fetch data from disk to volatile memory/RAM
   C. To store updates to the data
   D. To process the data before returning it to the application
   E. All of the above
8. An online table reorganization is preferred when
   A. Access to database tables is critical while reorg is occurring
   B. Data needs to be clustered perfectly
   C. More storage space is required
   D. Index rebuild is required
   E. All of the above
9. When is statistics collection recommended?
   A. When the data is fragmented
   B. After many selects on the data
   C. After many updates on the data
   D. All of the above
   E. None of the above
10. An incremental backup backs up:
   A. All the data
   B. Only the modified data after the last full backup

C. Only the modified data after the full or incremental backup

D. Only the modified data after any kind of backup

E. None of the above

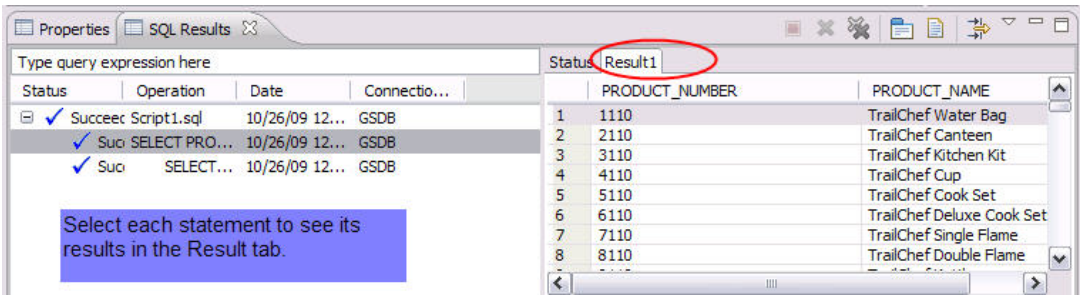

Figure 4.20- Results of multiple SQL statements in a single script

4.4 Summary

In this chapter we described some basic tasks that developers will use frequently, including creating a Data Development project, building queries, and creating SQL and XQuery scripts. With this fundamental background, you are now prepared to do more advanced tasks, such as creating stored procedures, user-defined routines, and Data Web Services, described in subsequent chapters.

4.5 Review questions

1. What is the difference between the Data Project Explorer view and the Data Source Explorer views?
2. List and describe the DB2 application process settings that you can configure when creating a Data Development project.
3. Describe some of the different launch points Data Studio provides you to run SQL scripts.
4. What are the major features that differentiate the SQL builder and SQL editor components?
5. What type of XML artifacts can you have in a Data Development Project?
6. The Data Project Explorer is part of which Data Studio perspective?

