

Streaming Analytics at ING

Bas Geerdink & Martijn Visser, 13 September 2017



Bas Geerdink

IT Chapter Lead at ING Analytics

Academic background: A.I. & Informatics

Working in IT since 2004

At ING since 2013



@bgeerdink

<https://nl.linkedin.com/in/geerdink>

Martijn Visser

Product Owner / Data Analyst at ING

Academic background: ICT & Law

Working with data since 2008

At ING since 2013



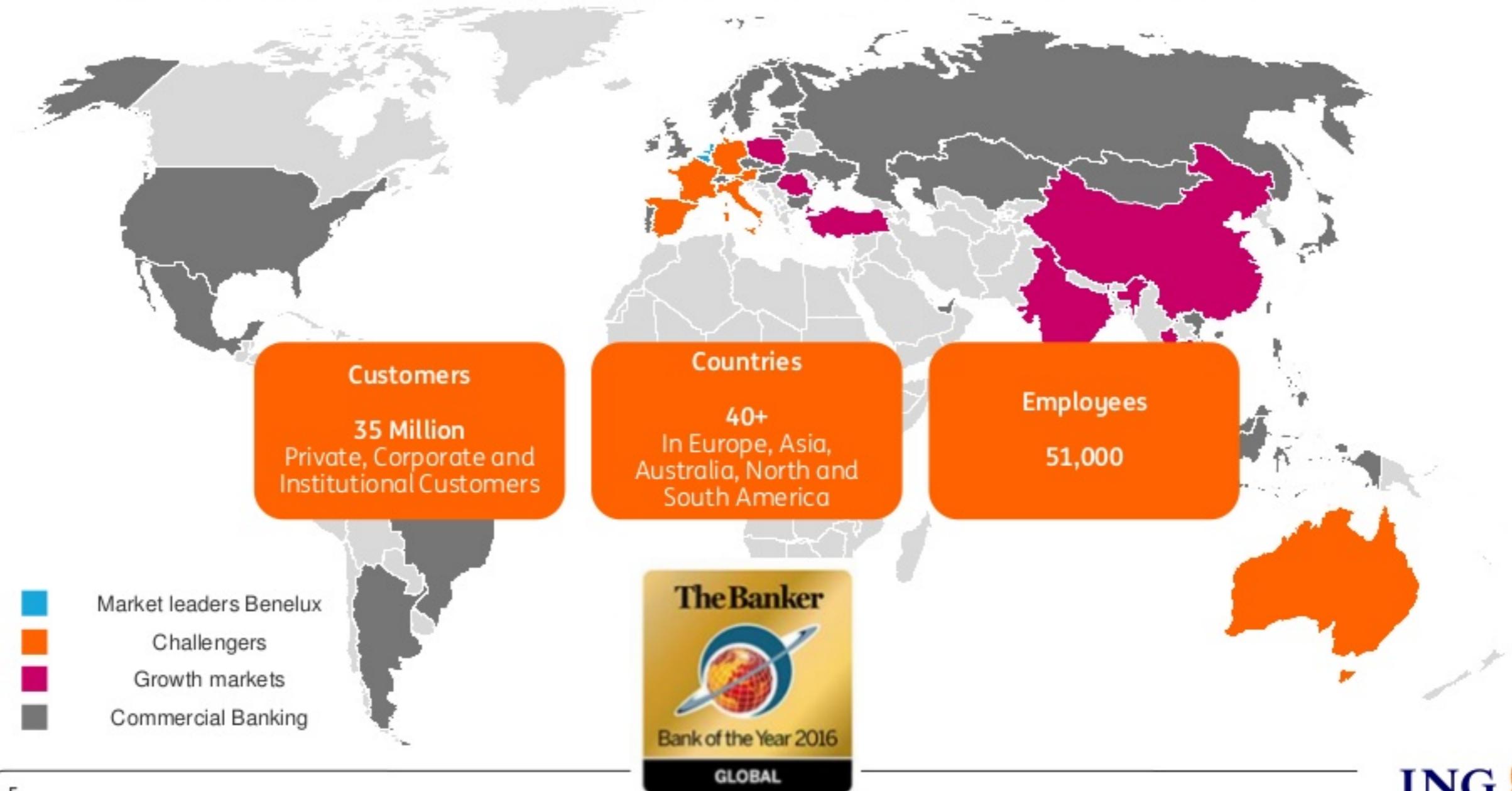
@martijnvisser82

<https://nl.linkedin.com/in/martijnvisser1982>

Strategy



ING is a top financial enterprise, operating since 1881



ING's Think Forward strategy requires a data-driven approach

Purpose



Empowering people to stay a step ahead in life and in business.

