IoT Data Analytics using Python

Learn how to use Python to collect, analyze, and visualize IoT data

M S Hariharan

www.bpbonline.com

First published: 2024

Published by BPB Online
WeWork
119 Marylebone Road
London NW1 5PU

UK | UAE | INDIA | SINGAPORE

ISBN 978-93-5551-575-9

www.bpbonline.com

Dedicated to

*My beloved mother, **A.C. Kavery***
and
*My sister **Uma** and niece **Pavithra***

About the Author

M S Hariharan, is a seasoned professional in the field of IT with over 25 years of experience. He holds a Masters in Data Science from Liverpool John Moores University, UK. Throughout his career, he has specialized in areas such as digital transformation, system integration, and automation.

Currently, he serves as the Principal Director at LTIMindtree, where he leads the development of the NxT Data platform , focusing on building innovative products. Prior to joining LTIMindtree, he spent six years leading Digital Transformation initiatives at AstraZeneca, contributing to their success in implementing effective solutions. Additionally, he also held the position of Lead Solution Architect in Hewlett-Packard Software Professional Services for more than 8 years.

His diverse experience and expertise has allowed him to contribute to numerous projects and drive impactful transformations within the industry. He is passionate about leveraging technology to drive business growth and optimize operations.

About the Reviewers

- **Niharika Panda** is a distinguished expert in the realm of the Internet of Things (IoT), IoT communication protocols, wireless sensor networks, and machine learning. With a profound skillset and a trail of accomplishments, Niharika has significantly contributed to the advancement of these fields. She is proficient in Python programming and applies her coding prowess in Machine Learning and data analysis for IoT protocols. Throughout her journey, Niharika has garnered remarkable achievements and certifications. She has successfully completed certifications in Python from prestigious platforms like Coursera, and in the realm of IoT, she holds certificates from NPTEL provided by IIT Kharagpur and Lema Labs, IIT Madras. Niharika holds a B.Tech in Information Technology and a M.Tech in Computer Science and Engineering and is currently pursuing a Ph.D. in the field of Internet of Things and Data Science at Amrita School of Computing, Bengaluru. This extensive educational background demonstrates her commitment to continuous learning and staying at the forefront of technological advancements.

- **Kumar Shashank** is a knowledgeable and discerning expert in the field of Data Science and Machine Learning. With an extensive background encompassing six years of industry experience, including three years as a dedicated Data Scientist, his insights carry significant weight. His expertise covers a wide range of BI and reporting tools, as well as mastery of ML/AI algorithms, IIoT, SQL, and Python with essential libraries like Sklearn, TensorFlow, Numpy, Pandas, Matplotlib, and Seaborn.

  He brings a critical eye to the evaluation of technical books, analyzing the clarity, depth, practicality of the content and its relevance to the industry standards. His impartial assessments provide valuable guidance to aspiring data scientists and ML practitioners, ensuring they invest their time in books that truly enrich their knowledge. With excellent communication skills and a collaborative mindset, Shashank’s contributions extend beyond written critiques, making them a respected authority in the field and a valuable asset to the technical community.

Acknowledgement

I would like to express my deepest gratitude to the individuals who have provided their unwavering support during the creation of this book. First and foremost, I would like to thank my mother for her constant encouragement and my sister and niece for their unwavering support. Without their presence and encouragement, this book would not have been possible. It saddens me that my mother is not here to witness the publication of this book, but her spirit lives on in my work.

I am immensely grateful for the knowledge and experiences gained from my current and previous organizations. In particular, I would like to extend my heartfelt appreciation to my current organization, LTIMindtree, for providing me with the wonderful opportunity to delve into the development and implementation of IoT Data Analytics. Their support and guidance have been invaluable.

I would also like to extend my heartfelt thanks to Mr. Mahesh Mahajan, my mentor and leader, for his guidance and mentorship throughout this journey. His expertise and encouragement have been instrumental in shaping this book.

