

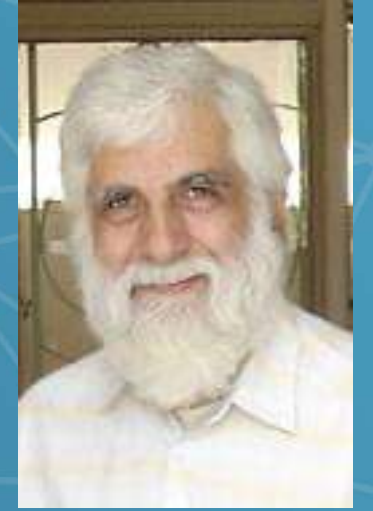


# How to Build a Career in Data Science

Ehsan Vahedi

Senior Data & Applied Scientist, Microsoft

July 27<sup>th</sup>, 2021



*In loving memory of:*  
*Professor Mohsen Shiva (- 2008)*  
*Professor Caro Lucas (1949 – 2010)*

# Biography:

Ehsan Vahedi is a Senior Data & Applied Scientist at Microsoft, Seattle, WA, USA. He joined Microsoft in 2016, has served as a Microsoft technical interviewer since 2018, and has been a data science technical lead since 2019. Prior to Microsoft, he was a Senior Research Scientist at BBTv, a subsidiary of RTL media group in Vancouver, BC, Canada, from 2012 to 2016. Ehsan received his M.Sc. degree in Electrical and Computer Engineering from the University of Tehran, Tehran, Iran in 2008 under the supervision of Professor Mohsen Shiva and Professor Reza Zoroofi. He received his Ph.D. degree in Electrical and Computer Engineering from the University of British Columbia (UBC), Vancouver, BC, Canada in 2013. His thesis was on Security and Privacy of RFID systems, under the supervision of Professor Ian F. Blake and Professor Rabab K. Ward. Ehsan has co-authored two US patents, one book chapter and over 30 conference and journal papers, and has been an IEEE Senior Member since 2016.

## Disclaimer:

All views expressed during this talk are my own based on the observations I had, and do not represent the opinions of any entity whatsoever with which I have been, am or will be affiliated.

# Agenda:

- Evolution of data science and its future prospect
- Data Scientist vs. Applied Scientist vs. Machine Learning Engineer (and other) roles
- Startups vs. large corporates, Eng vs. PM culture
- Data Science roles at:
  - Microsoft, Amazon and Apple
  - Facebook and Google
  - Twitter and Netflix
- How to build a career in Data Science
- Career growth in Data Science
- Useful pointers and links, personal experiences and sample projects
- **Q&A**



# Evolution of Data Science

# A Little Bit of History:

- • In 1962, John Tukey wrote about a shift in the world of statistics, predicted the transition from the classic “statistician” role and envisioned the modern “data analyst” role.
- In 1974, Peter Naur authored the *Concise Survey of Computer Methods*, and invented the term “Data Science” for the first time.
- ...
- In 2002, the first journal dedicated to Data Science was published by International Council for Science.
- • In 2006, Hadoop 0.1.0, an open-source, non-relational database, was released.
- • In 2008, the title, “Data Scientist” became a buzzword, and eventually a part of the IT language. DJ Patil (LinkedIn) and Jeff Hammerbacher (Facebook) are given credit for initiating its use as a buzzword.
- In 2009, the term NoSQL was reintroduced.
- In 2011, James Dixon suggested using Data Lakes rather than Data Warehouses.
- • In late 2012, Harvard Business Review published its famous [article](#) “Data Scientist: The Sexiest Job of the 21st Century”. In 2013, IBM shared statistics showing 90% of the data in the world had been created within the last two years.
- Starting 2014, there was an immense increase in seminars and conferences devoted to Data Science and Big Data. Data Science proved itself to be a source of profit and became a critical part of modern corporate culture.
- • In 2015, Google’s speech recognition experienced an amazing performance improvement of 49% using Deep Learning techniques.
- • In 2018 (IMO), Data Science started diverging into more specialized roles: Research Data Scientist, Machine Learning Engineer, Product Analyst, Applied Scientist to name a few, and the list goes on.
- • Future Prospect: SDE level coding skills and building production level solutions is the new essential qualification (example: Amazon Applied Scientist and Facebook Research Data Scientist).

# Interest over Time:

● Data Scientist  
Search term

+ Compare

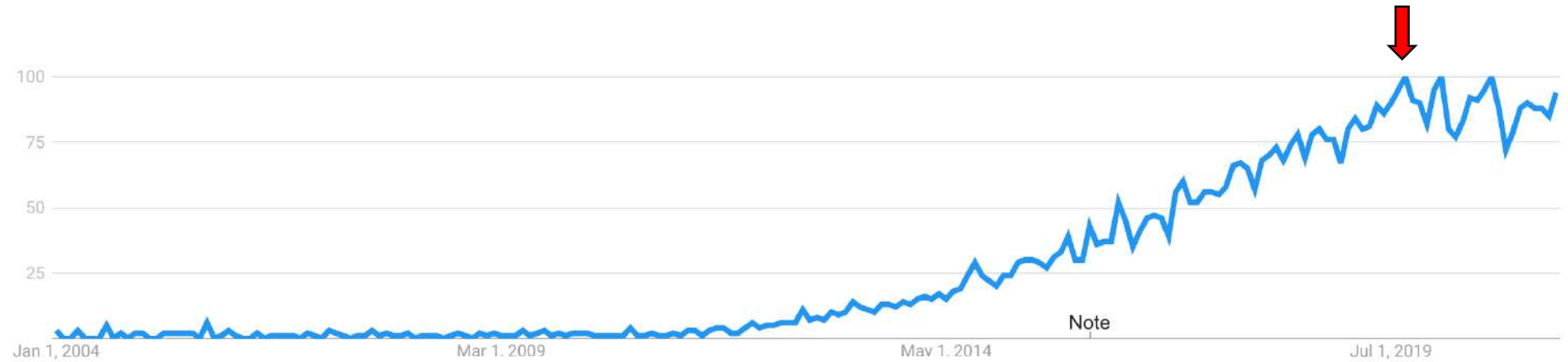
Worldwide ▾

2004 - present ▾

All categories ▾

Web Search ▾

Interest over time ?



Source: [Google Trends](https://trends.google.com/trends/)



# Differences Between Tech Companies



# Expectations from Data Scientists:

## • **Size:** Startups (and Medium) vs. Large Corporates

- **Hiring:** Central Standard vs Team/Org-based
- **Culture:** Engineering/SDE-based vs. Product/PM-based Corporates

## • **Technical Expectations**

- Work Permit and Sponsorship
- Agility, Stability
- Main Advantages of Smaller Companies

**RESPONSIBILITIES**

- ✓ Apply your expertise in quantitative analysis and the presentation of data to see beyond the numbers and understand how our users interact with our growth products.
- ✓ Work as a key member of the product team to solve problems and identify trends and opportunities.
- ✓ Inform, influence, support, and execute our product decisions and product launches.
- ✓ Set KPIs and goals, design and evaluate experiments, monitor key product metrics, understand root causes of changes in metrics.
- ✓ Exploratory analysis to discover new opportunities: understanding ecosystems, user behaviors, and long-term trends
- ✓ Identifying levers to help move key metrics.

**MINIMUM QUALIFICATIONS**

- ✓ 3+ years experience doing quantitative analysis, preferably for a web or mobile company
- ✓ Fluency in SQL
- ✓ Core statistical knowledge
- ✓ Proven experience leading data-driven projects from definition to execution: defining metrics, experiment design, communicating actionable insights.
- ✓ Able to mentor and guide junior data scientists
- ✓ Experience leading analytics teams - advantage
- ✓ Python, R or any other scripting language - advantage

Facebook Data Scientist – Product Analytics

**About The Role**

- We are working on:
  - Search Ranking
  - Recommender System
  - Computational Advertising Ranking
  - Information Retrieval
  - Learning to RankDeep
  - Learning to MatchQuery
  - UnderstandingIntent Understanding

**About You**

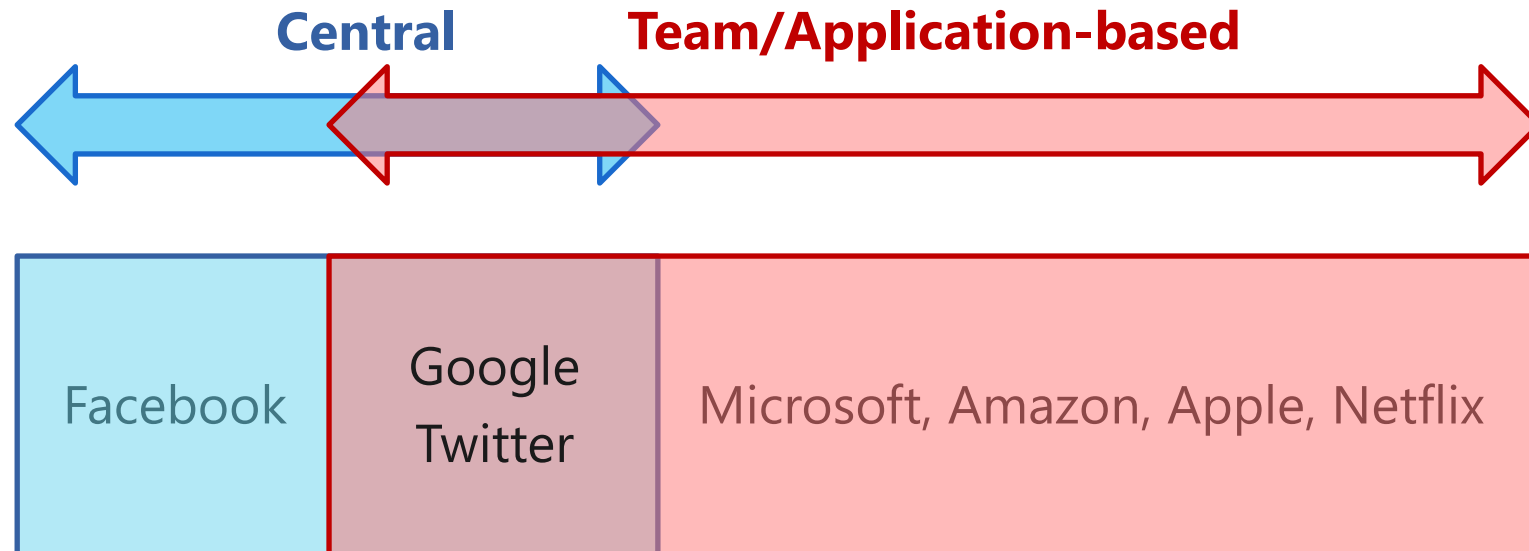
- You share our values (below) and are looking for a company that has a solid mission.
- You have strong analytical and quantitative skills. You are familiar with techniques in Recommender Systems, Information Retrieval, Computational Advertising, Search Ranking, Learning To Rank, Deep Learning to Match or related fields.
- You have a Ph.D. degree or Master degree in Computer Science, Machine Learning, or related fields.
- You have strong technical and programming skills. You are familiar with relevant technologies and languages (e.g. Python, Java, Scala, C++ and etc.). You have experience in or desire to learn Hadoop/Spark related Big Data technologies.
- You have demonstrated the capability to review and write technical papers.
- You can contribute to research that can be applied to products.
- You have the ability to quickly prototype ideas and solve complex problems by adapting creative approaches.
- You are a strong collaborator and communicator and you make the engineers around you grow and learn.

