accenture
Latvia ATC

Тестирование в мире BIG DATA

by Konstantin Pletenev

GROW
CONFI
DENTLY

SQA DAYS#23

BIG DATA IS NOT ABOUT THE DATA

THE REVOLUTION IS NOT THAT THERE'S MORE DATA AVAILABLE

THE REVOLUTION IS THAT WE KNOW WHAT TO DO WITH THIS DATA NOW

—GARY KING
PROFESSOR OF HARVARD UNIVERSITY

accenture
Latvia ATC

65 МИЛЛИОНОВ ПОСТОВ В СОЦИАЛЬНЫХ СЕТЯХ КАЖДЫЙ ДЕНЬ

1 ЧАС, 16 МИНУТ КАЖДЫЙ ДЕНЬ ТРАТИТСЯ НА ПРОСМОТР ВИДЕО С МОБИЛЬНОГО УСТРОЙСТВА

100 МИЛЛИОНОВ ПОДПИСЧИКОВ НА СТРИМИНГОВЫЙ СЕРВИС NETFLIX

SQA® DAYS#23

accenture
Latvia ATC

ЧТО ТАКОЕ
BIG DATA

SOA® DAYS#23

ЧТО НЕ BIG DATA

accenture
Latvia ATC

- НЕ о традиционных данных, таких как документы и базы данных
- НЕ просто еще одно обозначение «Большого количества данных»
- НЕ только о данных

SQA® DAYS#23

ЧТО ЖЕ BIG DATA

accenture
Latvia ATC

- **BIG DATA** это большой объем данных обрабатываемый на ежедневной основе
- **BIG DATA** это то, что мы можем делать с этими данными
- **BIG DATA** нужна для принятия лучших и правильных стратегических решений

SQA DAYS#23

3 **BIG DATA** ПРОЕКТА В ACCENTURE ЗА 2 ГОДА

accenture
Latvia ATC

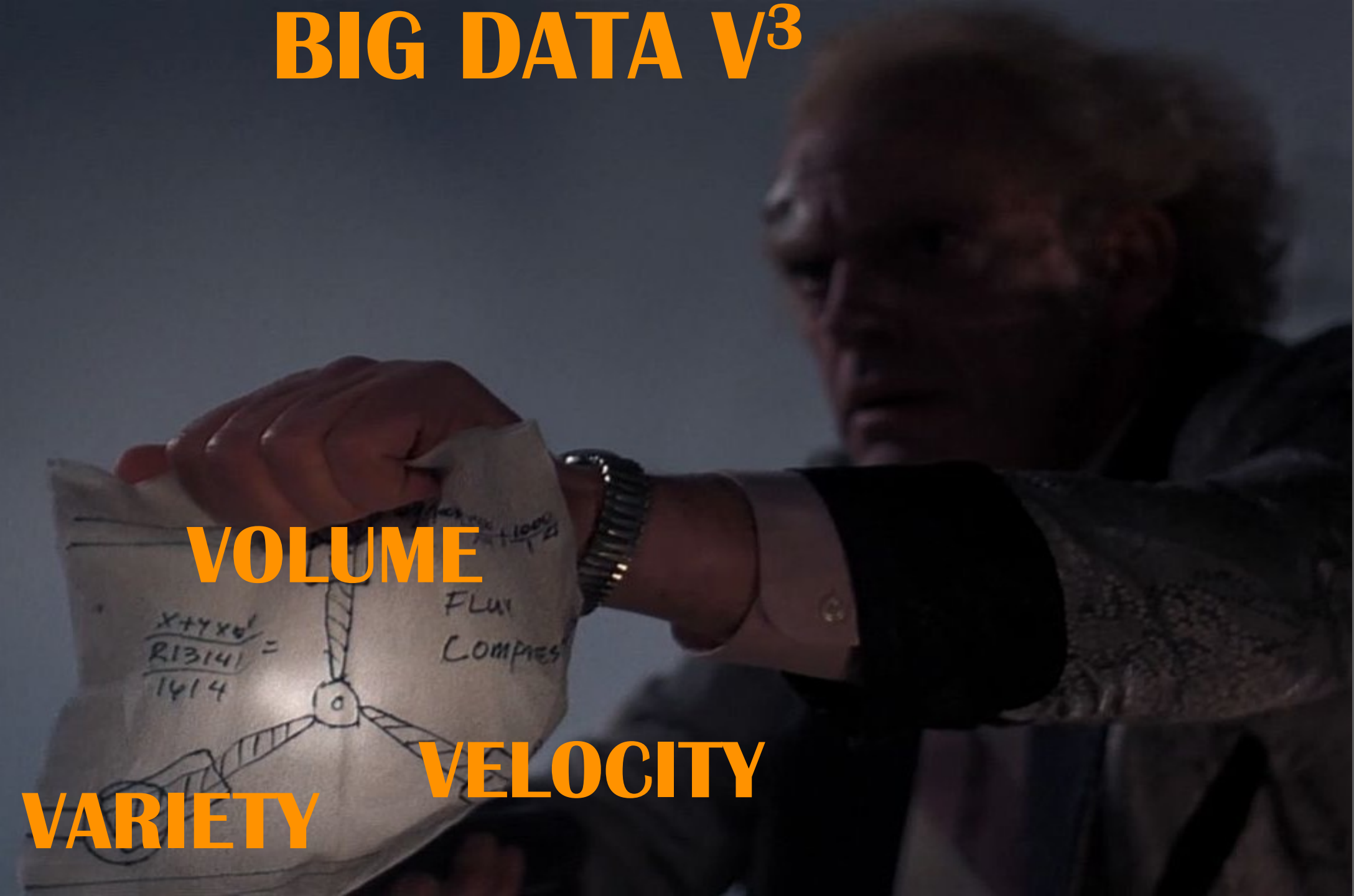
1. **ОБРАБОТКА ДАННЫХ О ПОЛЬЗОВАТЕЛЬСКОЙ АКТИВНОСТИ И ИХ СЕРИАЛИЗАЦИЯ**

2. **СИСТЕМА ПРЕДОСТАВЛЕНИЯ ВИДЕО КОНТЕНТА И ПОДСЧЕТ KPI ДЛЯ БИЗНЕС-АНАЛИТИКИ**

3. **ОБРАБОТКА И ПОДГОТОВКА ДАННЫХ О ПОЛЬЗОВАТЕЛЬСКОЙ АКТИВНОСТИ ДЛЯ DATA SCIENTISTS**

SQA DAYS#23

BIG DATA V^3



accenture
Latvia ATC

SOA® DAYS#23

V¹ - VOLUME

FACEBOOK STATISTIC

accenture
Latvia ATC

- 100 PB НА ОДНОМ HDFS КЛАСТЕРЕ
- 300 TB ДАННЫХ ОБРАБАТЫВАЮТСЯ В HIVE КАЖДЫЙ ЧАС
- 500 TB НОВЫХ ДАННЫХ ДОБАВЛЯЮТСЯ В ХРАНИЛИЩЕ КАЖДЫЙ ДЕНЬ

https://www.facebook.com/notes/facebook-engineering/under-the-hood-hadoop-distributed-filesystem-reliability-with-namenode-and-avata/10150888759153920/

