



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
<https://doi.org/10.1007/978-1-4842-6669-4>

ISBN-13 (electronic): 978-1-4842-6669-4

Copyright © 2021 by Deep Mehta

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

Trademarked names, logos, and images may appear in this book. Rather than use a trademark symbol with every occurrence of a trademarked name, logo, or image we use the names, logos, and images only in an editorial fashion and to the benefit of the trademark owner, with no intention of infringement of the trademark.

The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

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 Authorxix

About the Technical Reviewerxxi

Acknowledgmentsxxiii

Introductionxxv

Part I: Splunk Architecture, Splunk SPL (Search Processing Language), and Splunk Knowledge Objects 1

Chapter 1: An Overview of Splunk 3

 Overview of the Splunk Admin Exam 3

 Structure..... 4

 Requirements 4

 Blueprint..... 6

 An Introduction to Splunk 8

 The History of Splunk 8

 The Benefits of Splunk 9

 The Splunk Architecture 9

 Installing Splunk 13

 Installing Splunk on macOS..... 13

 Installing Splunk on Windows..... 16

 Adding Data in Splunk..... 20

 Summary..... 25

 Multiple-Choice Questions 25

 Further Reading 26

TABLE OF CONTENTS

Chapter 2: Splunk Search Processing Language..... 27

 The Pipe Operator 28

 Time Modifiers 28

 Understanding Basic SPL..... 29

 Search Language Syntax..... 29

 Boolean Operators in Splunk..... 30

 Syntax Coloring in SPL 31

 Sorting Results..... 31

 Sort..... 31

 Filtering Commands 32

 where 32

 dedup..... 33

 head..... 33

 tail..... 34

 Reporting Commands 34

 top 34

 rare 34

 history 35

 table..... 35

 stats..... 36

 untable 38

 chart 38

 timechart 39

 Filtering, Modifying, and Adding Fields 40

 eval 40

 Rex..... 47

 lookup..... 48

 Field..... 49

 Grouping Results..... 49

 Transaction 50

Summary.....	50
Multiple-Choice Questions	51
References.....	52
Chapter 3: Macros, Field Extraction, and Field Aliases	53
Field Extraction in Splunk	54
Regular Expressions.....	54
Delimiters	57
Macros	58
Create a Macro Using Splunk Web	58
Create a Macro Using the .conf File	60
Field Aliases in Splunk.....	62
Setting up Field Aliases	62
Splunk Search Query	67
Summary.....	71
Multiple Choice Test Questions	72
References.....	73
Chapter 4: Tags, Lookups, and Correlating Events	75
Splunk Lookups	75
Looking up Table Files	77
Lookup Definitions.....	78
Automatic Lookups.....	79
Splunk Tags.....	81
Create Tags in Splunk Using Splunk Web	82
Tag Event Types in Splunk Web	83
Reporting in Splunk	85
Creating Reports in Splunk Web	85
Report Acceleration in Splunk	88
Scheduling a Report in Splunk	90

TABLE OF CONTENTS

- Alerts in Splunk..... 92
 - Create Alerts in Splunk Using Splunk Web 92
 - Cron Expressions for Alerts 94
- Summary..... 96
- Multiple-Choice Questions 97
- References..... 99
- Chapter 5: Data Models, Pivot, and CIM 101**
 - Understanding Data Models and Pivot..... 102
 - Datasets and Data Models..... 102
 - Creating Data Models and Pivot in Splunk 102
 - Event Actions in Splunk 112
 - GET Workflow Actions..... 112
 - Search Workflow Action..... 114
 - Common Information Model in Splunk..... 117
 - Defining CIM in Splunk 117
 - Summary..... 121
 - Multiple-Choice Questions 121
 - References..... 123
- Chapter 6: Knowledge Managers and Dashboards in Splunk..... 125**
 - Understanding the Knowledge Manager’s Role in Splunk 125
 - Globally Transferring Knowledge Objects 126
 - Enabling Knowledge Object Visibility..... 128
 - Restricting Read/Write Permissions on an App 129
 - Orphaned Knowledge Objects 130
 - Dashboards..... 132
 - Static Real-Time Dashboards 133
 - Creating a Dashboard 137
 - Adding a Report to a Dashboard..... 139
 - Dynamic Form-based Dashboards..... 140
 - Adding a Radio Button Using XML 140
 - Adding a Time Modifier Using XML..... 142