Etsy Data Scientist

[Source](#)

# Expectations from Data Scientists:

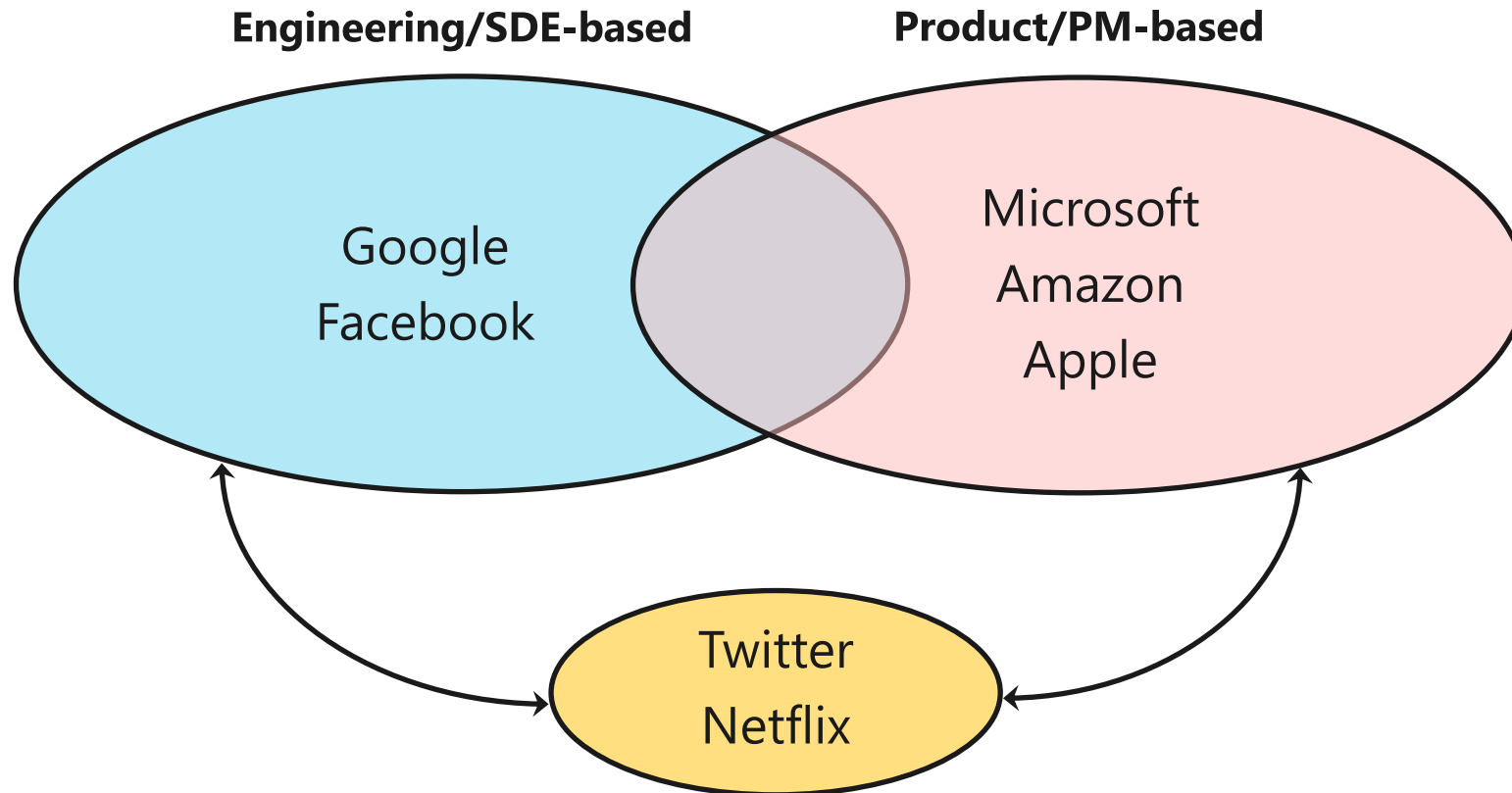
- **Size:** Startups (and Medium) vs. Large Corporates
- **Hiring:** Central Standard vs Team/Org-based
- **Culture:** Engineering/SDE-based vs. Product/PM-based Corporates



Does it have any impact on you? YES!

# Expectations from Data Scientists:

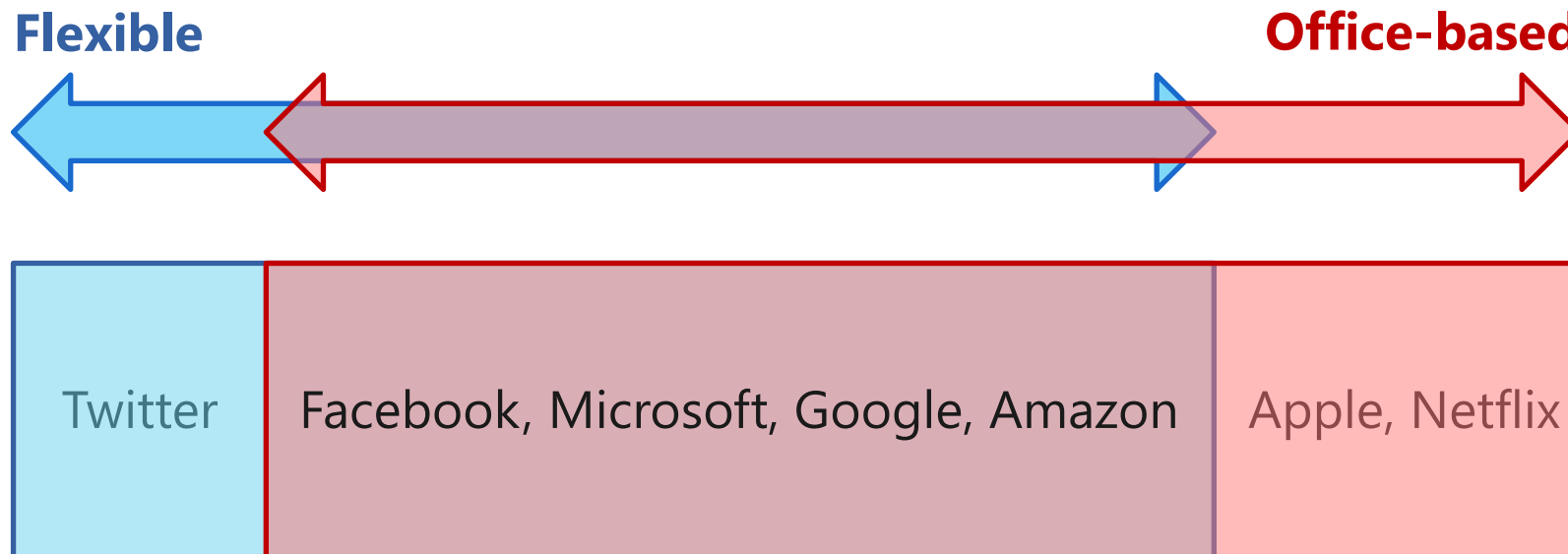
- **Size:** Startups (and Medium) vs. Large Corporates
- **Hiring:** Central Standard vs Team/Org-based
- **Culture:** Engineering/SDE-based vs. Product/PM-based Corporates



How does this impact you?

# Post Pandemic Era and Remote Work:

- Companies have different cultures and priorities, and as a result, very different approaches towards WFH.
- Outsourcing more IT jobs to Canada, Europe and other countries is (IMHO) a byproduct of the post pandemic concerns.



## Microsoft

- **Hiring Levels:** 59 - 80
- **Culture:** PM-based
- **Hiring Process:** Team-based
- **JD:** Specific and Detailed

## Amazon

- **Hiring Levels:** L4 – L10
- **Culture:** PM-based
- **Hiring Process:** Team-based
- **JD:** Specific and Detailed

## Google

- **Hiring Levels:** L3 – L10
- **Culture:** Eng-based
- **Hiring Process:** Central
- **JD:** General

## Facebook

- **Hiring Levels:** E3 – E9
- **Culture:** Eng-based
- **Hiring Process:** Central
- **JD:** General (mostly)

