



# NLP for Research in Pharma Industry

Anish Mahapatra | 7 January, 2022

# Agenda

---

01 Introduction

02 The Data Lifecycle

03 Understanding Proof of Concept (PoC) versus Production for NLP

04 NLP for Research Architecture

05 The Data Science Process Flow

06 Key Takeaways

# Introduction



Senior Data Scientist 🧐



LinkedIn: [www.linkedin.com/in/anishmahapatra](https://www.linkedin.com/in/anishmahapatra)



Medium: [anishmahapatra.medium.com](https://anishmahapatra.medium.com)



Website: [www.anishmahapatra.com](https://www.anishmahapatra.com)



GitHub: [www.github.com/anishmahapatra](https://www.github.com/anishmahapatra)

Q. How to get into Data Science?

A. Simple, ask a real-world Data Scientist.



# The Data Lifecycle

01

## Data Access

Identify List of Data Owners,  
Request Access,  
Maintain Single POC,  
Understand current process

02

## Data Silos to Data Warehouse

Make Data Dictionary,  
Make ER Diagram,  
Form Analytical Data frame

03

## Data Analytics & Modelling

Exploratory Data Analysis,  
Business Analytics,  
Heuristic, Statistical & ML Models.  
Experimentation

06

## Business Decisions, Impact Analysis

Model Evaluation,  
Monitoring Model Performance and Drift,  
Data Governance Policies

05

## Data Visualization & Dashboarding

Dashboards to aid decision-making,  
Decide concurrency, cadence and security

04

## Feature Engineering

Building new features,  
Combining external data,  
Building a feature store

# PoC versus Production

## Proof of Concept

01



### Idea

Every great project starts with an idea.  
Project idea, funding, team composition

02



### Experiment

Quick and effective experiments can help prove/disprove hypotheses

03

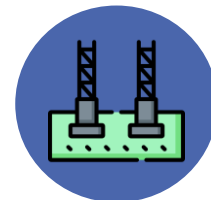


### Iterate

Building and understanding multiple ways to tackle a problem are the core of a good proof of concept

## Production

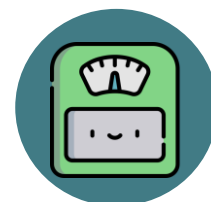
01



### Strong Foundation

Production projects generally start off with proof that it is worth scaling

02



### Scale & Test

Multiple components such as concurrency, infrastructure, integration have to be tested as projects scale