To all those who have provided their hidden support, whether through words of encouragement, feedback, or moral support, I extend my heartfelt thanks. Your contributions have made a significant impact on the completion of this book.

Once again, I extend my deepest appreciation to everyone who has played a part in making this book a reality. Thank you for your continued support and belief in my work.

Preface

Welcome to the world of IoT data analytics, a realm where Python developers can unlock the full potential of the Internet of Things (IoT) and harness the power of advanced analytics. In this book, **IoT Data Analytics using Python** we embark on an enlightening journey to equip Python developers with the skills and knowledge necessary to excel in the realm of IoT data analytics. Throughout this book, we will be using real-world datasets, providing practical insights and hands-on experience that bridge the gap between theory and application. Additionally, this book is also well suited for anyone wanting to learn Python as they embark on their journey into IoT data analytics. Join us as we explore the intersection of Python programming and IoT data analytics, paving the way for innovative solutions and data-driven decision-making.

Chapter 1: Necessity of Analytics Across IoT - In this chapter, we explore the essentials of IoT analytics, the impact of the Internet of Things and the Industry Internet of Things, and the revolution brought about by Industry 4.0. We delve into the critical role of IoT data analytics in driving digital transformation and discuss the components of a data pipeline for analytics. Python is introduced as the go-to language for IoT Data Analytics.

Chapter 2: Up and Running with Data Analytics Fundamentals - This chapter serves as a comprehensive guide to Data Analytics fundamentals. We explore various data analysis methods and techniques, enabling Python developers to perform data analysis using different frameworks. From the SEMMA, KDD, and CRISP-DM frameworks to industry-specific frameworks, we provide a solid foundation for analytical endeavors.

Chapter 3: Setting Up IoT Analytics Environment - In this chapter, we delve into the practical aspects of setting up an IoT Data Analytics environment. We discuss the rationale behind choosing Python as the language of choice and guide readers through the installation and configuration of essential tools such as Python IDE, Apache Kafka, MQTT, and PostgresSQL. Python packages are also explored, along with Python basics, data analysis, wrangling, and visualization.

Chapter 4: Managing Data Pipeline and Cleaning - This chapter delves into the crucial task of managing data pipelines and ensuring data cleanliness. We explore different types of IoT data formats and equip you with the necessary skills to handle diverse data formats efficiently. Delving into real-time data streaming, data pipeline management, and the concept of IoT Dataflow, we empower the reader to build robust and scalable data pipelines. Furthermore, we unveil the world of IoT simulators and guide you through their implementation using

Python. With a focus on data cleansing and transformation, we provide you with the tools and techniques to ensure data accuracy and quality.

Chapter 5: Designing Data Lake and Executing Data Transformation - This chapter introduces the concept of designing a data lake and executing data transformation. We uncover the intricacies of data lakes, their role in managing real-time IoT data streaming, and building effective data architectures. Additionally, we emphasize the importance of building Key Performance Indicators (KPIs) and metrics objectives for extracting meaningful insights from IoT data. We will use Python to implement a simple and effective data lake.

Chapter 6: Implementing Descriptive Analytics Using Pandas - This chapter focuses on implementing descriptive analytics using the powerful Pandas library. We explore techniques for descriptive data analysis and time series analysis, enabling Python developers to uncover patterns and trends in time-stamped IoT datasets. Testing methods for time series data are also discussed.

Chapter 7: Time Series Forecasting and Predictions - This chapter takes us into the realm of time series forecasting and predictions. We explore techniques such as data smoothing, lag identification, and autocorrelation, empowering Python developers to make accurate predictions. Python libraries such as Tsfresh and AutoTS are introduced for advanced feature extraction and high-accuracy forecasting. Additionally, we showcase storing wind turbine predictions and constructing an Analytical Base Table (ABT) for efficient analysis.

