# Splunk Certified Study Guide

Prepare for the User, Power User, and Enterprise Admin Certifications

—

Deep Mehta

Apress®

# Splunk Certified Study Guide

## Prepare for the User, Power User, and Enterprise Admin Certifications

**Deep Mehta**

Apress®

*Splunk Certified Study Guide*

Deep Mehta
Printserv, Mumbai, India

ISBN-13 (pbk): 978-1-4842-6668-7 ISBN-13 (electronic): 978-1-4842-6669-4
https://doi.org/10.1007/978-1-4842-6669-4



Managing Director, Apress Media LLC: Welmoed Spahr
Acquisitions Editor: Divya Modi
Development Editor: Matthew Moodie
Coordinating Editor: Divya Modi

Cover designed by eStudioCalamar

Cover image designed by Pixabay

Distributed to the book trade worldwide by Springer Science+Business Media New York, 1 New York Plaza, New York, NY 10004. Phone 1-800-SPRINGER, fax (201) 348-4505, e-mail orders-ny@springer-sbm.com, or visit www.springeronline.com. Apress Media, LLC is a California LLC and the sole member (owner) is Springer Science + Business Media Finance Inc (SSBM Finance Inc). SSBM Finance Inc is a **Delaware** corporation.

For information on translations, please e-mail booktranslations@springernature.com; for reprint, paperback, or audio rights, please e-mail bookpermissions@springernature.com.

Apress titles may be purchased in bulk for academic, corporate, or promotional use. eBook versions and licenses are also available for most titles. For more information, reference our Print and eBook Bulk Sales web page at http://www.apress.com/bulk-sales.

Any source code or other supplementary material referenced by the author in this book is available to readers on GitHub via the book's product page, located at www.apress.com/978-1-4842-6668-7. For more detailed information, please visit http://www.apress.com/source-code.

Printed on acid-free paper

*This book is dedicated to the late Mr. Amit Mahendra Mehta.*
*This book would not have been possible without his blessing. Where I am today is because of him.*

# Table of Contents

# About the Author

**Deep Mehta** is an AWS Certified Associate Architect (ongoing), Docker Certified Associate, Certified Splunk Architect, and Certified Splunk User, Power User, and Admin. He has worked with the Splunk platform since 2017 and has related consulting experience in the telecommunication, aviation, and health care industries. In addition to his passion for big data technologies, he loves playing squash and badminton.

# About the Technical Reviewer

**James Miller** is an innovator and senior project lead and solution architect with 37 years of extensive design and development experience in multiple platforms and technologies. He leverages his consulting experience to provide hands-on leadership in all phases of advanced analytics and related technology projects. His work includes recommendations on process improvement, report accuracy, adoption of disruptive technologies, enablement, insight identification, statistics for data science, predictive analytics, big data visualization, Watson analytics, and implementing and mastering Splunk.

# Acknowledgments

I am grateful to all my readers for choosing this book. I am also grateful to my parents, Paresh and Rupa Mehta, and my family and friends for their support.

I would like to thank Sanket Sheth of Elixia Tech. for providing the dataset, and a very special thanks to Neha Doshi for extending her support and providing guidance.

# Introduction

Splunk is a software technology for monitoring, searching, analyzing, and visualizing machine-generated data in real time. This book discusses the roles of a Splunk admin and explains how Splunk architecture can be efficiently deployed and managed. It covers everything you need to know to ace the Splunk exams. The book is written to be used interactively and includes practice datasets and test questions at the end of every chapter.

This book is divided into four modules, and three modules are dedicated to the exam.

The first module comprises six chapters dedicated to passing the Splunk Core Certified User exam and the Splunk Core Certified Power User exam. It covers installing Splunk, Splunk's Search Processing Language (SPL), field extraction, field aliases and macros in Splunk, creating Splunk tags, Splunk lookups, and invoking alerts. You learn how to make a data model and prepare an advanced dashboard in Splunk.

The second module is dedicated to the Splunk Enterprise Certified Admin exam and consists of four chapters. It covers Splunk licenses and user role management, Splunk forwarder configuration, indexer clustering, Splunk security policies, and advanced data input options.

The third module also focuses on the Splunk Enterprise Certified Admin exam, but its chapters teach admins to troubleshoot and manage the Splunk infrastructure.

In the fourth module, you learn how to set up Splunk Enterprise on the AWS platform, and you are introduced to some of the best practices in Splunk.

At the end of every chapter is a multiple-choice test to help candidates become more familiar with the exam.

# PART I

# Splunk Architecture, Splunk SPL (Search Processing Language), and Splunk Knowledge Objects

CHAPTER 1

# An Overview of Splunk

Splunk is a software technology for monitoring, searching, analyzing, and visualizing machine-generated data in real-time. This tool can monitor and read several types of log files and store data as events in indexers. It uses dashboards to visualize data in various forms.