Customer Promise



Clear and Easy



Anytime, Anywhere



Empower



Keep Getting Better

Strategic Priorities



Creating a differentiating customer experience

1. Earn the primary relationship
2. Develop analytics skills to understand our customers better
3. Increase the pace of innovation to serve changing customer needs
4. Think beyond traditional banking to develop new services and business models

Enablers



Simplify & Streamline

Operational Excellence

Performance Culture

Lending Capabilities

We're making a shift from batch to real-time



Secure and reliable



Predictive



Omnichannel



Actionable



Personal



Relevant

We're building a Streaming Analytics platform with high-end capabilities

- **High performance & low latency** to manage and process users' events.
- **Scalability & availability** to build new features, scenarios and integration with other external systems.
- **Fault tolerance mechanism** to ensure high quality output with strong information consistency.
- **Online machine learning** and **business rules management** capabilities.



Main use cases for a Streaming Analytics Platform

- Fraud detection:
 - Detect possible fraudulent transactions
 - Identify possible money laundering accounts
- Actionable Insights:
 - Small financial insights and *robotized* advise
 - Customer-defined alerts and push notifications
 - Offer chat/call when customers don't finish their customer journey
 - Marketing and commercial offerings ('Next Best Action')



Future scenarios

- Wisdom of the crowd:
 - making use of similar customers to make better predictions
- Online learning:
 - making use of customer feedback in real-time to train models
- Beyond banking:
 - enabling IoT
 - robo advise



Use case: Insights



Do people still need Banks, or just Banking?



Alibaba, the most valuable retailer, has no inventory



Uber, the world's largest taxi company owns no fleet



AirBnB, the largest accommodation provider owns no real estate

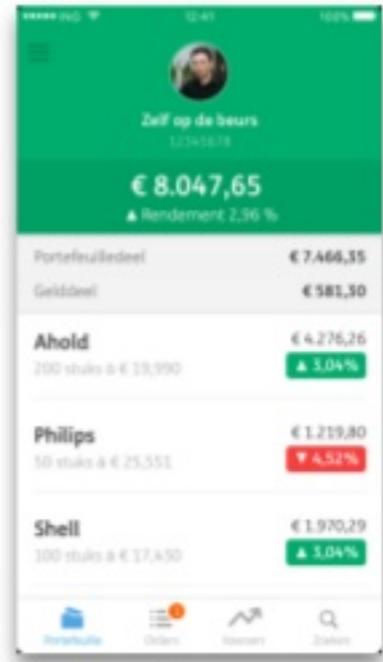
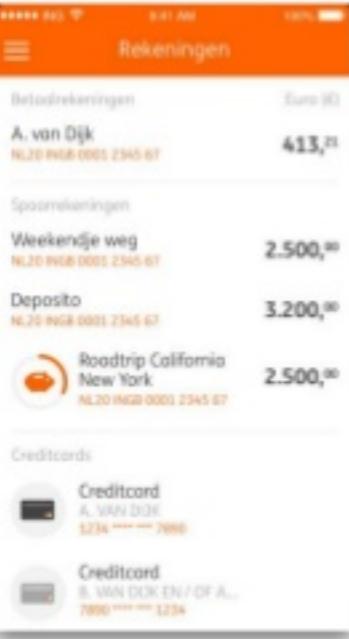
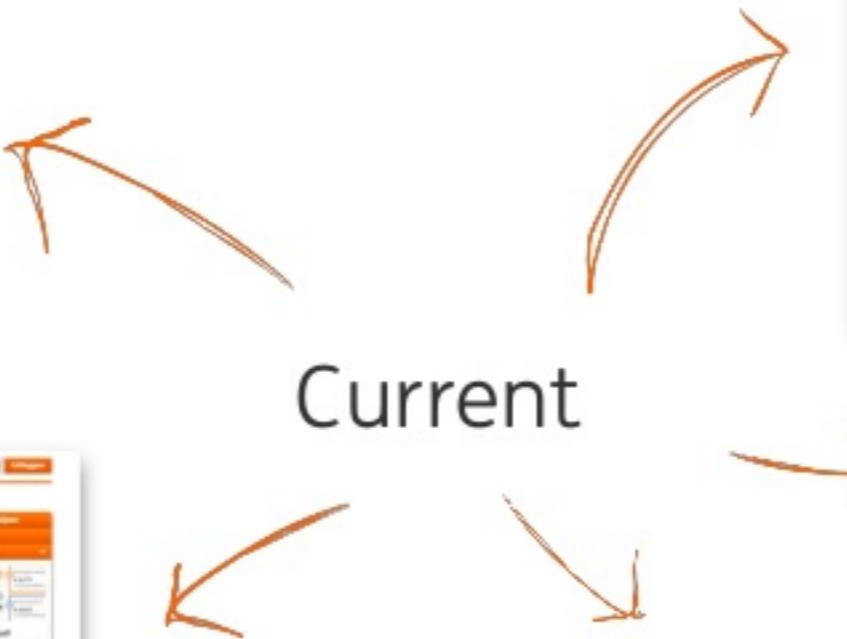


