DATA[‡]AI SUMMIT BY Sector databricks

Engineering Data Data Lakehouse Systems for the Next Ten Years of Growth

Adi Polak

Data & AI Technology Strategist

Databricks 2023



About me – Adi Polak

Thesis in Machine Learning and Cyber Security Author: Scaling Machine Learning book Work experience: Fortune 500 and startups Data & AI technology Strategist



O'REILLY'

Scaling Machine Learning with Spark

Distributed ML with MLlib, TensorFlow, and PyTorch

Adi Polak



amzn.to/3WgHQvd



Our 20 min together

* How NOT to build data Lakehouse systems

- Overcoming CAP Theorem
- Boil the Ocean
- Build Rube Goldberg Machines
- Don't Budget for Complexity
- Don't use Managed capabilities and in house knowledge



Overcoming CAP Theorem



Practical aspects of CAP Theorem

In designing a database, we can choose two of the three

- Consistency
- Availability
- Partition tolerance



Practical aspects of CAP Theorem

Distributed systems: provide two of three properties simultaneously:

Consistency, Availability, & Partition tolerance..





Boil the Ocean



Lambda architecture



Abstraction Levels

Hardware Operating system

Software framework

Cluster

System

Architecture

Build Rube Goldberg Machines



STROLES THROUGH A CACTUS FIELD IN HIS BARE FEET, AND SCREAMS OUT AN IDEA FOR A SELF-OPERATING NAPKIN.

As you raise spoon of soup (A) to your Mouth it pulls string (B), thereby Jerking LADLE (C) which throws cracker (D) past parrot (E). Parrot Jumps After cracker and perch(F) tilts, upset ting seeds (G) into Pail(H). Extra weight in Pail pulls cord(I) which opens and lights automatic cigar lighter (I), setting off sky-rocket (K) which causes sickle (L) to cut string (M) and allow pendulum with Attached Napkin to swing Back and forth thereby wiping off your. CHIN.

AFTER THE MEAL, SUBSTITUTE A HARMONICA FOR THE NAPKIN AND YOU'LL BE ABLE TO ENTERTAIN THE GUESTS WITH A LITTLE MUSIC.



"Self-Operating Napkin"

But didn't budget for complexity...







Emerging LLM App Stack



Forget to use Managed services and in house knowledge

My team has specific skills and we want to build in house generative AI platform

Work with lowcode to expend abilities and Turn AI problems into a data engineering problem



Sum it up

- CAP
- Boil the Ocean
- Build Rube Goldberg Machines
- Don't forget to:
 - Budget complexity
 - Existing skills & skills planning
 - Managed services



Thank you!

Linkedin.com/in/polak-adi/ Twitter: @adipolak