



Hello!

I am Miguel Rios Berrios

I lead Consumer Data Science at Twitter

@miguelrios

Scaling **Data Science** teams





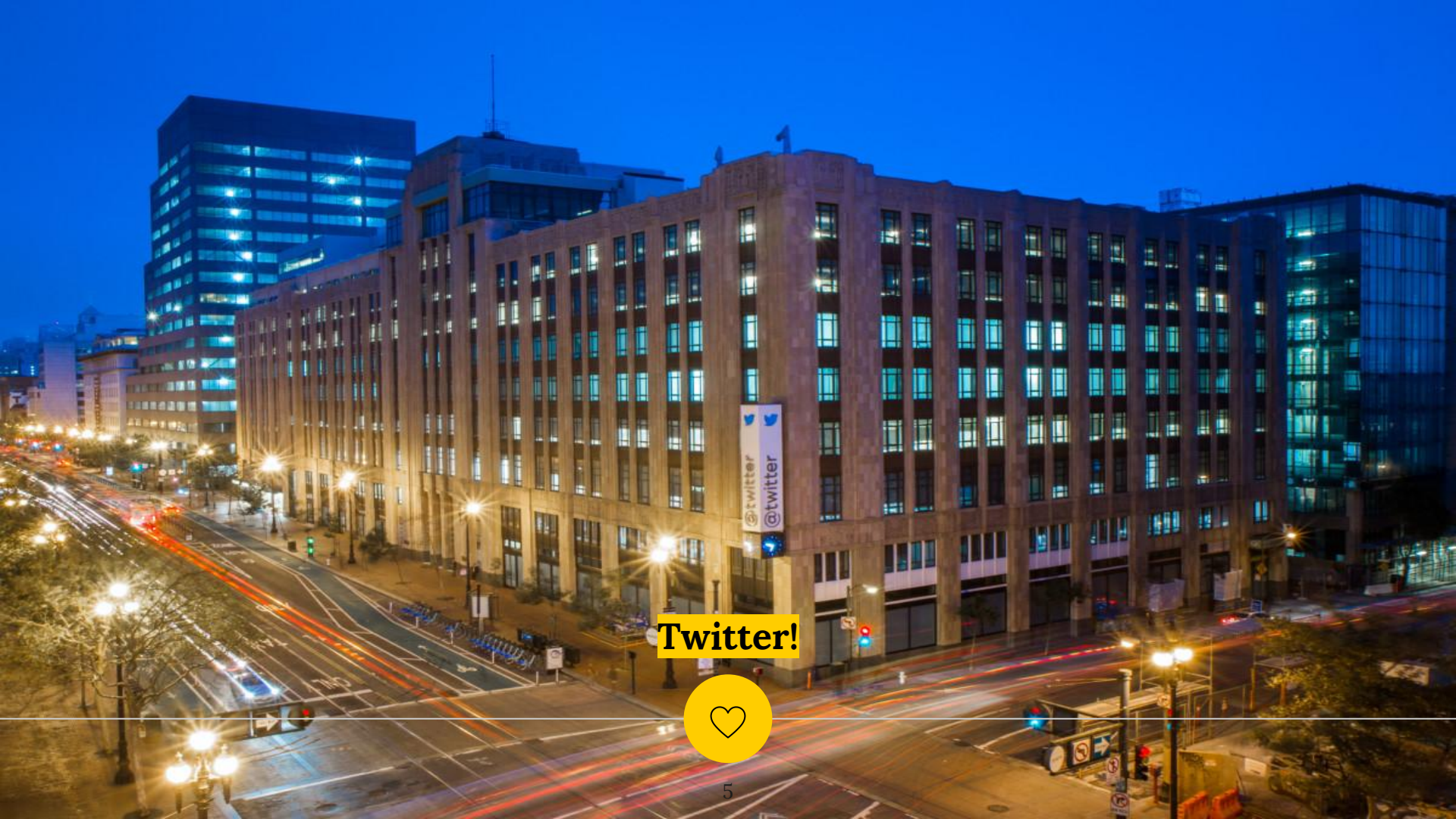
Before we start

A little bit of info about me

A complex network visualization on a black background. It features a dense web of lines, primarily green and red, forming a large, curved, and somewhat symmetrical shape. The lines are thin and numerous, creating a textured, almost brush-stroke-like appearance. The green lines are more numerous and form the main body of the shape, while the red lines are more prominent in the upper and right portions. The overall effect is one of dynamic, interconnected data.

Data Visualization



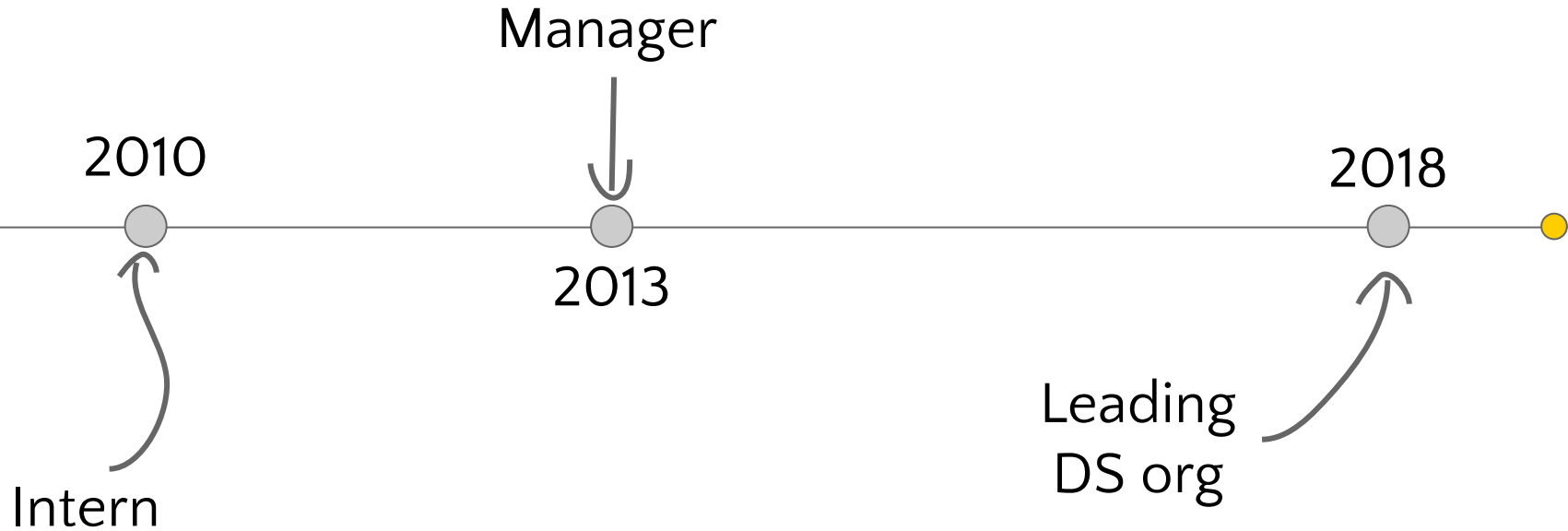


Twitter!





My Timeline at Twitter





Twitter's Timeline

2010



150

employees

+4000

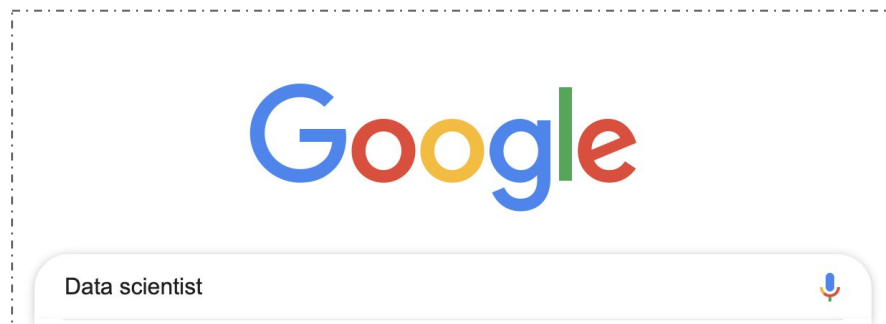
employees



2019



“Data Scientist” Timeline



2010

2019



Before I go on: **some caveats**

- This talk is based on one data point
- Well, more like a 9-year longitudinal study
- I will present insights out of my own experience (and opinion)
- YMMV – industry category, cultural differences, company size, etc. *Take this with a grain of salt.*



What is Data Science?

I won't spend much time on this...

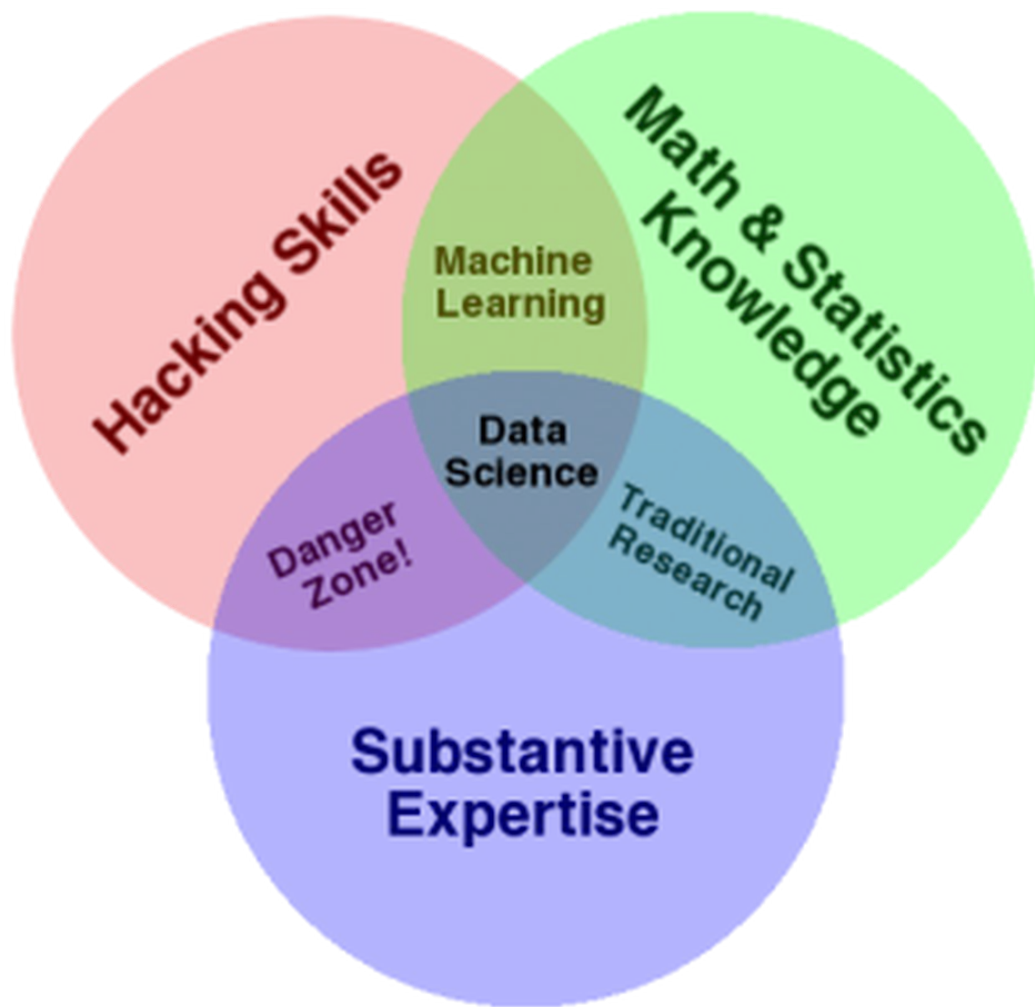


Data science

From Wikipedia, the free encyclopedia

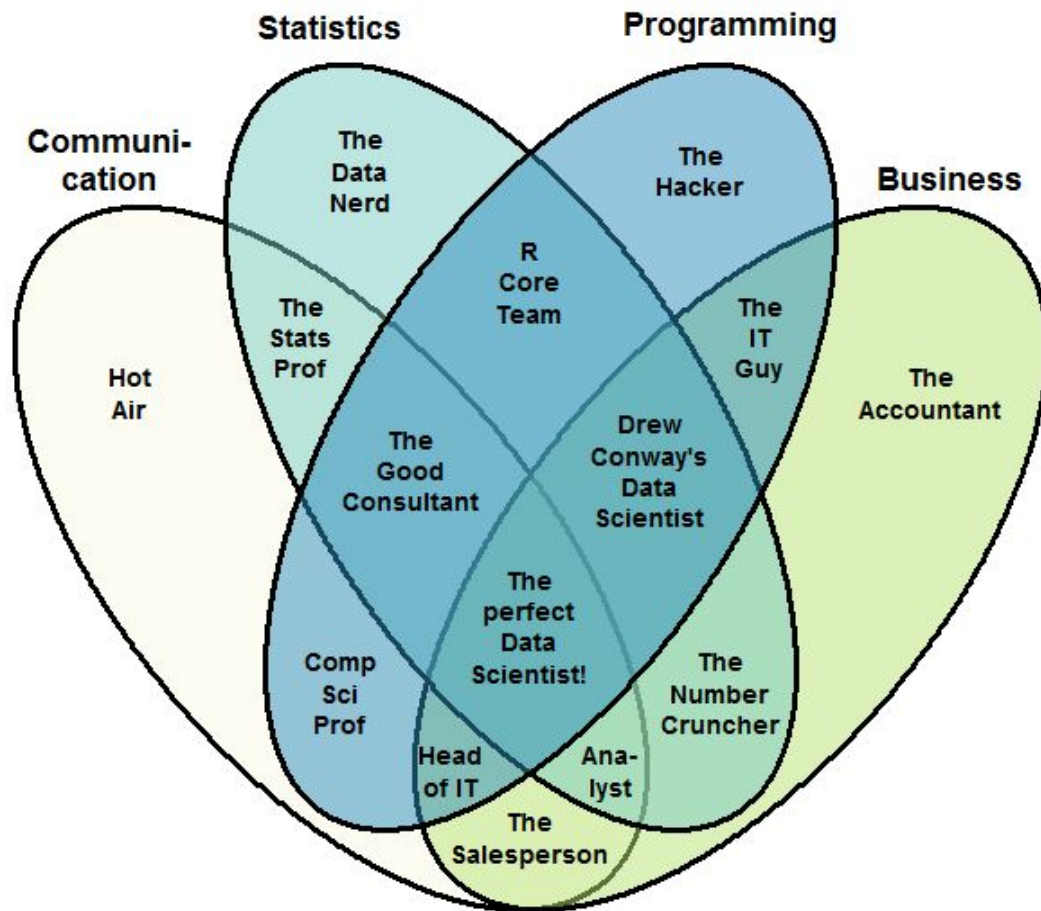
Not to be confused with [information science](#).

