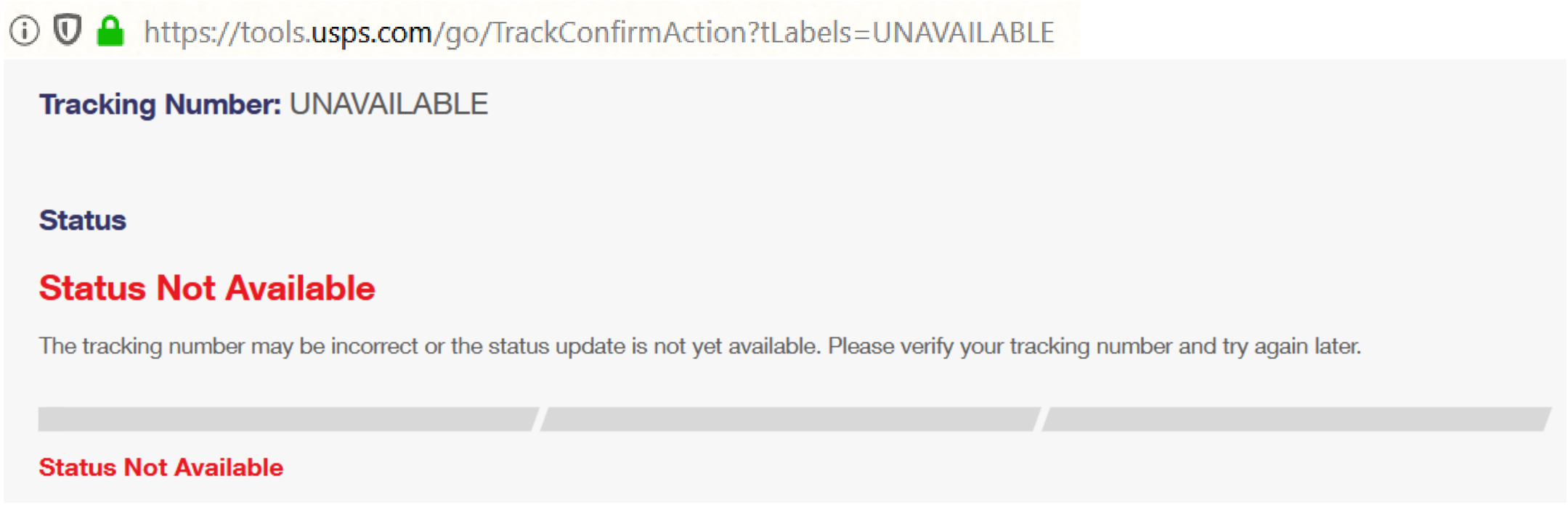
DataArt

Typical bugs in technical architecture

Denis Tsyplakov
Solution Architect // DataArt

“We mailed your card to the address you gave us. The tracking number assigned is UNAVAILABLE. You can use your tracking number at www.USPS.com in the Quick Tools Tracking section.”



Possible reason

DataArt

I do not know for sure, but probably there was something like this

```
public String getTackingNumber(String id){
    try{
        // do some steps to get tracking number
        // 1 - step 1
        // 2 - this step could throw IOException
        // 3 - step 3
        return "some tracking number"
    }catch (IOException ex){
        log.warn("Error {} happened while trying to get tracking number for
id {}",ex,id);
        return "UNAVAILABLE";
    }
}
```

Why do I want to speak about this topic?

DataArt

- Past several years I audited lots of projects
- Obviously if somebody pays you money for project audit – this project has problems (we call them issues ☺).
- Some problems are unique
- But some problems are very common
- I want to speak about common bugs

Bug hierarchy

DataArt

- Bugs in code: +1/-1; NPE; wrong method parameters; etc
- **Bugs in technical architecture**
- Bugs in system architecture
- Bugs in product design
- Bugs in business strategy

What is technical architecture?