This chapter discusses the basics of Splunk, including its history and architecture, and delves into how to install the software on local machines. You see the layout of the Splunk Enterprise Certified Admin exam. And, you learn how to add user data and a props.conf file, and you learn the process of editing timestamps, which is useful in the later chapters. A few sample questions are at the end of the chapter.

Summing it up, this chapter covers the following topics.

- An overview of the Splunk Enterprise Certified Admin exam
- An introduction to Splunk
- The Splunk architecture
- Installing Splunk on macOS and Windows
- Adding data to Splunk

## Overview of the Splunk Admin Exam

A Splunk Enterprise Certified Admin is responsible for the daily management of Splunk Enterprise, including performance monitoring, security enhancement, license management, indexers and search heads, configuration, and adding data to Splunk. The following are the areas of expertise that the exam tests.

- Splunk deployment
- License management
- Splunk applications




D. Mehta, *Splunk Certified Study Guide*, https://doi.org/10.1007/978-1-4842-6669-4_1

- Splunk configuration files
- Users, roles, and authentication
- Adding data
- Distributed searches
- Splunk clusters
- Deploying forwarders with forwarder management
- Configuring common Splunk data inputs
- Customizing the input parsing process

In the next section, you learn about the admin exam's structure.

## Structure

The Splunk Enterprise Certified Admin exam is in multiple-choice question format. You have 57 minutes to answer 63 questions and an additional 3 minutes to review the exam agreement, totaling 60 minutes. The passing score for this exam is 75%. The exam's registration fee is $120 (USD). Refer to `www.splunk.com/pdfs/training/Splunk-Test-Blueprint-Admin-v.1.1.pdf` for more information.

The exam questions come in three formats.

- **Multiple choice**: You must select the option that is the best answer to a question or to complete a statement.
- **Multiple responses**: You must select the options that best answer a question or completes a statement.
- **Sample directions**: You read a statement or question and select only the answer(s) that represent the most ideal or correct response.

## Requirements

The Splunk Enterprise Certified Admin exam has two prerequisites. You must first pass the following exams.

- The Splunk Core Certified Power User exam
- The Enterprise System Administration and Splunk Data Administration courses

Four courses support these exams. The learning flow is shown in Figure 1-1.

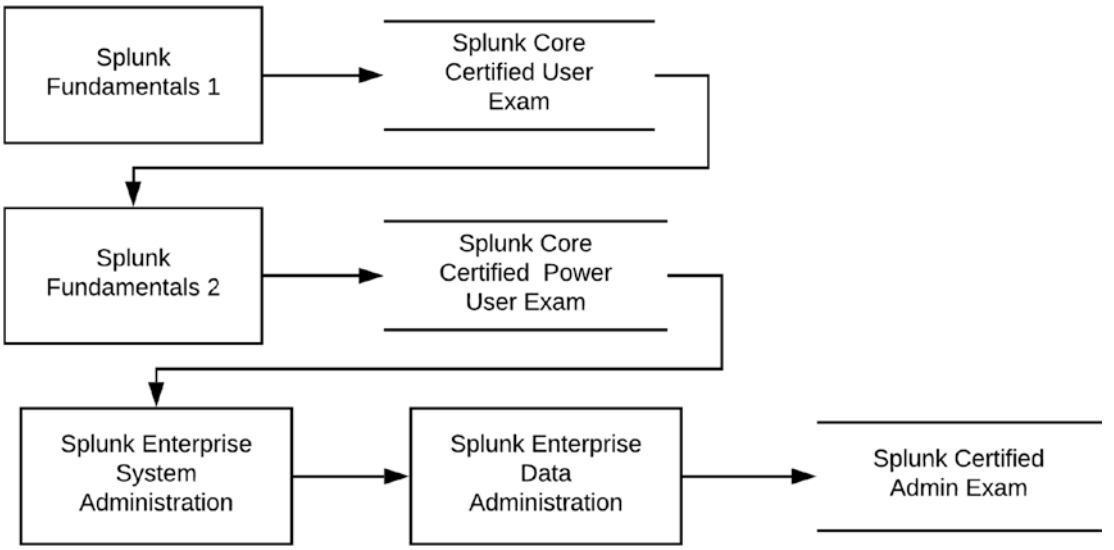


***Figure 1-1.*** *Splunk exam prerequisites*

- **Splunk Fundamentals 1** is offered to students in two ways: e-learning or instructor-led. This course introduces you to the Splunk platform.
- **Splunk Core Certified User Exam** tests your knowledge of and skills in searching, using fields, creating alerts, using lookups, and creating basic statistical reports and dashboards.
- **Splunk Fundamentals 2** is an instructor-led course on searching and reporting commands and creating knowledge objects.
- The **Splunk Core Certified Power User exam** tests the knowledge and skills required for SPL searching and reporting commands and building knowledge objects, using field aliases and calculated fields, creating tags and event types, using macros, creating workflow actions and data models, and normalizing data with the Common Information Model.

