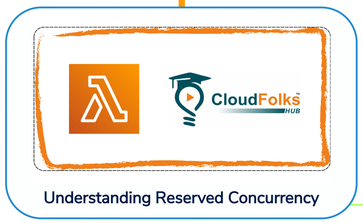

Understanding Reserved Concurrency

1 Introduction

1. This represents the maximum number of concurrent instances allocated to your function
2. When a function has reserved concurrency, no other function can use that concurrency
3. Reserved concurrency is useful for ensuring that your most critical functions always have enough concurrency to handle incoming requests
4. Configuring reserved concurrency for a function incurs no additional charges

2 Realistic Business Example

1. Let's say you run an e-commerce website
2. You have 3 Lambda functions
   1. FunctionA – PlaceOrderFunction
      1. 👉 Handles user purchases and payment
      2. 👉 It's very critical – if this fails or slows down, you lose money and customer trust
   2. FunctionB – SendEmailFunction
      1. 👉 Sends order confirmation emails
      2. 👉 Important, but can tolerate delay
   3. FunctionC – GenerateReportFunction
      1. 👉 Generates admin sales reports every hour
      2. 👉 Not urgent at all
3. ❌ Problem Without Reserved Concurrency
   1. Although AWS sets a default soft limit of 1000, in some accounts (especially new or less-used accounts), AWS might start with a lower limit, like 400
   2. You can request an increase to the full 1000 or more, based on your use case
   3. In our example, the account-level concurrency limit is 400
   4. Suddenly, it's the end of the month and FunctionC starts generating heavy reports
   5. At the same time, FunctionB is sending thousands of promotional emails
   6. They together consume 350 concurrency
   7. Now only 50 slots are left for FunctionA
      1. 👉 Customers trying to place orders experience delays or errors
      2. 👉 Your business loses revenue and customer satisfaction drops
4. ☑ Solution – Use Reserved Concurrency
   1. You configure FunctionA → Reserved Concurrency = 300
      1. FunctionA will always have up to 300 executions available
      2. Even if the other functions get overloaded, they can't touch those 300
      3. FunctionA can continue to process orders smoothly during peak hours
   2. 🎯 Why This Matters
      1. Even if reports are delayed or emails are queued — it's okay. But if customers can't place orders, the business suffers
      2. That's why we protect critical functions using Reserved Concurrency