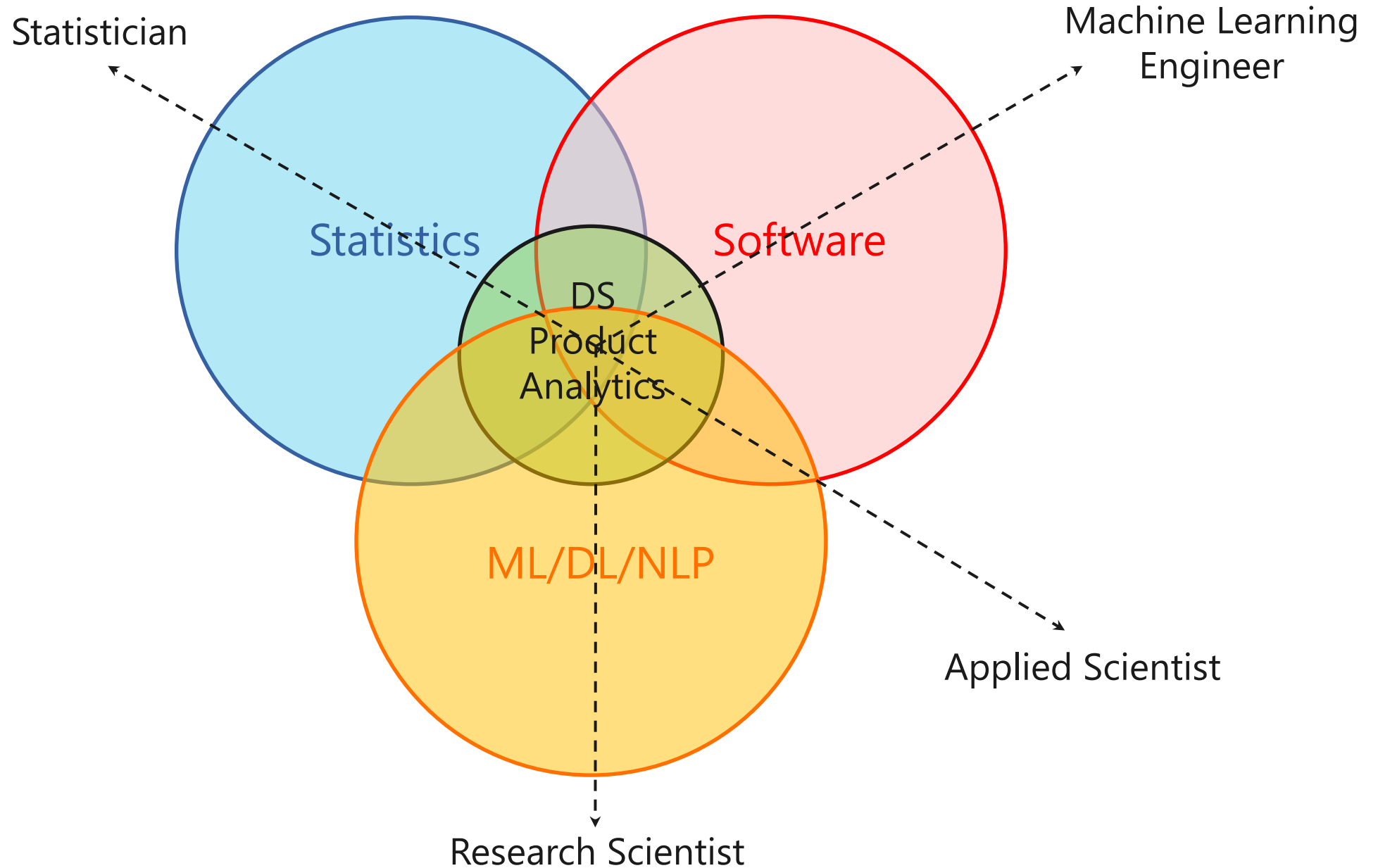


# Data Science Roles and their Required Skillset

# Branches of Modern Data Science:

- Data Engineer
- Data Analyst / Business Analyst
- Statistician / Decision Scientist / Core Data Scientist
- Data Scientist
- Product Analyst
- Quantitative Analyst
- Research Data Scientist
- NLP Scientist
- Research Scientist
- Applied Scientist
- Machine Learning Engineer

# Data Science Skillset:

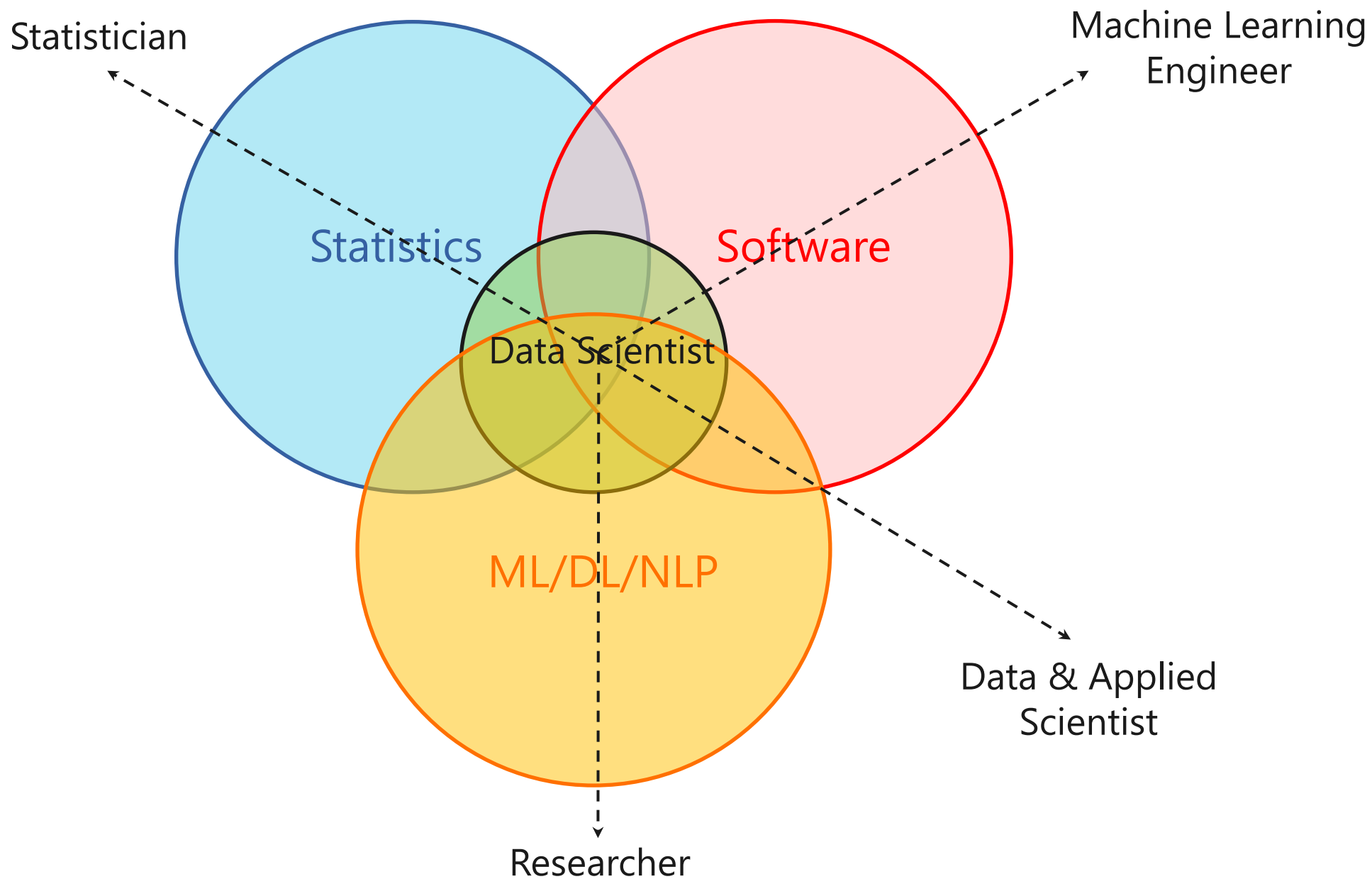




The image features a solid blue background with a subtle, light-colored network pattern of interconnected nodes and lines. The word "Microsoft" is centered in a white, sans-serif font.

**Microsoft**

# Microsoft DS Jobs:



# Microsoft DS- Related Roles:

- **Data & Applied Scientist**

- JD Keywords: Python, R, SQL, Machine Learning, A/B Experimentation (DS), Statistics, Hypothesis Testing, Product, Metrics (DS), Inference, Communication Skills, Data Acquisition, E2E Pipelines, Azure, Spark, ...
- Internal Tools: Scope/C#, Cosmos, Kusto, Azure ML, Azure DL, PowerBI, ...

- **Researcher**

- JD Keywords: Machine Learning, Deep Learning, NLP, Research, Publication, KDD, NIPS, ICML, CVPR, ...
- MSR Organization

- **Machine Learning Engineer**

- JD Keywords: Distributed Programming, Scalable Production Code, Algorithms, System Design, C++, Java, Azure, API, Debugging, Docker, PowerShell, ...

- **Statistician, Data Engineer, Data Analyst, ...**

Hiring Levels: 59 – 80 (what does it mean?)

# [Sample JD] Microsoft Data & Applied Scientist:

We are seeking people who have experiences which reflect the qualifications below:

## Required Qualifications:

- 5+ years of industry experience with machine learning algorithms for at least one of the following: classification, regression, clustering, reinforcement learning, dimensionality reduction. Expertise in one or more application domains of NLP, computer vision, time series modeling, including some knowledge of the concepts of taking projects to scale in the cloud.
- 5+ years of experience in research and development of ML/AI based solutions and systems.
- Skilled in one of the following programming languages Python, Java, C#, C++.
- Fluency in English.

## Preferred Qualifications:

- Experience with Numpy, Pandas, Scikit-learn. Most of our work uses Python ML stack. Being also comfortable with R is a big plus.
- Experience with agile development practices and Git version control. We build reproducible systems.
- Experience with one or more of the DNN frameworks like TensorFlow or PyTorch. Many of our projects involve deep learning models.
- Domain experience in one or more of the following domains: financial services, retail, marketing, health care, manufacturing, media, or telecommunications.
- Ability to communicate in business and technical settings. We need to be able to build models and explain to others how and why they work.
- Enjoy travel and comfortable with regional travel up to 20%. We meet customers where they are, both technically and geographically. (Nobody is expected to travel for business until the COVID-19 pandemic is under control.)
- Some experience with Spark, SQL, Graph stores, or NoSQL stores would be helpful. Many of our customers have a lot of data.
- MS or PhD in Computer Science, Electrical Engineering, Statistics, Operations Research or equivalent technical field.

## Plus:

- Statistics
- Experimentation

# [Sample JD] Microsoft Researcher:

## Responsibilities

The mission of our group is to make fundamental contributions to advancing the **state-of-the-art** in speech, computer vision, and language technology both within Microsoft and in the external research community. As a researcher, you will:

- Collaborate with world-class researchers and engineers to develop cutting-edge algorithms to create innovative applications and solutions, solving real-world challenges.
- Conduct influential **research published in top-tier venues** such as CVPR, ICCV, ECCV, ICASSP, Interspeech, ASRU, SLT, ACL, EMNLP, NAACL, ICLR, NIPS, SIGDIAL.
- Experiment with methods to improve and fine-tune state-of-the-art deep learning models
- Develop prototype software based on research ideas to stimulate new thinking, and to aid the transition of research into products. Contribute with research ideas, as it complements existing research concepts and algorithms.
- Collaborate with our development team, often serving as a bridge between the research team and the product group.

## Qualifications

### Required Qualifications:

- **Ph. D. in Computer Science** (or related discipline) or Masters and 2+ years relevant experience
- 6+ years of research or engineering experience in speech, computer vision, natural language processing, machine learning, optimization, or other subfields of artificial intelligence.
- 2+ years of hands-on experience in deep Learning using frameworks such as **PyTorch, or TensorFlow**.

# [Sample JD] Microsoft Machine Learning Engineer:

## Responsibilities

- Engage directly with key partners to understand and implement complex inferencing capabilities including managing large scale multi-tenant inferencing platform.
- Anticipate, identify, assess, track and mitigate project risks and issues in a fast-paced start up like environment.
- Drive thought leadership, architecture and high scale, complex graph topologies, high throughput, low latency inferencing.