SQA DAYS#23

V² - VELOCITY

NETFLIX STATISTIC

accenture
Latvia ATC

- **37% ВСЕЙ** ПРОПУСКНОЙ СПОСОБНОСТИ ИНТЕРНЕТА В СЕВЕРНОЙ АМЕРИКЕ
- АНАЛИЗ ПОСТУПАЮЩИХ ДАННЫХ В **РЕАЛЬНОМ ВРЕМЕНИ**
- **100,000** СЕРВЕРНЫХ ИНСТАНЦИЙ

https://www.linkedin.com/pulse/netflix-dominates-fixed-internet-video-bandwidth-mobile-jeff-nelson

SQA® DAYS#23

V^3 - VARIETY

INSTAGRAM STATISTIC

accenture
Latvia ATC

- **300 MB** ИЗОБРАЖЕНИЙ ЗАГРУЖАЕТСЯ КАЖДУЮ МИНУТУ
- **4.7 МИЛЛИОНА** ЛАЙКОВ КАЖДЫЙ ДЕНЬ
- **1 МИЛЛИОН** ПРОСМОТРОВ СТОРИС В ДЕНЬ

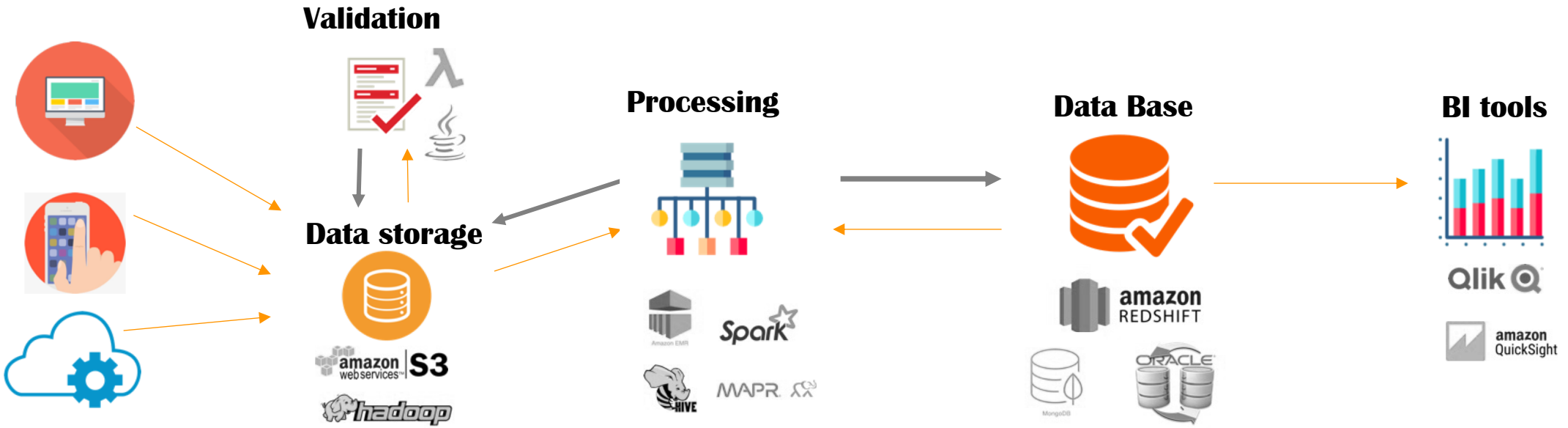
https://www.forbes.com/sites/bernardmarr/2018/03/16/the-amazing-ways-instagram-uses-big-data-and-artificial-intelligence/#1524dcf25ca6