Facebook, the most popular media owner, creates no content

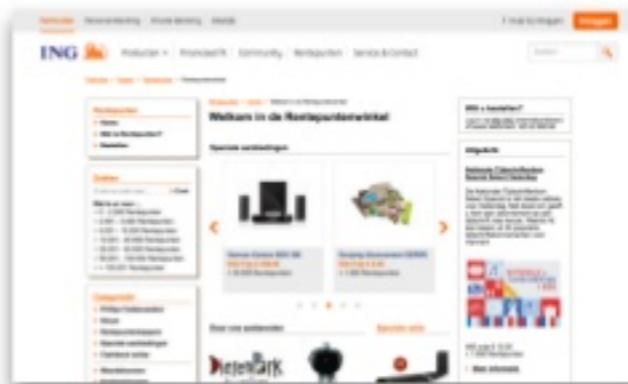
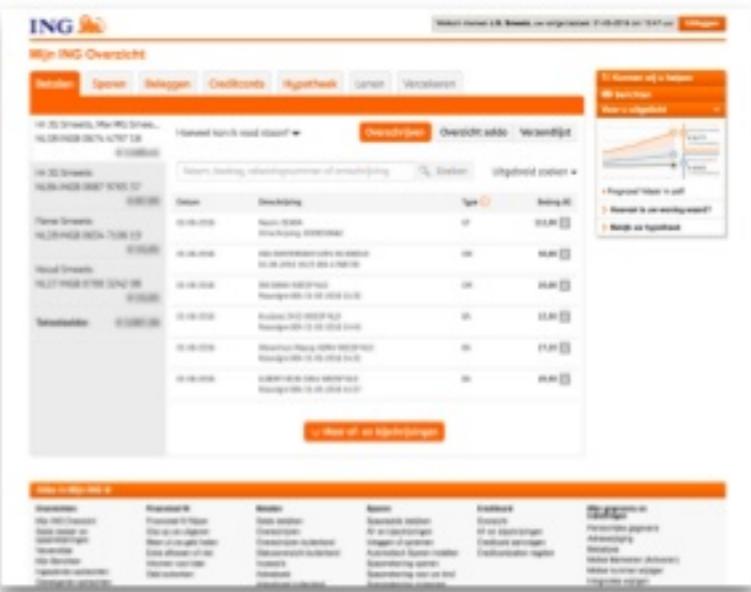
Whoever controls the *customer experience controls the customer value.*

We want our environment to be the front and center of customer financial needs.

Problem: apps & site going everywhere



Current

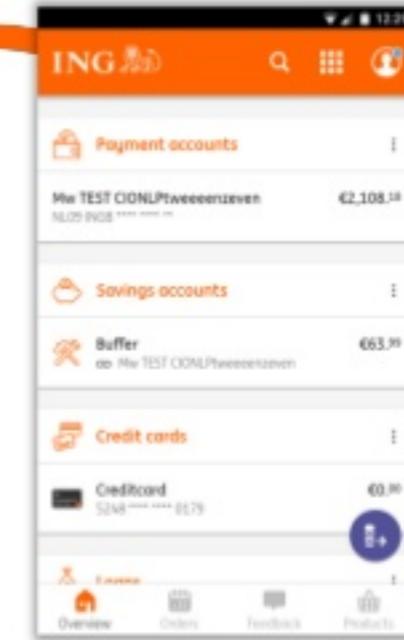


One app to rule them all... ??

