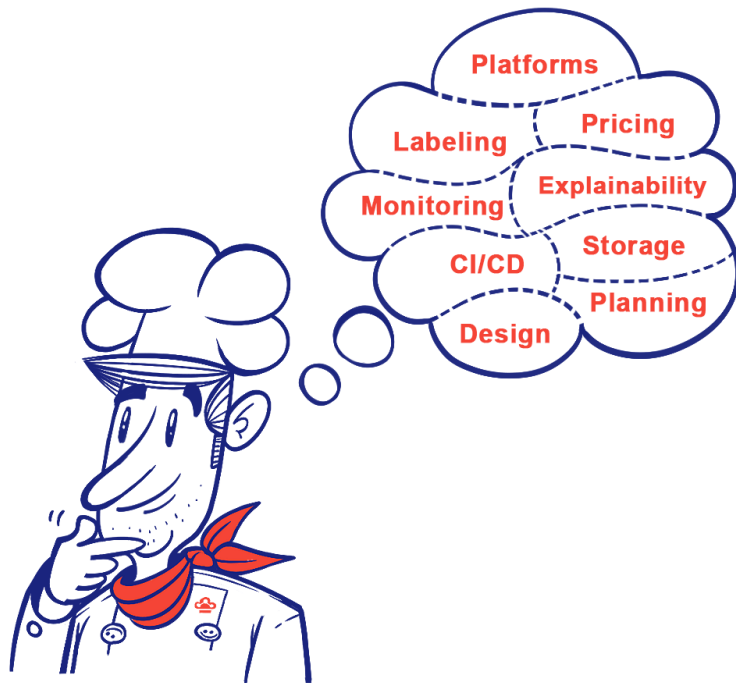


10 THINGS

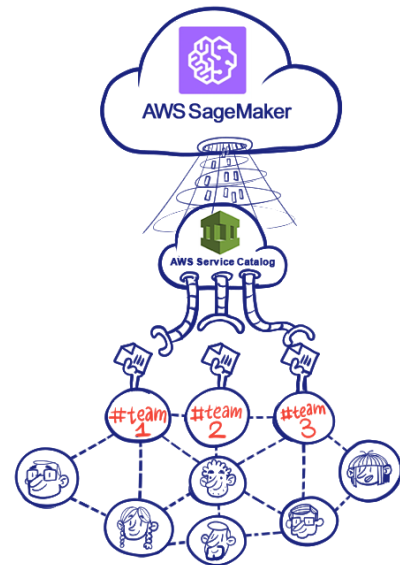
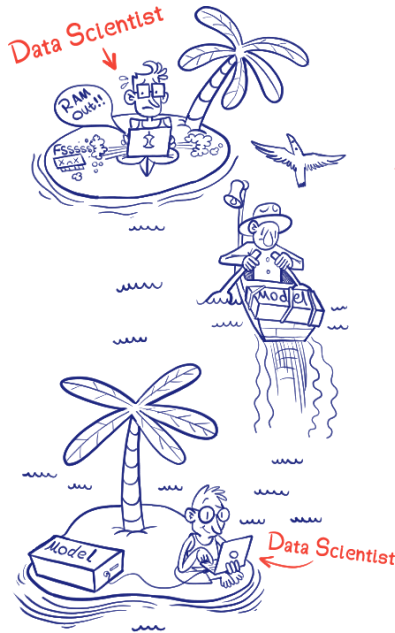
We Wish We Knew
Before Venturing into
Real Data Science
On AWS

BY:



1 How to efficiently build a self-service data science platform for easy collaboration

- Automatically provisioning pre-configured privileges according to your standards
- Using portfolios to manage access without cutting tickets to infra team
- Set constraints centrally and apply them consistently across teams
- Replicable governed data science environments



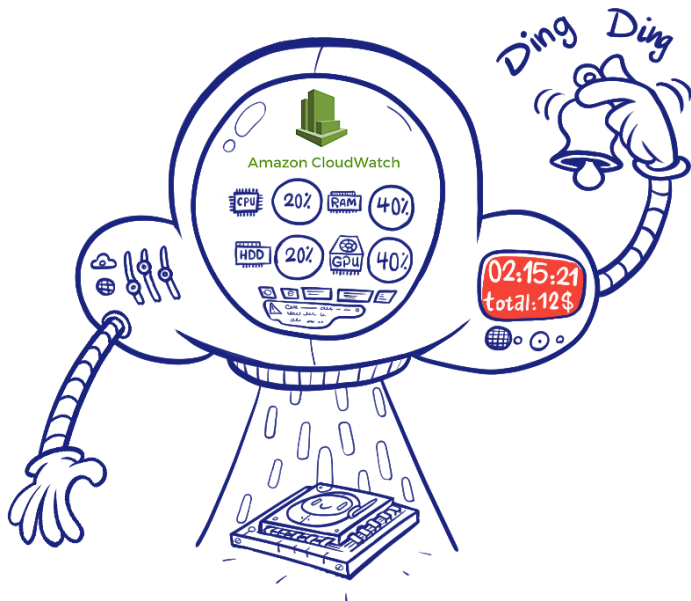
2 How to save costs with AWS special pricing plans