SQA® DAYS#23

ПРИМЕР BIG DATA PIPELINE

accenture
Latvia ATC

Ежедневная обработка / Batch Pipeline

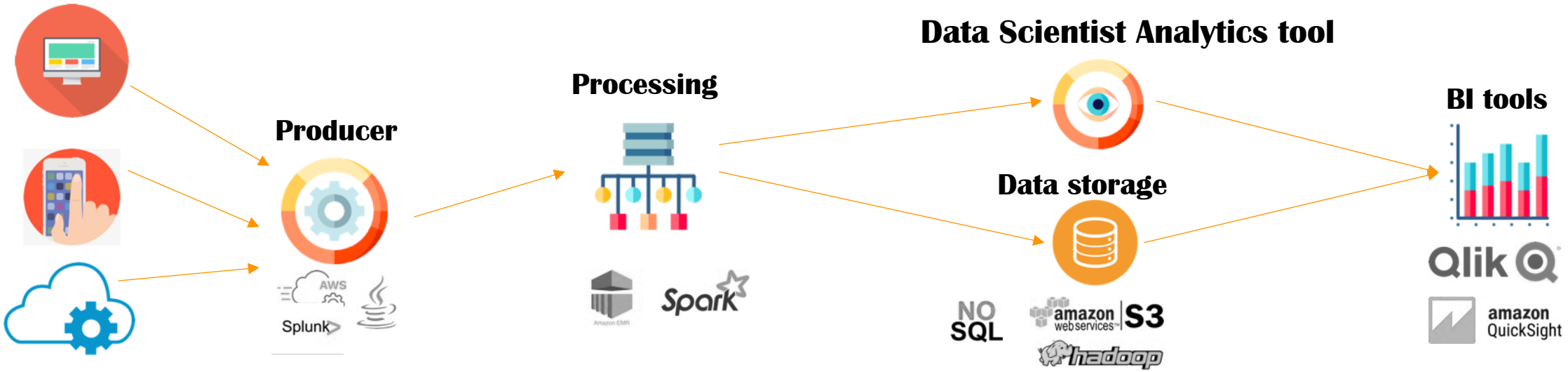


SQA DAYS#23

ПРИМЕР BIG DATA PIPELINE

accenture
Latvia ATC

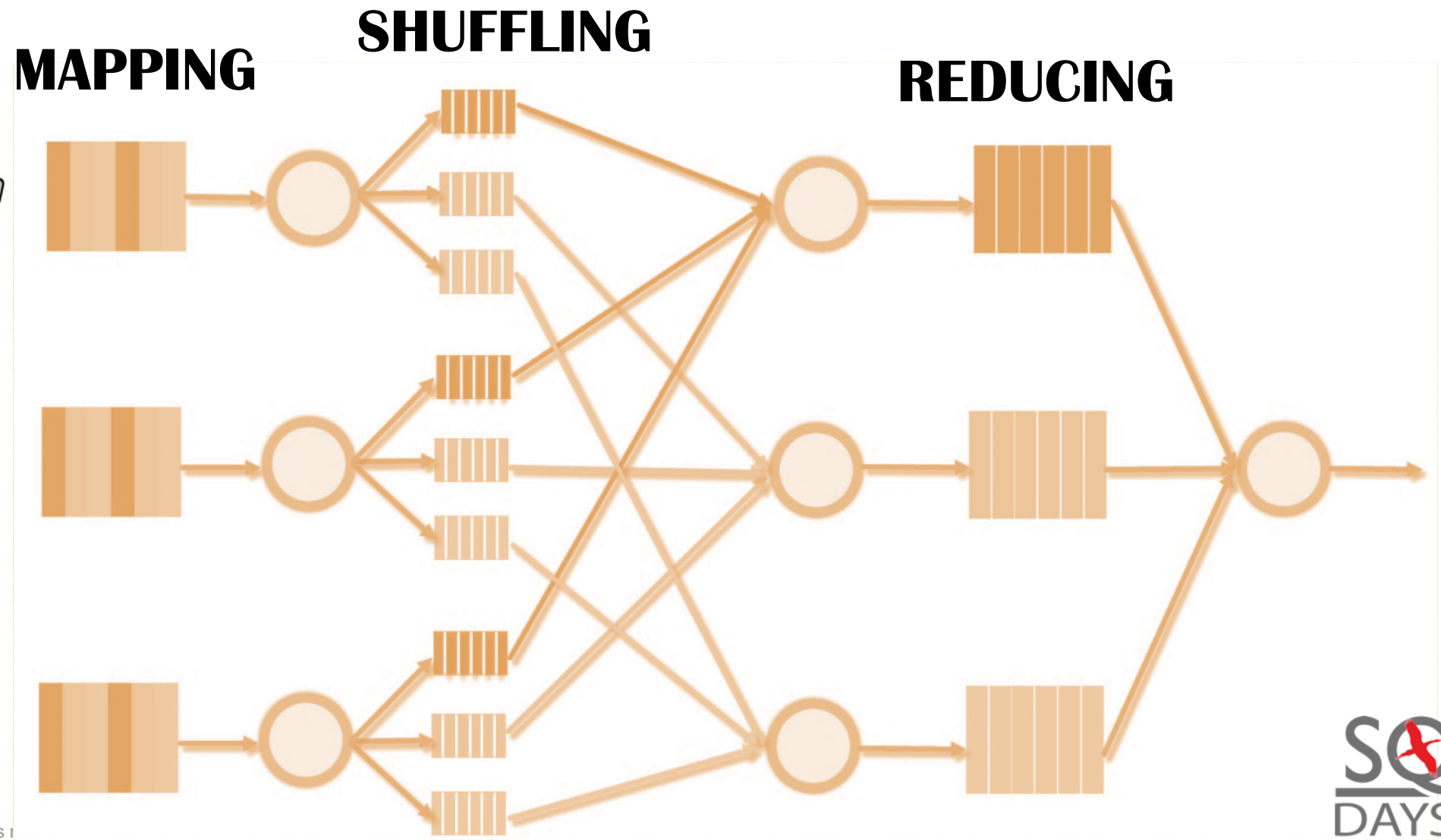
Обработка в реальном времени / Real Time Pipeline



SQA DAYS#23

BIG DATA MAP REDUCE

accenture
Latvia ATC



hadoop

Spark

Splunk>

talend*
*open integration solutions

SQA DAYS#23

accenture
Latvia ATC

ТЕСТИРОВАНИЕ BIG DATA VS ТРАДИЦИОННОЕ ТЕСТИРОВАНИЕ

SQA DAYS#23

Область	Традиционное тестирование	BIG DATA тестирование
Данные	Тестировщики работают с структурированными данными	Тестировщики работают одновременно и с структурированными данными и с неструктурированными данными
	Подходы к тестированию определены и проверены временем	Подходы к тестированию сфокусированы больше на R&D
	Тестировщики могут применять "Sampling" техники для тестирования	"Sampling" техника в Big data не работает или требует очень больших затрат
	Возможно ручное тестирование	Возможно только автоматизированное тестирование

accenture
Latvia ATC

SQA DAYS#23

accenture
Latvia ATC

Область	Традиционное тестирование	BIG DATA тестирование
Инфраструктура	Испытуемая система может быть запущена локально	Требует отдельного окружения для запуска из-за объема данных и количества файлов (HDFS)
Инструменты валидации	Тестировщики имеют множество инструментов, начиная от Excel и заканчивая инструментами UI автоматизации	Нет специальных инструментов именно для тестирования, тестировщики используют тот же набор инструментов, что и разработчики (MapReduce, HIVE)
	Инструменты могут быть использованы без специальных знаний	Требуется знание программирования и прохождения тренингов по работе с данными