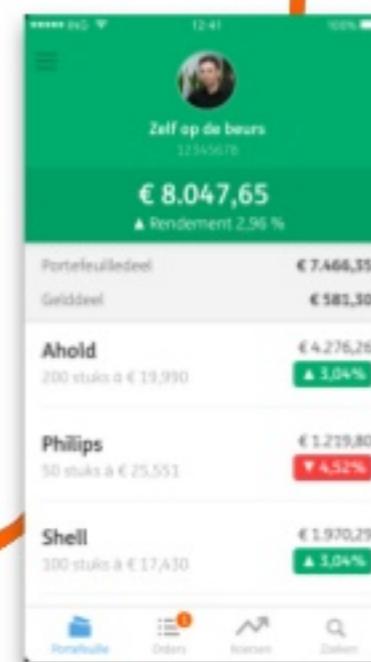
The screenshot shows the Google Play Store interface with the search bar set to "Search". The main category is "Apps". On the left, there's a sidebar with "My apps" and "Shop" sections, and a "Games", "Family", and "Editors' Choice" section below it. The main content area displays a grid of ING mobile banking apps from different countries:

| Region | App Name | Description | Rating |
|--------|--------------------|-------------------------|--------|
| NL | ING Bankieren | ING Nederland | ★★★★★ |
| TR | ING Mobil | ING BANK | ★★★★★ |
| PL | Moje ING mobile | ING Bank Śląski S.A. | ★★★★★ |
| BE | ING Smart Banking | ING Belgium NV | ★★★★★ |
| RO | ING HomeBank | ING BANK N.V. AMSTERDAM | ★★★★★ |
| ES | ING InsideBusiness | ING Nederland | ★★★★★ |
| | ING DIRECT España | ING BANK NV, Sucursal | ★★★★★ |
| .be | ING ActivePay | ING Belgium NV | ★★★★★ |
| .de | ING Smart Banking | ING Belgium NV | ★★★★★ |
| DE | ING-DiBa Banking | ING-DiBa | ★★★★★ |
| | ING Corporate Card | ING Nederland | ★★★★★ |
| CZ | ING Direct France | ING Bank France | ★★★★★ |
| | ING Bank CZ | ING Bank CZ | ★★★★★ |
| | ING Australia Bank | ING Australia | ★★★★★ |
| | ING-DiBa Banking | ING-DiBa | ★★★★★ |
| LU | ING | ING | ★★★★★ |
| | ING DIBa | ING DIBa | ★★★★★ |
| | ING | ING | ★★★★★ |
| | ING | ING | ★★★★★ |

Design for One Experience



Omnichannel



Formula for creating compelling digital experience

AI + UI + I

- **Actionable Insights** rather than data – but select problems that are easy for machines but difficult for people e.g. forecasting
- **Real time alerts** can be the real differentiators e.g. when a card auth is received, when a customer makes a payment that he has made a few times previously
- **Hierarchy of alerts** – absolutely tell me when things are going wrong but occasionally celebrate with me too
- **UI adapts to confidence** in AI. E.g. The message differs when we have high confidence in forecast
- **UI also needs to adapt to customer behaviours** e.g. easy login when a customer logs in to the same device
- **The key role of AI is to help me stay a step ahead in life** – so UI needs to make it easy to find items to act upon
- **Digital experience is personalised** to what's important to me
- **Make the benefit of using your system clear** to me immediately rather than making me discover the benefit
- **I expect consistent experience Everywhere** e.g. once I dismiss an action online, I don't want to be prompted for it on mobile again
- **I am used to certain way of working** e.g. I am used to settings being in one place, and I am used to certain interaction patterns. So reduce my learning curve by reusing those