Adding a Drop-Down Menu Using XML	145
Adding a Link List Using XML	147
Using the User Interface for Input	150
Summary.....	152
Multiple-Choice Questions	152
References.....	154
Chapter 7: Splunk User/Power User Exam Set	155
Questions	155
Summary.....	160
Part II: Splunk Data Administration and System Administration	161
Chapter 8: Splunk Licenses, Indexes, and Role Management	163
Buckets	163
How Does a Bucket Work?	164
How Search Is Performed in Buckets	165
Understanding journal.gz, .tsidx, and Bloom Filters.....	166
How Do Search Functions Work?	166
Splunk Licenses.....	167
Changing a License Group in Splunk.....	168
Managing Splunk Licenses	169
License Masters and Slaves	169
Adding a License in Splunk	171
License Pooling.....	172
Creating a License Pool	172
Managing Indexes in Splunk.....	172
Creating an Index in Splunk.....	173
User Management.....	177
Adding a Native User	177
Defining Role Inheritance and Role Capabilities.....	179

TABLE OF CONTENTS

Summary..... 181

Multiple-Choice Questions 182

References..... 184

Chapter 9: Machine Data Using Splunk Forwarder and Clustering 185

 Splunk Universal Forwarder..... 186

 Configuring Splunk Indexer to Listen to Data for Universal Forwarder 186

 Configuring Windows Splunk Forwarder 187

 Splunk Universal Forwarder Using Windows..... 187

 Splunk Universal Forwarder Using .msi 188

 Configuring Linux Splunk Forwarder 189

 Splunk’s Light and Heavy Forwarders 190

 Splunk Heavyweight Forwarder 191

 Splunk Light Forwarder 192

 Forwarder Management..... 193

 Configuring Forwarder Management..... 193

 Configuring the Forwarder Management Client 195

 Splunk Indexer Clusters 195

 Configuring Indexer Clusters 196

 Splunk Lightweight Directory Access Protocol (LDAP)..... 201

 Creating an LDAP Strategy 202

 Mapping LDAP Group to Splunk Roles..... 205

 Splunk Security Assertion Markup Language (SAML)..... 206

 Configuring Splunk SAML..... 206

 Map SAML to User Roles 209

Summary..... 210

Multiple-Choice Questions 210

References..... 212

Chapter 10: Advanced Data Input in Splunk	213
Compress the Data Feed	214
Indexer Acknowledgment	215
Securing the Feed	215
Queue Size	216
Monitor Input	217
Monitor Files	217
Monitor Directories	218
Monitor Files and Directory Using Splunk Web	218
Monitor File and Directory Using inputs.conf	220
Scripted Input	221
Scripted Input Using Splunk Web	222
Scripted Input Using inputs.conf file	224
Network Input	226
Add Network Input Using Splunk Web and Deploy It to the Forwarder	226
Modify Network Input Using .conf Files	228
Pulling Data Using Agentless Input	231
HTTP Input Using Splunk Web	231
Configure HTTP Event Collector in Splunk	234
Configure HTTP Input Using .conf File	235
Configure HTTP Event Collector in Splunk Using .conf File	236
Parse Data in Splunk Using HTTP Event Collector	237
Summary	238
Multiple-Choice Questions	238
References	241
Chapter 11: Splunk's Advanced .conf File and Diag	243
Understanding Splunk .conf files	243
props.conf	244
indexes.conf	246
transforms.conf	246
inputs.conf	247