Data science is a [multi-disciplinary](#) field that uses scientific methods, processes, algorithms and systems to extract [knowledge](#) and insights from structured and unstructured [data](#).^{[1][2]} Data science is the same concept as [data mining](#) and [big data](#): "use the most powerful hardware, the most powerful programming systems, and the most efficient algorithms to solve problems".^[3]





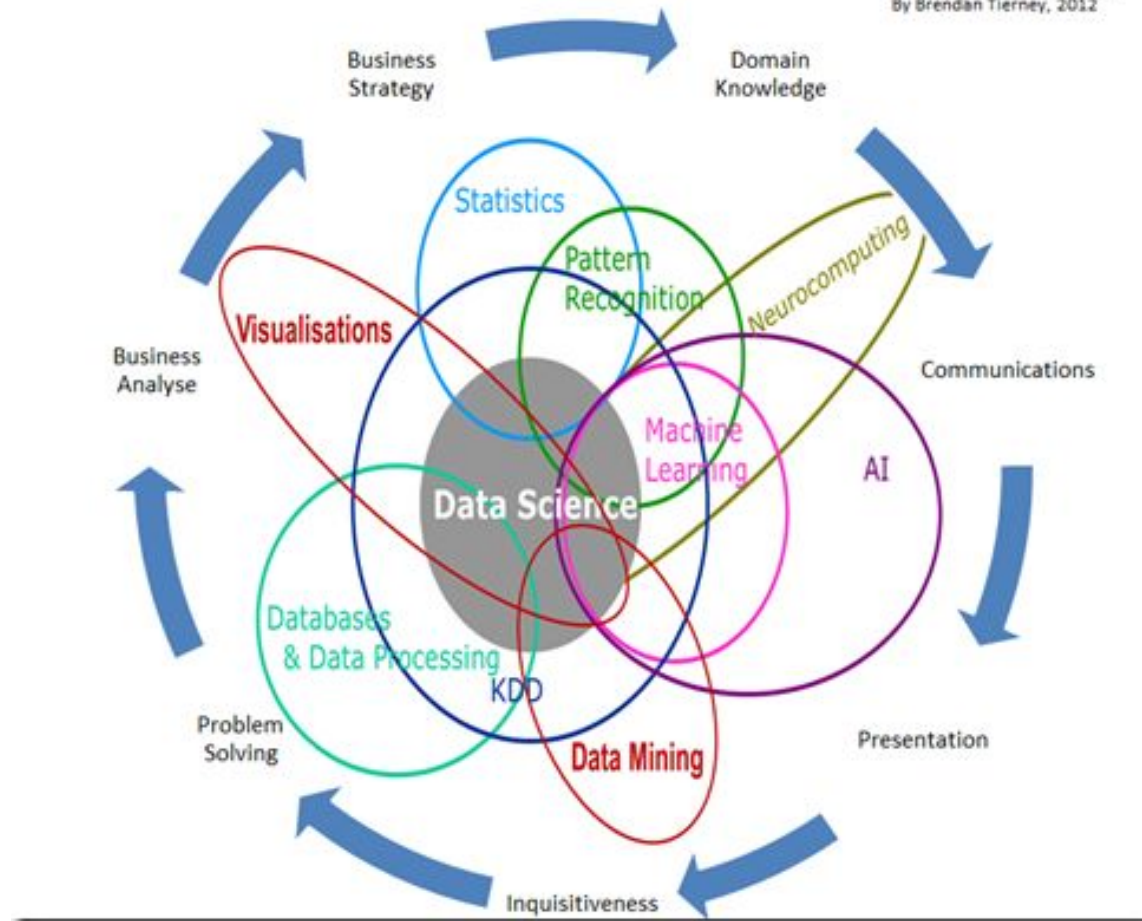
The Data Scientist Venn Diagram





Data Science Is Multidisciplinary

By Brendan Tierney, 2012



[Follow](#)[ATEST](#) [DEV](#) [AI](#) [LEARN BLOCKCHAIN](#) [UPCOMING PODCAST GUESTS](#) [BLOCKCHAIN EXPERTS](#) [CRYPTO LIFE](#)[Cassie Kozyrkov](#)[Follow](#)

Chief Decision Intelligence Engineer, Google. ❤️ Stats, ML/AI, data, puns, art, theatre, decision science. All views are my own. twitter.com/quaesita

Aug 18 · 7 min read

What on earth is data science?

The quest for a useful definition

**Krista Seiden**[@kristaseiden](#)[Follow](#)

A Data Scientist is a Data Analyst who lives in California [#eMetrics](#) [#measure](#)

1:57 PM - 6 Mar 2012

7 Retweets 1 Like



1



7



1

**O'REILLY®****Safari**[Sign In](#)[START FREE TRIAL](#)

Doing Data Science by Cathy O'Neil, Rachel Schutt

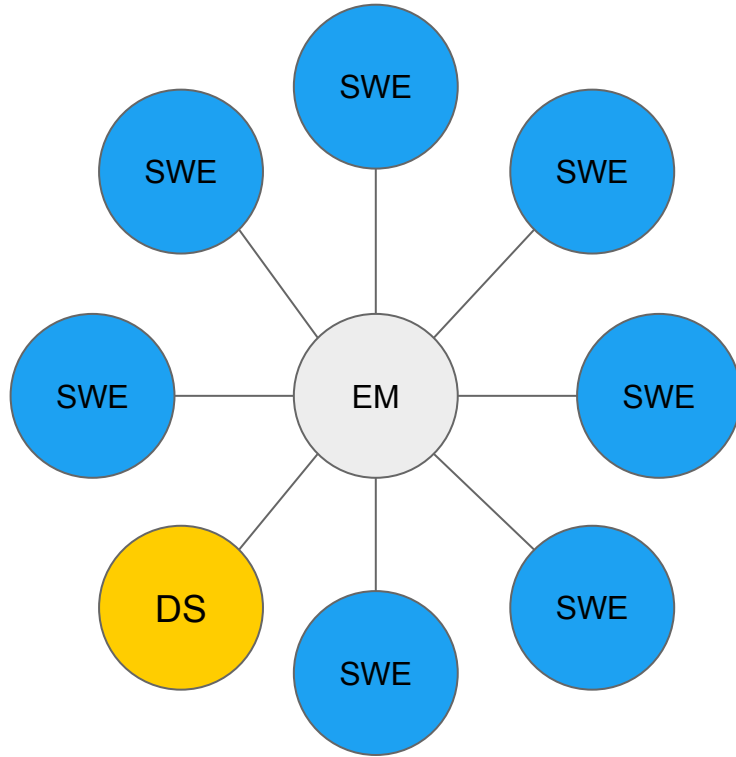
Chapter 1. Introduction: What Is Data Science?

1

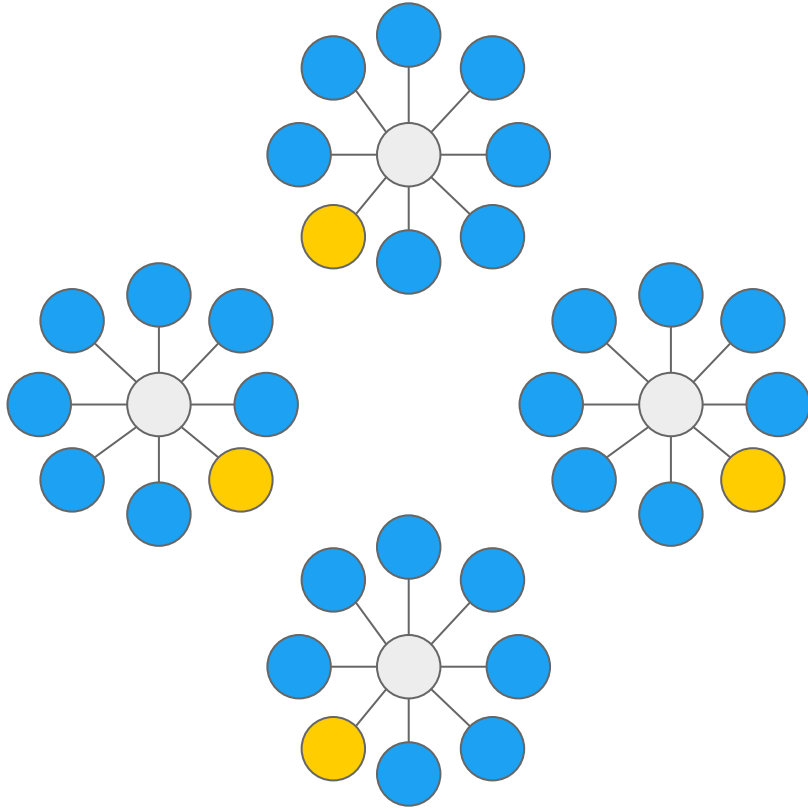
Structuring teams

Where data scientists fit in an engineering-driven organization

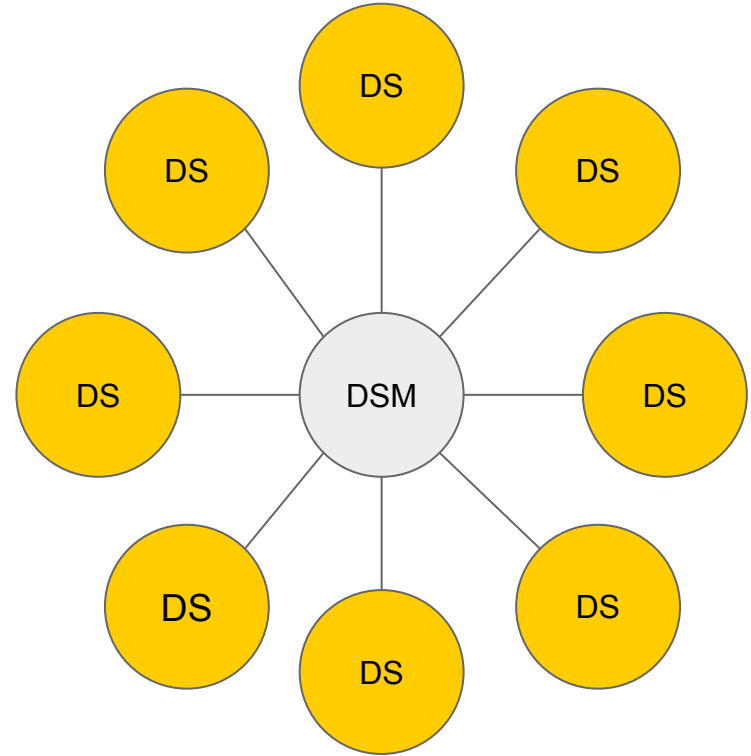
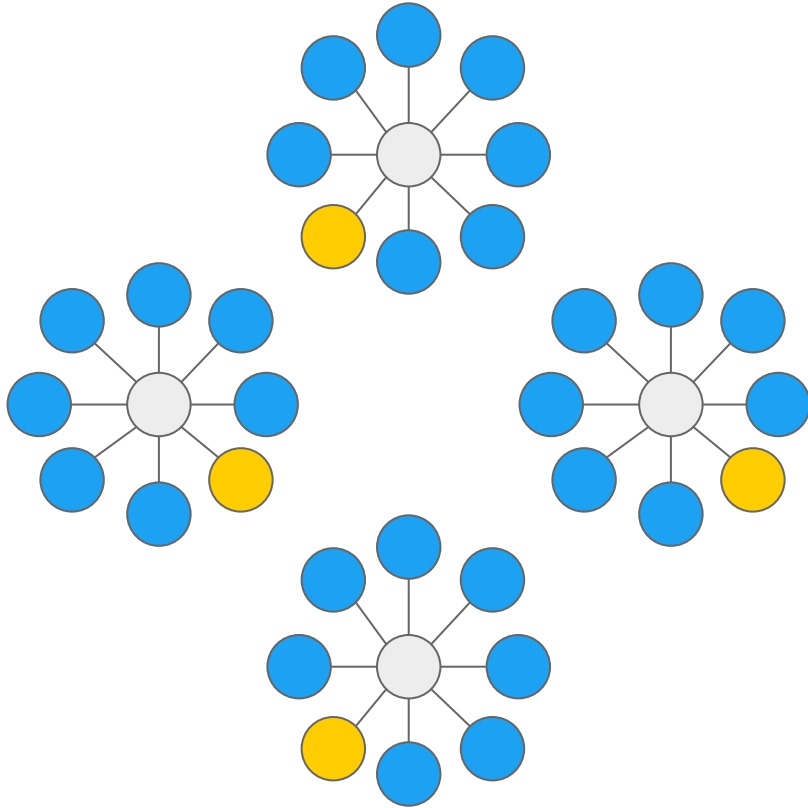
Embedded vs. Centralized data science practices



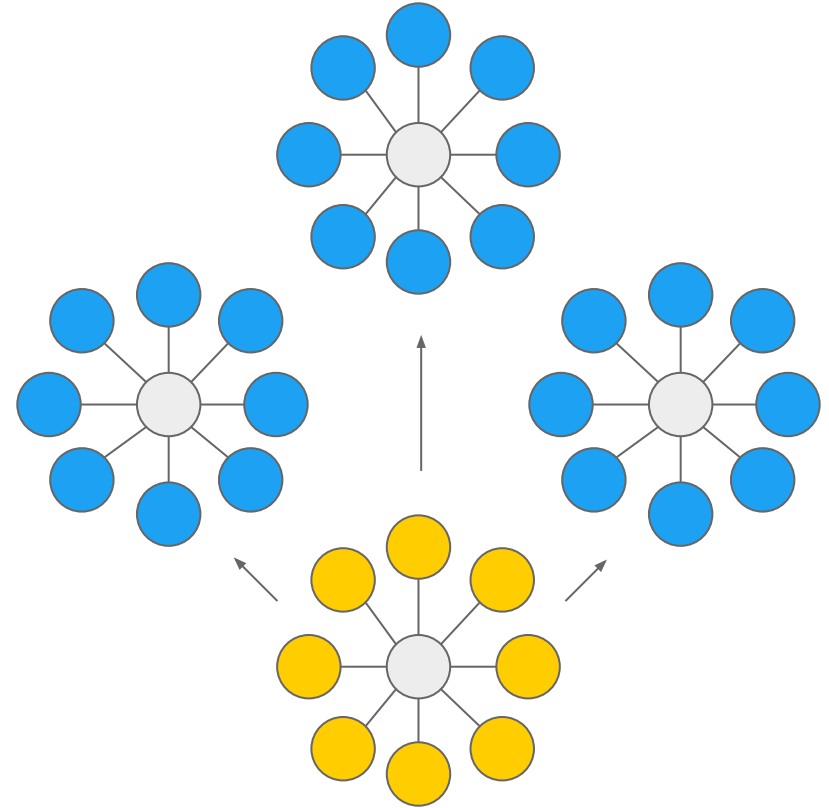
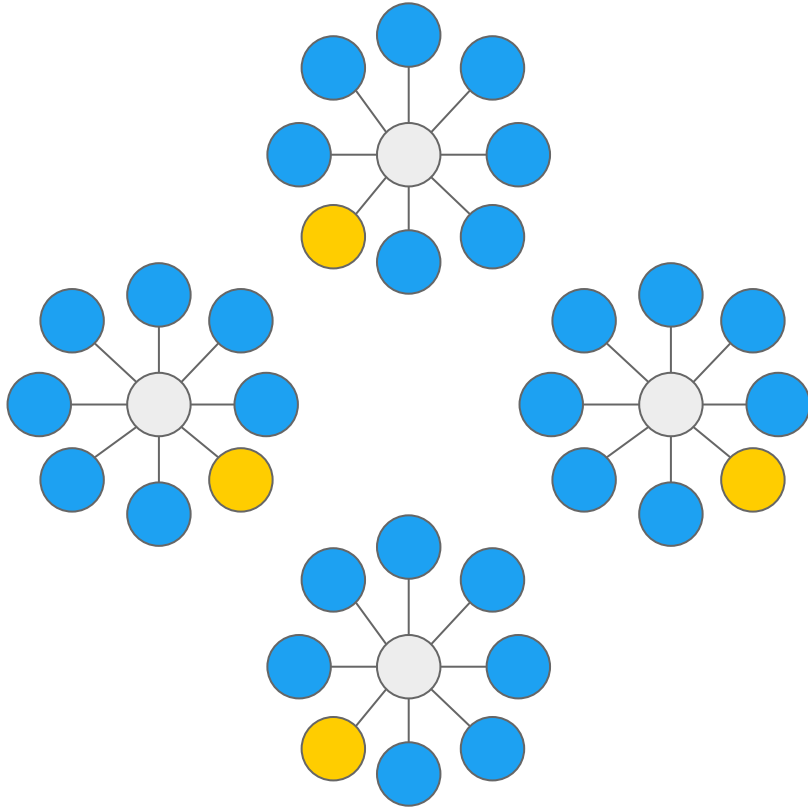
Embedded vs. Centralized data science practices