Chapter 8: Monitoring and Preventive Maintenance - In this chapter, we delve into the critical aspects of monitoring and preventive maintenance in the IoT landscape. We explore condition monitoring, Condition-Based Maintenance (CBM), corrective maintenance, preventive maintenance, and predictive maintenance. The chapter also covers text mining product manuals and automating the creation of maintenance tickets, optimizing maintenance processes.

Chapter 9: Model Deployment on Edge Devices - This chapter takes us on a journey of model deployment on edge devices. We introduce edge computing and analytics, simulators for IoT systems, and provide a step-by-step guide to installing and configuring edge devices and FastAPI. Model building, reuse, and deployment on edge devices are explored, along with the concepts of continuous learning and adaptive learning.

Chapter 10: Understanding Edge Computing with MicroPython - This chapter unravels the intricacies of understanding edge computing with MicroPython. We explore the concepts of edge computing, edge analytics, and the edge platform. Data flow from edge to cloud and diverse use cases for edge analytics are discussed. With a focus on MicroPython, developers can invoke machine learning models and leverage the power of edge computing.

Chapter 11: IoT Analytics for Self-driving Vehicles - In this final chapter, we present a compelling case study that showcases the application of IoT Data Analytics in the context of self-driving vehicles. By leveraging the CRISP-DM framework, we dive deep into the business aspects of self-driving vehicles and demonstrate how to perform end-to-end IoT Data Analytics using Python.

Throughout this book, our aim is to provide Python developers with a comprehensive understanding of IoT Data Analytics and its practical application across industries and use cases. Whether you are a Python developer looking to enhance your skills or someone seeking to learn Python through the lens of IoT Data Analytics, this book will serve as a valuable resource.

Code Bundle and Coloured Images

Please follow the link to download the
Code Bundle and the ***Coloured Images*** of the book:

https://rebrand.ly/a3237f

The code bundle for the book is also hosted on GitHub at **https://github.com/bpbpublications/IoT-Data-Analytics-using-Python**. In case there's an update to the code, it will be updated on the existing GitHub repository.

We have code bundles from our rich catalogue of books and videos available at **https://github.com/bpbpublications**. Check them out!

Errata

We take immense pride in our work at BPB Publications and follow best practices to ensure the accuracy of our content to provide with an indulging reading experience to our subscribers. Our readers are our mirrors, and we use their inputs to reflect and improve upon human errors, if any, that may have occurred during the publishing processes involved. To let us maintain the quality and help us reach out to any readers who might be having difficulties due to any unforeseen errors, please write to us at :

errata@bpbonline.com

Your support, suggestions and feedbacks are highly appreciated by the BPB Publications' Family.

Piracy

If you come across any illegal copies of our works in any form on the internet, we would be grateful if you would provide us with the location address or website name. Please contact us at **business@bpbonline.com** with a link to the material.

If you are interested in becoming an author

If there is a topic that you have expertise in, and you are interested in either writing or contributing to a book, please visit **www.bpbonline.com**. We have worked with thousands of developers and tech professionals, just like you, to help them share their insights with the global tech community. You can make a general application, apply for a specific hot topic that we are recruiting an author for, or submit your own idea.

Reviews

Please leave a review. Once you have read and used this book, why not leave a review on the site that you purchased it from? Potential readers can then see and use your unbiased opinion to make purchase decisions. We at BPB can understand what you think about our products, and our authors can see your feedback on their book. Thank you!

For more information about BPB, please visit **www.bpbonline.com**.

Join our book's Discord space

Join the book's Discord Workspace for Latest updates, Offers, Tech happenings around the world, New Release and Sessions with the Authors:

https://discord.bpbonline.com

Table of Contents

Chapter 1

Necessity of Analytics Across IoT

Introduction

This chapter covers the basic concepts of **Internet of Things** (**IoT**) and how the industry 4.0 revolution impacts our life, economy, and technology advancements. It focuses on how the IoT industry is connected to Industry 4.0 revolution and the role it plays in industry automation. We will investigate important use cases that solve key business problems in industries such as manufacturing, construction, Oil and gas, railroad, pharmaceutical, and other infrastructure developments. We will cover applications for IoT Data Analytics in **Environment, Social, and Governance** (**ESG**), where we will discuss how IoT Data Analytics can help fight global warming and help us achieve carbon emission reduction. In the smart city use cases, we will discuss how the automation of infrastructure and water distribution is improving our lives. In the case of medical and Health Sciences, and pharmaceuticals, we will discuss how it is helping patients, doctors, drug manufacturers, and hospitals to manage their processes more efficiently to help save lives. We will also touch upon other industries and areas where IoT Data Analytics is helping our society in a better way.