    A. Database Development perspective

    B. Data perspective

    C. Database Debug perspective

    D. Resource

    E. None of the above.

7. Which objects can be created in a Data Development perspective?

    A. SQL scripts, SQL stored procedures, COBOL stored procedures, Web services

B. SQL scripts, SQL stored procedures, Java stored procedures, COBOL stored procedures

C. SQL scripts, SQL stored procedures, Java stored procedures, Data Web Services

D. SQL stored procedures, Java stored procedures, COBOL stored procedures, Data Web Services

E. All of the above

8. Which tool or tools can be used to create SQL scripts?

   A. SQL Editor

   B. SQL and XQuery Editor, SQL Query Builder

   C. Database Object Editor, SQL Query Builder

   D. Routine Editor, SQL and XQuery Editor

   E. None of the above

9. Which of the following represents all types of SQL statements that can be created using the SQL Query Builder?

   A. SELECT, INSERT, UPDATE, DELETE

   B. SELECT, INSERT, UPDATE, JOIN, FULLSELECT, WITH

   C. SELECT, INSERT, UPDATE, DELETE, FULLSELECT, WITH

   D. SELECT, INSERT, UPDATE, DELETE, JOIN, FULLSELECT, WITH

   E. None of the above.

10. In which Data Studio view can you see the results of executing your SQL statements?

    A. Data Source Explorer

    B. Project Explorer

    C. SQL Outline

    D. SQL Results

    E. None of the above

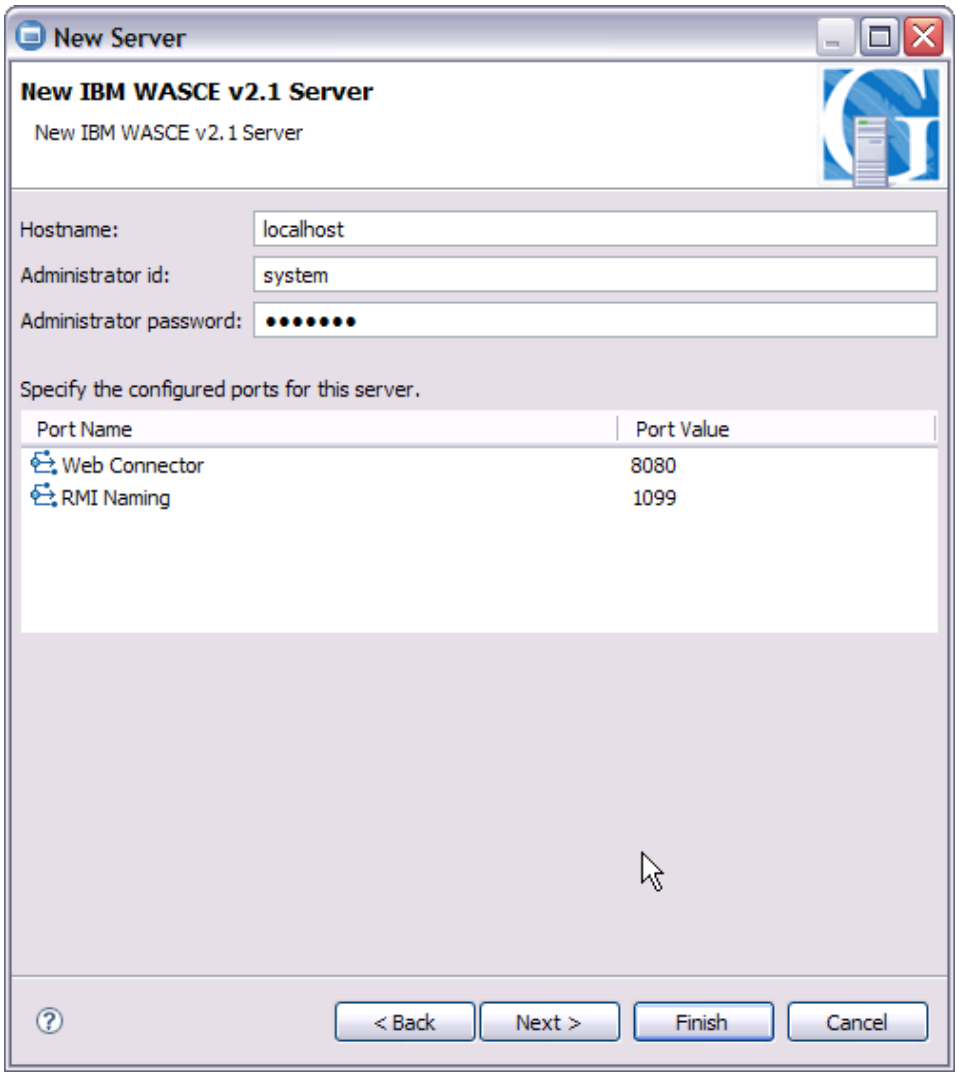

Figure 6.7 – Configuring the server connectivity information

The Administrator ID and Administrator Password are the credentials of the WAS CE admin user. By default, the Administrator ID is “system” and the password is “manager”. You might need the Administrator ID and Administrator Password at a later time when you try to log on to the Administration console from within or outside of Data Studio.

The Web Connector defines the TC/IP port for the HTTP protocol, which, by default, is **8080**.

The Remote Method Invocation (RMI) Naming defines the port that is used by Data Studio to perform administrative tasks at the application server. By default, this port

5. The approach of creating a Data Web Service based on existing database logic is called _____ development.
6. You create a new Data Web Service in:
   A. A Data Design project
   B. A Data Development project
   C. The Data Source Explorer
   D. SQL and XQuery Editor
   E. None of the above
7. Business logic for a Data Web Service can be provided by:
   A. SQL procedures
   B. XQuery statements
   C. SQL statements
   D. All of the above
   E. None of the above
8. Which transport protocol is used with a Data Web Service?
   A. FTP
   B. RMI
   C. HTTP
   D. SMTP
   E. None of the above
9. What is the Web Services Explorer used for?
   A. Browsing the Web
   B. Editing XML files
   C. Testing Web services
   D. Browsing the file system on a remote server
   E. All of the above
10. What are the three major steps in the development of a Data Web Service?
   A. Design, develop, deploy
   B. Create, deploy, test
   C. Model, develop, test
   D. Design, model, deploy

E. None of the above

7

Chapter 7 – Developing user-defined functions

In this chapter you will learn how to develop user-defined functions (UDFs) using Data Studio.