- **Splunk Enterprise System Administration** is instructor-led and designed for system administrators responsible for managing the Splunk Enterprise environment. The course teaches fundamental information for Splunk license managers, indexers, and search heads.
- **Splunk Enterprise Data Administration** is instructor-led and designed for system administrators responsible for adding remote data to Splunk indexers. The course provides fundamental information on Splunk forwarders and methods.
- The **Splunk Enterprise Certified Admin** exam tests your knowledge of and skills in managing various components of Splunk Enterprise, including license management, indexers and search heads, configuration, monitoring, and adding data to Splunk.

Modules 2 and 3 of this book focus on the Splunk Enterprise system administration and data administration exams.

## Blueprint

The Splunk Enterprise Certified Admin exam has 17 sections, described as follows.

- **Section 1: Splunk Admin Basics (5%)** This section focuses on identifying Splunk components.
- **Section 2: License Management (5%)** This section focuses on identifying license types and understanding license violations.
- **Section 3: Splunk Configuration Files (5%)** This section focuses on configuration layering, configuration precedence, and the Btool command-line tool to examine configuration settings.
- **Section 4: Splunk Indexes (10%)** This section focuses on basic index structure, types of index buckets, checking index data integrity, the workings of the indexes.conf file, fish buckets, and the data retention policy.
- **Section 5: Splunk User Management (5%)** This section focuses on user roles, creating a custom role, and adding Splunk users.

- **Section 6: Splunk Authentication Management (5%)** This section focuses on LDAP, user authentication options, and multifactor authentication.
- **Section 7: Getting Data In (5%)** This section focuses on basic input settings, Splunk forwarder types, configuring the forwarder, and adding UF input using CLI.
- **Section 8: Distributed Search (10%)** This section focuses on distributed search, the roles of the search head and search peers, configuring a distributed search group, and search head scaling options.
- **Section 9: Getting Data In—Staging (5%)** This section focuses on the three phases of the Splunk indexing process and Splunk input options.
- **Section 10: Configuring Forwarders (5%)** This section focuses on configuring forwarders and identifying additional forwarder options.
- **Section 11: Forwarder Management (10%)** This section focuses on deployment management, the deployment server, managing forwarders using deployment apps, configuring deployment clients, configuring client groups, and monitoring forwarder management activities.
- **Section 12: Monitor Inputs (5%)** This section examines your knowledge of file and directory monitor inputs, optional settings for monitor inputs, and deploying a remote monitor input.
- **Section 13: Network and Scripted Inputs (5%)** This section examines your knowledge of the network (TCP and UDP) inputs, optional settings for network inputs, and a basic scripted input.
- **Section 14: Agentless Inputs (5%)** This section examines your knowledge of Windows input types and the HTTP event collector.
- **Section 15: Fine-Tuning Inputs (5%)** This section examines your knowledge of the default processing during the input phase and configuring input phase options, such as source type fine-tuning and character set encoding.

- S**ection 16: Parsing Phase and Data (5%)** This section examines your knowledge of the default processing during parsing, optimizing, and configuring event line breaking, extraction of timestamps and time zones from events, and data preview to validate event created during the parsing phase.
- **Section 17: Manipulating Raw Data (5%)** This section examines your knowledge of how data transformations are defined and invoked and the use of transformations with props.conf and transforms.conf and SEDCMD to modify raw data.

# An Introduction to Splunk

The word *splunk* comes from the word *spelunking*, which means to explore caves. Splunk can analyze almost all known data types, including machine data, structured data, and unstructured data. Splunk provides operational feedback on what is happening across an infrastructure in real time—facilitating fast decision-making.

Splunk is commonly thought of as "a Google for log files" because, like Google, you can use Splunk to determine the state of a network and the activities taking place within it. It is a **centralized log management tool**, but it also works well with structured and unstructured data. Splunk **monitors, reports, and analyzes real-time machine data**, and indexes data based on timestamps.

## The History of Splunk

Splunk was founded by Rob Das, Erik Swan, and Michael Baum in October 2003. It grew from a small startup company to one of the biggest multinational corporations for security information and event management (SIEM) tools. Before Splunk, a business needing to troubleshoot its environment had to rely on the IT department, where a programmer wrote scripts to meet needs. This script ran on top of a platform to generate a report.

As a result, companies didn't have a precise way to discover problems deep inside their infrastructure. Splunk was created to deal with this issue. Initially, Splunk focused on analyzing and understanding a problem, learning what organizations do when something goes wrong, and retracing footprints.

The first version of Splunk was released in 2004 in the Unix market, where it started to gain attention.