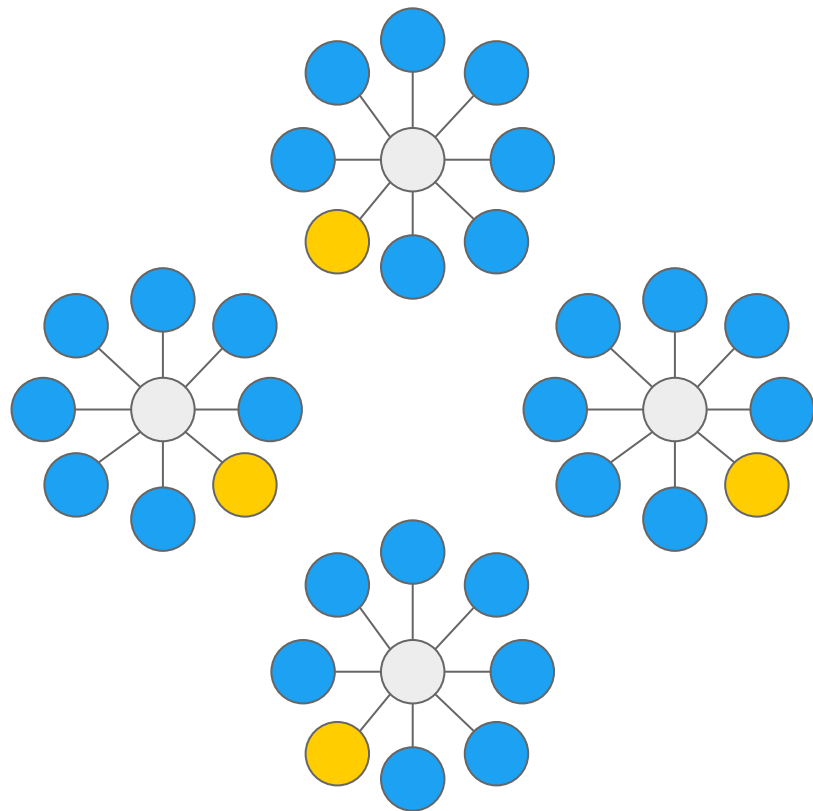
Embedded vs. Centralized data science practices



Embedded vs. Centralized data science practices



Embedded pros and cons



Pros

- Dedicated data science resourcing
- Alignment between DS and the rest of the team
- One roadmap, fewer dependencies
- Data science has a more natural “seat at the table”

Cons

- Rigid resourcing (harder to move DS between teams)
- Barriers for collaboration between data scientists
- Manager may not have domain knowledge (typically a EM)
- Risk of Data Science being a support or service to eng. team

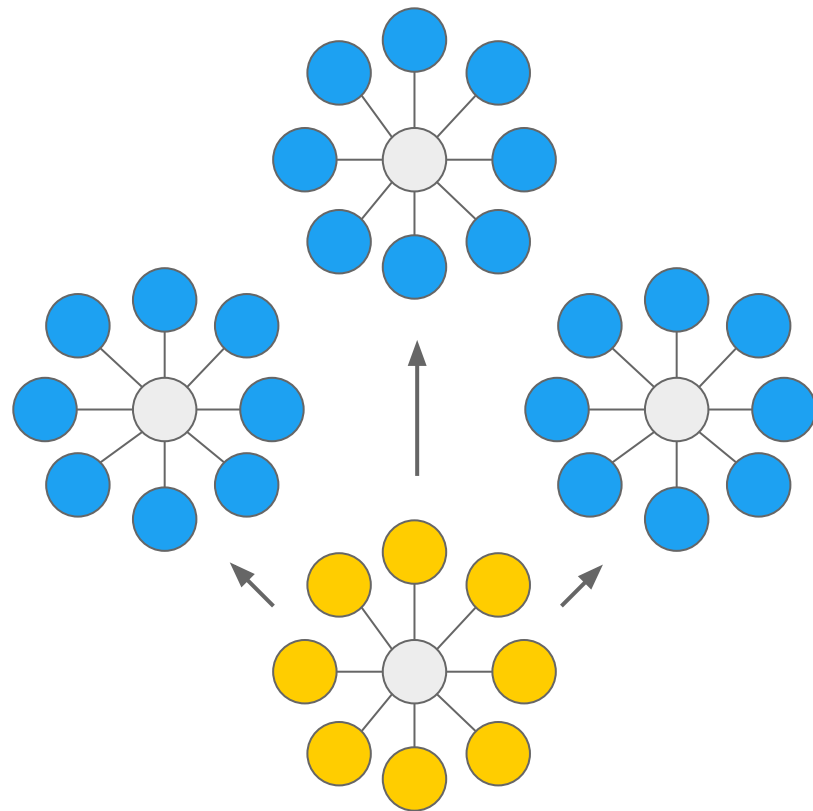
Centralized pros and cons

Pros

- Data scientists working together (collaboration and knowledge sharing)
- DS manager has domain knowledge (better career dev)
- Resources can be rebalanced to meet customer demand
- Advocacy for better and consistent tech (tooling, datasets, etc)

Cons

- Coordination between teams (DS and stakeholder) becomes more complicated
- In eng. centric orgs the DS teams need to influence eng roadmap
- Risk of data science work not being aligned with product
- Company needs to support one more function

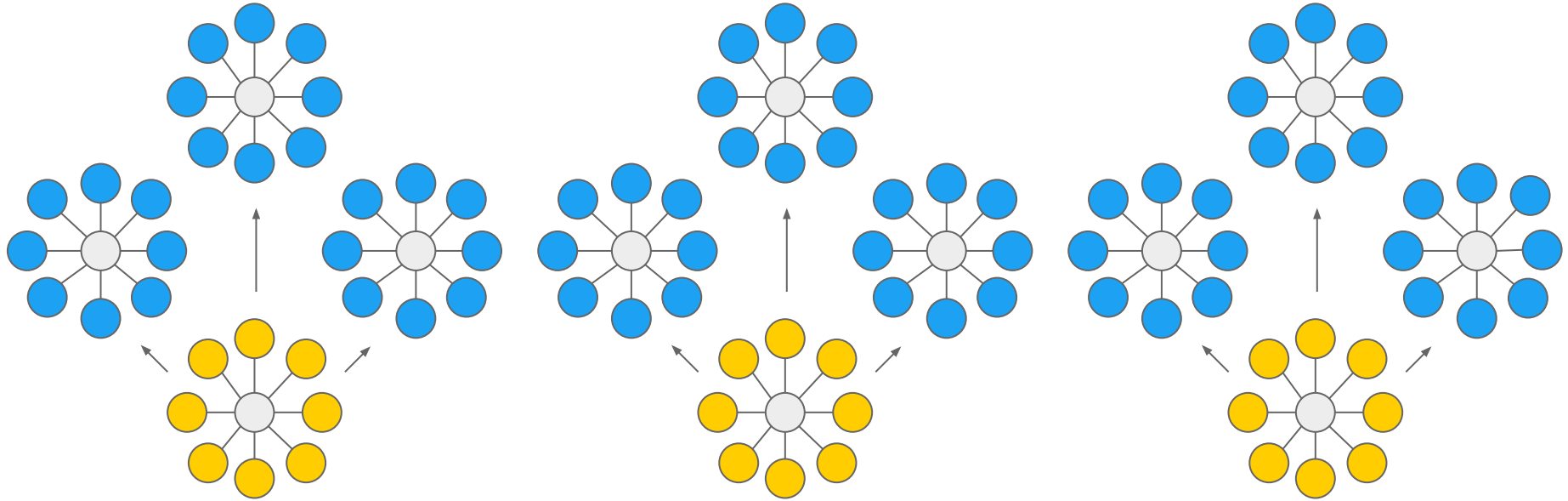




Can we do better?

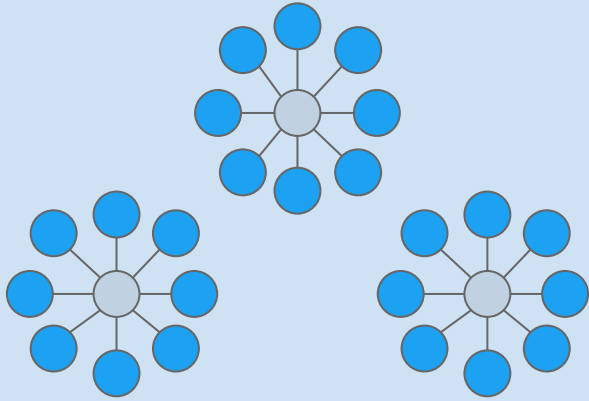
Maybe we can design a structure with the best out of both words?

Centralized org with Embedded teams

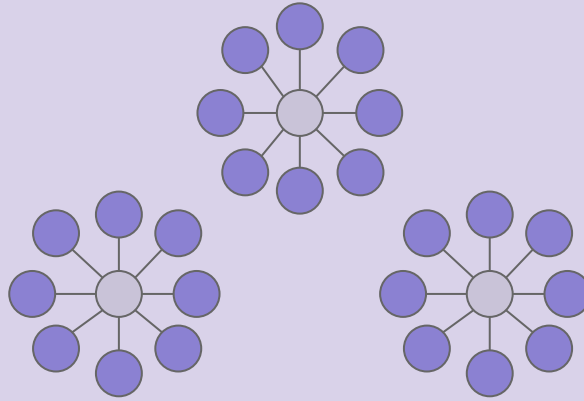


Centralized org with Embedded teams

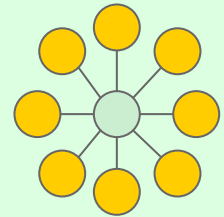
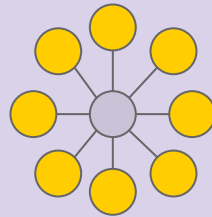
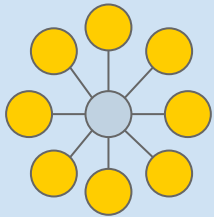
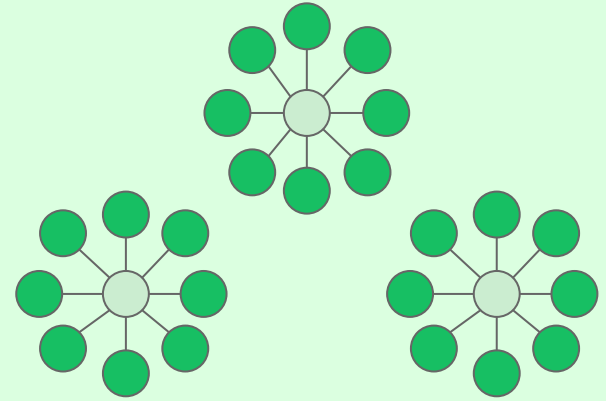
Growth Eng



Product Eng

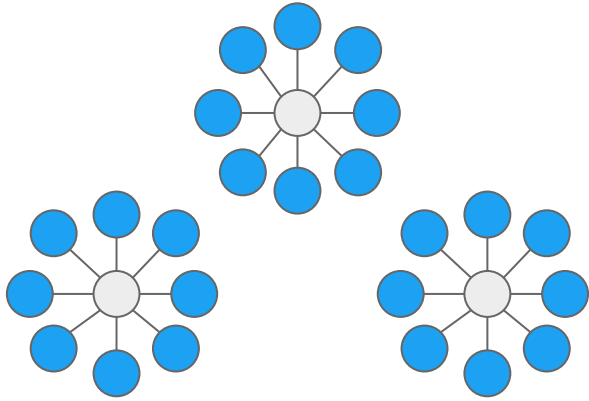


Health Eng

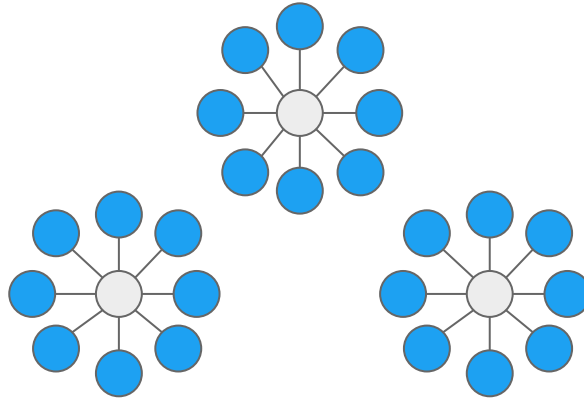


Centralized org with Embedded teams

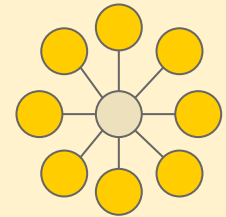
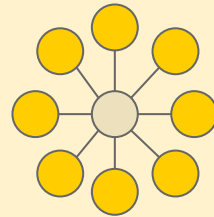
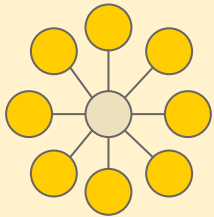
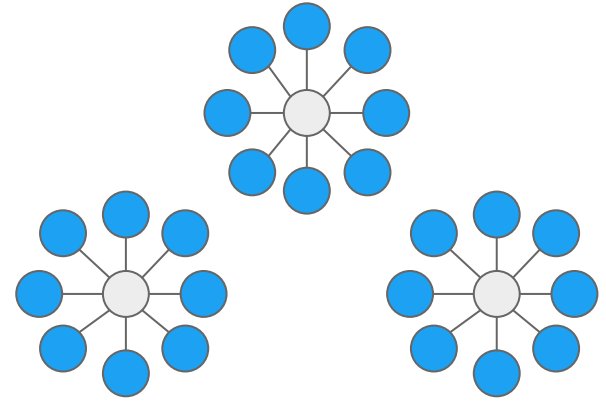
Growth Eng



Product Eng



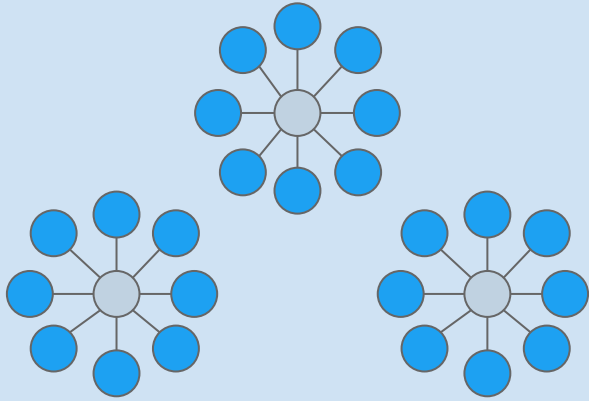
Health Eng



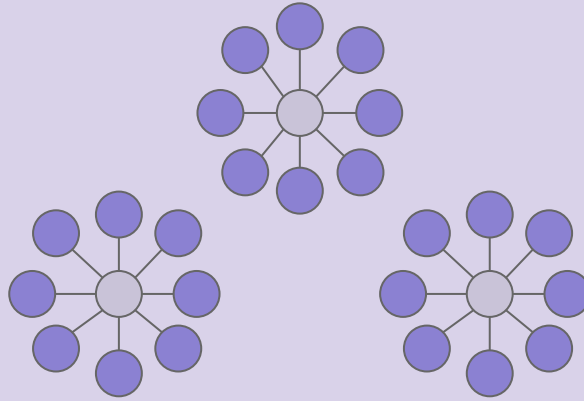
Data Science Org

Centralized org with Embedded teams

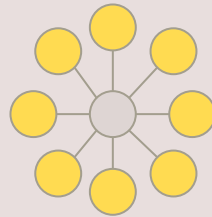
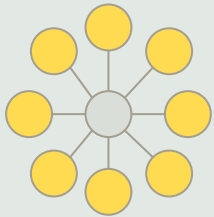
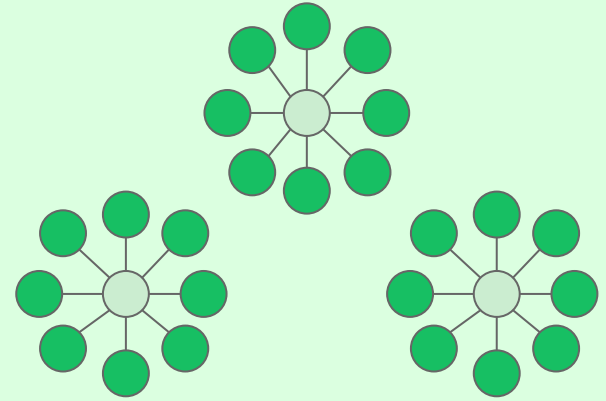
Growth Eng