7.1 Developing user-defined functions: The big picture

Like stored procedures, UDFs encapsulate reusable code. Because UDFs can be used in SQL statements, it lets you extend the SQL language with your own logic. For example, you might want to create a function that encapsulates the logic to calculate a tax in your country when a value is given as input, or you can create a function that extracts information from an XML document and returns it in a tabular form that can be used to perform a JOIN with another table.

As you can see, UDFs provide a very flexible way to extend your application.

For scalar UDFs developed in Data Studio, the supported languages are: SQL and PL/SQL. For table UDFs, the supported languages are: SQL and OLE DB. The DB2 server supports other languages for UDFs, such as Java and C++, but those are not supported in Data Studio tooling.

Note:

Although Data Studio tooling lets you create PL/SQL functions for DB2 projects, PL/SQL support is not available in DB2 Express-C, which is the edition of DB2 used for the examples in this book. If you wish to develop PL/SQL functions, you'll need to upgrade to a different edition of DB2 that contains PL/SQL support.

function, but the development of scalar functions in Data Studio follows the exact same steps.

7.5 Exercise

As an exercise for this chapter, create a table UDF that returns the name and schema of all functions that have the qualifier equal to the value passed as a parameter to the function.

7.6 Review questions

1. Explain some of the advantages of using UDFs for database application development.
2. What is the difference between a scalar UDF and a table UDF?
3. Describe a scenario where you would use a UDF instead of a stored procedure
4. Describe a scenario where you would use a UDF instead of plain SQL statements.
5. What is an aggregate UDF?
6. What languages are supported for user-defined function development in a Data Development project associated with a DB2 Linux, UNIX and Windows connection?
   A. SQL, PL/SQL
   B. SQL, OLE DB, PL/SQL
   C. SQL, Java, OLE DB, PL/SQL
   D. SQL, OLE DB
   E. None of the above
7. What result type or types are supported for SQL user-defined functions in Data Studio?
   A. scalar, list
   B. table, list
   C. scalar, table
   D. scalar, table, list
   E. None of the above
8. Which editor can be used to edit user-defined functions in Data Studio?
   A. SQL and XQuery Editor
   B. Data Object Editor
   C. Routine Editor
   D. Database Table Editor

E. All of the above

9. What type of statement or statements can make up the body of a user-defined function?

   A. SQL statement

   B. SQL statement, SQL expression

   C. SQL expression

   D. SQL expression, regular expression

   E. All of the above

10. Where can you see the results of running a UDF?

    A. Console

    B. SQL editor

    C. SQL Results View

    D. Data Source Explorer

    E. None of the above

8.2.5 Summary of capabilities

The following table lists product capabilities and whether they are included with the Optim products described in this section. For more detail, you'll want to look at the Features and Benefits section of each product Web page.

Function	Data Studio	Optim Develop-ment Studio	Optim Database Administr-ator	InfoSphere Data Architect	Optim Query Tuner
Create overview diagram	Yes	Yes	Yes	Yes	Yes
Create and view physical models	No	Yes	Yes	Yes	No
Compare and sync physical models	No	No	Yes	Yes	No
Create logical models, compare and sync logical models. Create privacy attributes. Create glossary models. Volumetric modeling for capacity planning.	No	No	No	Yes	No
Basic database object management (create, drop, alter)	Yes	Yes	Yes	No	No
Command support, and task assistants for utilities	Yes	No	Yes	No	No
Script creation, dependency analysis for complex schema changes	No	No	Yes	No	No

Function	Data Studio	Optim Develop-ment Studio	Optim Database Administr-ator	InfoSphere Data Architect	Optim Query Tuner
Basic data development (SQL procedures and UDFs query building)	Yes	Yes	Yes	Yes	Yes
Data Web Services development*	Yes	Yes	No	No	No
XML development*	Yes	Yes	No	Yes	No
Visual Explain (visual diagram of data access path)	Yes	Yes	Yes	Yes	Yes
Advanced Java database development: pureQuery APIs, advanced tooling for dependency analysis and hot spot analysis	No	Yes	No	No	No
Advanced Web services (deploy on WebSphere DataPower, JMS, support)	No	Yes	No	No	No
Query tuning advisors, tools and configuration.	No**	No	No	No	Yes

Table 8.1 – Capabilities in some Optim products.

