

# Pivotal

A NEW PLATFORM FOR A NEW ERA

# Greenplum Architecture, Administration, and Implementation

Pivotal

© 2015 Pivotal Software, Inc. All rights reserved.

2

Welcome to the Greenplum Architecture, Administration, and Implementation.

Copyright © 1996, 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015 EMC Corporation. All Rights Reserved. EMC believes the information in this publication is accurate as of its publication date. The information is subject to change without notice.

THE INFORMATION IN THIS PUBLICATION IS PROVIDED “AS IS.” EMC CORPORATION MAKES NO REPRESENTATIONS OR WARRANTIES OF ANY KIND WITH RESPECT TO THE INFORMATION IN THIS PUBLICATION, AND SPECIFICALLY DISCLAIMS IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Use, copying, and distribution of any EMC software described in this publication requires an applicable software license.

EMC2, EMC, Data Domain, RSA, EMC Centera, EMC ControlCenter, EMC LifeLine, EMC OnCourse, EMC Proven, EMC Snap, EMC SourceOne, EMC Storage Administrator, Acartus, Access Logix, AdvantEdge, AlphaStor, ApplicationXtender, ArchiveXtender, Atmos, Authentica, Authentic Problems, Automated Resource Manager, AutoStart, AutoSwap, AVALONidm, Avamar, Captiva, Catalog Solution, C-Clip, Celerra, Celerra Replicator, Centera, CenterStage, CentraStar, ClaimPack, ClaimsEditor, CLARiiON, ClientPak, Codebook Correlation Technology, Common Information Model, Configuration Intelligence, Configuresoft, Connectrix, CopyCross, CopyPoint, Dantz, DatabaseXtender, Direct Matrix Architecture, DiskXtender, DiskXtender 2000, Document Sciences, Documentum, eInput, E-Lab, EmailXaminer, EmailXtender, Engenuity, eRoom, Event Explorer, FarPoint, FirstPass, FLARE, FormWare, Geosynchrony, Global File Virtualization, Graphic Visualization, Greenplum, HighRoad, HomeBase, InfoMover, Infoscape, Infra, InputAccel, InputAccel Express, Invista, Ionix, ISIS, Max Retriever, MediaStor, MirrorView, Navisphere, NetWorker, nLayers, OnAlert, OpenScale, PixTools, Powerlink, PowerPath, PowerSnap, QuickScan, Rainfinity, RepliCare, RepliStor, ResourcePak, Retrospect, RSA, the RSA logo, SafeLine, SAN Advisor, SAN Copy, SAN Manager, Smarts, SnapImage, SnapSure, SnapView, SRDF, StorageScope, SupportMate, SymmAPI, SymmEnabler, Symmetrix, Symmetrix DMX, Symmetrix VMAX, TimeFinder, UltraFlex, UltraPoint, UltraScale, Unisphere, VMAX, Vblock, Viewlets, Virtual Matrix, Virtual Matrix Architecture, Virtual Provisioning, VisualSAN, VisualSRM, Voyence, VPLEX, VSAM-Assist, WebXtender, xPpression, xPresso, YottaYotta, the EMC logo, and where information lives, are registered trademarks or trademarks of EMC Corporation in the United States and other countries.

All other trademarks used herein are the property of their respective owners.

© Copyright 2015 EMC Corporation. All rights reserved. Published in the USA.

Revision Date: May 2015

Revision Number: MR-1CN-GRNADM.4.3.4.



## Copyright

Copyright © 2016 Pivotal Software, Inc. All rights reserved. This manual and its accompanying materials are protected by U.S. and international copyright and intellectual property laws.

Pivotal products are covered by one or more patents listed at <http://www.pivotal.io/patents>.

Pivotal is a registered trademark or trademark of Pivotal Software, Inc. in the United States and/or other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies. The training material is provided “as is,” and all express or implied conditions, representations, and warranties, including any implied warranty of merchantability, fitness for a particular purpose or noninfringement, are disclaimed, even if Pivotal Software, Inc., has been advised of the possibility of such claims. This training material is designed to support an instructor-led training course and is intended to be used for reference purposes in conjunction with the instructor-led training course. The training material is not a standalone training tool. Use of the training material for self-study without class attendance is not recommended.

These materials and the computer programs to which it relates are the property of, and embody trade secrets and confidential information proprietary to, Pivotal Software, Inc., and may not be reproduced, copied, disclosed, transferred, adapted or modified without the express written approval of Pivotal Software, Inc.