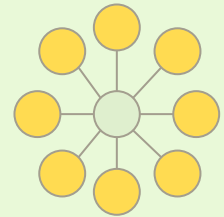
Product Eng



Health Eng

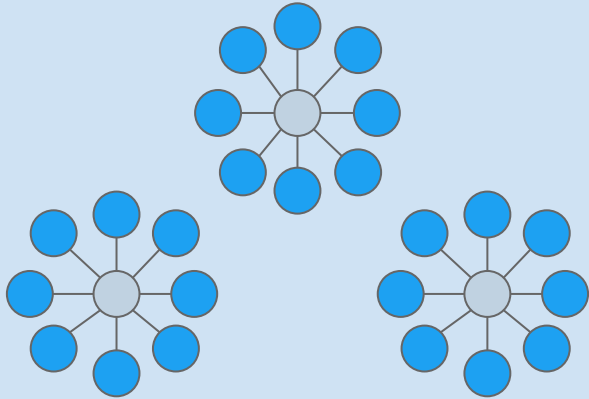


Data Science Org

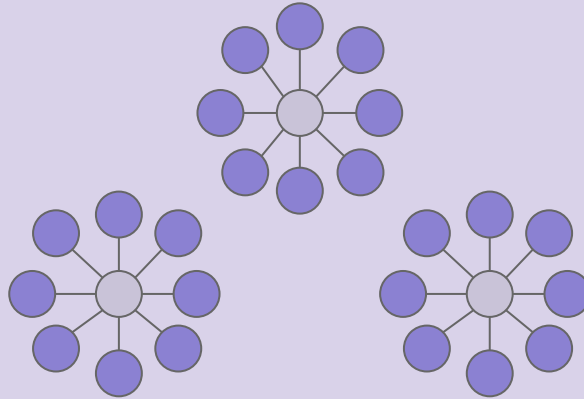


Centralized org with Embedded teams

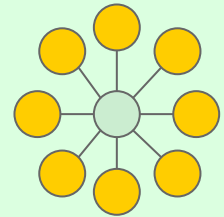
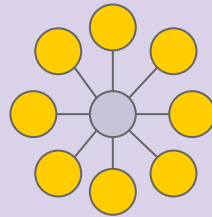
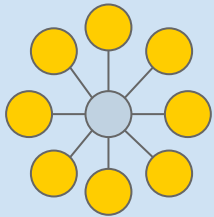
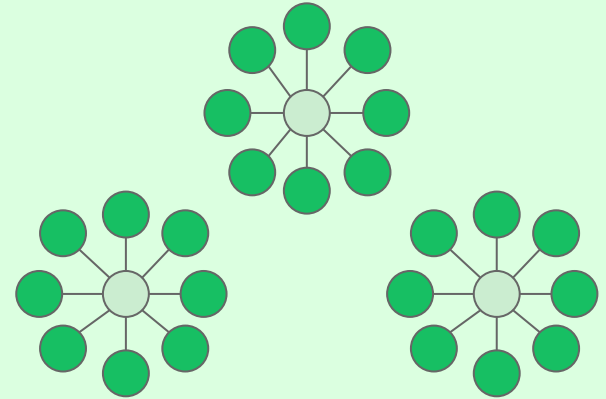
Growth Eng



Product Eng

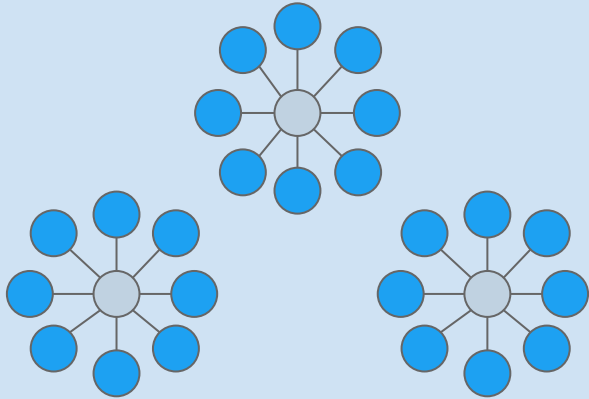


Health Eng

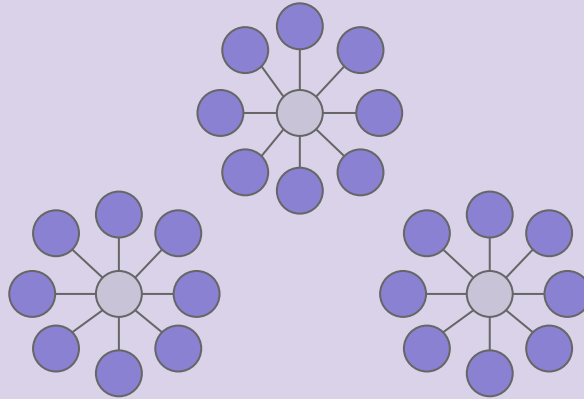


Centralized org with Embedded teams

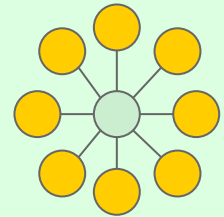
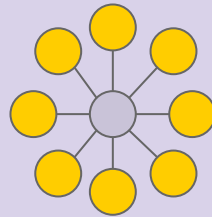
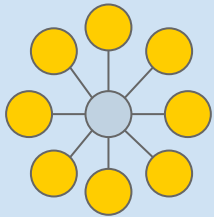
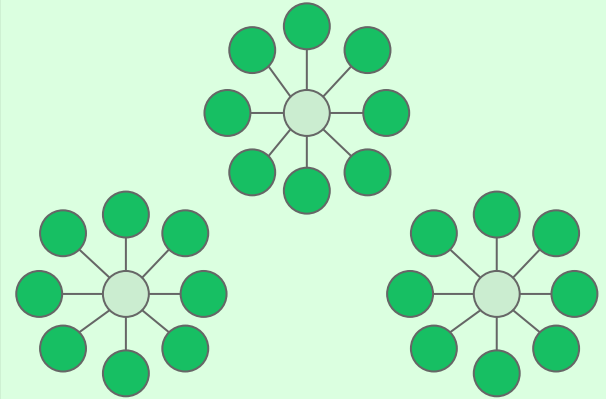
Growth Eng



Product Eng

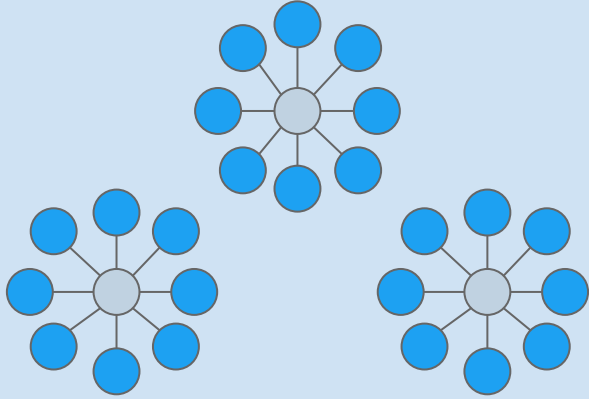


Health Eng

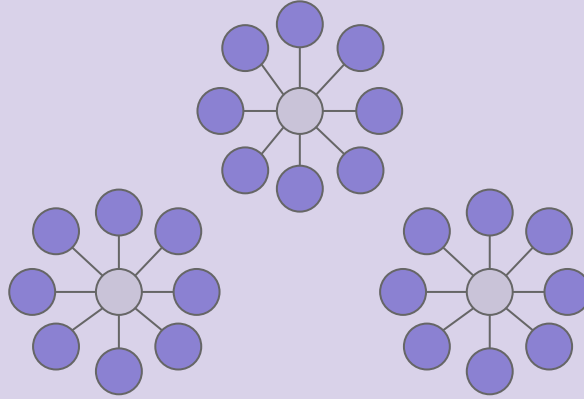


Centralized org with Embedded teams

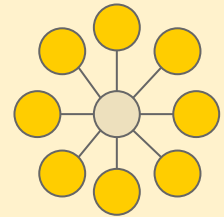
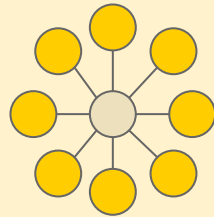
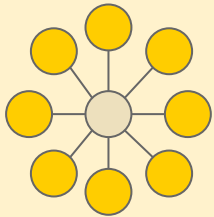
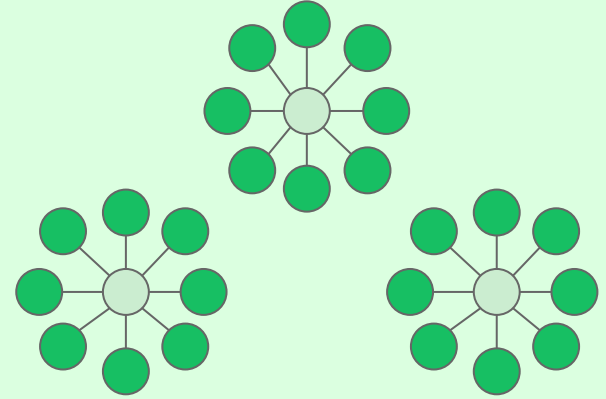
Growth Eng



Product Eng



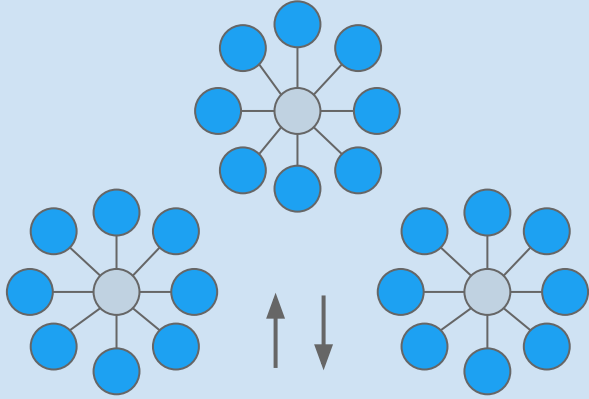
Health Eng



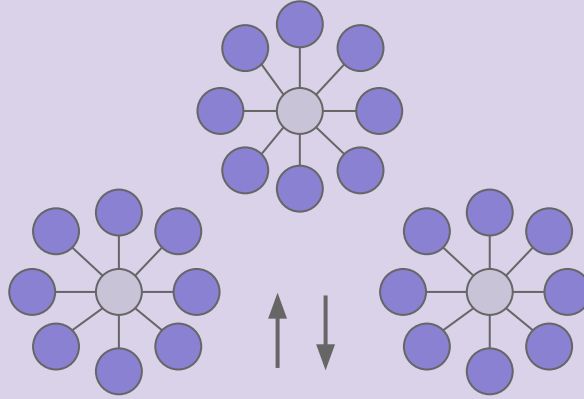
Data Science Org

Centralized org with Embedded teams

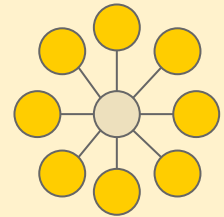
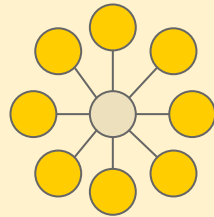
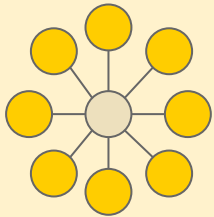
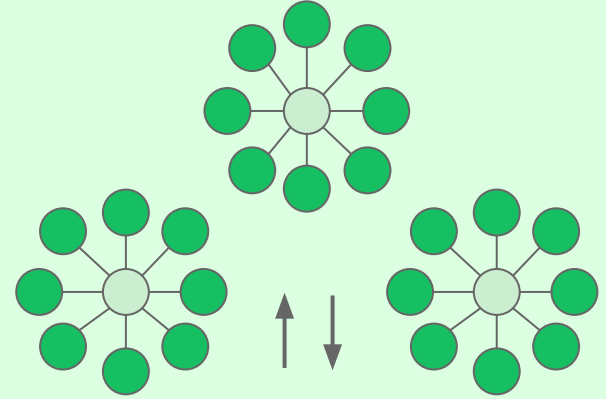
Growth Eng



Product Eng



Health Eng



Data Science Org



Benefits of this **hybrid** structure

- Centralized **processes** for the DS function (e.g. hiring, onboarding, career development)
- Common **resources**, **frameworks**, and **tech** for data scientists regardless of their team
- Organizational **alignment** between data scientists
- Work **alignment** between data scientists and their product and engineering counterpart



Challenges

- Everyone has at least **two** teams – their DS team and the eng / product counterpart they work with.
 - Risk of meeting and planning overload
 - Which is their “main” team?
- Risk of mismatch of expectation between DS leadership and product / eng teams’ leadership

2

Organizing DS work at scale

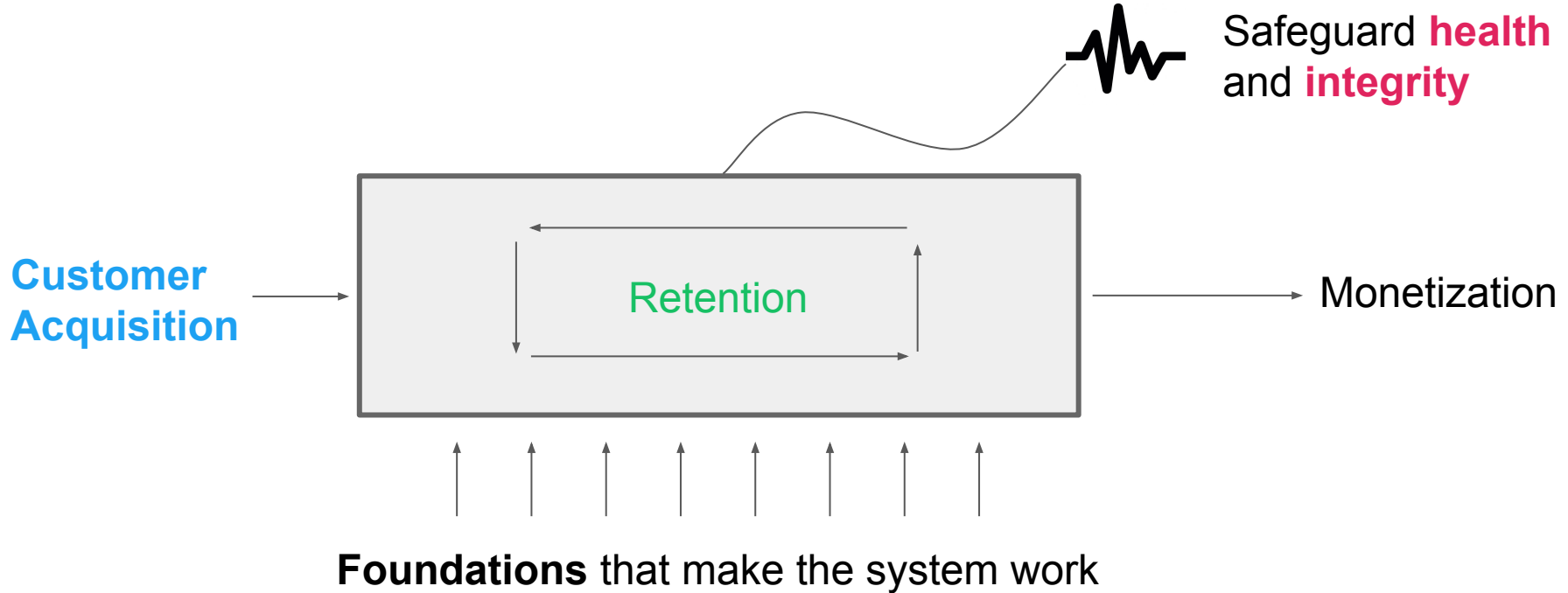
Finding alignment with other DS teams and stakeholders



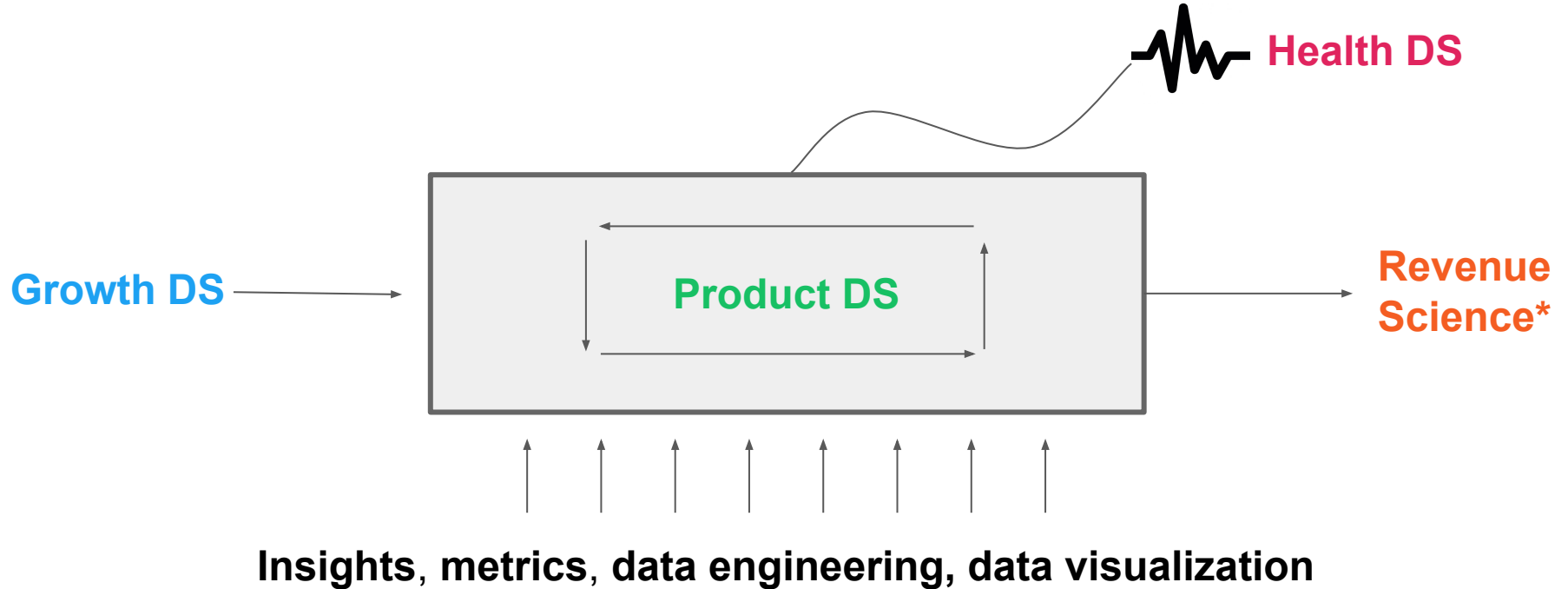
Problem

- In the previous example we had 3 small DS teams (-20 people total)
- How to scale this hybrid org structure to ~100 Data Scientists?
- Quick answer: create more layers of abstraction (grouping teams together)
- Follow Up question: how? What would be the grouping criteria?