TABLE OF CONTENTS

outputs.conf	248
deploymentclient.conf	250
Setting Fine-Tuning Input.....	250
Custom Source Types Using Splunk Web	251
Custom Source Types Using props.conf	252
Anonymizing the Data	253
props.conf to Anonymize Data with a sed Script.....	254
props.conf and transforms.conf to Anonymize Data with Regular Expressions	255
Understanding Merging Logic in Splunk	256
Configuration File Precedence.....	257
Splunk .conf Files Location	257
Configuration Merging Logic	258
Debugging Configuration Files	260
Example: Btool for Troubleshooting a Configuration File	260
Creating a Diag	261
Creating a Diag in Splunk	261
Summary.....	263
Multiple-Choice Questions	263
Reference.....	265
Chapter 12: Splunk Admin Exam Set.....	267
Questions	267
Summary.....	272
Part III: Advanced Splunk	273
Chapter 13: Infrastructure Planning with Indexer and Search Head Clustering	275
Capacity Planning for Splunk Enterprise.....	276
Dimensions of a Splunk Enterprise Deployment	276
Disk Storage for Splunk Enterprise	277

Configuring a Search Peer	278
Configuring a Search Peer from Splunk Web	278
Configure Splunk Search Peer from the .conf File	279
Configure Search Peer from Splunk CLI	279
Configure a Search Head	280
Configuring a Search Head Using Splunk Web	282
Configure Splunk Search Head Using .conf file	282
Configuring a Search Head from Splunk CLI	283
Search Head Clustering	283
Search Head Cluster Captain	285
The Role of Captains	285
Captain Election	285
Configure Search Head Cluster Using CLI in Splunk	286
Multisite Indexer Clustering	287
Configure Multisite Indexer Clustering Using .conf Files	288
Configure Splunk Multisite Indexer Clustering Using CLI	290
Splunk Validated Architectures (SVAs)	294
Designing Splunk Validated Architectures	294
Splunk Architecture Practices	303
Use Case: Company XYZ	303
Splunk Data Inputs	305
Splunk Index Calculation	307
Splunk Total Disk Size	308
Splunk User Planner	309
Hardware and Splunk Scaling Considerations	310
Disk Size Calculation	311
Summary	313
Multiple-Choice Questions	313
References	315

Chapter 14: Troubleshooting in Splunk 317

 Monitoring Console 317

 Single Instance Deployment Monitoring Console 318

 Multi-Instance Deployment Monitoring Console 318

 Monitor System Health with Monitoring Console 321

 Configure Forwarder Monitoring for the Monitoring Console 323

 Log Files for Troubleshooting 324

 The metrics.log File 325

 Pipeline Messages..... 327

 Queue Messages 328

 Thruput Messages..... 328

 Tcpout Connection Messages..... 329

 udpin_connections Messages..... 329

 bucket_metrics Messages 330

 Job Inspector 331

 Job Inspector Example Query..... 331

 Troubleshooting License Violations 334

 Violation Due to an Improper Connection Between License Master and Slave Node..... 334

 Troubleshooting Deployment Issues 335

 Troubleshooting Splunk Forwarders..... 336

 Troubleshooting Splunk Indexers 336

 Troubleshooting Clustering Issues 337

 Multi-Search Affinity..... 338

 More Bucket Issues 338

 Summary..... 338

 Multiple-Choice Questions 339

 References..... 341

Chapter 15: Advanced Deployment in Splunk	343
Deploying Apps Through the Deployment Server.....	343
Create App Directories and View Apps in Forwarder Management.....	344
Redeploy Apps to the Client.....	346
App Management Issues.....	347
Creating a Server Group Using ServerClass.conf.....	347
Configure a Server Class Using Forwarder Management.....	348
Deploy Configuration File Through Cluster Master.....	349
Managing Indexes on Indexer Using Master Node	349
Deploy App on Search Head Clustering.....	352
Configure the Deployer to Distribute Search Head Apps	352
Load Balancing	355
Configure Static Load Balancing in Splunk Forwarder Using outputs.conf	356
Configure a Static Load Balancer by Time.....	356
Specify Load Balancing from Splunk CLI.....	357
Indexer Discovery	357
Configure Indexer Delivery	357
SOCKS Proxy	360
Configure SOCKS Proxy	360
Summary.....	361
Multiple-Choice Questions	362
References.....	363
Chapter 16: Advanced Splunk	365
Managing Indexes.....	365
Configure Event Indexes.....	366
Configure Metrics Indexes.....	367
Remove Indexes and Index Data for Managing Indexes.....	368
Configure Index Parallelization for Managing Indexes	368
Manage Index Storage	369
Move the Index Database	369