DataArt

Technical architecture is a set of architectural decisions you make when you design an executable unit (service, application etc). This usually includes:

- Choice of technology
- Build tools and infrastructure
- Class & package hierarchy;
- Layers of data processing;
- Typical dataflow & call chains
- Data structure design

Example: requirements

DataArt

In terms of our car rental system, we need to design a service that will do the following

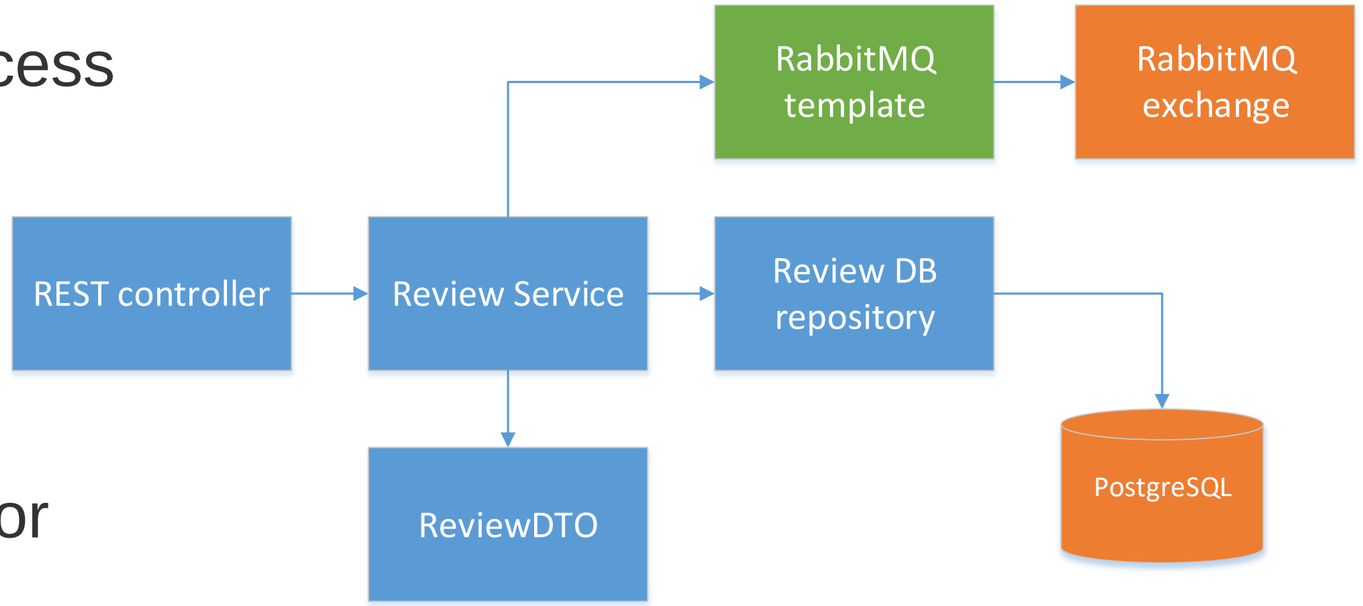
- Store user reviews for different cars models in database;
- Provide list of reviews via REST API in a paginated form
- In case if review has rating <2 – send notification to a RabbitMQ exchange

Nonfunctional requirements: we have <100 models. Typically we have 10 reviews per day. We need to return first (most recent) 3 reviews for each model in 1 second.

Example: possible solution

DataArt

- Spring Boot 2.x based service.
- Spring Boot starter web
- PostgreSQL as a main DB, access via Spring Data JDBC
- Spring AMPQ
- Spring Caching
- Cache 3 most recent reviews for each model in memory



How to design right technical architecture?

DataArt

Technical architecture could be different, you could have different priorities and have different believes. On the slide above we can use:

- RxJava
- Kotlin
- R2DBC
- JBoss

It is difficult to give an advice that will be good for all teams in all cases.

But, based on failures that I saw it is easy to give an advice, what you must not do.

Top 7 bugs in technical architecture

Incorrect error handling

DataArt

Every statement in the code could throw an error, this could be:

- Any input/output operation. Input/output is a synonym to Exception
- Issues with operational system
- Bugs in your code, library code or issue on JVM -- java.lang.UnknownError - «something bad happend, I am not sure what exactly»

One of the most popular ways of adding severe bugs – «exception suppressing» i.e. intercept an error, write log message and continue execution with some random data.