Product as a system

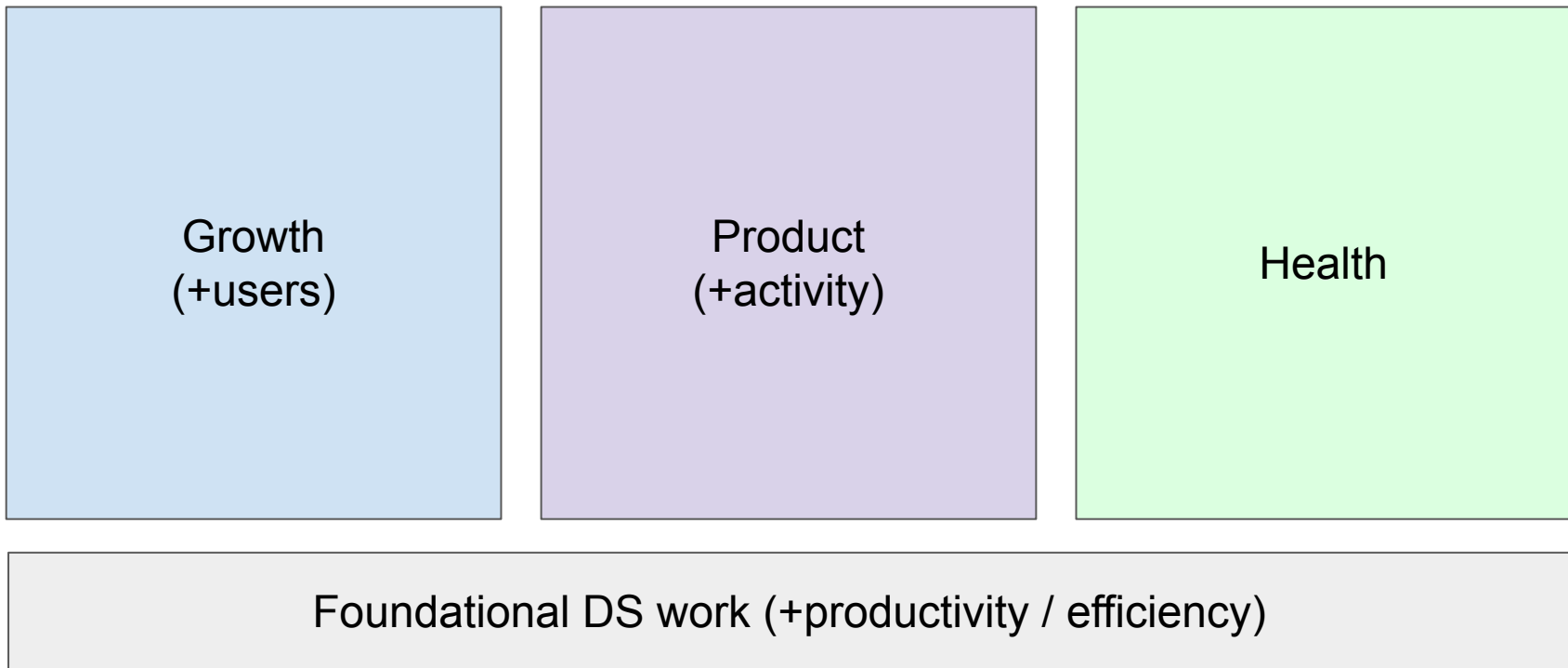


A product as a system





Structure of a big Data Science org





Product features supported by pillar

- ❖ Sign up
- ❖ Onboarding
- ❖ Marketing
- ❖ SEO
- ❖ Business Development
- ❖ Notifications
- ❖ Email
- ❖ **Key markets**
- ❖ **Global modeling**
- ❖ **Client Analytics**

- ❖ Events / Moments
- ❖ Search
- ❖ Trends
- ❖ Recommendations
- ❖ Tweets
- ❖ Home Timeline
- ❖ Conversations
- ❖ Direct Messages
- ❖ Video

- ❖ Health Measurement
- ❖ Health Product
- ❖ White Hat Research
- ❖ Ad hoc investigations
- ❖ Health/security research
- ❖ Scaling identification of misinformation

Data Engineering / Metrics / Data Visualization / Experimentation

** At Twitter these areas are supported from our Singapore office #JoinTheFlock*

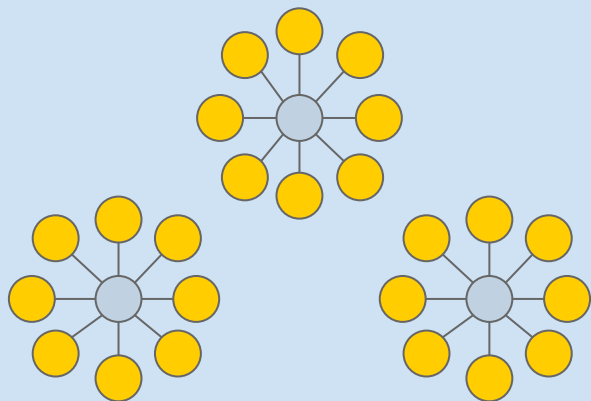
3

Keeping your teams together

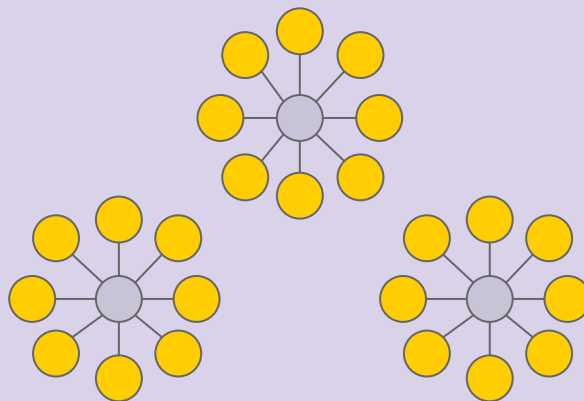
Build teams, not working groups

~100 people DS organization

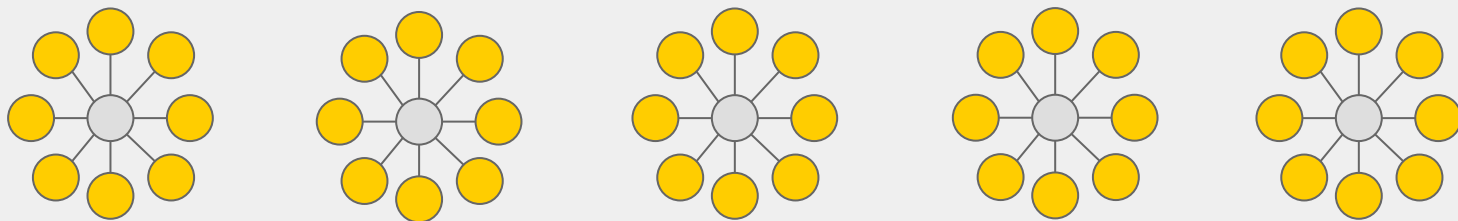
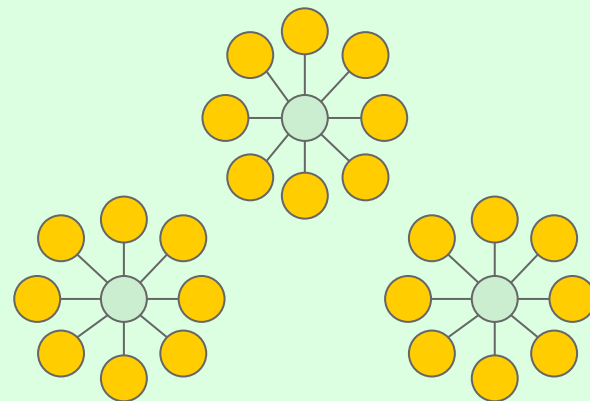
Growth DS



Product DS



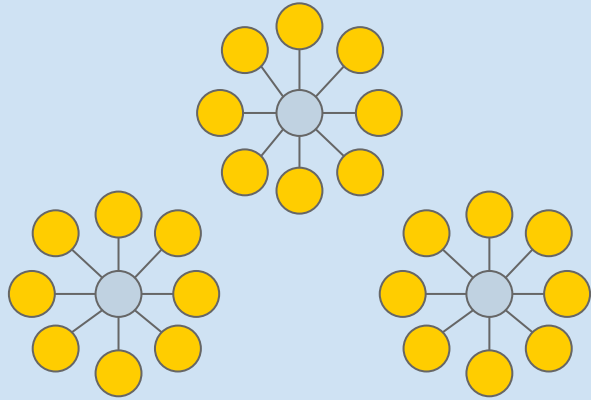
Health DS



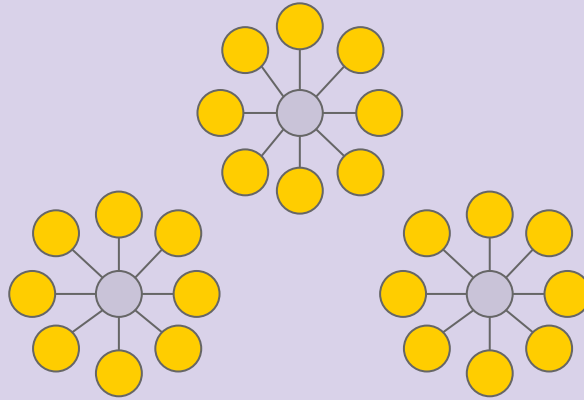
Foundational DS

Team charters - what is the reason a team exists?

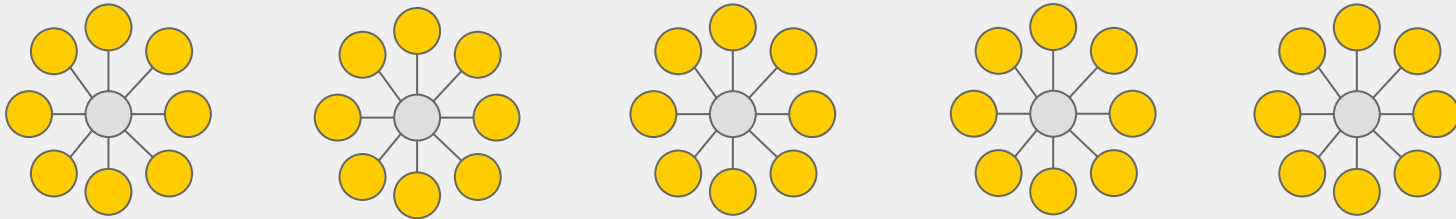
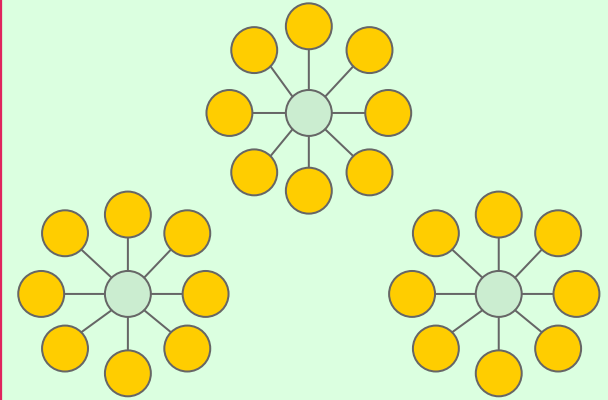
Growth DS



Health DS



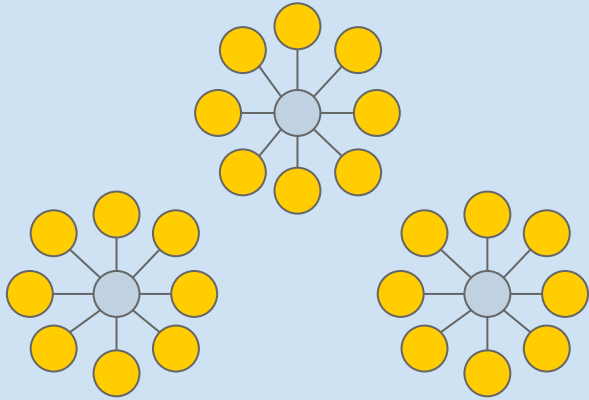
Product DS



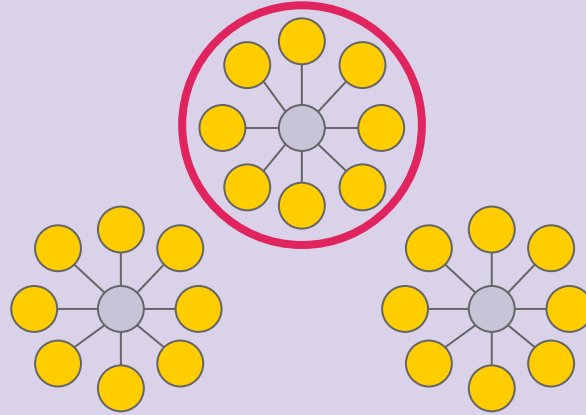
Foundational DS

Team charters - what is the reason a team exists?