- On-demand for temporary usage
- Reserved for predictable usage
 - Up to 64% discount for 1-3 year usage commitment
- Spot instances for time-insensitive usage
 - Up to 90% cheaper
- Ask AWS for private pricing: HUGE discount for committed SageMaker customers



3 } Crucial role of monitoring with CloudWatch

- Avoiding overshoots by choosing the proper hardware based on the actual usage
- Tracking training procedures on a unified dashboard
- Monitoring the health of products
- Keeping track of costs to avoid surprises
- Getting alarmed when things go wrong



4 Demystifying black box models through Bias Detection and Model Explainability

- Identifying bias in your dataset before training
- Shedding light on complex models to see how they work
- SageMaker Clarify does this in an easy and scalable way



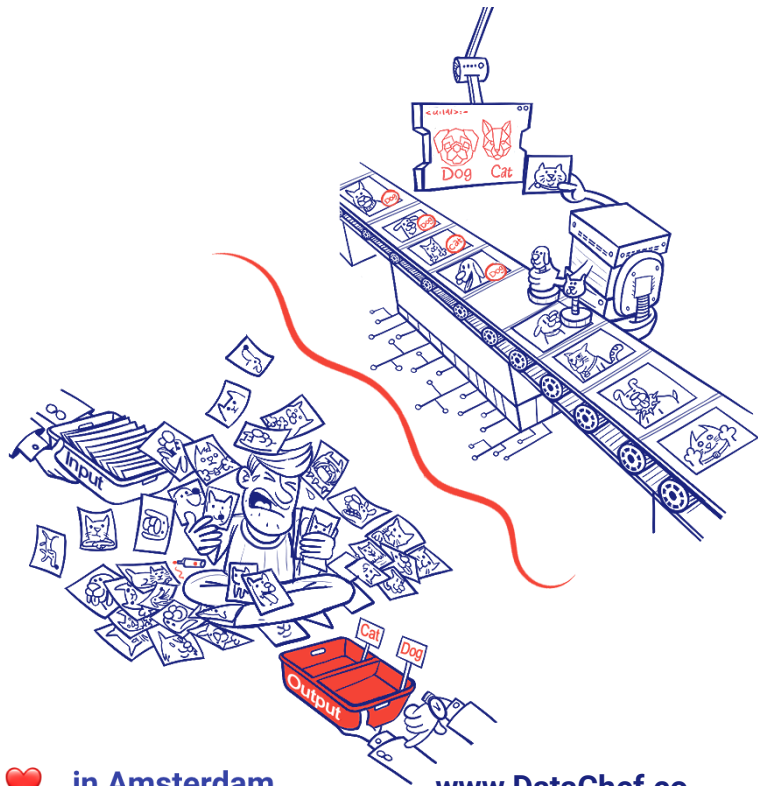
5) There is a good chance that others have already done it!

- Solutions and algorithms delivered by other companies! Ask your AWS partners for pointers.
- Highly reliable recipes tested in real-world and production environments.
- Reducing resource costs with readily available products.
- Pay as you go
- Saving time and money on human resources



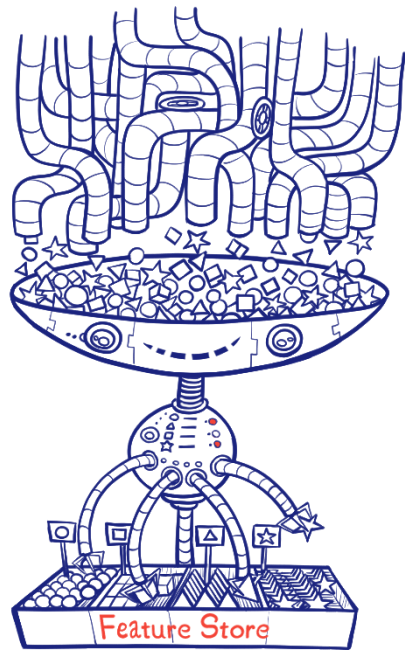
6 Facilitated labeling using SageMaker Ground Truth

- Providing a simple user interface for labeling different types of data
- Defining internal/external workforce for labeling and educating them
- Pre-labeling your data using ML (learn as you go)
- Automatically organizing the datasets, ready to be fed to algorithms



7 } Build your features once, and use them for a lifetime.