Structure

In this chapter, we will cover the following topics:

- Internet of Things and Industrial Internet of Things
- Industrial Revolution and Industry 4.0

- IoT Data Analytics
- IoT Data Analytics for Digital Transformation
- Hardware Devices for IoT Data Analytics
- Data Pipeline for Analytics
- Python: The Go-to Language for IoT Analytics

Objectives

By the end of this chapter, you will be able to understand the basic concepts of IoT, Industry IoT, various equipment used in the industry IoT, and IoT Data Analytics. You will learn about industrial use cases for IoT Data Analytics and IoT hardware devices for analytics. In the last part of the chapter, you will learn about the data pipeline for analytics and the programming language for IoT Data Analytics.

Internet of Things and Industrial Internet of Things

The **Industrial Internet of Things** (**IIoT**) is revolutionizing various industries by enabling efficient data monitoring and analysis. For instance, consider the application of IIoT in the wind turbine industry. Wind turbines are equipped with advanced sensors that continuously collect data on various parameters, such as wind speed, temperature, and turbine performance. These sensors generate a constant stream of data that provides valuable insights into the turbine's behavior and efficiency.

However, due to the massive volume of data generated, it becomes impractical to store all this information locally within the turbine. Instead, the data is transmitted to the cloud through the internet. In the cloud, this data is processed, analyzed, and stored in vast databases that have virtually unlimited storage capabilities. Engineers and technicians can access this data remotely and in real-time, enabling them to monitor the turbines' performance, detect anomalies, and identify potential maintenance needs.

Similarly, we can provide multiple use cases and examples from other industries such as self-driving vehicles, smart cities, smart buildings, and many others on the impact of IIoT. These industries leverage IIoT technologies to collect and analyze data from various sources, enabling them to optimize operations, enhance efficiency, and improve overall performance. IIoT is a powerful tool that empowers industries to make data-driven decisions, improve productivity, and drive innovation.

Before diving into our main objective of analytics in the IoT world, let us quickly understand some of IoT's basic concepts and background:

- **Demand for IoT and Analytics**: There are several predictions regarding how IoT and IoT Data Analytics will grow in the coming years. *Gartner*, leading research, and advisory company, predicts that the number of connected IoT devices worldwide will reach over 20 billion by 2025, a significant increase from the 8 billion devices recorded in 2018. In addition, another report by Gartner forecasts that by 2030, a staggering 90% of all data will be generated by IoT devices. As a result, the market for IoT Data Analytics is expected to continue to grow, as organizations aim to extract value from the massive amount of data generated by these devices. Another report from the McKinsey Global Institute predicts that by 2026, there will be a need for more than 4 million data and analytics professionals worldwide. This includes those with skills in IoT Data Analytics. This highlights the significance of learning IoT Data Analytics, as it is becoming an increasingly vital area of expertise in the field. With this projection, there is a growing opportunity for professionals with the necessary skills in IoT Data Analytics. Therefore, acquiring proficiency in this area is essential to meet the rising demand and reap the benefits of a rapidly growing industry.