It is important to understand why this software was developed. The following section discusses Splunk's many useful benefits.

## The Benefits of Splunk

Splunk offers a variety of benefits, including the following.

- Converts complex log analysis report into graphs
- Supports structured as well as unstructured data
- Provides a simple and scalable platform
- Offers a simple architecture for the most complex architecture
- Understands machine data
- Provides data insights for operational intelligence
- Monitors IT data continuously

## The Splunk Architecture

The Splunk indexer works in a specified manner in a set architecture (see Figure 1-2).

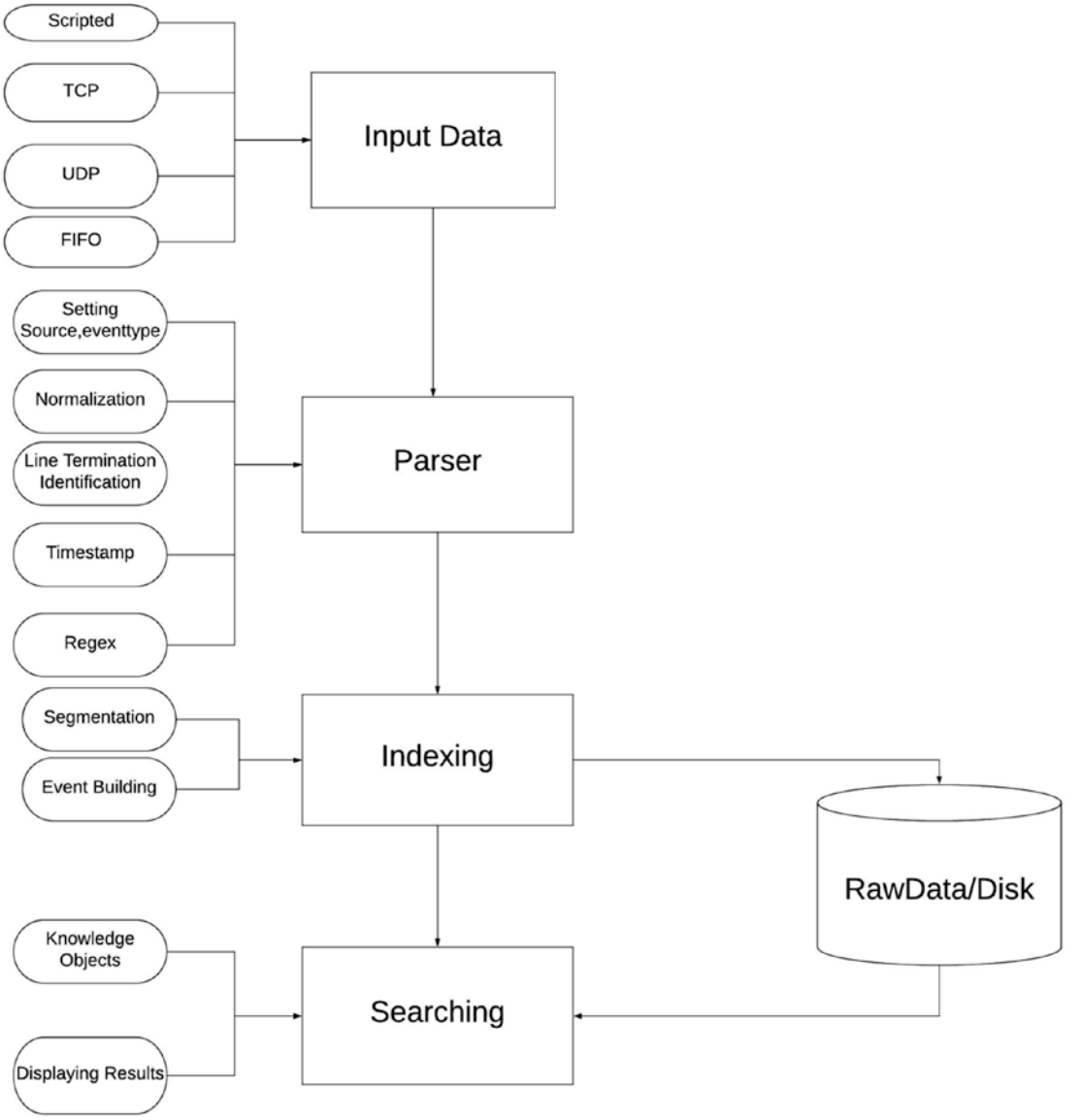


***Figure 1-2.*** *Splunk architecture diagram*

Let's parse this diagram and introduce its components.

- **Input data**: This is the first phase of onboarding data. There are several methods to bring data into Splunk: it can listen to your port, your REST API endpoint, the Transmission Control Protocol (TCP), the User Datagram Protocol (UDP), and so on, or use scripted input.