SQA DAYS#23

ШАГИ К УСПЕХУ

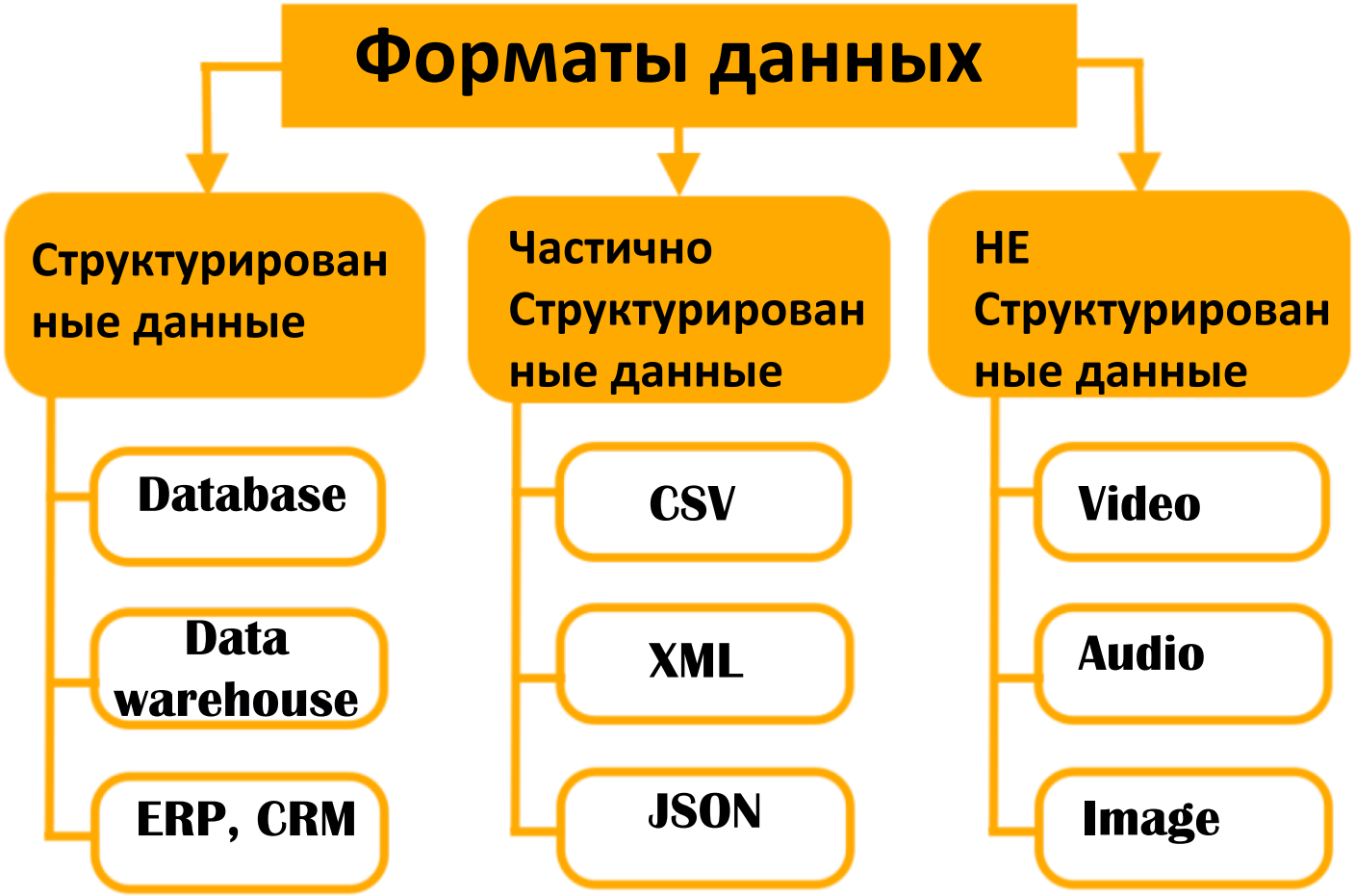
accenture
Latvia ATC

1. ПОНЯТЬ СВОИ ДАННЫЕ
2. ПОНЯТЬ СВОЙ ПРОЦЕСС (BATCH TIME/REAL TIME)
3. АВТОМАТИЗИРОВАТЬ, АВТОМАТИЗИРОВАТЬ И ЕЩЕ РАЗ АВТОМАТИЗИРОВАТЬ
4. ПРИГОТОВЬТЕСЬ К БОЛЬШОМУ ОБЪЕМУ НАГРУЗОЧНОГО ТЕСТИРОВАНИЯ