## Course Overview

Description	The Greenplum Architecture, Administration, and Implementation course introduces students to the Greenplum environment, consisting of the Greenplum Database and supported systems. Students are introduced to fundamental concepts on data warehousing, business intelligence, and how Greenplum helps to solve business problems in managing and analyzing Big Data. Students will install, configure, and manage the Greenplum database system by evaluating logical models and business requirements to determine the best physical design for a Greenplum database.
Audience	The intended audience for this course includes Greenplum database implementers and designers, system administrators for the Greenplum environment, Greenplum database administrators, and database developers.
Prerequisites	To understand the content and successfully complete this course, a student must have an understanding of: <ul style="list-style-type: none"><li>• Basic UNIX or Linux commands</li><li>• SQL Syntax</li><li>• Fundamental relational database concepts</li></ul>

Pivotal

© 2015 Pivotal Software, Inc. All rights reserved.

### Course Overview

The Greenplum Architecture, Administration, and Implementation course provides you with the foundation you need to interact with, manage, and develop for your Greenplum environment, as an administrator, implementer, or developer. You will gain insight into the benefits that Greenplum offers to your organization in managing and analyzing your data.

Once you install and configure Greenplum, you will manage the data by evaluating logical models and business requirements to determine the best physical design for your database. You will consider specific performance impacts to retrieving your data based on how your data is stored within the database.

To successfully complete this course, you should have:

- A working knowledge of basic UNIX or Linux commands to help you navigate the Linux environment
- Some basic understanding of SQL and SQL syntax
- Fundamental relational database concepts

## Course Objectives

Upon completion of this course, you should be able to:

- Gain a basic understanding of data warehousing concepts and be able to describe PostgreSQL to develop a foundation for understanding the Greenplum solution.
- Learn about the Greenplum architecture and hardware solutions to support the architecture so that they understand how to properly implement a solution based on their company's business needs.
- Build and implement a Greenplum software solution to manage the Greenplum environment and database through a variety of Greenplum utility tools and PSQL.
- Learn PostgreSQL and Greenplum specific SQL features and functions to manage data and optimize SQL query performance in a Greenplum database.

Pivotal

© 2015 Pivotal Software, Inc. All rights reserved.

## Course Objectives

On completion of the course, you should be able to:

- Identify and explain data warehousing concepts and define what PostgreSQL is as well as its relationship to Greenplum
- Identify the major components within the Greenplum architecture and highlight the importance of certain hardware solutions that support the architecture
- Build and implement a Greenplum software solution in to manage the Greenplum database, database instances, and data using a variety of Greenplum utility tools and the PSQL client
- Use both PostgreSQL and Greenplum-specific SQL commands and functions to manage data and optimize SQL query performance in a Greenplum database

# Agenda

## Modules/Lessons

## Labs

Day 1	<ul style="list-style-type: none"> <li>Greenplum Fundamental Concepts               <ul style="list-style-type: none"> <li>The Basics of Data Warehousing</li> <li>Greenplum Concepts, Features, and Benefits</li> <li>Greenplum Architecture – Shared Nothing and MPP</li> <li>Greenplum Product Overview</li> </ul> </li> <li>Database Installation and Initialization               <ul style="list-style-type: none"> <li>Systems Preparation and Verification</li> <li>Greenplum Database Initialization</li> </ul> </li> <li>Greenplum Database Tools, Utilities, and Internals               <ul style="list-style-type: none"> <li>Using PSQL Client and Greenplum Utilities</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>Lab Preparation and Initialization</li> <li>Greenplum Product Overview</li> </ul>
	<ul style="list-style-type: none"> <li>Greenplum Database Tools, Utilities, and Internals (Cont)               <ul style="list-style-type: none"> <li>Pivotal Greenplum Command Center</li> <li>Greenplum Database Server Configuration</li> <li>Greenplum Database Internals</li> </ul> </li> <li>Defining and Securing the User Database               <ul style="list-style-type: none"> <li>Data Definition Language</li> <li>Data Manipulation Language and Data Query Language</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>Systems Preparation and Verification</li> <li>Pivotal Greenplum Database Initialization</li> <li>Using the PSQL Client and Greenplum Utilities</li> </ul>

Pivotal

© 2015 Pivotal Software, Inc. All rights reserved.

