



THE MAKING OF AMAZON PRIME NOW

“Amazon is nimble and modern:
How hard could it be?”

Tisson Mathew
CTO, Alignment Healthcare

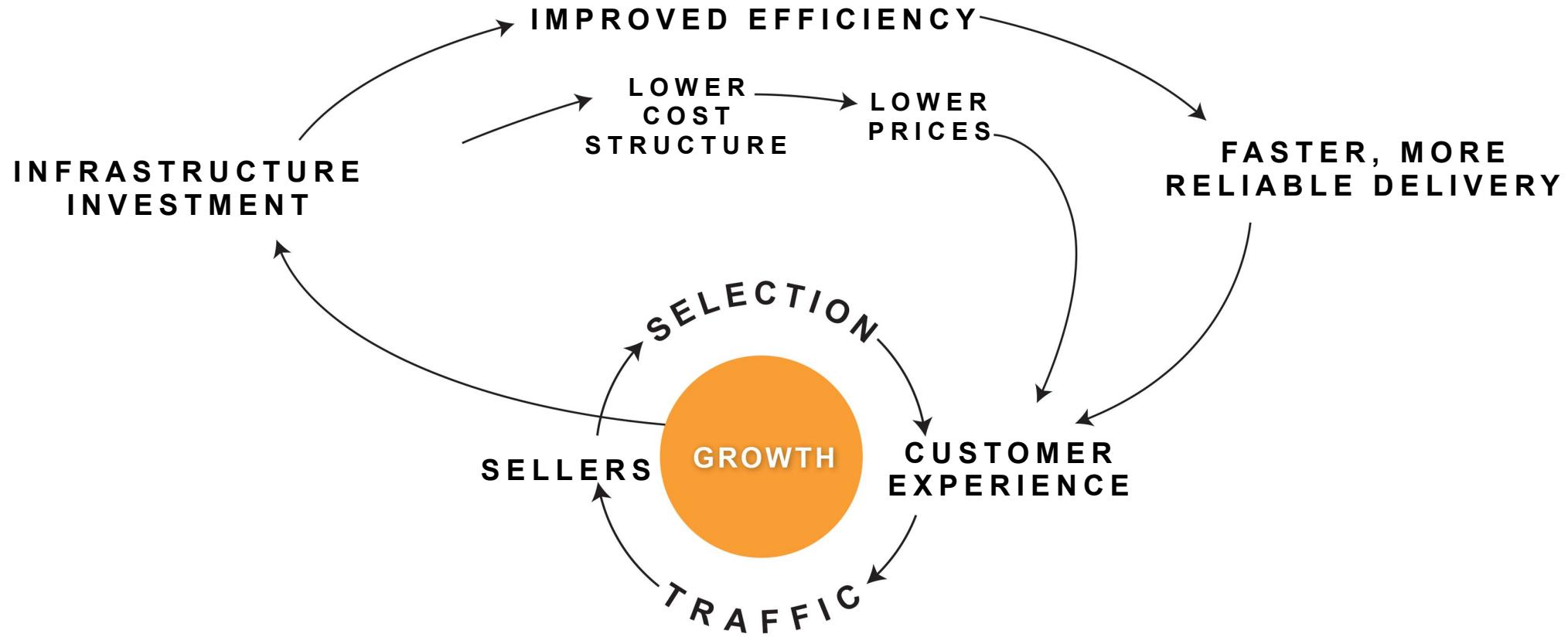
Former Director
Amazon Logistics Technology
Amazon.com

WHAT IS AMAZON PRIME NOW?

Prime Now offers household items and essentials you need every day plus the best of Amazon, with FREE 2-hour delivery.