SQA DAYS#23

ПОНЯТЬ СВОИ ДАННЫЕ

accenture
Latvia ATC



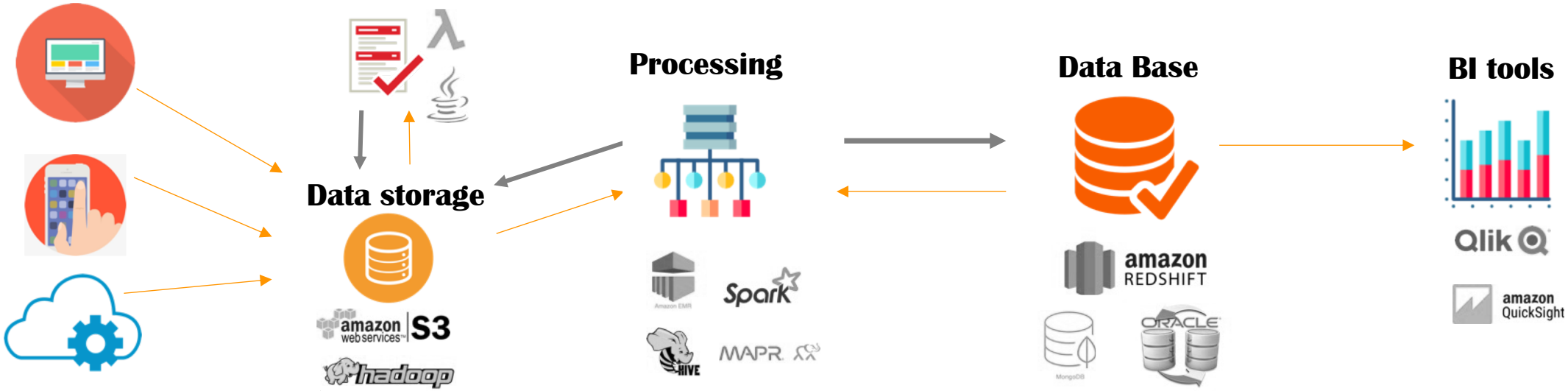
SQA DAYS#23

ПОНЯТЬ СВОИ ПРОЦЕССЫ

accenture Latvia ATC

Ежедневная обработка / Batch Pipeline

1. ГЕНЕРАЦИЯ ДАННЫХ
2. ПРОВЕРКА ДОСТОВЕРНОСТИ ДАННЫХ
3. ПРОВЕРКА ОБРАБОТКИ ДАННЫХ
4. ПРОВЕРКА ВЫВОДА



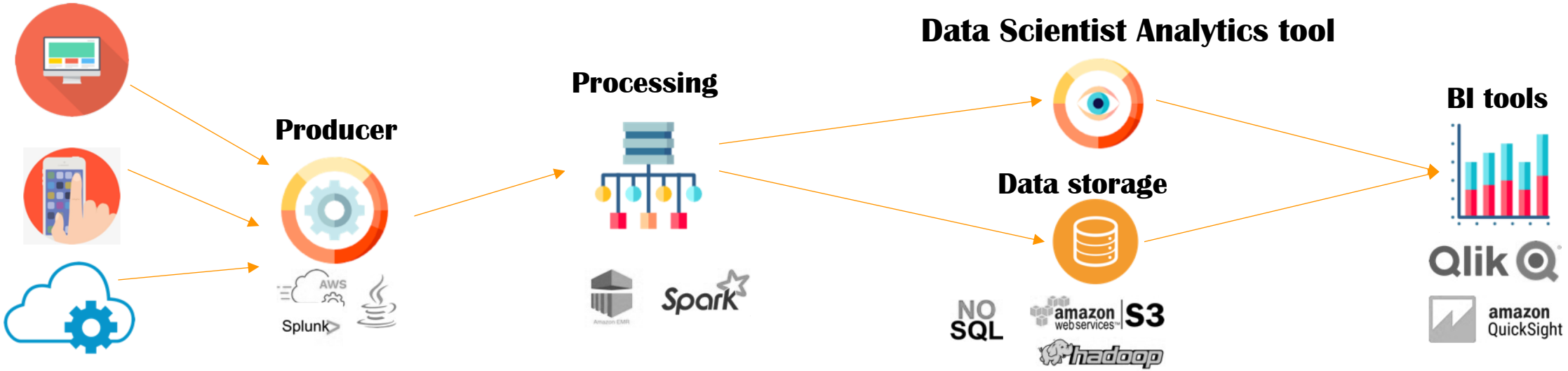
SQA DAYS#23

ПОНЯТЬ СВОИ ПРОЦЕССЫ

accenture Latvia ATC

Обработка в реальном времени / Real Time Pipeline

1. СИМУЛЯЦИЯ АКТИВНОСТЕЙ
2. АНАЛИЗ СКОРОСТИ ОБРАБОТКИ ДАННЫХ
3. АНАЛИЗ МЕТРИК
4. ПРОВЕРКА ВЫВОДА

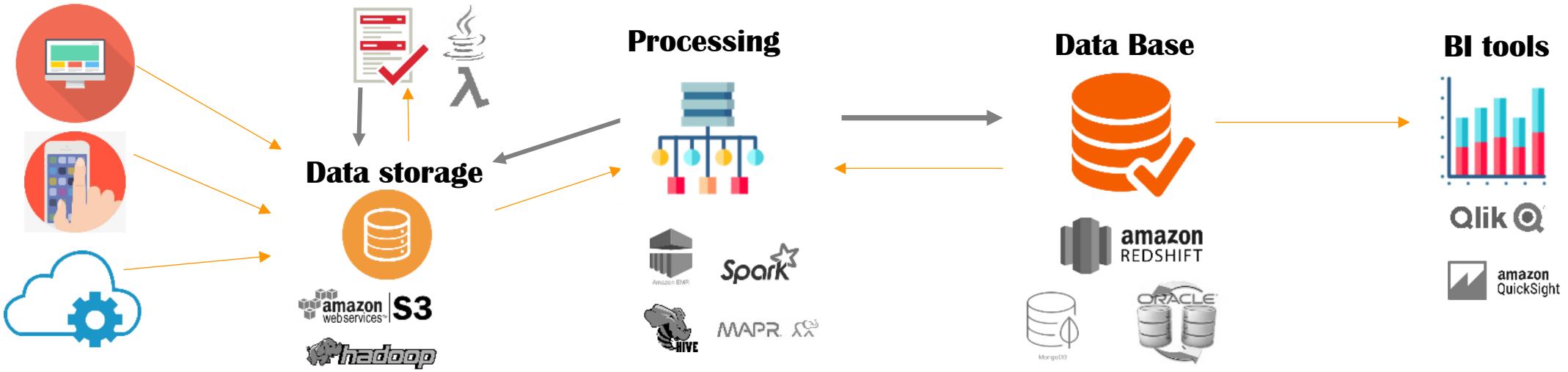


SQA DAYS#23

TOOLSET ДЛЯ ТЕСТИРОВАНИЯ

accenture
Latvia ATC

Batch Pipeline



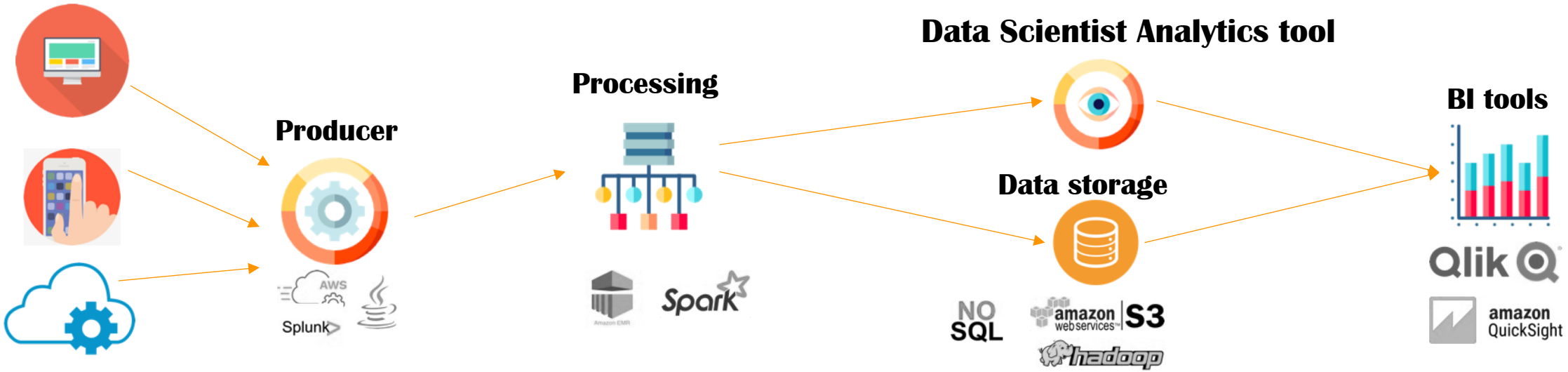
Data generation	Acceptance testing	Data validation testing	UI testing
talend* *open integration solutions	Java	talend* *open integration solutions	POSTMAN
cucumber	talend* *open integration solutions	cucumber	Java
Java	cucumber	Java	Se
		ZEPHYR	cucumber