* Indicates that capability is not included in Data Studio stand-alone package.

** As of Version 2.2.0.1, Data Studio stand-alone does include single-query index advisor.

B. InfoSphere Data Architect

C. Optim Database Administrator

D. Optim Development Studio

E. None of the above

1. The five phases of the data lifecycle are: design, develop, deploy, operate, optimize and govern. Governance is the aspect that needs to be considered across all phases of the lifecycle.
2. The design phase is most is most concerned with translating business requirements into a physical database representation? The main product for doing this is InfoSphere Data Architect, perhaps in conjunction with other modeling tools such as Rational Software Architect.
3. The develop lifecycle is when data access layers, database routines, and data services created, tested, and tuned. The primary products associated with this phase are Optim Development Studio, Optim Query Tuner and Optim Test Data Management.
4. Optim Database Administrator contains change management capabilities.
5. Optim Query Tuner helps DBAs and developers improve SQL queries so that they can perform faster.
6. The answer is C. Although metadata exchange is a key implementation approach to integrated tooling, it is not a goal of using the integrated data management approach. The goals are to reduce risk, improve collaboration among roles, and to improve efficiency of development and deployment.
7. The answer is E. All of the listed products include the ability to create and debug SQL procedures.
8. The answer is A, B, and D.
9. The answer is D. Although Java data access should be developed with efficiency and performance in mind, the optimize phase of the lifecycle generally reflects activities around optimizing existing applications and resources.
11. The answer is D. Optim Development Studio provides developers with capabilities that help them create high performance data access layers for Java applications