## Qualifications

### Required:

- 5+ years of software development experience with expertise in **Java, C# or C++**
- Bachelor's or Master's degree in Computer Science or a related field

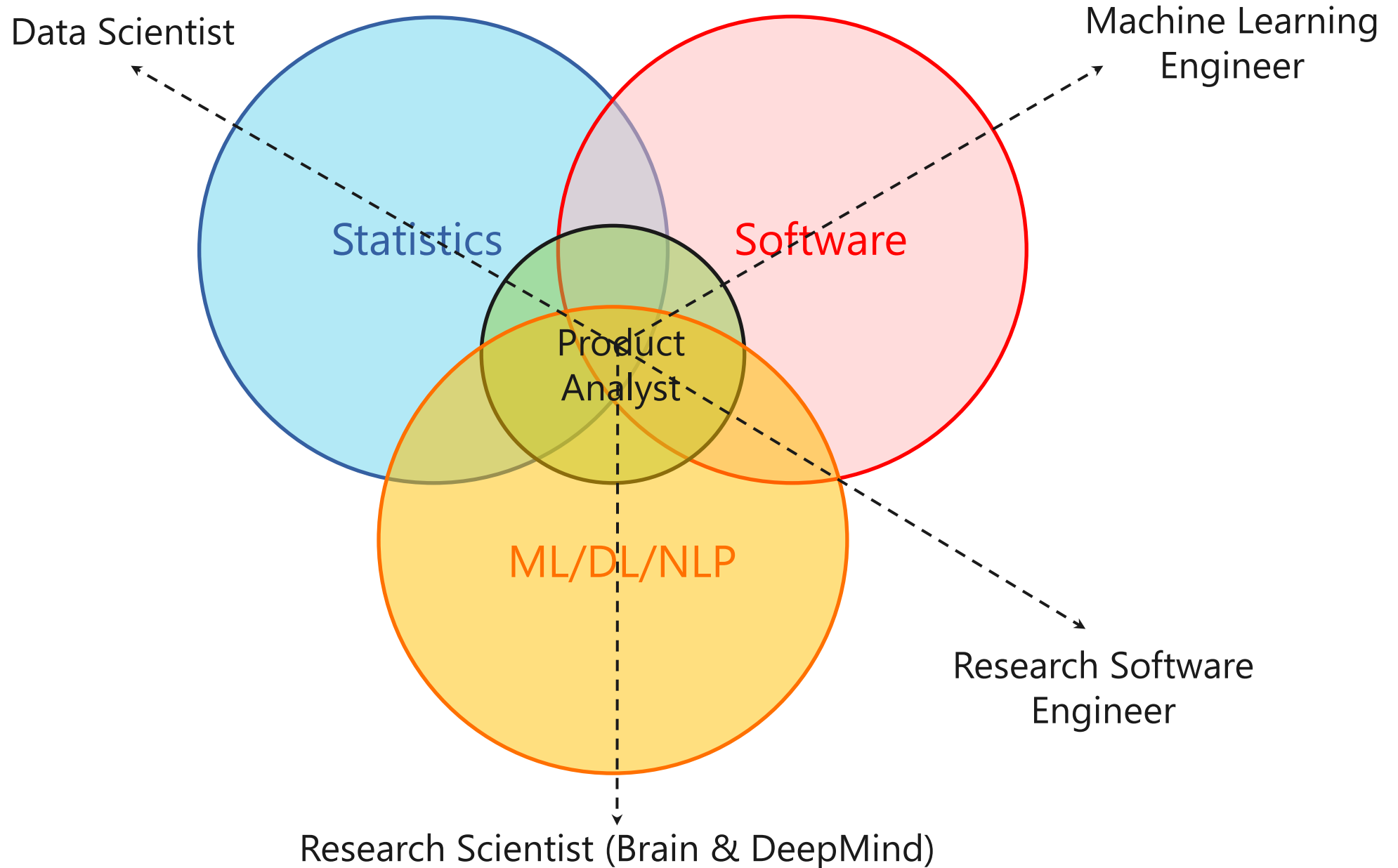
### Preferred:

- Experience in **distributed computing and architecture.**
- Platform and capability development experience in Azure or other Cloud technologies strong plus.
- Experience in developing low latency systems.
- Experience in developing and operating high scale, reliable online services.
- Knowledge and experience in **OSS, Docker, Kubernetes, .NET, C#, java, python** or equivalent programming languages
- Experience with **ML platforms** and frameworks
- Good communication, collaboration skills and a great team player



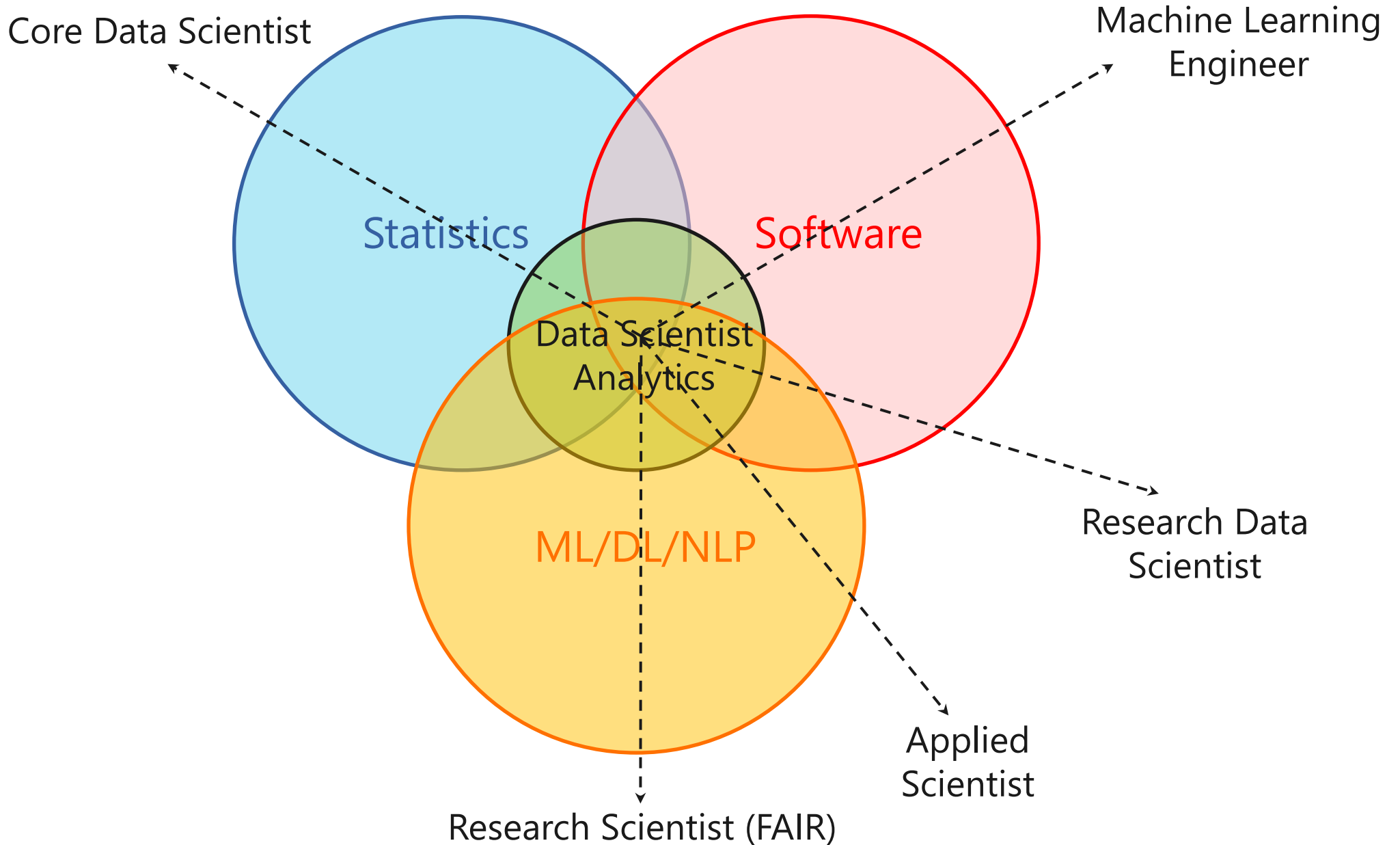
# Other Companies

# Google DS Jobs:





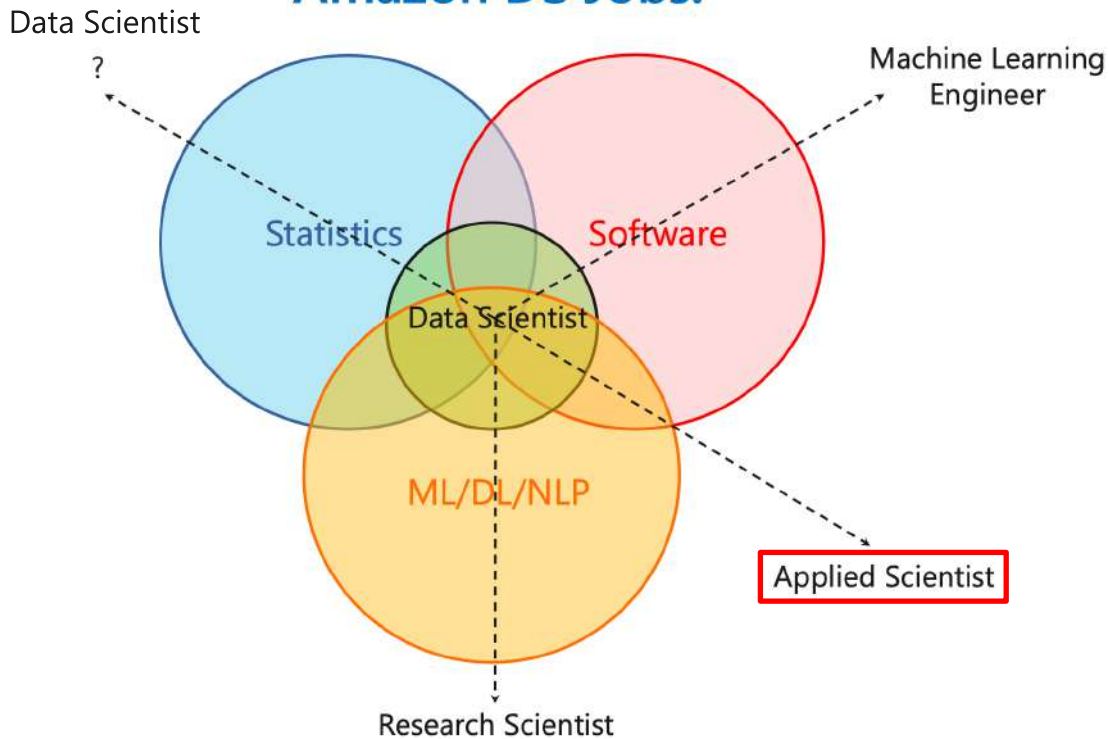
# Facebook DS Jobs:



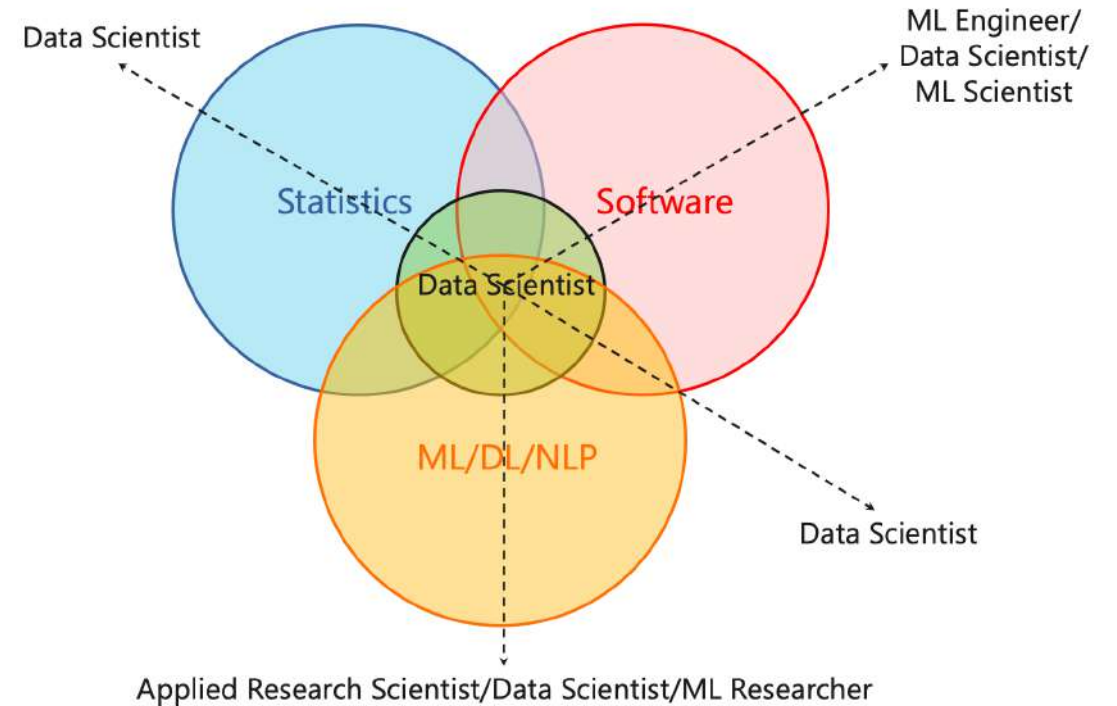
# Amazon and Apple:

- Amazon and Apple are both Product/PM-based companies and very similar to Microsoft in terms of the expectations from data scientists, definition of roles and “team-based” hiring process.
- Amazon has clear definitions of (and distinctions between) Data Scientist and Applied Scientist roles.

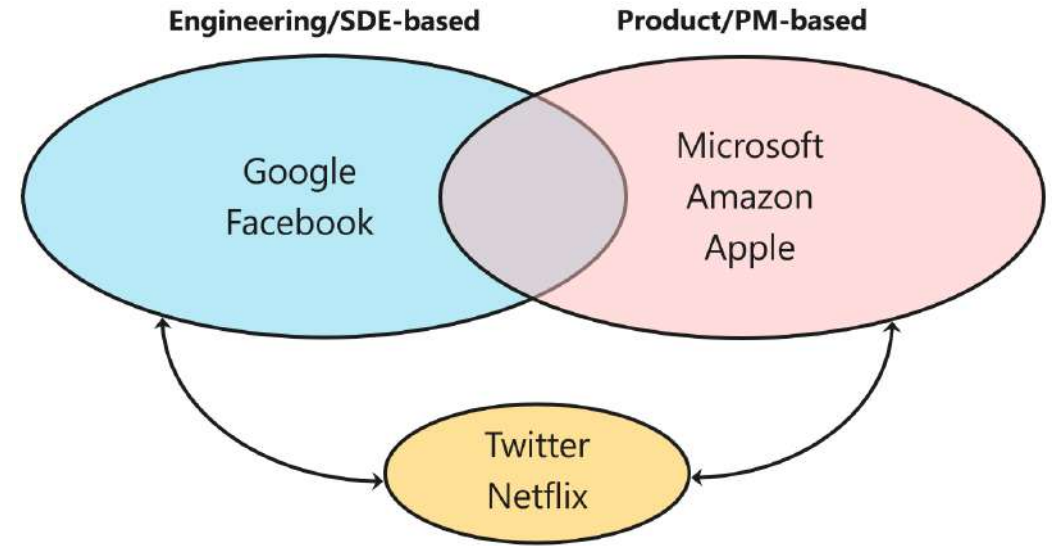
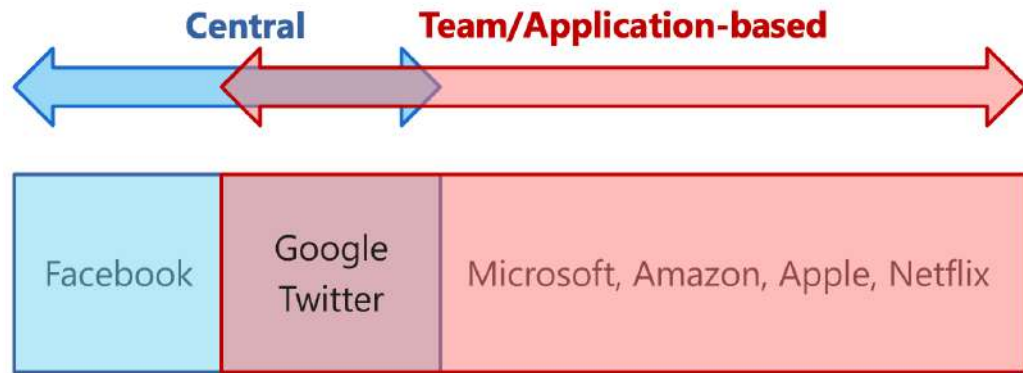
## Amazon DS Jobs:



## Apple DS Jobs:



# Twitter and Netflix:



Let's briefly talk about:

- Culture, WFH policy and office locations, Income, etc.
- Interview Process



# Questions to Think About, Hiring Process, Common Mistakes, ...

# Questions to Think About Carefully:

- Why so many people with engineering and science background are transitioning to data science roles? Is it really because of what [HBR](#) says?
- Why the churn and burnout rates are so high among data scientists? Fantasy vs. Reality? Link [1](#) and [2](#) and [3](#)
- Is it the right job for me? If yes, should I aim and plan for big corporates or smaller companies?
- What are the priorities of top companies? How these priorities impact the hiring process and my chance of getting an offer?
- Do we need a PhD for data science jobs?
- **There are so many topics to learn and cover and they are all evolving very fast, where should I start?**
- How important is the behavioral interview? How much time do I need to spend to prepare for it?
- What are the advantages and disadvantages of working for small (large) companies?
- The “X” factor! What can I do about it?

# Hiring Process:

- **Internship Application:**

- Don't underestimate the power of internship.
- How to get an interview for internship?
- What to expect during the interview?
  - Screening Interview
  - Onsite (1 - 2 Interviewers)
  - Expectations
- What to expect during the internship?
- What's the main advantage of going through internship?

- **Full-time Application:**

- How to get an interview for a full-time job?
  - Moving from small companies to large corporates
  - Moving from one large corporate to another large corporate
  - Moving from a large corporate to a small company
- What to expect during the interview?
  - Recruiter
  - Hiring Manager
  - Screening Interview
  - Onsite (5 – 14 Interviewers)
  - Expectations

**Getting an Offer:** Compensation Package, Negotiation, Rank, location, Immigration Sponsorship, etc.

# Common Mistakes to Avoid:

- **Referral:**

- Adding someone on LinkedIn and immediately asking for referrals
- Not knowing an employee at all and asking for referrals
- Asking for referrals for jobs that you are not qualified, don't have the skillset or the required experience
- Not paying attention to internal levels, or mixing levels at small and large companies

- **Resume:**

- Too many pages and/or too wordy
- Too many “in-depth” and “non-overlapping” skills
- Sci-Fi
- Listing something that you haven't done, or a skill that you have very shallow knowledge of it
- Listing papers for non-research jobs

- **Interview:**

- It's good to have backbone, but don't be arrogant!
- Not paying attention to behavioral questions at Amazon and Apple.
- Not doing your research on the company culture and products.
- Not thinking out loudly, not asking clarifying questions, going down through a rabbit hole

# Growing Your Career as a Data Scientist:

- **IC Roles:**

- Advantages and disadvantages
- You need to know many things to get the first job, but don't get trapped with being a "generalist"
- Come out of your comfort zone and explore new topics
- Find your own niche topic, invest in that and become "the specialist" in that topic
- Constantly pay attention to trends, evolving technologies and changing tools, otherwise...

- **People Management Roles:**

- Microsoft/Google/Apple vs Facebook/Amazon
- Advantages vs. disadvantages

**Question:** How to plan for and get a promotion?



# IC vs Manager Roles:



## Hamed S. Neshat

Senior Manager (ML/DS/AI lead), Research Data Science at Facebook

[View full profile](#)

Hamed S. Neshat replied to Seyyed Ali Towliat's comment on this



### Hamed S. Neshat • 1st

Senior Manager (ML/DS/AI lead), Research Data Science at Fa...

2w • Edited • ↻

۱- امروز کمی وقت گذاشتم و به صفحه همکاران ایرانی متصل مستقیم و غیر مستقیم در لینکدین نگاه کردم. تعداد زیادی از دوستانی که در ایران هستند، مدیر عامل و یا مدیر ارشد هستند، و بین آن ها بسیار کسانی هستند که اخیرا از دانشگاه فارغ التحصیل شدند. ۲- چند سال پیش در سفری که به ایران داشتم، با برخی از دوستان و همکاران که در شرکت های مختلف حوزه فناوری اطلاعات مشغول بودند جلسه ای دوستانه داشتیم؛ و همه دوستان از نبود "نیروی کار قوی" گلایه می کردند.

به نظر میاد مشاهده دوم از تاثیرات مشاهده اول است. گذراندن مدارج ترقی و شروع کردن از سمت های پایه، نه تنها در آینده شغلی می تواند تاثیر مثبت داشته باشد، بلکه دید وسیع تر و واقع گرایانه تری نسبت به تاسیس شرکت های جدید و همچنین سمت های مدیریتی خواهد داشت. فرهنگ "سمت گرایی" در کنار سابقه منفی ما در "مدرک گرایی"، در کوتاه مدت و بلند مدت بسیار مخرب اند.

پی نوشت: شرکت های بزرگ حوزه فناوری اطلاعات (تجربه دست اول از مایکروسافت، فیسبوک و گوگل)، ارزش بیشتری برای مهندسان ارشد قائل هستند به نسبت مدیران؛ و تمام تلاششان را در جهت راضی نگه داشتن مهندسان می کنند.