SQA DAYS#23

TOOLSET ДЛЯ ТЕСТИРОВАНИЯ

accenture
Latvia ATC

Real Time Pipeline



Activity simulation

talend
*open integration solutions
Gatling
LOAD TESTING
Scala
Java

Performance testing

Gatling
LOAD TESTING
Scala
Splunk
AWS

Acceptance testing

Java
talend
*open integration solutions
cucumber
ZEPHYR

UI testing

POSTMAN
Java
Se
cucumber

SQA DAYS#23

BDD СЦЕНАРИИ ДЛЯ ПРИЁМОЧНОГО ТЕСТИРОВАНИЯ

accenture
Latvia ATC

```
@Daily @VOD @Guest @Content @1day
Scenario: Total views of the guest users by content type = '%VOD%'
  Given following data was ingested and aggregated
    |partition_date |action_type   |user_id    |platform_type|count|content_type|
    |@date          |accessesviews|@guest_1   |@platform_0  |  1  |content_1   |
    |@date          |accessesviews|@guest_2   |@platform_0  |  1  |content_2   |
    |@date          |accessesviews|@guest_4   |@platform_0  |  1  |content_1   |
    |@date          |accessesviews|@guest_3   |@platform_0  |  1  |content_1   |
    |@date          |accessesviews|@user_id_1 |@platform_1  |  1  |content_2   |

  When I execute following  <query>
    """
    SELECT views
    FROM agg_views_daily
    WHERE partition_date = '@date' and user_type like '@guest' and contentype like '@content_1'
    """
  Then I expect to receive following results
    | views   |
    |    3    |
```


SQA DAYS#23

BDD СЦЕНАРИИ ДЛЯ ВАЛИДАЦИИ ДАННЫХ

accenture
Latvia ATC

```
@Negative @Mandatory
Scenario: Check existing of mandatory fields for Join Flow
  Given Generate Join csv with this data
    | TENANTID                | USERNAME      | UNIQUECONTRACT | TIMESTAMP      |
    | mandatory_fields_for_Join |               | 1000000000000  | 20160301000001 |
    | mandatory_fields_for_Join | test_username |                | 20160301000001 |
    | mandatory_fields_for_Join | test_username | 1000000000000  |                |

  When I execute following query
  """
    SELECT  REASON
    FROM    JOIN_REJECTED
    WHERE   BODY like '%mandatory_fields_for_Join%'
    ORDER BY REASON
  """
  Then I expect to receive following results
    | reason                                    |
    | Mandatory field 'timestamp' is empty      |
    | Mandatory field 'uniquecontract' is empty |
    | Mandatory field 'username' is empty       |
```


SQA DAYS#23

BDD И DDT ДЛЯ ГЕНЕРАЦИИ ДАННЫХ

accenture
Latvia ATC

inde	group	id#	partition_dat	user_type	sessionid	action_type	cour	platform_typ	currenc	revenu
287	agg_purchases	purchases (7)	date+2	user_id_1	session_0	purchase	1	platform_0	EUR	12
288	agg_purchases	purchases (7)	date+2	user_id_8	session_0	purchase	2	platform_0	EUR	5
289	agg_purchases	purchases (7)	date+2	user_id_9	session_0	purchase	1	platform_0	USD	5
290	agg_purchases	purchases (7)	date+2	user_id_11	session_11	purchase	1	platform_11	EUR	2

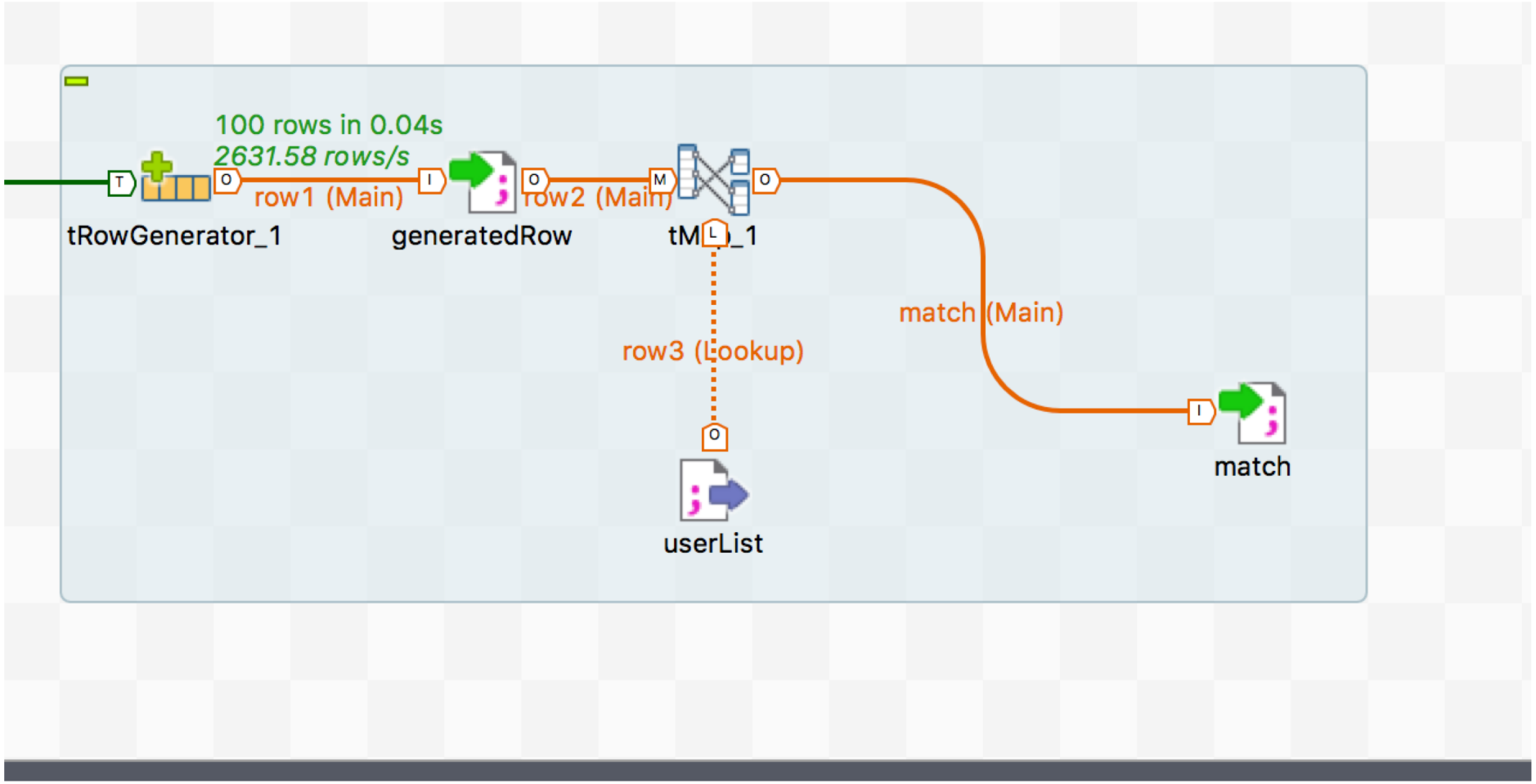
```
@Revenues @Daily @Platform @Currency
Scenario: A_162,  Revenues ( USD, GBP, EUR ) by currency = (EUR and USD) and platform and state for daily agg
  Given following data was ingested and aggregated
    | partition_date | action_type | user_id     | region     | platform_type | currency | revenue | count |
    | @date+2        | purchase    | @user_id_1  | @region_1  | @platform_0   | EUR      | 12      | 1     |
    | @date+2        | purchase    | @user_id_8  | @region_8  | @platform_0   | EUR      | 5       | 1     |
    | @date+2        | purchase    | @user_id_8  | @region_8  | @platform_0   | EUR      | 5       | 1     |
    | @date+2        | purchase    | @user_id_9  | @region_9  | @platform_0   | USD      | 5       | 1     |
    | @date+2        | purchase    | @user_id_11 | @region_11 | @platform_11  | EUR      | 2       | 1     |
```