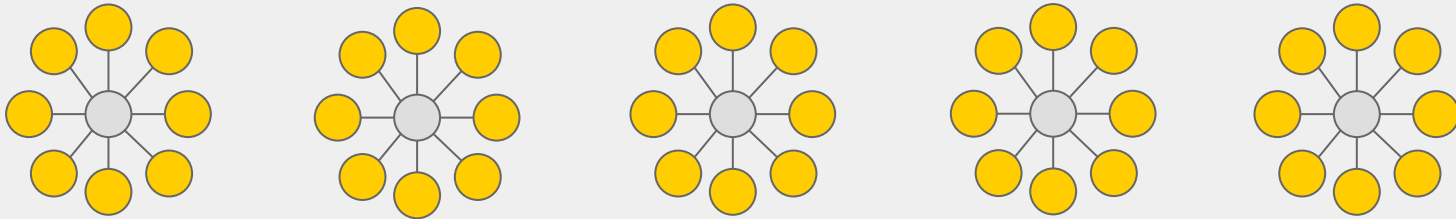
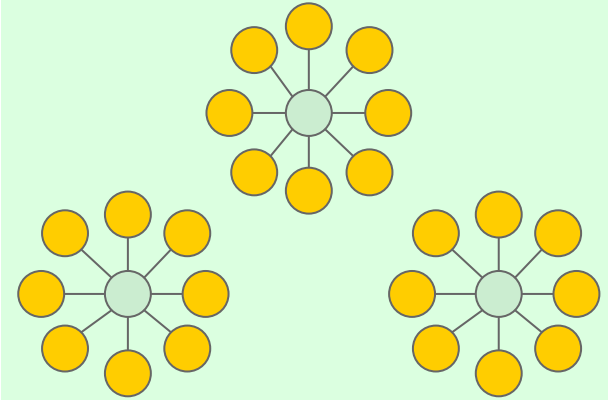
Growth DS



Health DS



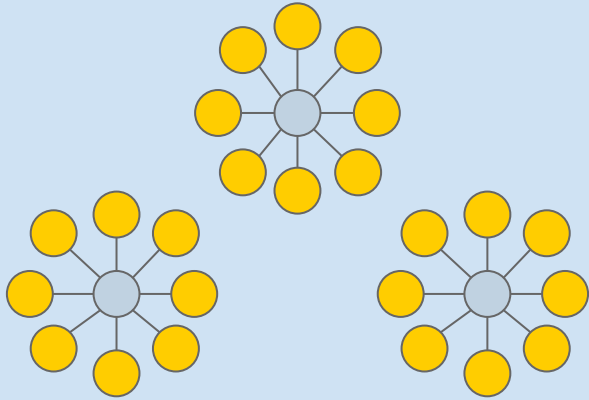
Product DS



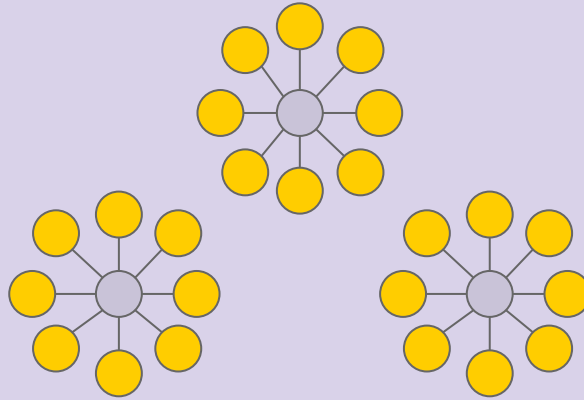
Foundational DS

Team charters - what is the reason a team exists?

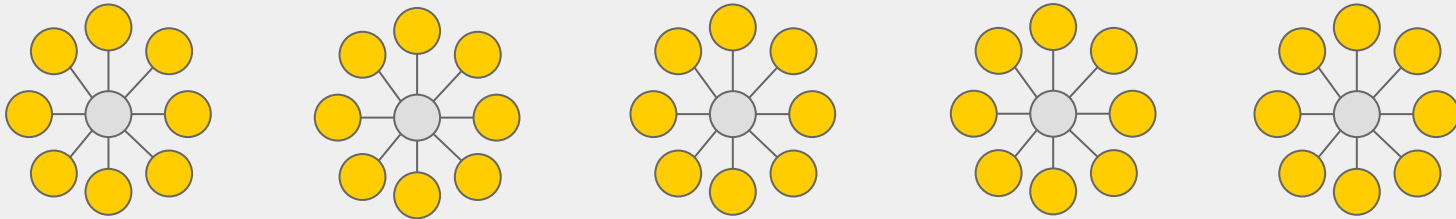
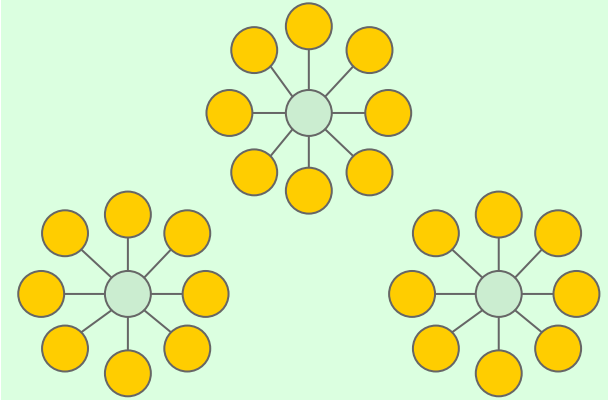
Growth DS



Health DS



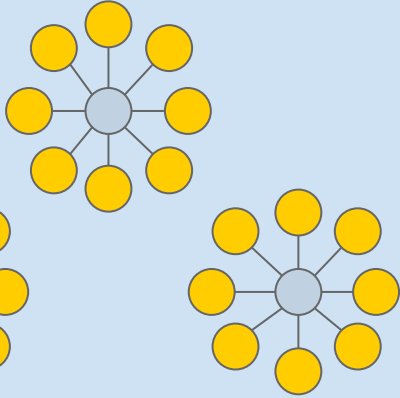
Product DS



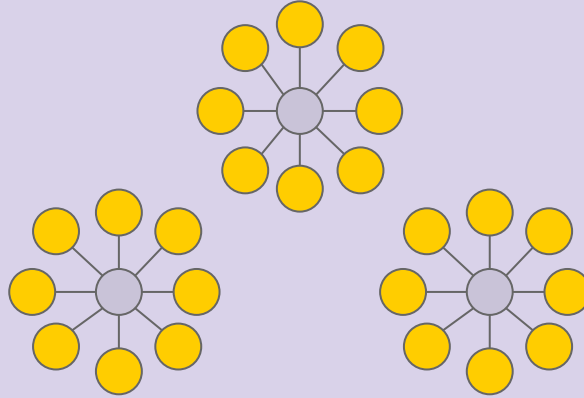
Foundational DS

“Swimlanes” – clear differentiation between teams

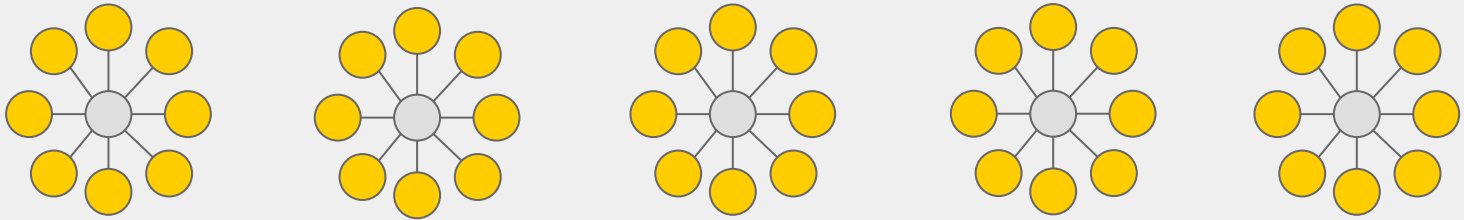
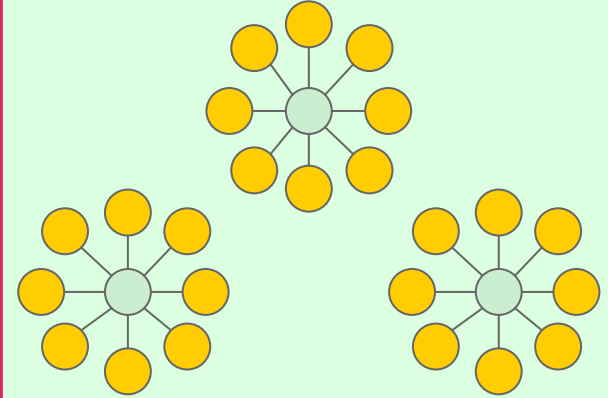
Growth DS



Health DS



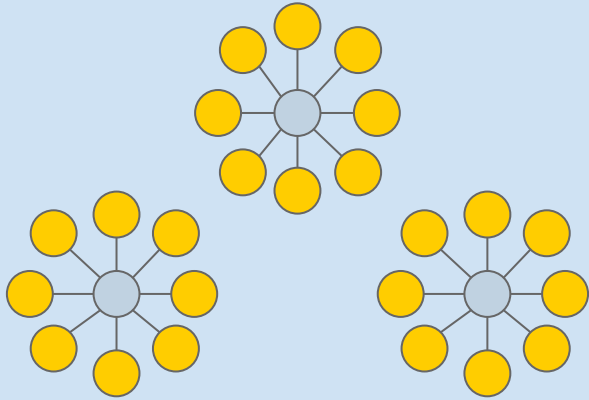
Product DS



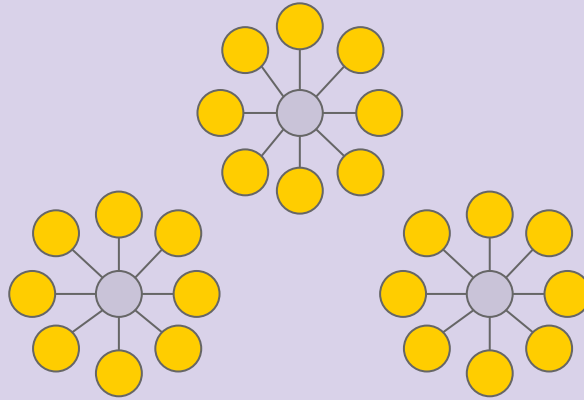
Foundational DS

Working agreement - what to expect from other teams?

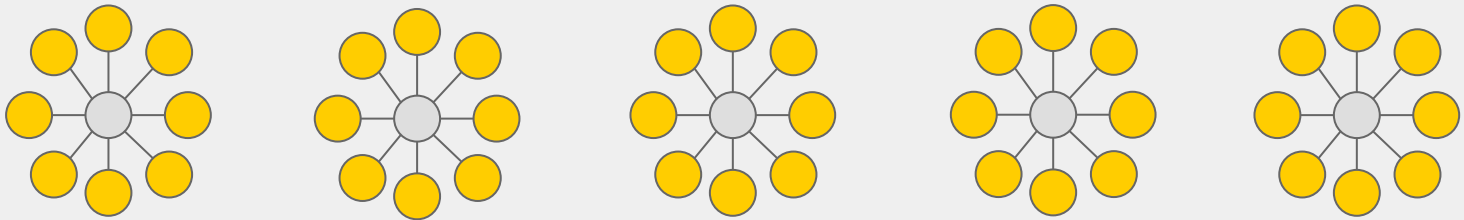
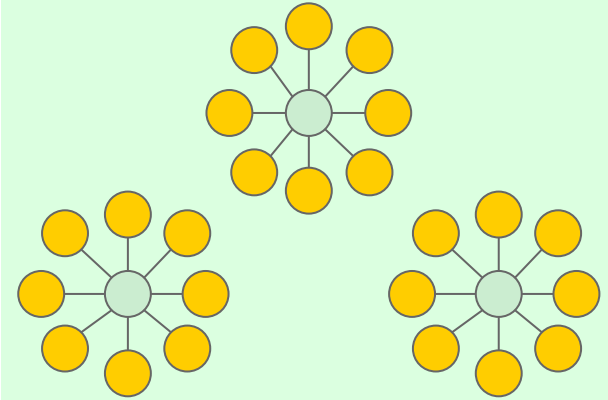
Growth DS



Health DS



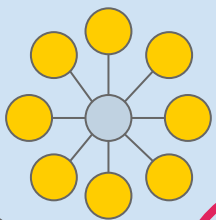
Product DS



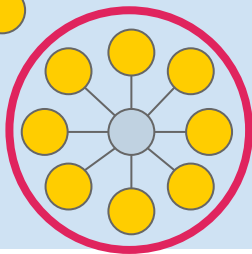
Foundational DS

Working agreement - what to expect from other teams?

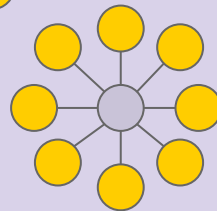
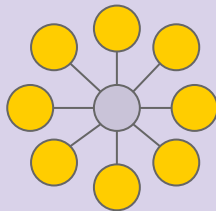
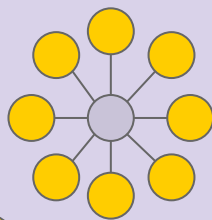
Growth DS



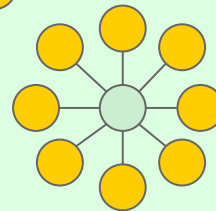
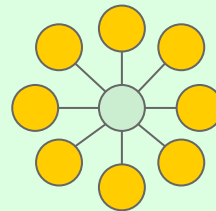
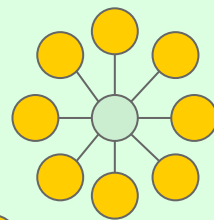
Notifications
DS team



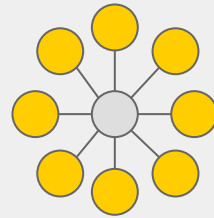
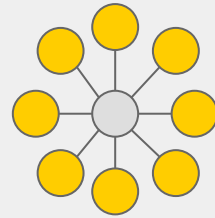
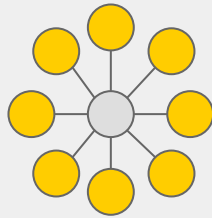
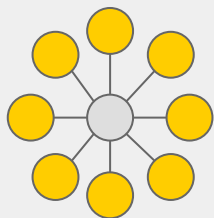
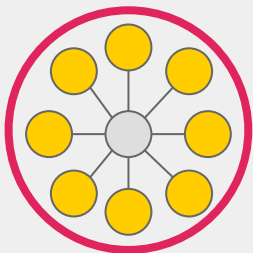
Health DS



Product DS



Data
eng.
team



Foundational DS

Working agreement - what to expect from other teams?