## Agenda

The course is a mixture of lecture and lab over a period of five days. The course is divided into eight modules which define specific high-level related concepts and activities for implementing and managing Greenplum. These modules are:

- **Greenplum Fundamental Concepts** – You will explore the concepts of data warehousing and how these concepts apply to Greenplum. You will then examine the features and benefits of implementing Greenplum in your environment. Finally, you will examine the concepts of shared nothing and MPP to understand how they help Greenplum to scale as well as it does.
- **Database Installation and Initialization** – A solid hardware solution greatly improves your Greenplum queries. You will take an in-depth look at the Greenplum product itself and examine these hardware solutions. You will prepare a system for Greenplum installation, perform the installation, and verify that all requirements are met before proceeding with initializing your Greenplum database. A variety of tools will allow you to connect to and manage the Greenplum environment. Once you learn how to use these tools, you will manage your Greenplum database instance.
- **Greenplum Database Tools, Utilities, and Internals** – Commonly used tools and utilities are examined in this module. In this module, you will examine and use utilities that provide access to the database and configure, monitor, and manage various aspects of the database.
- **Defining and Securing the User Database** – On installing the database, you must have a solid understanding of the database objects you will manage and the syntax available for managing these objects and your data. You will manage workload resources to protect not only data, but also the resources within the database environment. To do that, you will secure the environment by creating appropriate privileges and roles.

# Agenda

	Modules/Lessons	Labs
Day 3	<ul style="list-style-type: none"> <li>Defining and Securing the User Database (Cont)               <ul style="list-style-type: none"> <li>Roles, Privileges, and Resources</li> </ul> </li> <li>Data Loading and Distribution               <ul style="list-style-type: none"> <li>Implementing Table Storage Models, Compression, and Tablespaces</li> <li>Data Loading</li> <li>Table Partitioning</li> </ul> </li> <li>Database Management and Archiving               <ul style="list-style-type: none"> <li>Managing the Greenplum Database</li> <li>Backups and Restores</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>Roles, Privileges, and Resources</li> <li>Controlling Access</li> <li>Table Management</li> <li>Data Loading</li> <li>Table Partitioning</li> <li>Managing the Greenplum Database</li> <li>Backups and Restores</li> </ul>
Day 4	<ul style="list-style-type: none"> <li>Data Modeling and Design               <ul style="list-style-type: none"> <li>Data Modeling</li> <li>Physical Design Decisions</li> </ul> </li> <li>Performance Analysis and Tuning               <ul style="list-style-type: none"> <li>JOIN Tables – Types and Methods</li> <li>Database Tuning</li> <li>Query Profiling</li> <li>Explain the Explain Plan – Analyzing Queries</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>Data Modeling</li> <li>Physical Design Decisions</li> <li>Database Tuning</li> <li>Explain the Explain Plan – Analyzing Queries</li> </ul>

Pivotal

© 2015 Pivotal Software, Inc. All rights reserved.

## Agenda (continued)

- **Data Loading and Distribution** – Once the database has been created, you will load the data. You must make decisions on how the data will be distributed. You will therefore need to understand what table partitioning is and how it can aid you.
- **Database Management and Archive** – You will learn commands that will aid you in managing the database instance as well as creating data backups and restoring the data as necessary.
- **Data Modeling and Design** – While earlier modules examined data loading and some data distribution techniques, you will more closely examine how to model your data and make key physical design decisions which will impact how the data is distributed in your environment.
- **Performance Analysis and Tuning** – Another in-depth subject, this module takes a closer look at specific commands and methods that can impact how your data is retrieved. You look at methods of examining your queries so that you develop appropriate queries to maximize performance gained based on how your data is distributed. You examine how to improve statistical analysis of your data to improve the overall query performance. You also examine whether or not to index your tables, implement OLAP methods in your queries, create PostgreSQL functions, and apply specific performance tuning tips for your data and your queries.

# Agenda

	Modules/Lessons	Labs
Day 5	<ul style="list-style-type: none"><li>• Performance Analysis and Tuning (Cont)<ul style="list-style-type: none"><li>• Improve Performance with Statistics</li><li>• Indexing Strategies</li></ul></li><li>• Developing Reports Using Advanced SQL<ul style="list-style-type: none"><li>• Advanced Reporting Using OLAP</li><li>• PostgreSQL Functions</li><li>• Advanced SQL Topics and Performance Tips</li></ul></li></ul>	<ul style="list-style-type: none"><li>• Improve Performance with Statistics</li><li>• Indexing Strategies</li><li>• Advanced Reporting Using OLAP</li><li>• PostgreSQL Functions</li></ul>

Pivotal

© 2015 Pivotal Software, Inc. All rights reserved.

## Agenda (continued)

- **Developing Reports Using Advanced SQL** – Analysts use a combination of custom ad-hoc queries and Business intelligence tools to generate reports. The module highlights the various OLAP methods available to generate reports as well as how to build functions to access commonly used methods.



This slide is intentionally left blank.