New Experience

AI + UI + I

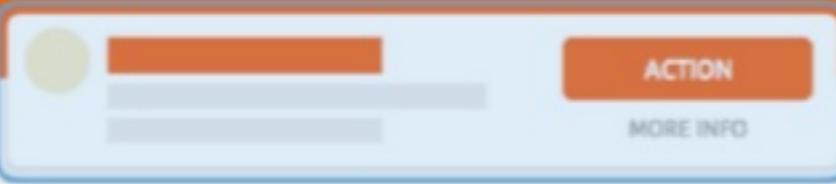
Everywhere

Personalized

Actionable

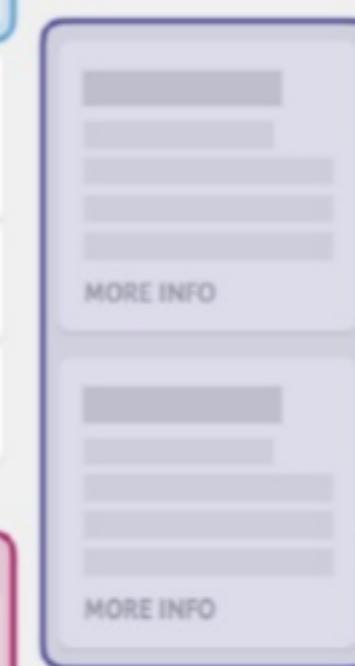
Based on
Material

-
-
-



Rekeningen

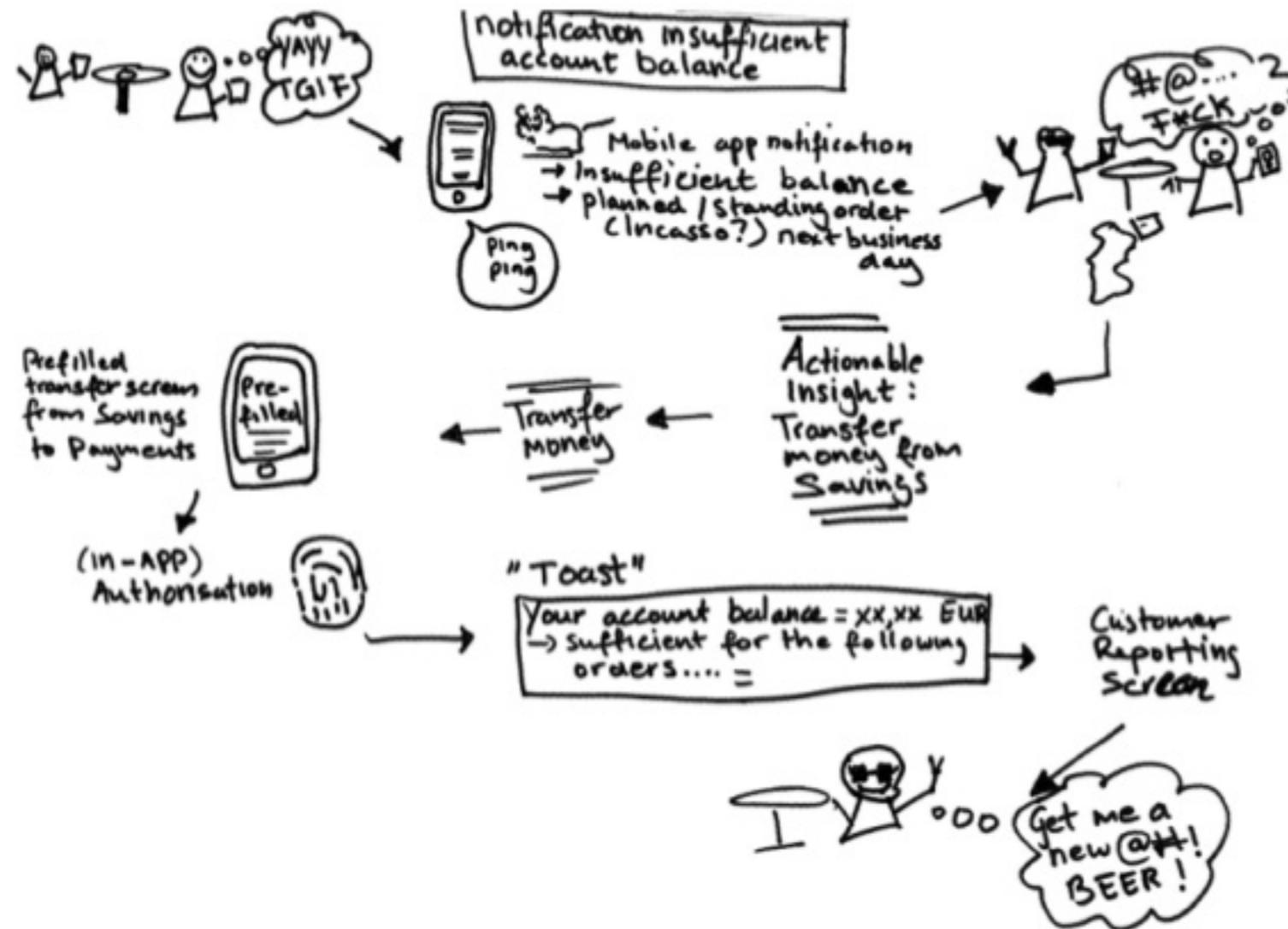
-
-
-