- **Definition of Internet of Things:** In simple terms, any devices or physical objects that connect to the external environment via sensors to generate and transfer data via networks for further usage. Here, we used various terms such as devices and sensors, in the preceding definition. Let us go through each of the terms to understand their role in IoT implementation:
- **Devices:** Any physical object with multiple sensors to serve a purpose is called a device. In our previous example, the wind turbine is a device. The following are multiple devices serving different purposes:
    - **Energy management systems:** The device with sensors to manage and control power usage. Air condition control systems or smart sensors to turn on the lights based on human movement are examples of energy management systems.
    - **Industrial control systems:** The systems used to produce material in the manufacturing assembly. It controls the production process and contains when to switch on the machines.
    - **Vehicle tracking systems:** The device to track the movement of vehicles in the supply chain business. This helps in optimizing routes and managing the delivery of materials.
    - **Process control systems:** The devices used in process-centric industries such as Pharmaceutical or Chemical, where processing is critical for the product and safety.
- **Sensors**: Any physical object which deducts the external environment conditions and changes by measuring them constantly. The following are examples of sensors:
    - **Motion sensors:** The sensors which monitor the environment for any movement of the object to trigger an action, such as turning on/off the lights, or triggering an alarm when an object is detected.
    - **Proximity sensors:** The sensors which monitor the presence of objects near to specific location to trigger an action, such as opening a door when a person is near the door.

- **Pressure sensors:** Pressure sensors are used for detecting pressure in a system like a boiler or steam engine to trigger an action based on the level of pressure detected. We can set thresholds to trigger an alarm if the pressure is too much for the system. This will prevent accidents due to unexpected pressure levels in industrial boilers.
- **Humidity sensors:** These sensors can monitor the humidity level of the air and can trigger an action based on the surrounding humidity levels, such as turning on/off the air conditioning system.

There are many devices and sensors which can help automate the generation of useful information about the environment or system to manage the expected outcome. We will go through various IoT devices and sensors in the later part of this book to understand their role in different industries.

Industrial Revolution and Industry 4.0

The **Industrial Revolution** is used for elaborating the history of our advancements in the industries such as manufacturing, automobile, travel, and transportation. To understand the full context of the industrial revolution, let us go through its history, in brief, to understand the significance of the present industrial revolution and how IoT and related technologies are transforming economies and our lives in general.

First Industrial Revolution

The first industrial revolution started in the 18^{th} century and matured in the early 19^{th} century. It is also called the Agricultural Revolution. The advancement in the production of goods using machines to reduce hard labor is the major contribution of the first industrial revolution. The invention and development of steam engines to transform the transportation industry, mechanical spinning jenny, and power loom to transform the textile industry are a few examples. Then advancement in the production of coal, Iron, mines, and other materials are the key contribution of this Industry 1.0. The great progress in human civilization and expansion of urbanization happened because of the large production of goods using new machines invented during this period. This resulted in new factories, city expansion, economic growth, and further science and technology advancement. The first industrial revolution paved a new way for human civilization and further advancement of science.

Second Industrial Revolution

The second industrial revolution started as a continuation of the first revolution in the 19^{th} century and continued in full momentum until the 20^{th} century. During this period, major technological changes happened which came to be known as the period of technological revolution. With the discovery of new energy sources such as natural gas and petroleum, there was a significant shift in the coal industries that replaced major energy consumption. The inventions such as the telephone and telegraph become a paradigm shift in communication and information sharing. There emerged a major industry that revolutionized our transportation called the **automotive industry**. Along with these industries, there are significant progress made in the chemical, textile, and material industries due to the discovery of petroleum and natural gas. Humanity made huge progress in this period regarding transportation, communication, and social and economic conditions.

Third Industrial Revolution

The digital revolution started at the start of the current century that changed the speed and mode of communication. The emergence of computers, the internet, and mobile technology marked a new era in human civilization. In general, the third industrial revolution expanded inventions and discoveries in every aspect of technology, industry, and society. We re-defined the existing industries that were part of previous revolutions and created new ones. This has transformed productivity, the economy, and social condition with significant automation in every industry. A remarkable shift happened in the telephone, information, and communication industries with the adoption of computers, the internet, and mobile phones. With the growth of **Information and Communication Technology** (**ICT**), our social and economic conditions are radically changed. The emergence of mobile devices and internet technology marked a new phase in the technological revolution, leading to the development of smart technologies such as the Internet of Things, robotics, and artificial intelligence. These advancements in technology have paved the way for the current industrial revolution and continue to drive innovation in the digital landscape.