80% of project I am audited – had such bug.

How to fix, part I?

DataArt

Basically there are two valid options:

1. In case of any error – shutdown a service instance (it must comply with Crash-only-design), so container orchestrator (K8S) will launch new fresh instance.

Pros:

- Easy to implement
- Your exception processing code will not throw new exception while processing old exception

Cons:

- Not always applicable, what if you code is an orchestrator? Or your code controls automobile breaks?
- Such approach could cause cascade micro-service system blackout wave – case when shutdown of one secondary service causes restart of dozens of primary services.

How to fix, part II?

DataArt

2. Analyze exception root cause and mitigate it in an optimal way. E.g. in an example above – what could a mitigation way in case if RabbitMQ call throws an error?

Pros:

- Applicable to any application
- Much more resource wise and eco friendly
- Does not cause side effects

Cons:

- Requires much more time to implement
- You could make an error in error processing code that will throw an error
- In general requires higher programming skills and understanding of computer science

Do not get me wrong in 99.9% of all cases I recommend to use approach #2

Ignoring data integrity contract

DataArt

Not all possible combinations of data in DB are consistent and meaningful from an application standpoint. Especially if application is using more then one data storage. In this cases transactions are not an ultimate answer

Example – when you rent a car application need to

1. Make a record in the main DB
2. Charge money from client's credit card
3. Do some financial reporting
4. Inform parking IoT system that this car is rented and could leave parking
5. Notify driver that car must be delivered to a client
6. Arrange insurance

What if we are not on a happy path and our program stopped (due to broken CPU) at step N? Are we in consistent state? What will happen with a client if steps 5 and 6 are omitted?

How to fix?

DataArt

Main point – application flow will be interrupted at some random point. There is no way to design 100% error free application, just because hardware could always fail at some random point.

Analyze data state in case when business transaction sequence was stopped at some random point. What is error price, who will die?

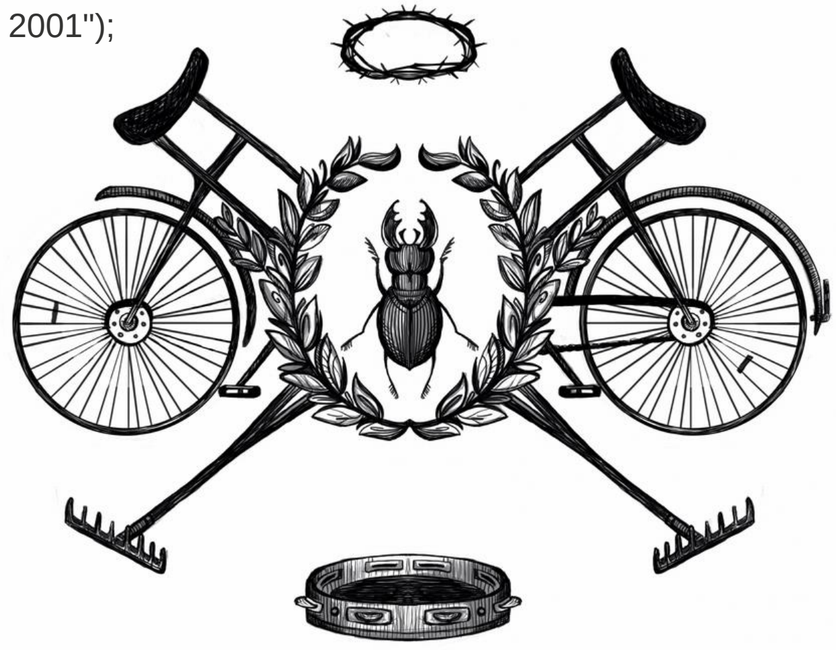
What is application transaction model. What data will rollback automatically what you need to fix manually?

In case of queues – does application has “1 or many” or “0 or 1” delivery model? How to fix “0” and “many” cases?