[See translation](#)

👍 🌱 🕒 113 • 10 comments



# Wrapping Up

# A Few Words of Advice:

- Find a good mentor!
- Companies have very different (and sometimes contradicting) cultures and values, don't take this lightly.
- Be curious and proactive, invest in learning new skills and tools, and know your value.
- It requires wisdom, courage and willpower to give up when it is necessary.
- In the end, do not choose a job or industry because it has a higher average salary or because it seems more prestigious to others. It does not matter if your title is Data Scientist or ML Engineer or Product Analyst, you can make a good and comfortable living out of all these jobs.
- *"Life is not fair, get used to it."* – Bill Gates
- There is a big difference between "industry" and "academia".
- Where you are today is the outcome of what you planned and executed 5 years ago, where you will be in 5 years depends on what you plan today.
- You are not a tree! If you don't like where you are, move.
- Employees don't leave companies, they leave managers!
- Develop a thick skin! Be patient, be patient, and be patient!

# What Else?

- Be honest with yourself, what are your strengths, weaknesses, passion, and priorities in life?
- Decide what you want to achieve, be realistic and think “long term”.
- Make sure you understand what skills you need to succeed.
- Do your homework (LeetCode, Product, Statistics, A/B, ML) diligently and consistently.
- Talk to people from the same company, discipline, group and team and make sure you understand the culture, focus, requirements and priorities.
- Don't get caught up with what you cannot control (ex.: the X factor), instead focus on what you can control (ex.: research about the team and interview loop, improving your coding skills, etc).
- For companies like Amazon and Apple, be mindful of the behavioral loop and make sure to practice enough and get ready for it.
- If you got an onsite interview, don't leave anything up to chance! Make sure you understand the interview loop, do your homework and prepare for it.
- Learn how to come out of your comfort zone.
- Have backbone and be humble.

# Some useful pointers to check out:

- **StatQuest:** Basic and fundamental concepts, link [1](#) and [2](#) Plus **Penn State** Stat 501 to 510
- **MOOCs:** Udacity, edX, Coursera, Deeplearning.ai, DataCamp, Udemy
- **Research Arms of Tech Giants:**
  - Microsoft Research ([link](#))
  - Facebook FAIR ([link](#))
  - Google Brain ([link](#)) and DeepMind ([link](#))
  - Amazon Research ([link](#))
- **General Machine Learning:** Stanford ([link](#)), Coursera ([link](#)), e-Book ([link](#))
- **General Deep Learning:** e-Book ([link](#)), Programs on Coursera and Udacity
- **NLP:** Oxford ([link](#)), Stanford ([link](#)), Michigan ([link](#))
- **SHAP:** link [1](#) and [2](#) and [3](#) and [4](#)
- **LightGBM:** Very efficient regression and classification tool ([link](#))
- **Causal Impact:** link [1](#) and [2](#)
- **Time Series and Forecasting:** Prophet ([link](#)) and Kats ([link](#))
- **GAMLSS:** Generalized Additive Models for Location, Scale and Shape ([link](#))
- **DoubleML:** link [1](#) and [2](#) and [3](#) and [4](#)
- **Coding:** LeetCode ([link](#))
- **A/B:** Ron Kohavi's Book ([link](#)), Online Course ([link](#))
- **Conferences:** ICML, NIPS, KDD, CVPR, ...
- **Career Information:** levels.fyi ([link](#)), Blind ([link](#))



# Questions and Discussion

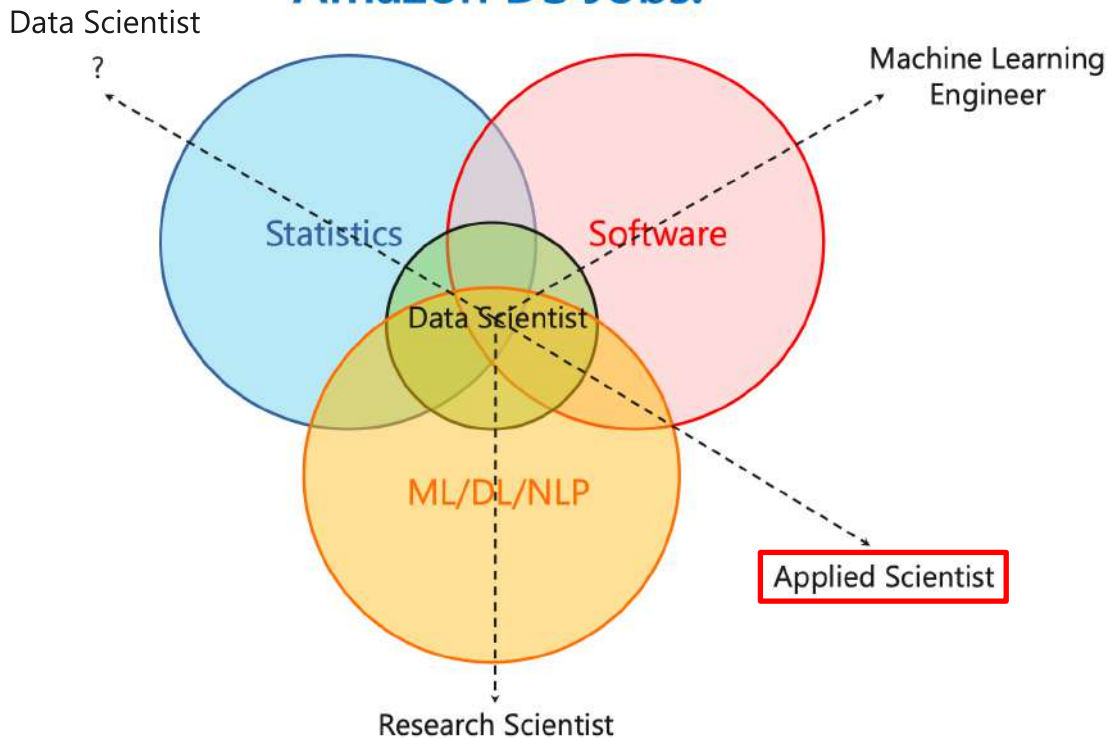


# Appendix

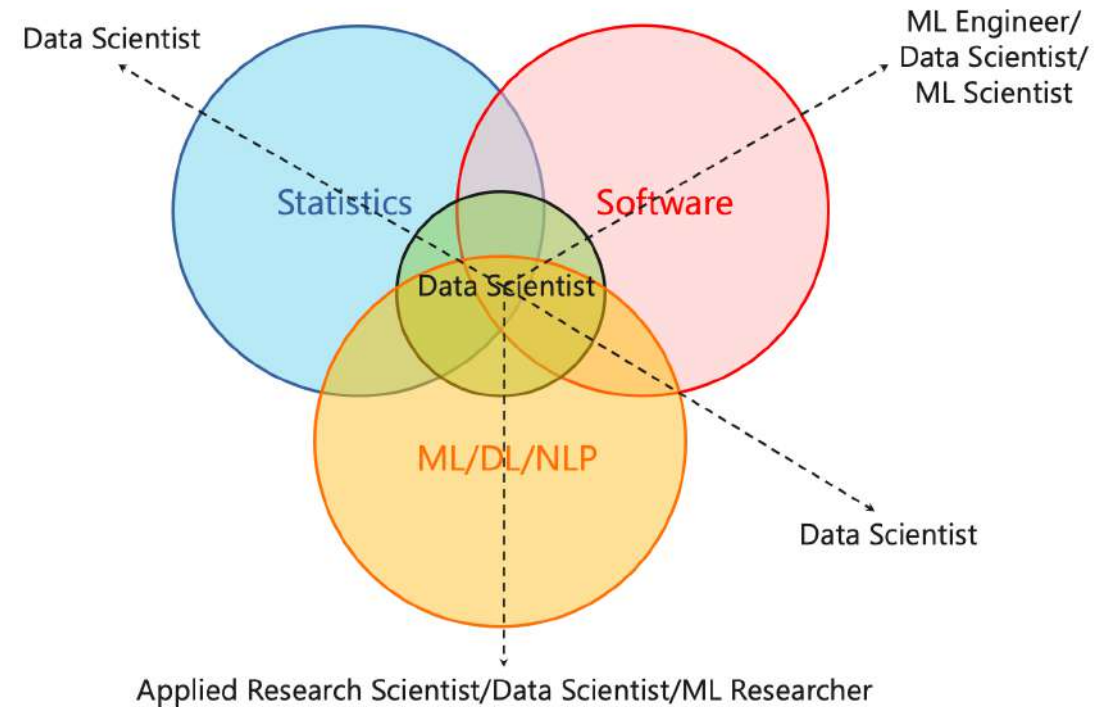
# Amazon and Apple:

- Amazon and Apple are both Product/PM-based companies and very similar to Microsoft in terms of expectations from data scientists, definition of roles and “team-based” hiring process.
- Amazon has clear definitions of (and distinctions between) Data Scientist and Applied Scientist roles.

## Amazon DS Jobs:



## Apple DS Jobs:





# [Sample JD] Apple Data Scientist:

- Job descriptions are general, can vary from pure statistics to pure machine learning or coding, all under “Data Scientist” title.
- Job descriptions and hiring process are completely “team-dependent”.

- Extensive knowledge in statistical methods and data mining.
- Excellent application skills in experimentation methodologies and causal inferences.
- Experience with common data science toolkits, such as R/Python and libraries like pandas, dplyr and ggplot.
- Proficiency in using query languages such as SQL and Hive.
- Experience with Big Data systems and distributed computing, such as Hadoop and Spark.
- Excellent communication skills with the ability to explain findings in layman terms and influence decision makers.
- Attention to data detail with regards to quality, transformation and potential impact.