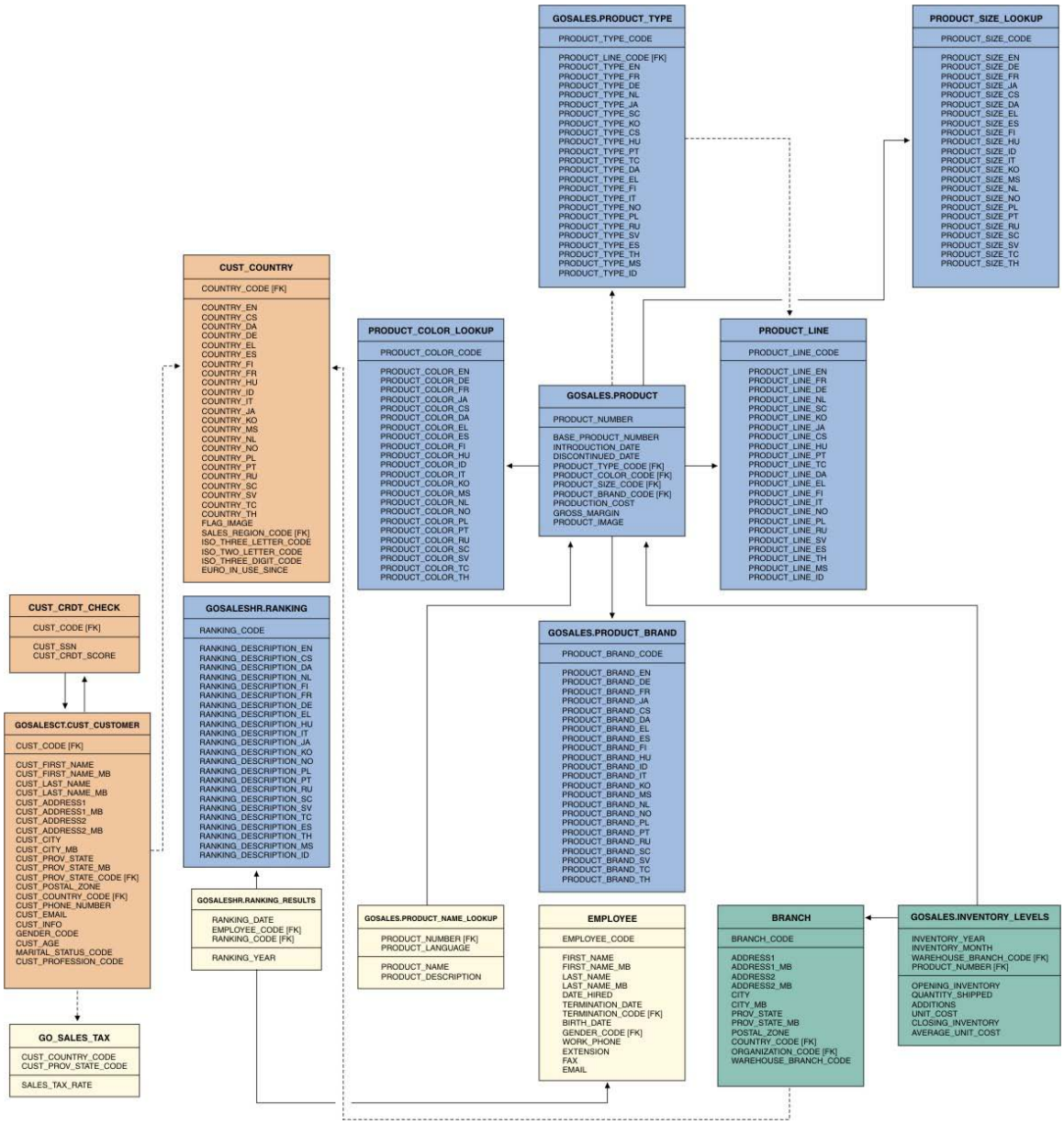

Figure D.1- Great Outdoors data model

D.2 Table descriptions

D.2.1 GOSALES schema

The GOSALES schema includes information about products and inventory.

D.2.1.1 GOSALES.BRANCH table

Row count: 29

The BRANCH table contains address information of each branch. Each branch has a collection of employees with different roles, including sales representatives operating from a regional base.

Not all branches have warehouses. The warehouse branch code is a repeating value of the branch code, identifying the regions covered by a particular warehouse.

D.2.1.2 GOSALES.INVENTORY_LEVELS table

Row count: 53730

This table shows inventory for all warehouses. Only 11 of the 29 branches have warehouses that maintain inventory.

D.2.1.3 GOSALES.PRODUCT table

Row count: 274

The company supplies sport gear for camping, climbing, and golfing. There are five product lines, further subdivided into 21 product types. There are a total of 144 unique products, or 274 products when including color and size.

D.2.1.4 GOSALES.PRODUCT_BRAND table

Row count: 28

Products of the same brand are associated by a style or price point.

D.2.1.5 GOSALES.PRODUCT_COLOR_LOOKUP table

Row count: 27

Product colors provide analysis by attribute. GO Accessories is the richest data source for attribute analysis including color and size.

D.2.1.6 GOSALES.PRODUCT_LINE table

Row count: 5

Row count: 900

The customer credit check table contains the credit scores of consumers that make online purchases.

D.2.2.3 GOSALESCT.CUST_CUSTOMER table

Row count: 31255

The customer table contains the name, address, and contact information of each customer. All customers in this table are online shoppers paying the retail price for items sold by the company or one of its partners.

D.2.2.4 GOSALESCT.GO_SALES_TAX table

Row count: 94

The Great Outdoors sales tax table contains sales tax rates at a country level, or state level if applicable.

Tax rates are for example only.

D.2.3 GOSALESHR schema

The GOSALESHR schema includes information used by the Great Outdoors company Human Resources department.

D.2.3.1 GOSALESHR.EMPLOYEE table

Row count: 766

The employee table contains the static information that repeats for each detail in the employee history table.

D.2.3.2 GOSALESHR.RANKING table

Row count: 5

The ranking dimension contains text descriptions of an employee's ranking. Ranking is done annually and is one of the following values:

Poor

Satisfactory

Good

Very good

Excellent

D.2.3.3 GOSALESHR.RANKING_RESULTS table

Row count: 1898

This fact table maintains ranking data for each employee. Rankings are published in the month of March based on the previous year.