TABLE OF CONTENTS

- Configure Maximum Index Size for Indexer Storage 370
- Set Limit for Disk Usage in Splunk 371
- Managing Index Cluster 372
 - Configuring Peer Node to Offline 372
 - Configure Splunk to Maintenance Mode Using Splunk CLI 373
 - Rolling Restart in Splunk Using Splunk CLI 374
 - Remove Excess Buckets Copies from the Indexer Cluster 375
 - Remove a Peer from Master’s List 375
- Managing a Multisite Index Cluster..... 376
 - Master Site in Multisite Index Cluster Fails 376
 - Restart Indexing in the Multisite Cluster After a Master Restart or Site Failure..... 377
 - Move a Peer to a New Site 377
- REST API Endpoints..... 378
 - Running Searches Using REST API 379
- Splunk SDK 381
 - Python Software Development Kit for Splunk 381
- Summary..... 385
- Multiple-Choice Questions 385
- References..... 387
- Chapter 17: Final Practice Set 389**
 - Questions 389
 - Summary..... 394
- Chapter 18: Setting up a Splunk Environment with AWS 395**
 - Amazon Web Services..... 395
 - Configuring an EC2 Instance Using the AWS Management Console..... 396
 - Configuring Splunk on an EC2 Instance 399
 - Deploying Multisite Index Clustering..... 401
 - Configuring a Cluster Master..... 401
 - Configuring a Slave Node 402

Deploying a Search Head	403
Configuring a Search Head	403
Configuring Search Head Clustering	404
Deploying Configurations	406
Configuring a Cluster Master	407
Deploying an App to Search Head Cluster Using Deployment Server	409
Configuring a Universal Forwarder for Indexer Discovery	411
Configuring a Universal Forwarder for Indexer Deployment	413
Configuring the Master Node, Deployment Server, and Search Head 4 to Send Internal Logs	416
Monitoring Distributed Environments	417
Adding a Search Peer to Monitor	417
General Setup for Distributed Environments	418
Conclusion	419
Index	421

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

- 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.

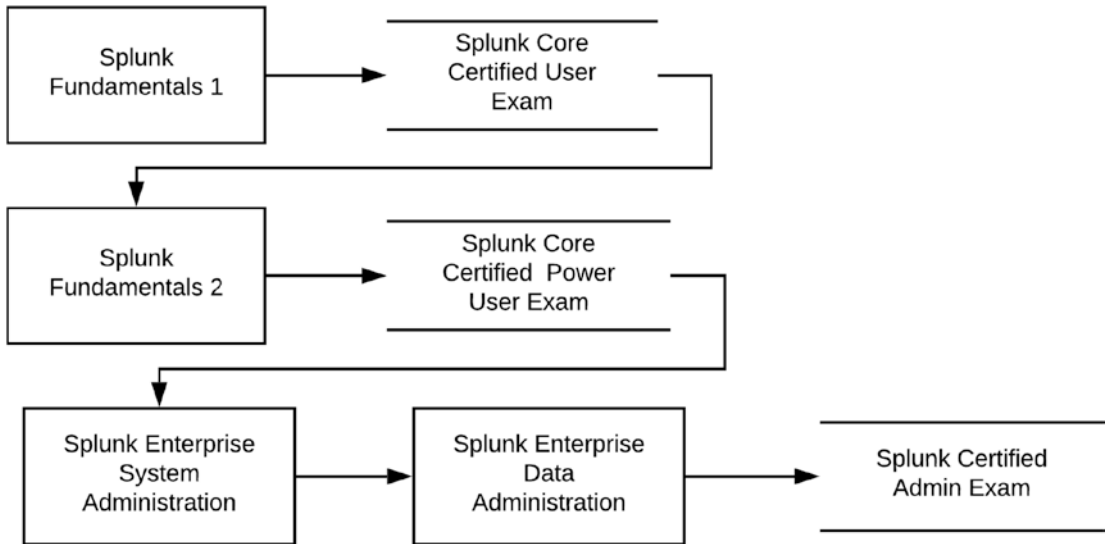


Figure 1-1. *Splunk exam prerequisites*

- **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.