**JD 1**, Title: Data Scientist

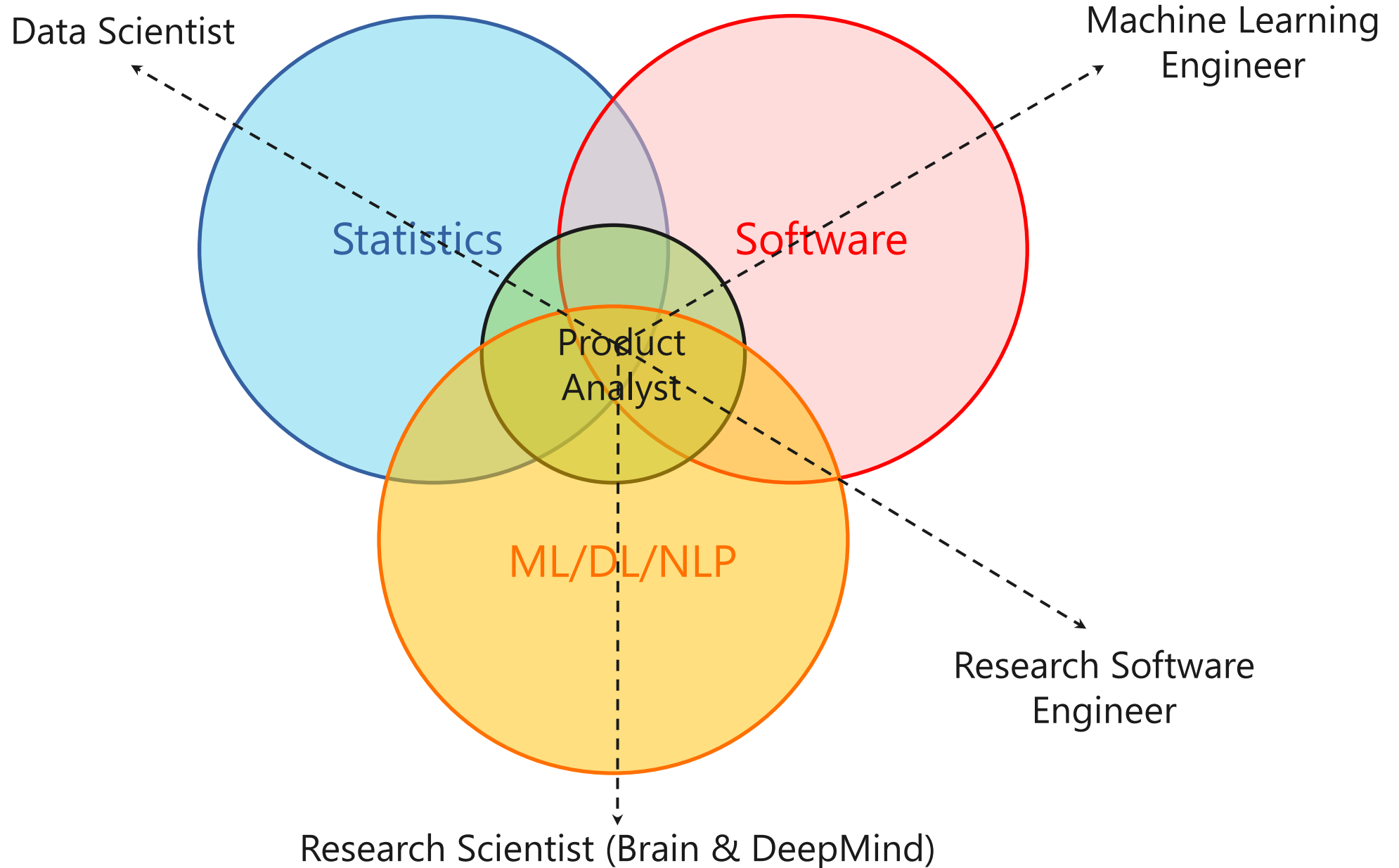
- Minimum of 2 years of experience in one or more of the following: deep learning and traditional ML modeling in NLP, information retrieval, recommendation and personalization systems.
- Minimum of 2 years of experience in algorithm design, modeling, or quantitative analysis.
- Minimum of 2 years of experience with big data systems (e.g. Spark and Hadoop) with TB to PB scale datasets.
- Minimum of 2 years of experience implementing end-to-end data science and machine learning projects. Python is strongly preferred.
- Experience with deep learning framework such as Pytorch, TensorFlow and Keras.
- Experience with optimization, approximation algorithms, distributed algorithm design, and hands-on implementation of these techniques.
- A background in computer science, information science, or similar quantitative field. Ph.D. is strongly preferred.

**JD 2**, Title: Data Scientist

The background of the image is a solid blue gradient. Overlaid on this is a complex network diagram consisting of numerous small, light-colored circular nodes connected by thin, light-colored lines. The nodes are scattered across the frame, and the lines form a dense, interconnected web of connections, resembling a social network or a data network. The overall aesthetic is clean, modern, and technical.

Google

# Google DS Jobs:



# Google DS- Related Roles:

- **Data Scientist**

- JD Keywords: Statistics, Python, R, SQL, Statistical Inference, (usually PhD in Statistics/Bioinformatics/Economics), A/B Experimentation, (LeetCode for Python and SQL), ...

- **Product Analyst**

- JD Keywords: Python, R, SQL, (MSc or PhD in CS/Eng), Experimentation, (LeetCode for Python and SQL), Product Sense, Metrics, Communication Skills, Statistics, ...

- **Machine Learning Software Engineer**

- JD Keywords: C++, Java, Python, Go, Algorithms, Systems Design, Distributed Systems, TensorFlow, ...

- **Research Scientist**

- JD Keywords: Python, C++, Publication, (usually PhD in CS)...

Hiring Levels: L3 – L10

# [Sample JD] Google Data Scientist:

## Minimum qualifications:

- Master's degree in a quantitative discipline (e.g., **Statistics, Bioinformatics, Economics**) or equivalent practical experience.
- 10 years of experience in data analysis or related field as a **Statistician**, Data Scientist, Computational Biologist, or bioinformatician.
- 5 years of experience in people management or leadership.
- Experience with statistical software (e.g., **R, Python**, MATLAB, pandas) and **database** languages.

## Preferred qualifications:

- **Doctorate's** degree in a quantitative discipline (e.g., **Statistics**, Operations Research, **Bioinformatics, Economics**, Computational Biology, Computer Science, Mathematics, Physics, Electrical Engineering, Industrial Engineering).
- 12 years of tech industry work experience as a statistician, bioinformatician, or data scientist. Experience in statistical data analysis, such as linear models **multivariate analysis, stochastic models, and sampling methods**.
- 8 years of leadership experience, including people management.
- Applied experience with machine learning on large datasets.
- Experience articulating business questions and using mathematical techniques to arrive at an answer using available data.
- Ability to select the right statistical tools given a data analysis problem.

# [Sample JD] Google Product Analyst:

## Minimum qualifications:

- Bachelor's degree or equivalent practical experience.
- 3 years of experience working with statistical packages (e.g. R, SAS, Stata, MATLAB, etc.).
- Experience articulating product questions, pulling data from datasets (i.e. SQL) and using statistics to arrive at an answer.

## Preferred qualifications:

- Master's degree in a quantitative field (e.g., Statistics, Computer Science, Engineering, Mathematics, Data Sciences).
- Experience in scripting with SQL, extracting large sets of data, and design of ETL flows.
- Experience in an inter-disciplinary/cross-functional field.
- Interest and aptitude in data, metrics, analysis, and trends and applied knowledge of measurement, statistics, and program evaluation.
- Ability to translate analysis results into business recommendations.
- Ability to speak and write in Japanese fluently.

# [Sample JD] Google ML Software Engineer:

## Minimum qualifications:

- 5 years of **coding** experience in one or more of the following languages: C, **C++, Java, or Python**.
- Experience with exploration and analysis across datasets (e.g. SQL, JavaScript, R, Python).

## Preferred qualifications:

- Experience utilizing Computer Vision
- Experience with Video Processing
- Modeling experience with TensorFlow

# [Sample JD] Google Research Scientist:

## Minimum qualifications:

- PhD degree in Computer Science or a related technical field, or equivalent practical experience
- Experience programming in Python, working with multi-processing libraries. Experience with deep learning frameworks (e.g. JAX)

## Preferred qualifications:

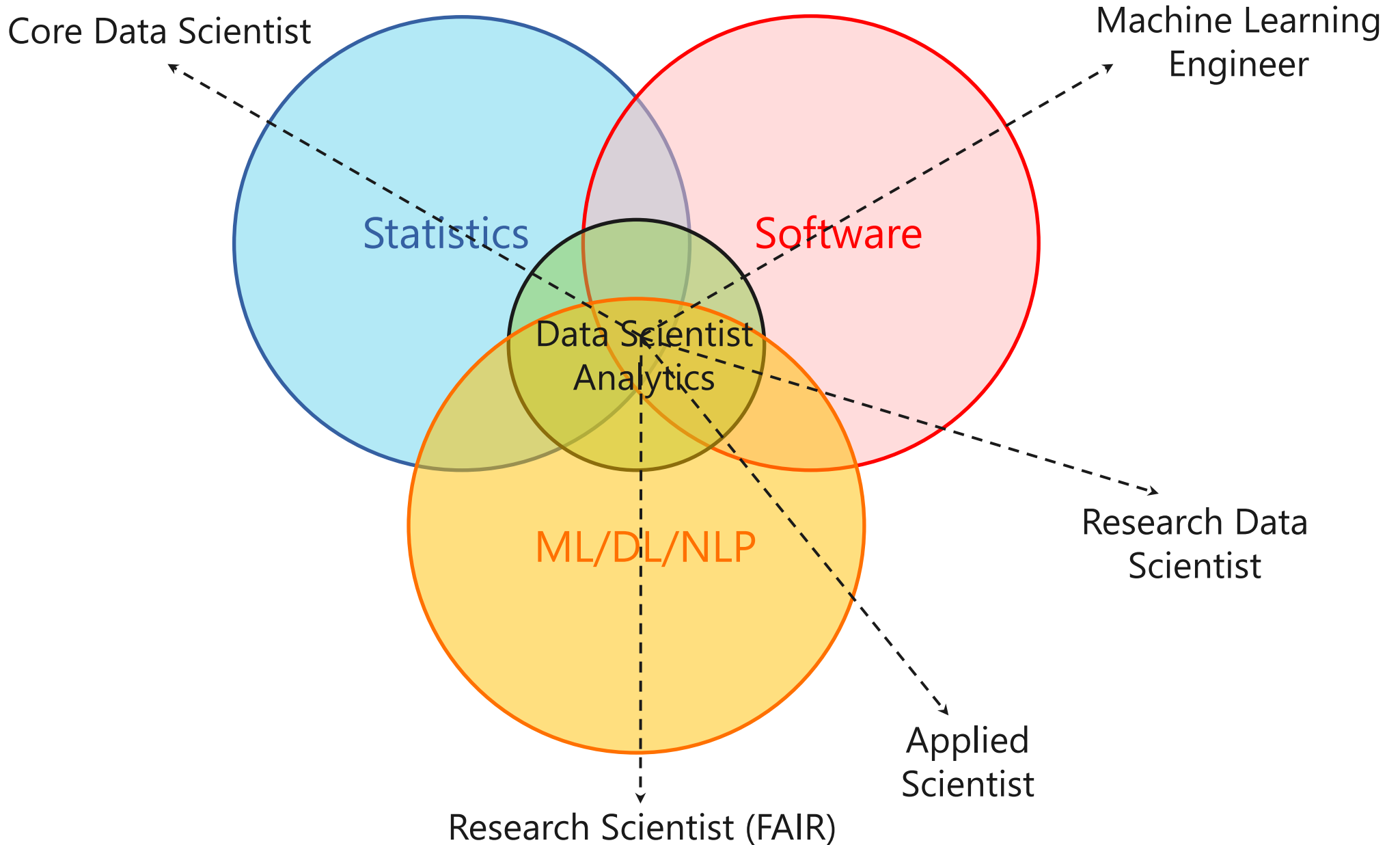
- Contributions to research efforts, including publications in journals such as Complex Systems, ALIFE, or conferences such as GECCO, or publication of articles related to complex systems topics (self-organization, cellular automata, collective intelligence, open-endedness, etc) in recognized machine learning or scientific venues such as NeurIPS, ICLR, ICML, AAAI, JMLR, CVPR, IROS, Nature, etc.
- Contributions to open-source projects that embody novel technologies related to complex systems; creation of interactive web demos or visualizations of complex systems; organizing workshops or other community-building efforts related to the field.
- Excellent written and verbal communication skills to disseminate research findings.