Fourth Industrial Revolution

The **Fourth Industrial Revolution (Industry 4.0)** is the current and ongoing progress in developing new digital technologies. Thanks to the discovery of digital and ICT industries in the early 20th century, we are now progressing with digital adoptions with Artificial Intelligence, Robotics, and the Internet of Things. The revolution impacted every aspect of human lives by adopting digital and smart technologies. We see a change in mobile devices and the integration of technologies in every aspect of our lives. This significant surge in technology adoption required us to expand our storage and processing of information. This enabled us to create new technologies such as Cloud Computing and Big Data. The storing and processing of large data have become an aftereffect of the wide adoption of digital technologies. This created new and advanced existing industries such as autonomous vehicles, smart cities, gene therapy, personalized medicines, advancements in robotics, and so on.

Another major contribution of the Industrial Revolution is in the field of nanotechnologies with the invention of nanomaterials and biodegradable plastics. These industries are making us advanced in our social and economic conditions and our environment sustainable.

IoT and Data Analytics plays a key role in Industry 4.0 as the demand for gathering and analyzing data from IoT devices is growing rapidly. Every industry is shifting to make their day-to-day process automated with the enablement of digital technologies. The need for expansion in terms of scale and quality dictates the need for adopting IoT and Analytics. Making advancements in this area are key to progress into the next industrial revolution of fully automated technologies which minimize human intervention in many aspects of industrial production.

IoT Data Analytics

Introduction to IoT Data Analytics: IoT Data Analytics is the process of acquiring, transforming, storing, analyzing, and visualizing data from IoT devices and sensors. The main objective of IoT Data Analytics is to extract insights from the data to enable business users to perform informed decisions and actions. We know from the previous section that IoT devices and sensors monitor the external environment to generate data at frequent time of interval. Because of this, we can see that the data generated by IoT devices and networks are in a large amount. Performing analysis on this large amount of data can benefit the business in optimizing their operations, improving the business outcome that leads to more profitability and expansion. New business models can be derived from the analysis of the large data. Before diving deeper into **IoT Data Analytics**, let us explore the key processes and technologies associated with it.

- **Data generation:** IoT devices, sensors and the network of devices monitor the environment based on the configured parameters and generate data continuously. These devices or sensors are tiny in nature with low storage and energy in the environment, which are required to operate. Hence, they may not have significant storage or processor facility within these IoT devices to perform the required data operations of collecting, storing, and transformation. Based on the preceding constrain in environment, we need a process for collecting data.
- **Data Collection:** In this step data is collected from the IoT devices, sensors and networks using push or pull method. Either the IoT devices send the data via push, or the server sends a request for collecting data via the pull method. There are various protocols used for the communication and transmission of data. Most frequently used protocols are, **Message Queuing Telemetry Transport** (**MQTT**), **Hypertext Transfer Protocol** (**HTTP**), **Constrained Application Protocol** (**CoAP**), and **Advanced Message Queuing Protocol** (**AMQP**). We will learn more about these protocols in the upcoming section.
- **Data storage:** The collected data needs to be stored in a secured, optimized, efficient way where extraction of data is possible anytime. The most common way for storing the data from the IoT devices will be on a centralized database. The location of these databases can be in the **Local Network (LAN)** or the cloud **Wide Area Network** (**WAN**). The cloud is the most used method for the operation since it is scalable, secured, efficient, and cost-effective. The choice of databases varies depending on application and analytics requirements. The most common databases are relational databases such as *Postgres* or *MySQL*. Since the data generated from the IoT devices are of time series in nature as the data generated is in frequent interval. Hence, time series relational databases could be a more suitable choice here. The time series databases such as