- **Section 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).

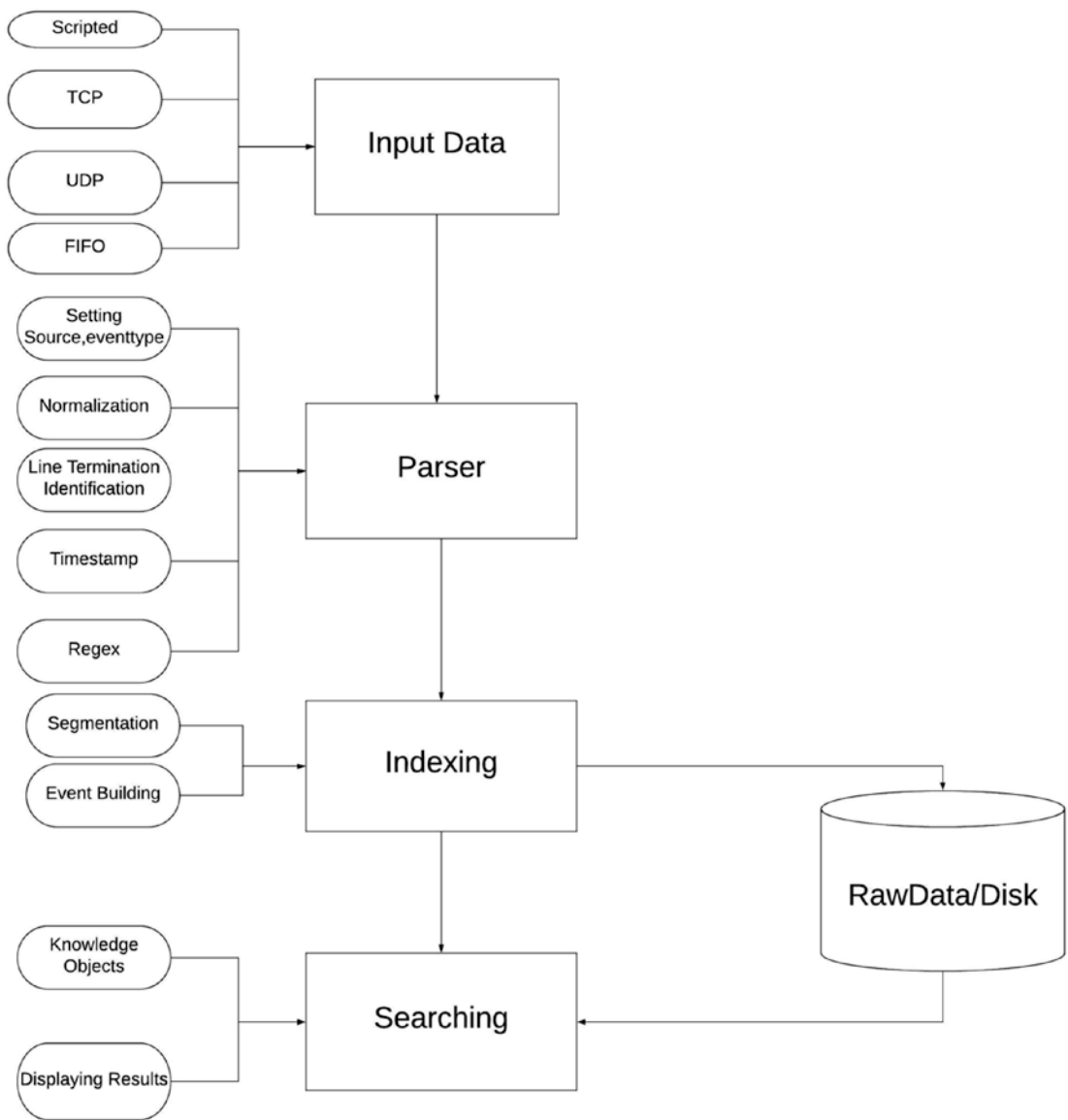


Figure 1-2. *Splunk architecture diagram*

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.