Growth DS

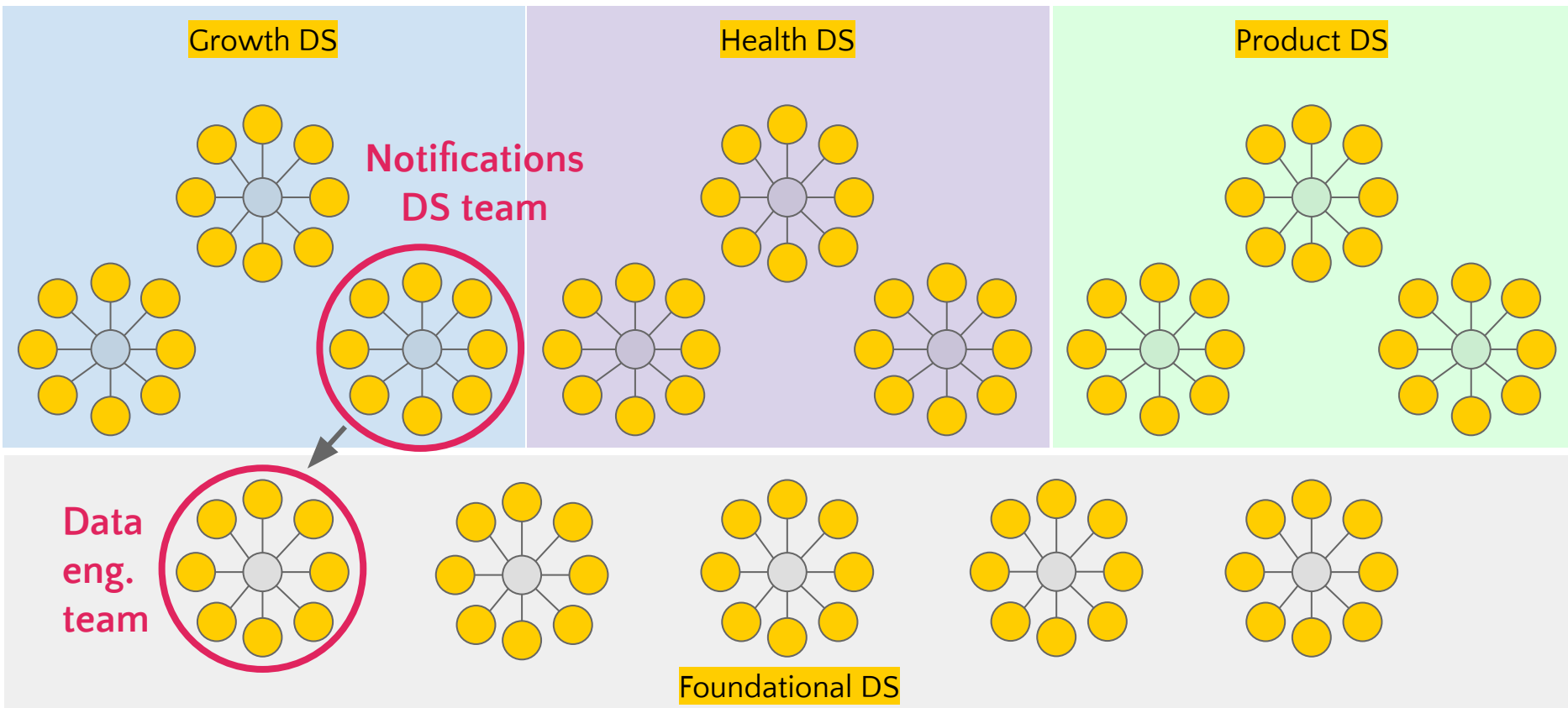
Health DS

Product DS

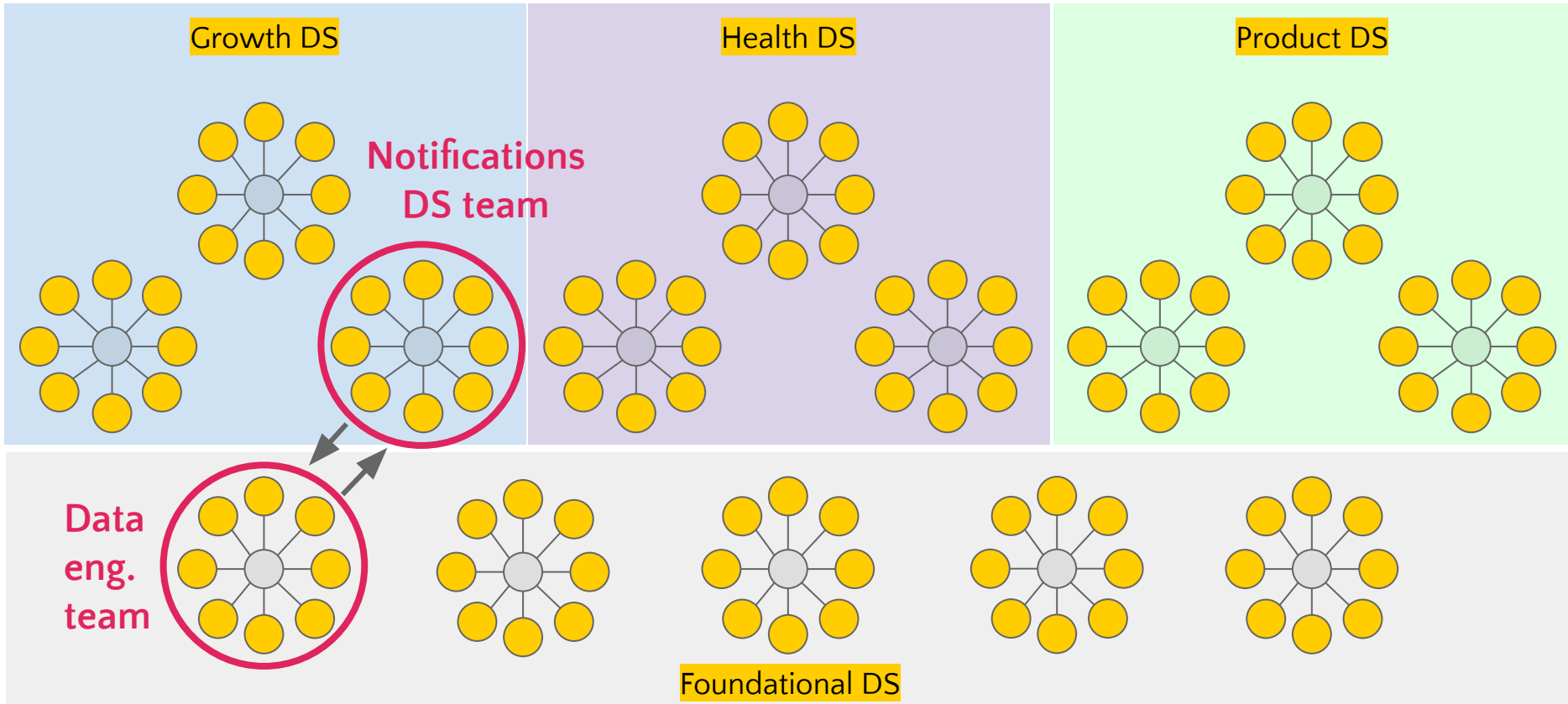
Notifications
DS team

Data
eng.
team

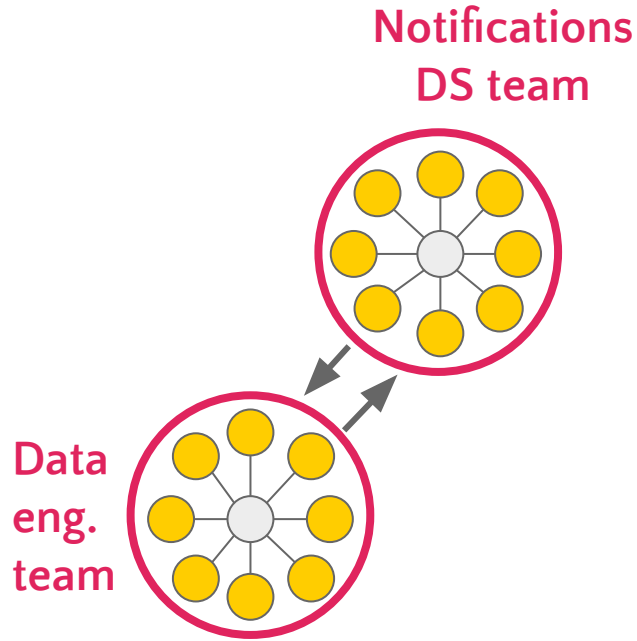
Foundational DS



Working agreement - what to expect from other teams?



Working agreement - what to expect from other teams?



- How does the data eng. team receives requests?
- What is the SLA for a dataset request?
- What would be the ownership structure for the dataset?
- On what basis this request will be prioritized?



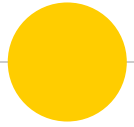
Create clear **communication channels**

- Have team meetings at all levels (org all-hands, team-level weekly or biweekly)
- Have recurrent sessions to review ongoing projects (we do monthly “data science reviews” for Growth, Product, Health, and Foundational)
- Have fun with each other – we do quarterly offsites and other activities. Other than having fun these settings foster interesting work conversations.



Build and strengthen your leadership team

- ◉ Your org's leadership should be operating as a **team**. In fact, this is their **first team**.
- ◉ Have a staff meeting. Keep a standing agenda ("how is hiring going?") and open it to everyone.
- ◉ Do leadership offsites and working sessions (we do them monthly – covering **one** specific topic)
- ◉ Make this team responsible for managing your org's relationship with stakeholders



Summary

Lessons learned, in short form



Summary: scaling data science teams

- Centralized or embedded? What about a hybrid approach where data scientists are under the same org but work very closely with eng and product?
- Align teams with objectives (create teams of teams if necessary, but have clear criteria to do so)
- Build the structures of your teams early on: team charters, “swimlanes”, working agreements, communication, and a strong leadership team.



We are hiring!

Data Scientists, Data Engineers, and Engineering managers in **Singapore** and U.S.A.



Thanks!

Any **questions** ?

You can find me at

- @miguelrios
- miguel@twitter.com



Instructions for use

EDIT IN GOOGLE SLIDES

Click on the button under the presentation preview that says "Use as Google Slides Theme".

You will get a copy of this document on your Google Drive and will be able to edit, add or delete slides.

You have to be signed in to your Google account.

EDIT IN POWERPOINT®

Click on the button under the presentation preview that says "Download as PowerPoint template". You will get a .pptx file that you can edit in PowerPoint.

Remember to download and install the fonts used in this presentation (you'll find the links to the font files needed in the [Presentation design slide](#))

More info on how to use this template at www.slidescarnival.com/help-use-presentation-template

This template is free to use under [Creative Commons Attribution license](#). You can keep the Credits slide or mention SlidesCarnival and other resources used in a slide footer.

Quotations are commonly printed
as a means of inspiration and to
invoke philosophical thoughts
from the reader.



“



This is a **slide title**

- Here you have a list of items
- And some text
- But remember not to overload your slides with content

Your audience will listen to you or read the content, but won't do both.



Big concept

Bring the attention of your audience over a key concept using icons or illustrations



You can also split your content

White

Is the color of milk and fresh snow, the color produced by the combination of all the colors of the visible spectrum.

Black

Is the color of coal, ebony, and of outer space. It is the darkest color, the result of the absence of or complete absorption of light.



In two or three columns

Yellow

Is the color of gold, butter and ripe lemons. In the spectrum of visible light, yellow is found between green and orange.

Blue

Is the colour of the clear sky and the deep sea. It is located between violet and green on the optical spectrum.

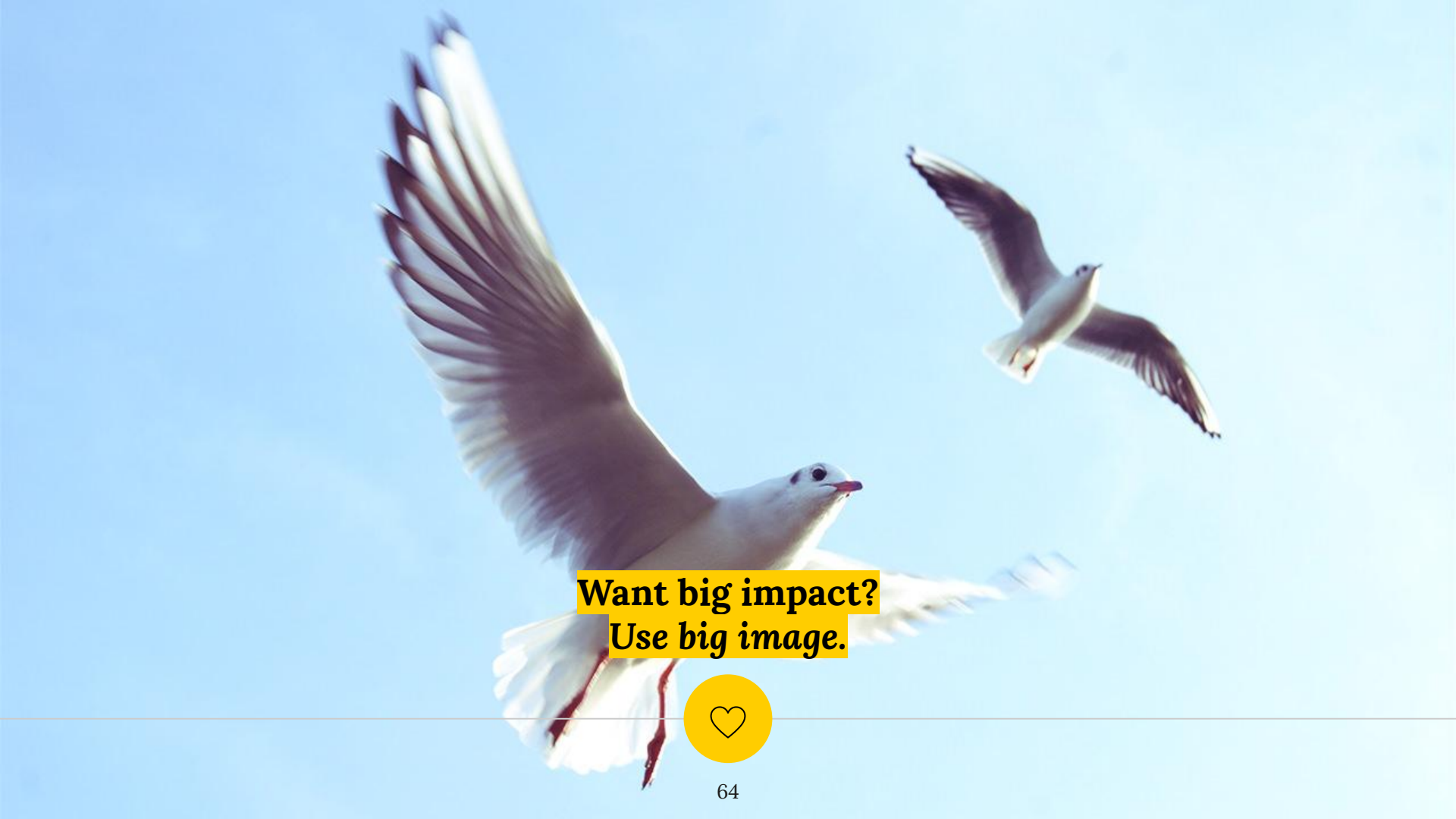
Red

Is the color of blood, and because of this it has historically been associated with sacrifice, danger and courage.



A picture is worth **a thousand words**

A complex idea can be conveyed with just a single still image, namely making it possible to absorb large amounts of data quickly.

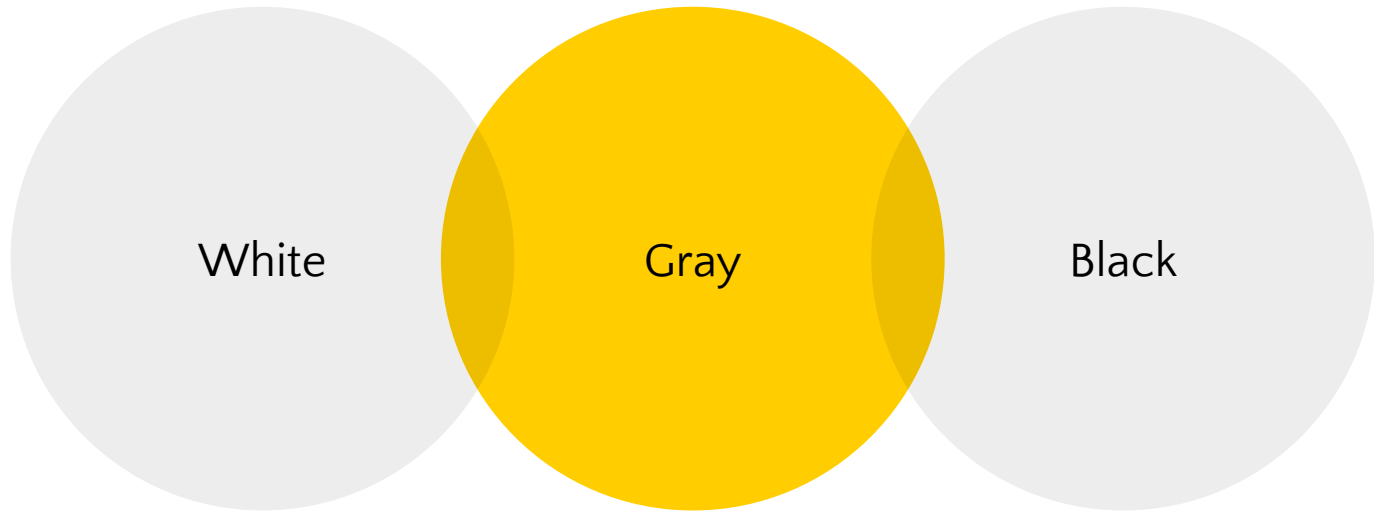


**Want big impact?
Use *big image*.**



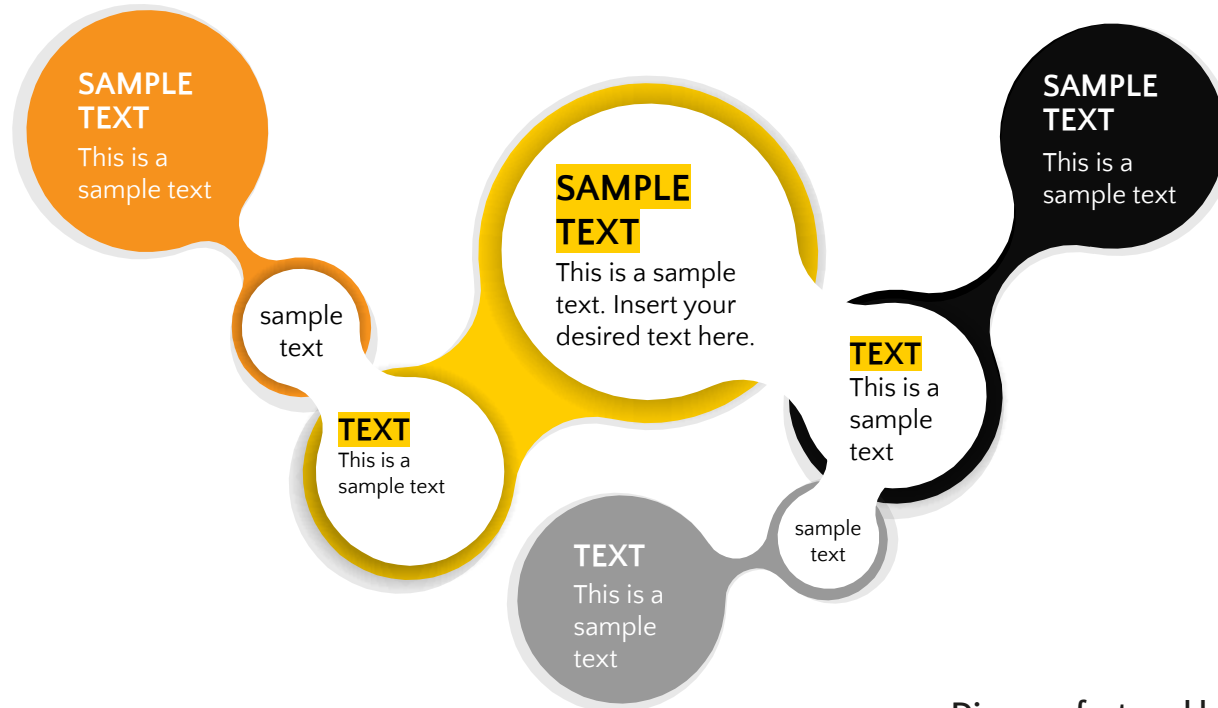


Use charts to explain your ideas





Or use **diagrams** to explain complex ideas





And tables to **compare data**

	A	B	C
Yellow	10	20	7
Blue	30	15	10
Orange	5	24	16



Maps



89,526,124

Whoa! That's a big number, aren't you proud?