Facebook

# Facebook DS Jobs:



# Facebook DS- Related Roles:

- **Data Scientist - Product**

- JD Keywords: Python, R, Hadoop, Presto, Hive, SQL, Product Analytics, A/B Experimentation, Statistics, Communication Skills, Metrics, ...

- **Research Data Scientist**

- JD Keywords: Python, R, Hadoop, Presto, Hive, SQL, Statistics, Forecasting, Algorithms, Time Series, Infrastructure, Production Code, (A/B Experimentation?), Optimization, Clustering, Classification, Regression, ...

- **Applied Scientist**

- JD Keywords: C++, Java, Python, Hadoop, Presto, Hive, SQL, Forecasting, Algorithms, Production Code, Optimization, Machine Learning, Deep Learning, ...

- **Machine Learning Engineer**

- JD Keywords: C++, Java, Python, System Design, Algorithms, Machine Learning, Hadoop, Presto, Hive, (PHP), ...

- **Research Scientist**

- JD Keywords: Python, C++, Publication, (usually PhD in CS)...

Hiring Levels: E3 – E9

# [Sample JD] Facebook Data Scientist - Product:

## Data Scientist - Instagram Stories Responsibilities

- Develop a framework to understand impact to both producers and consumers of Instagram Stories
- Set north star **metric** goals for Instagram Stories
- Present results and recommendations to Instagram leadership
- Ability to **communicate** the results of analyses with **product and leadership** teams to influence the strategy of the product
- Define problems and opportunities in a complex or ambiguous area

## Minimum Qualifications

- Bachelors/**Masters Degree with 4+ years (or PhD with 2+ years)** of experience working within an analytical role
- Experience in **querying** and manipulating raw datasets for analysis
- Understanding of statistics (**hypothesis testing, regressions** etc.)
- Experience working independently
- **Communicator experience explaining technical concepts and analysis implications to varied audiences and can translate business objectives into actionable analysis**

## Preferred Qualifications

- 7+ years of experience working in an analytics organization
- Experience with Hive, **R**, and **Python** preferred
- Experience presenting to executive level audiences preferred
- Experience driving **cross functional** team

# [Sample JD] Facebook Applied Scientist:

- Build pragmatic, scalable, and rigorous scientific solutions to drive enterprise productivity by leveraging or developing state of the art machine learning and optimization methodologies on top of Facebook's unparalleled data infrastructure
- Apply excellent communication skills in order to develop cross-functional partnerships and spread scientific best practices
- Be able to work both independently and collaboratively with cross-functional teams across the company, including scientists, engineers, designers, data engineers, and product managers to accomplish complex tasks and build scalable tools that deliver step-level gains to the enterprise
- Think creatively, proactively, and futuristically to identify new opportunities that will grow the enterprise's long-term roadmap and bring productivity gains for the enterprise

## Minimum Qualifications

- Ph.D. degree in computer science, operations research, electrical engineering, statistics, math, quantitative finance, physics, and related fields, or MS degree with 4+ years of experience developing and supporting models in production
- 2+ years experience in building models and developing algorithms for machine learning, statistics, mathematical programming, and simulation
- 2+ years experience in modeling and analyzing up to petabytes of data using R or Python
- Experience to initiating and driving applied research projects to completion with minimal guidance
- Experience communicating applied research work to technical and non-technical audience

## Preferred Qualifications

- Experience in lower level languages such as C++, Java
- Experience in scalable dataset assembly/data wrangling, such as Presto, Hive or Spark
- Familiarity with enterprise-wide application development life-cycle

# [Sample JD] Facebook Research Data Scientist:

## Research Data Scientist Responsibilities

- Build pragmatic **scalable**, and statistically rigorous solutions to large-scale web, mobile and data **infrastructure** problems by leveraging or developing state-of-the-art **statistical** and **machine learning** methodologies on top of Facebook's unparalleled data infrastructure
- Work cross-functionally to define problem statements, **collect data**, build analytical models and make recommendations
- Build and maintain data driven optimization models, **experiments**, **forecasting** algorithms, and machine learning models
- Leverage tools like **Python, R, Hadoop & SQL** to drive efficient analytics
- Communicate final recommendations and drive decision making

## Preferred Qualifications

- Advanced degree (Master's or PhD or Equivalent experience) in quantitative field
- Experience working with distributed computing tools (**Hadoop, Hive, Spark**, etc.)
- **Proficiency in algorithmic complexity**

## Minimum Qualifications

- Degree in quantitative field (e.g. Computer Science, Engineering, Mathematics, Statistics, Operations Research or other related field)
- 2+ years of industry or graduate research experience solving analytical problems and building models using quantitative, statistical or machine learning approaches
- Experience with Machine Learning, Statistics, or other data analysis tools and techniques
- Experience performing data extraction, cleaning, analysis and presentation for medium to large datasets
- Experience with at least one programming language (i.e. Python, R, Java, or C++)
- Experience writing SQL queries
- Experience with scientific computing and analysis packages such as **NumPy, SciPy, Pandas, Scikit-learn, dplyr, or ggplot2**
- Experience with statistics methods such as forecasting, time series, hypothesis testing, classification, clustering or regression analysis
- Experience with data visualization libraries such as **Matplotlib, Pyplot, ggplot2**
- Experience with machine learning libraries and packages such as **PyTorch, Caffe2, TensorFlow, Keras** or Theano

# [Sample JD] Facebook Machine Learning Engineer:

## Software Engineer, Machine Learning Responsibilities

- Consistently advance the state of Machine Learning for your problem, including setting and executing against roadmaps for 6-month+ timeframes and influencing business and product strategy.
- Define projects for other engineers to solve and achieve impact based on your direction.
- Own the full Machine Learning life cycle (<https://research.fb.com/the-facebook-field-guide-to-machine-learning-video-series/>) for a significant new product, including production quality.
- Code deliverables in tandem with the engineering team

## Preferred Qualifications

- MS degree in Computer Science or related quantitative field with 5+ years of machine learning related work or research, or PhD degree in Computer Science or related quantitative field
- Experience with filesystems, server architectures and distributed systems

## Minimum Qualifications

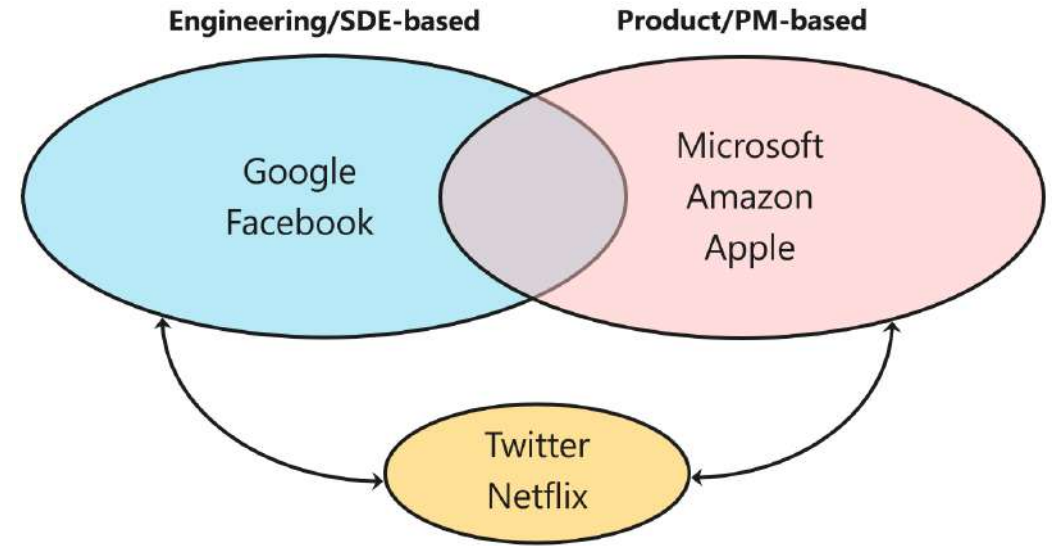
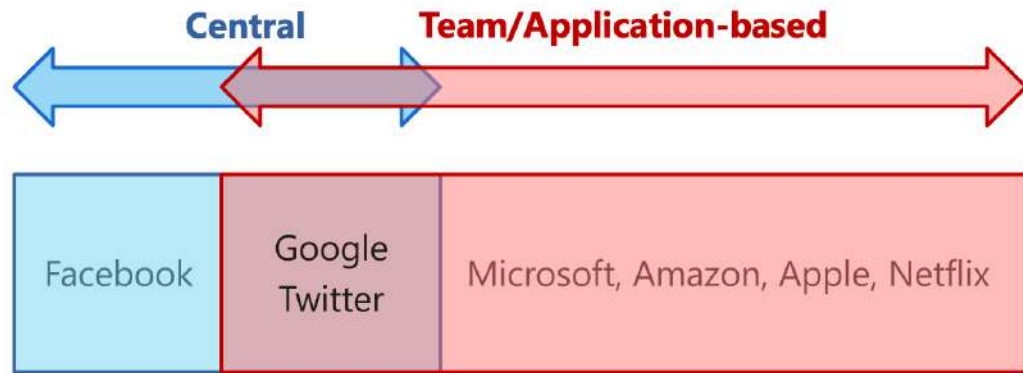
- PHD in Computer Science with 3+ years or Masters in Computer Science with 7+ years of experience in one or more of the following areas: machine learning, recommendation systems, pattern recognition, data mining or artificial intelligence
- Proven ability to translate insights into business recommendations
- Experience with Hadoop/HBase/Pig or MapReduce/Sawzall/Bigtable/Hive/Spark
- Expert knowledge in building and training machine learning models or developing production level Machine Learning products in Java, C++ or Python
- Expert with scripting languages such as Python, Perl, PHP, and/or shell scripts
- Experience developing and debugging in Java, C++ or similar
- Experience of demonstrating technical leadership working with teams, owning projects, defining and setting technical direction for projects.



# Twitter and Netflix



# Twitter and Netflix:



Let's briefly talk about:

- Culture, WFH policy and office locations, Income, etc.
- Interview Process