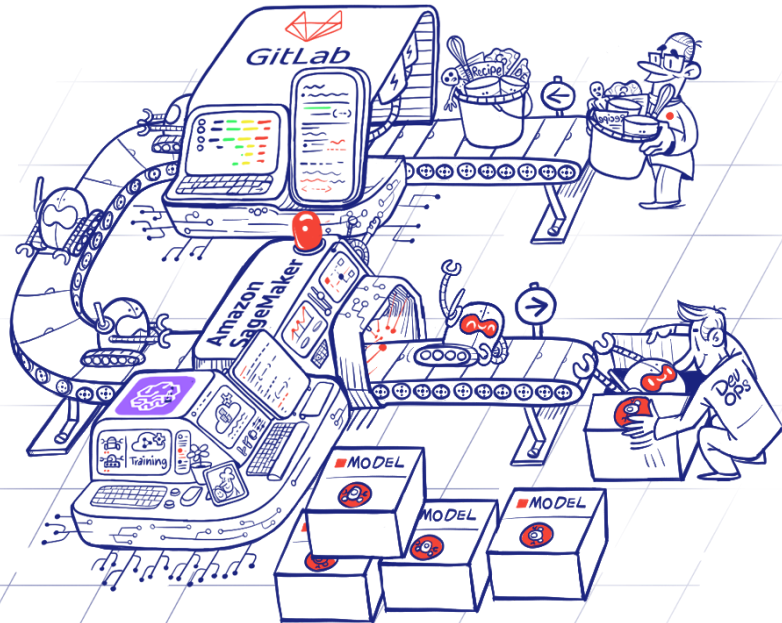
- Keeping track of feature versions as features evolve
- Reusing engineered features rather than redoing complex feature engineering procedures
- Ongoing access to the latest data by auto-pulling from various data sources
- Sharing with other teams via cross-account access
- Self-explanatory features through metadata, feature groups, descriptions, and tags.



8

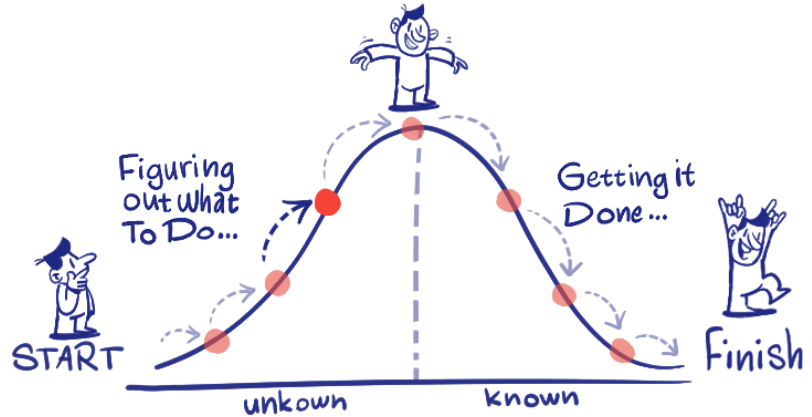
One SageMaker Project to rule them all

- CI/CD in data science and ML Engineering
- Building end-to-end pipelines in no-time
- Coordinating multiple AWS accounts on one project
- From Dev to Prod, with an expert's click



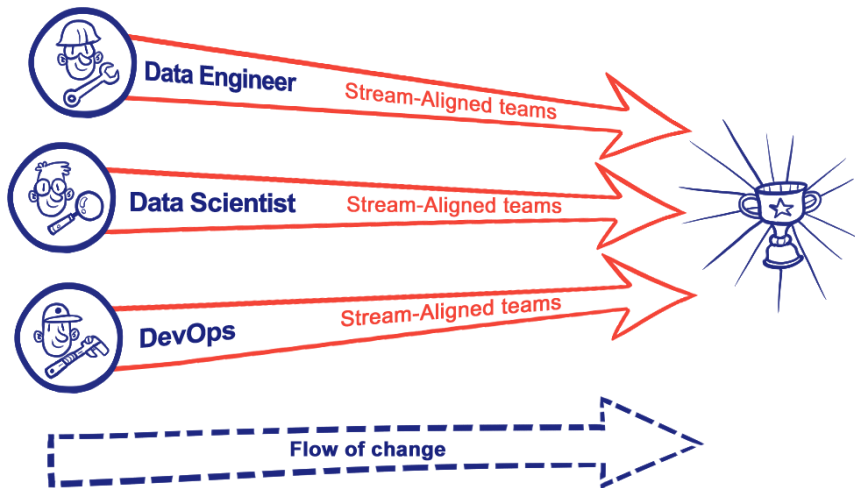
9) Shaping up: Making your expectations clear!

- Consistency over Crystal Ball predictions
- Being aware of rabbit holes
- The right level of abstraction
- Making a bet



10 All said: Team Design has the final say

- Don't let HR design interaction between teams
- Domain-Driven team up
- Building teams by workloads, not organizational chart roles
- Focus on minimizing hand-offs



Contact Us

~~~~~  
want to say hi?



[hi@datachef.co](mailto:hi@datachef.co)