SQA DAYS#23

TALEND OPEN STUDIO

talend*
*open integration solutions

accenture
Latvia ATC

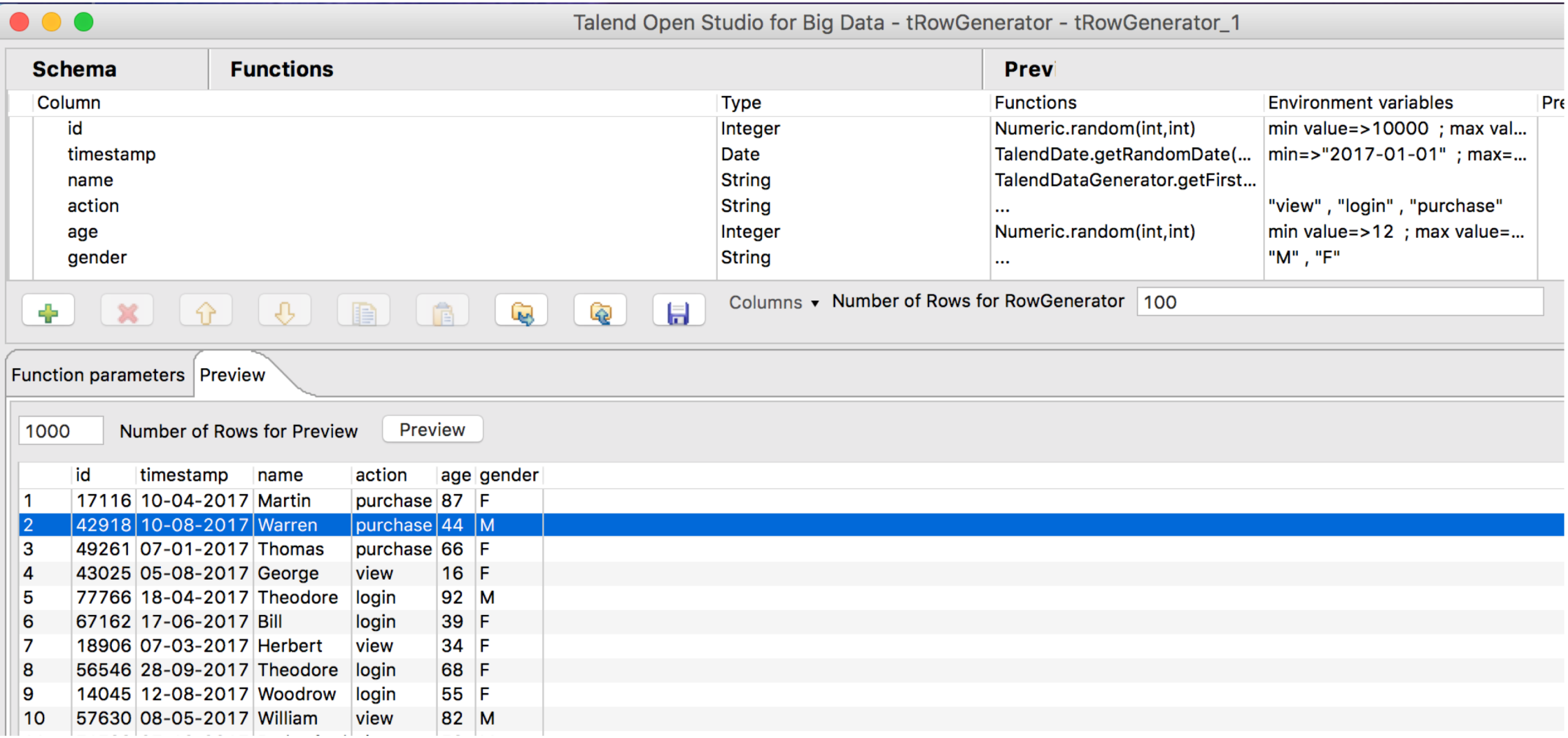


SQA DAYS#23

TALEND OPEN STUDIO

talend*
*open integration solutions

accenture
Latvia ATC

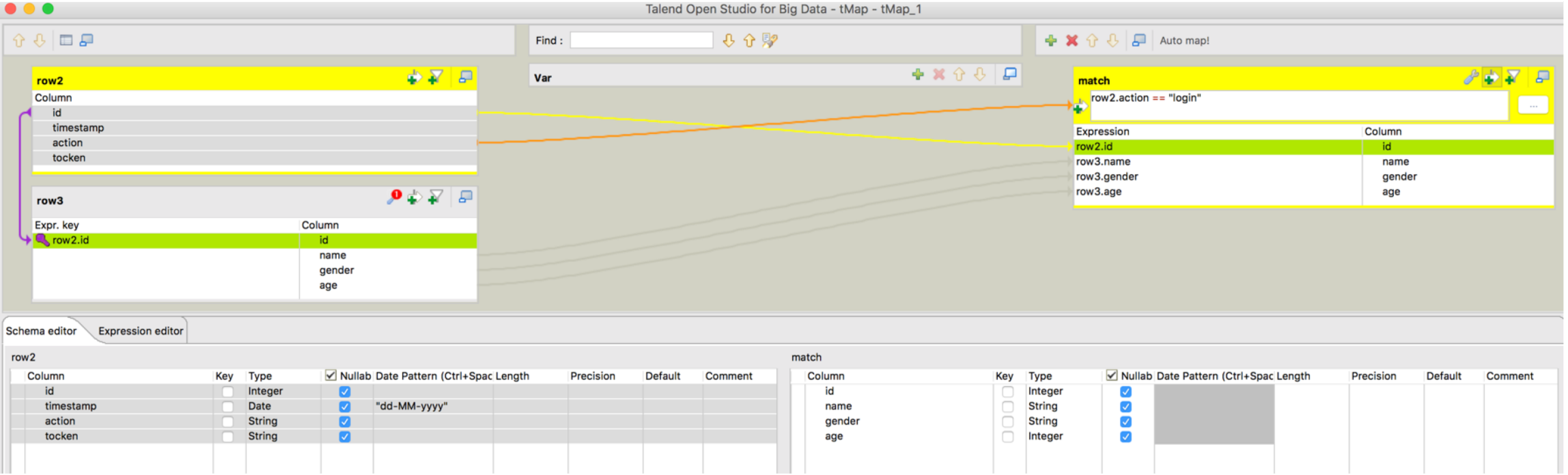


SQA DAYS#23

TALEND OPEN STUDIO

talend*
*open integration solutions

accenture
Latvia ATC

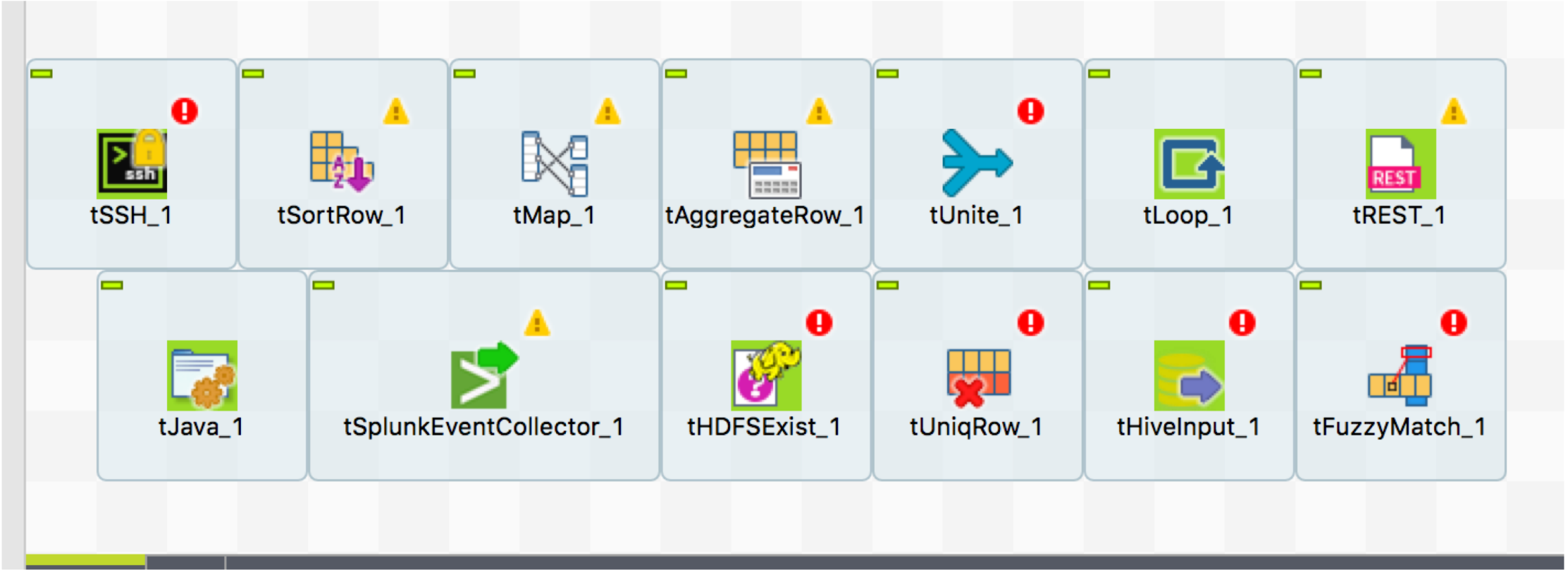


SQA DAYS#23

TALEND OPEN STUDIO

talend*
*open integration solutions

accenture
Latvia ATC



SQA DAYS#23

CONTINUOUS INTEGRATION

accenture
Latvia ATC

SQA DAYS#23

BDD ОТЧЕТЫ

accenture
Latvia ATC

Features Statistics

The following graphs show passing and failing statistics for features

Scenarios

	Steps						Scenarios			Features	
Feature	Passed	Failed	Skipped	Pending	Undefined	Total	Passed	Failed	Total	Duration	Status