Ignoring standard frameworks and libraries

```
try {
        String url = "jdbc:msql://localhost:1114/Demo";
        Connection conn = DriverManager.getConnection(url,"","");
        Statement stmt = conn.createStatement();
        ResultSet rs = stmt.executeQuery("SELECT Lname FROM Customers WHERE Snum = 2001");
        while ( rs.next() ) {
            String lastName = rs.getString("Lname");
            System.out.println(lastName);
        }
        conn.close();
    } catch (Exception e) {
        System.err.println("Got an exception! ");
        System.err.println(e.getMessage());
    }
```

2002 vs 2019



```
jdbcTemplate.queryForObject(
    "SELECT STREET_NAME FROM table WHERE ID=?", new Object[] { id }, String.class);
```

How to fix?

DataArt

Consider migration to a new framework that makes codebase smaller as roadmap item that decreases total cost of ownership. Smaller and cleaner codebase is a value itself.

Wordy, heavy weight code is a technical debt

On regular basis review possible evolution options

Overuse of 3rd party libraries

DataArt

Real case – usage of a library that brings 1500 classes to an application because developer needs to use literally 10 lines of code from this library.

Negative effects:

- Startup time impact
- Long classpath breaks gradle bootRun.
- Library dependency conflicts (libA-> libZ:1.8 | libB -> libZ:1.9)
- IDE starts slowly and consumes lots of memory
- Code insight in IDE returns zillions of similar classes
- Library assets could impact other libraries. E.g. one library could load default resources from other library.
- Wider attack surface for security attacks

How to fix?

DataArt

If it is really 10 line of isolated code – just rewrite it

If not, search for lighter library with a smaller dependency tree.

Store all data read from DB/file system in memory

DataArt

Typical bug flow:

- Read all data from file with `Files.readAllBytes(file);`
- Debug, test, all OK, package, rollout version
- Found, that on the production server on another side of the Globe typical size of a data file is 5Gb and typical VM has just 4Gb of RAM.

Or another case

- Implement pagination by reading all data in memory and the slicing data array to pages
- Works perfectly when you have one user and 10,000 rows in DB
- On prod you have 100 simultaneous user requests and 1,000,000 rows and you are in trouble.

How to fix?

DataArt

Know data specifics. How big file could be? Is it 1kb config file or 1Tb log file?

Understand common numbers related to business area. E.g. how many clients you will have in DB table day 1 and in 3 years. Is it 10 clients or 10,000,000 clients?

Each time you do I/O operation – ask yourself – what is min/avg/max number of bytes that would be processed with this instruction. At least approximately – are we talking about 1000 bytes or 1000 Gbs.

Build & Deploy is complex and fragile

- Application is not buildable with simple `gradle build` (or maven)
- Build is not stable. Yesterday it builds, today it is not.
- Build is fragile and depends on zillions of tricky settings that developer should do using 30 steps Readme.MD
- Build is not universal. E.g. application is buildable only on Ubuntu 16.04.6 LTS and you just have bought new MacBook Pro

How to fix?

DataArt

When you step beyond template provided by Spring Initializr you have a choice

1. Do it fast, but probably little bit hacky
2. Do it in right way, but slower

You need to remember that build complexity is a huge multiplicator to all developer efforts. I know a project where time from “git clone” to first “change line of code; build; run” is about a week.

Always check – “Would this build work in 3 years, for a junior developer in Australia, who just started his first day of work?”

Test system is a black box

- System has >1000 tests of all kinds (unit, functional, integrational) that provides 90% code coverage.
- You can run these tests against set of services defined in docker compose yaml.
- What functional and nonfunctional requirements these services really tests?
- Could you run these tests with any valid system installation or just with docker compose?
- What tests do you need to change when you change system functionality?

How to fix?

DataArt

Quality of tests is ratio:

> *(% coverage of functional and nonfunctional requirements) / (size of test source code)*

I.e. test code

- Must be compact, smart, transparent
- Must not be meaningless, bloated, extensive blackbox

In a reasonable amount of time new developer for a give functional requirement should be able to find – is it covered by some test, where test is located, how to updated the test

Do not write tests, just to write tests

Thank you!