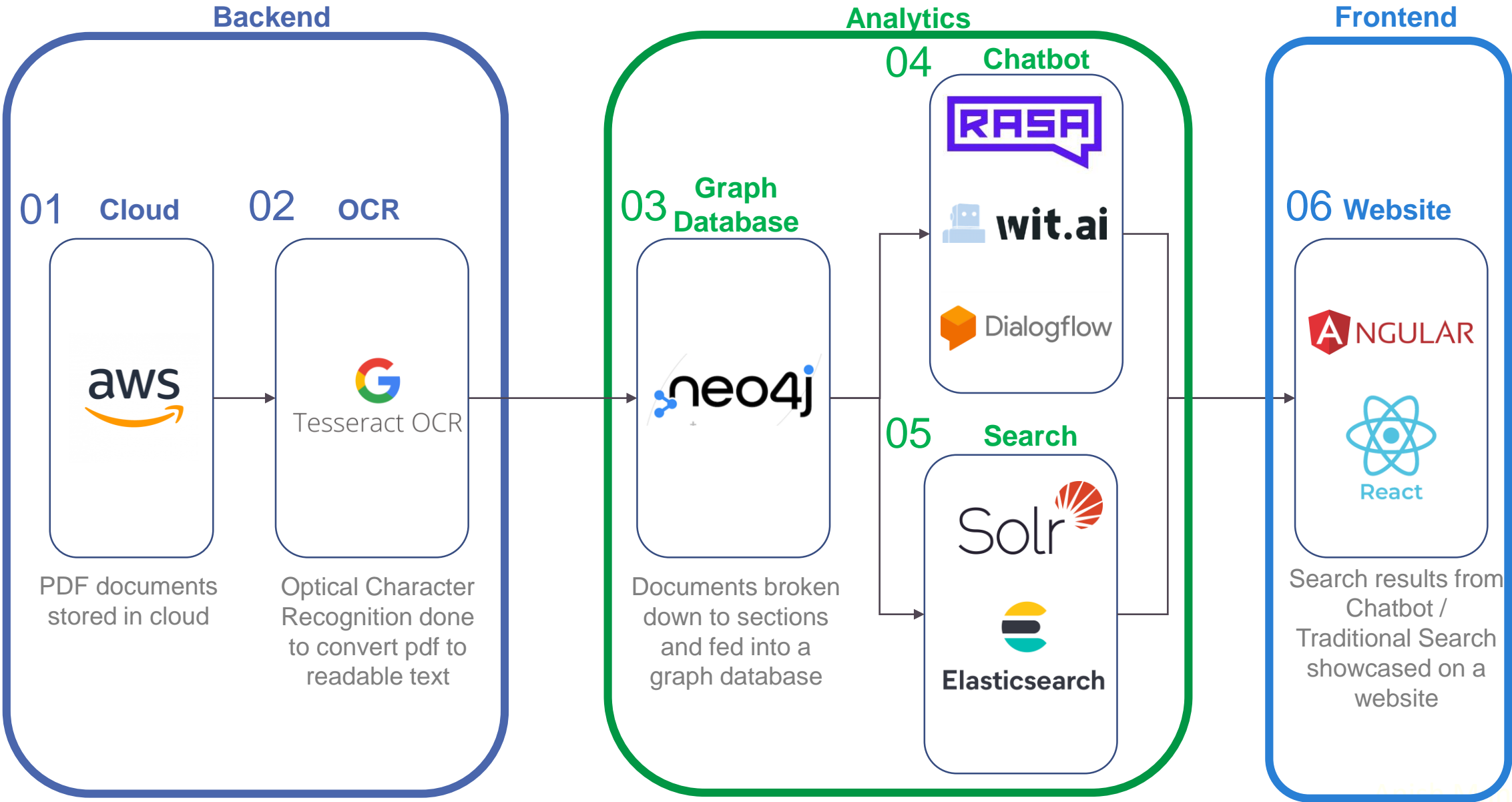
03



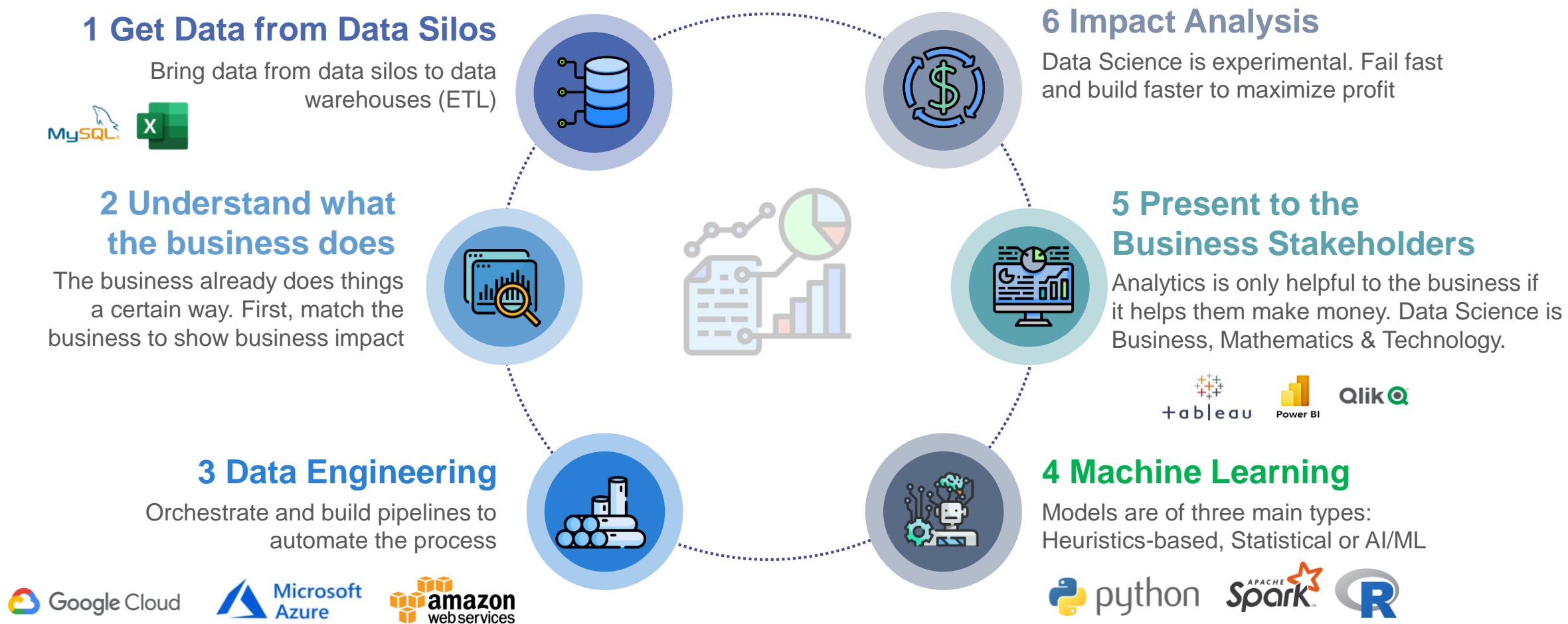
### Maintain & Improve

Large projects require maintenance and improvements based on feedback

# NLP for Research Architecture



# The Data Science Process Flow





# Key Takeaways

---

01 What happens to Data in Data Science

02 Working of PoC and Production Projects

03 The architecture for NLP

04 The Data Science Process Flow



# Career Road Map in Data Science

---

Anish Mahapatra

## Q&A Session



LinkedIn: [www.linkedin.com/in/anishmahapatra](https://www.linkedin.com/in/anishmahapatra)

Email: [anishmahapatra01@gmail.com](mailto:anishmahapatra01@gmail.com)

A man in a dark suit and light-colored tie is adjusting his tie with his hands. The background is dark and out of focus.

# NLP for Research in Pharma Industry

Anish Mahapatra | 4 January 2022

*Q. How to get into Data Science?*

A. Simple, ask a real-world Data Scientist.

# Thank you.

Email: [anishmahapatra01@gmail.com](mailto:anishmahapatra01@gmail.com)