A_20 Accesses	15	0	0	0	0	15	5	0	5	1s 323ms	Passed
A_21 Registered Accesses	15	0	0	0	0	15	5	0	5	328ms	Passed
A_25 Accesses	15	0	0	0	0	15	5	0	5	304ms	Passed
A_26 Accesses with View	12	0	0	0	0	12	4	0	4	22s 783ms	Passed
A_27 Accesses without View	30	0	0	0	0	30	9	0	9	12s 995ms	Passed
A_28 Failed Login	15	0	0	0	0	15	5	0	5	1s 790ms	Passed
A_29 Successful Login	30	0	0	0	0	30	10	0	10	3s 292ms	Passed
A_42 Users Access More Platforms	3	0	0	0	0	3	1	0	1	081ms	Passed
A_316 Lost Users	2	0	0	0	0	2	1	0	1	071ms	Passed
A_317 Reconnected Users	3	0	0	0	0	3	1	0	1	056ms	Passed
A_318 Loyal Users	3	0	0	0	0	3	1	0	1	064ms	Passed

SQA DAYS#23

accenture
Latvia ATC

Тестирование в мире BIG DATA

CONTACTS

Visit us @ Accenture.lv

Konstantin Pletenev
Test Automation Engineering Specialist
konstantin.pletenev@accenture.com

SQA® DAYS#23