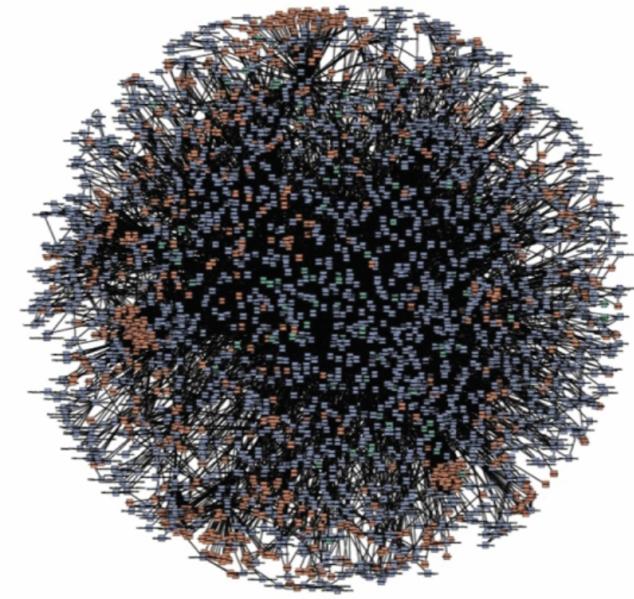


VIRTUOUS CYCLE



- **One-hour delivery**
 - Order using Prime Now app or website or Alexa, get your order at your doorstep in an hour like a “Cloud 7-11 store!”
 - Prime Now focuses on one-hour deliveries of different types of products, making it inherently local.
- **Online consumer ordering has 3 major functions**
 - Ordering, Fulfillment and Delivery
 - Built to handle orders in 2 to 5 days or more
- **What should we do?**
 - Build everything our own?
 - Leverage what we built?
 - Both?

- Service-oriented for a long-time
- Lots of micro-services talking to lots of services connected via APIs.
- Run by end-to-end intact teams, “2-pizza” teams (team size limited to two pizzas for a team meal)
- A team owns one or more services; no service should be bigger than what a two-pizza team can run
 - Teams control their own destiny (planning, Dev, Ops, QA and financial ownership)
 - Teams exist as part of larger orgs, such as website or transportation, for example
- Mostly no coordination between teams when deploying services
- Each team releases on its own schedule, resulting in millions of deployments annually and growing
- Lots and lots of monitoring & alerting



Amazon internal services architecture circa 2009

SOLUTION APPROACH #1

- **Just deliver really fast, super fast!!**
 - Build new services on delivery function, use existing capabilities in ordering and fulfillment.
 - To leverage existing services, you either use it as a black box or wait for yourself to be prioritized by the service team (“2-pizza team”).
 - Have to deal with 100s of service teams, all operating with full autonomy, many months of cycle time.
- **What to do?**
 - Army of program managers?
 - Steering committees?
 - Lots of email or even “Slack” channels?
 - Be really good friends with product and engineering managers?
 - Escalate?

PROBLEM WITH APPROACH #1

- **Ordering handoff is not fast enough**
 - Good to ship the same day, not good to ship in 15 minutes of the order
 - Cutting across two of the largest (and global), most powerful parts of business
 - Ordering
 - Fulfillment
- **Solution**
 - No changes to ordering
 - Integrate Fulfillment and Delivery functions
 - Steering committee and a rock-star TPM
 - Limit fulfillment changes to happy path only
 - Build a façade (or a composition service) in the delivery function

SOLUTION APPROACH #2

- **Evolution of Approach #1**
 - Integration tests of happy path (end to end), cutting across 100s of services
 - Monitoring and alerting
 - Test war rooms
- **“Façade” service became a mainstream service**
 - Decoupling
 - Provide caching
 - Dummy data provider
 - Acted as the request – response gateway to ordering, fulfillment and transportation
 - Further decoupling needed in delivery with crowdsourced transportation (“Amazon Flex”), good learning

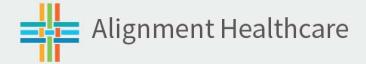
WHY SERVICE COMPOSITION?

- Too many fine-grained micro-services
- Green fields in enterprises are often limited
- Business functionalities often require to interact with more than one (micro) service
- Reality requires “Integration”
- Hybrid usage of micro-services and monolithic architecture is a reality
 - Need to integrate with existing legacy systems, proprietary systems, databases and SaaS
- Industry examples: Netflix API, Uber Edge Services, PayPal API Façade
- API gateways complement compositions (e.g., AWS API Gateway, Apigee, Azure API Management, etc.)

KEY LEARNINGS

- Pros and Cons of decoupled services
- End-to-End intact teams and DevOps are awesome!
- Enterprise-wide programs that cross cut many services have major business and engineering challenges
- How to minimize redundancy
- Utilize the right-sized teams of TPMs and Steering Committees
- Integration services (“mini services”) play a key part of MSA
- Understanding service granularities
- Technologies for building integration micro-services
- Single-threaded owner / leader for the program or initiative with direct access to the CEO

CURRENT STATUS OF PRIME NOW!



- 30+ US cities, UK and growing
- Became a launch pad to new businesses like Amazon Flex, Amazon Logistics, etc.
- 25K items across 25 categories, including things like household items, groceries, electronics, gifts, seasonal items, and more
 - Plus, select cities have restaurant and local store delivery
- Free two-hour delivery is one way of making the already hyper-convenient Amazon even better, to say the least
- Not unlike many of Amazon's innovations, this one feels particularly futuristic

Thank you!