89,526,124\$

That's a lot of money

185,244 users

And a lot of users

100%

Total success!





Let's review some concepts

Yellow

Is the color of gold, butter and ripe lemons. In the spectrum of visible light, yellow is found between green and orange.

Blue

Is the colour of the clear sky and the deep sea. It is located between violet and green on the optical spectrum.

Red

Is the color of blood, and because of this it has historically been associated with sacrifice, danger and courage.

Yellow

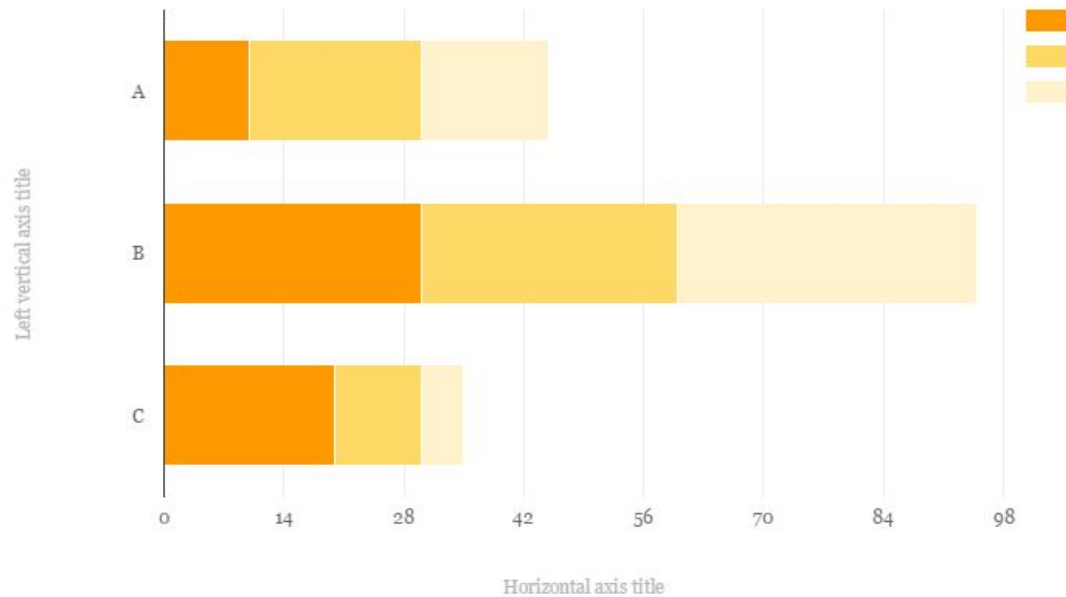
Is the color of gold, butter and ripe lemons. In the spectrum of visible light, yellow is found between green and orange.

Blue

Is the colour of the clear sky and the deep sea. It is located between violet and green on the optical spectrum.

Red

Is the color of blood, and because of this it has historically been associated with sacrifice, danger and courage.



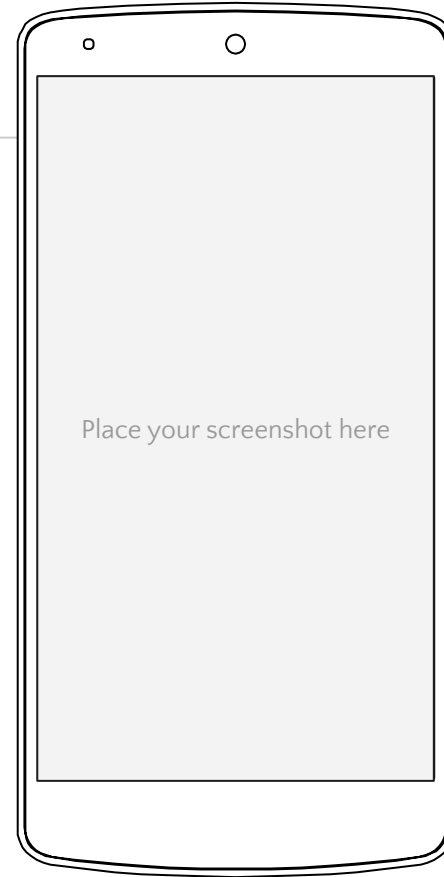
You can copy&paste graphs from Google Sheets





Android project

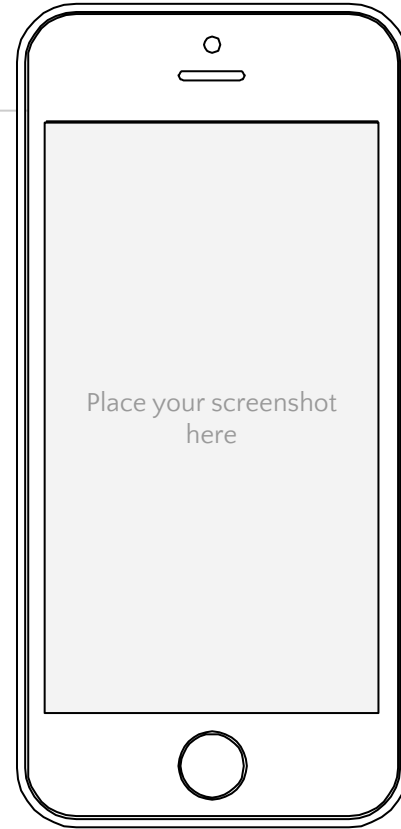
Show and explain your web, app or software projects using these gadget templates.





iPhone project

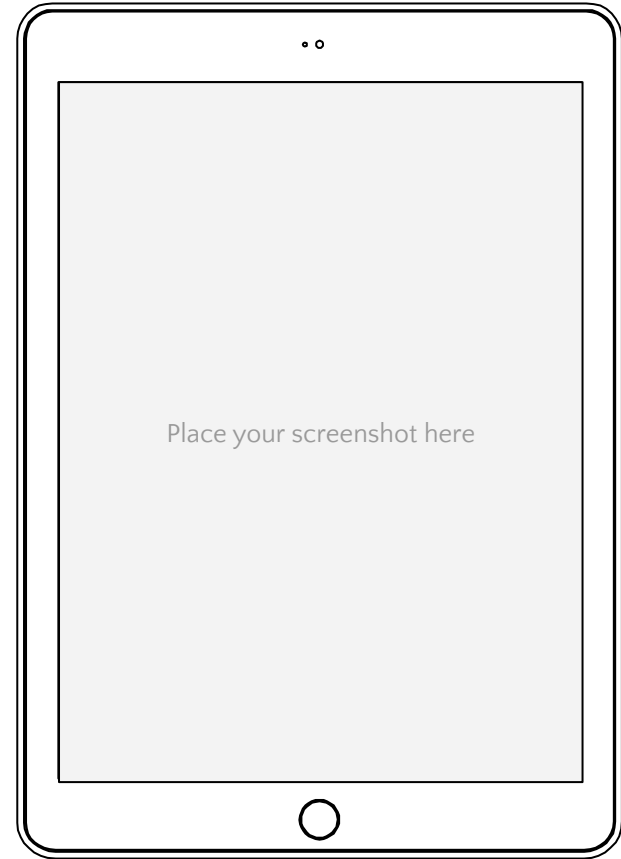
Show and explain your web, app or software projects using these gadget templates.





Tablet project

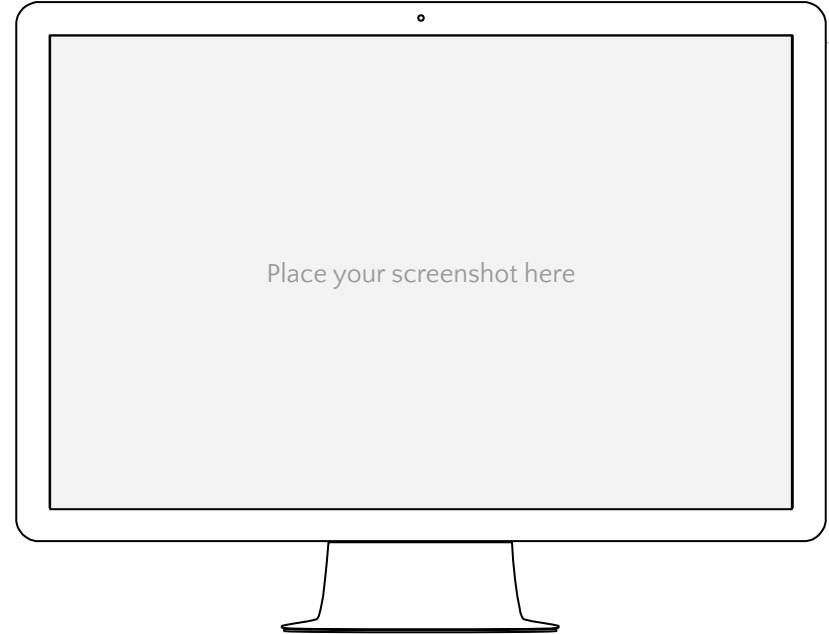
Show and explain your web, app or software projects using these gadget templates.





Desktop project

Show and explain your web, app or software projects using these gadget templates.





Thanks!

Any **questions** ?

You can find me at

- @username
- user@mail.me



Credits

Special thanks to all the people who made and released these awesome resources for free:

- Presentation template by SlidesCarnival
- Photographs by Unsplash



Presentation design

This presentation uses the following typographies and colors:

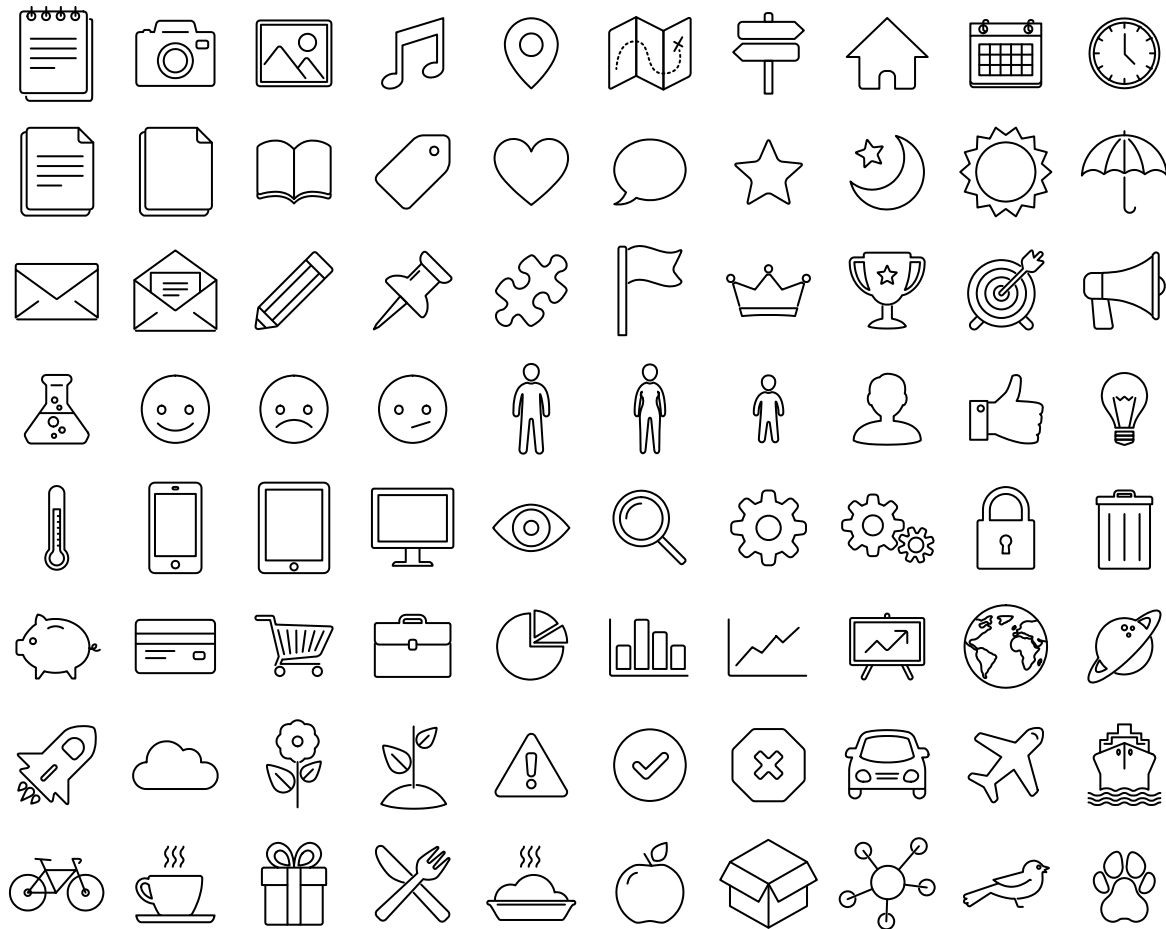
- Titles: **Lora**
- Body copy: **Quattrocento Sans**

You can download the fonts on these pages:

<https://www.fontsquirrel.com/fonts/lora>

<https://www.fontsquirrel.com/fonts/quattrocento-sans>

Yellow **#ffcd00** | Black **#000000** | Grey **#cccccc**



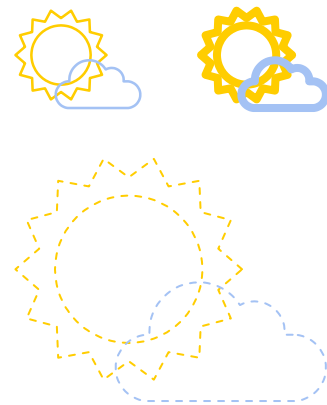
SlidesCarnival icons are **editable shapes**.

This means that you can:

- Resize them without losing quality.
- Change line color, width and style.

Isn't that nice? :)

Examples:



Now you can use any emoji as an icon!

And of course it resizes without losing quality and you can change the color.

How? Follow Google instructions

<https://twitter.com/googledocs/status/730087240156643328>



and many more...