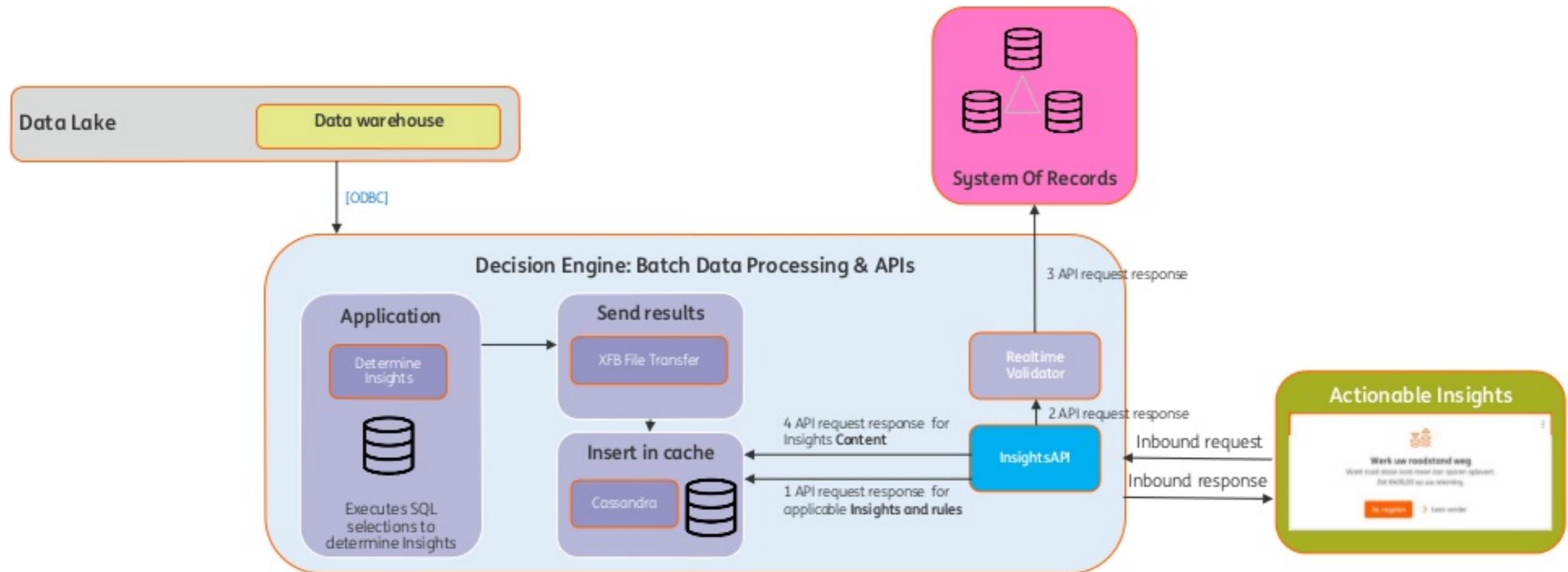
+ ADD PRODUCT



Actionable Insights



MVP: Batch implementation





Waar ben je naar op zoek?

Zoek



Overzicht

Opdrachten

Feedback

Producten



Werk je roodstand weg

Want rood staan kost meer dan sparen oplevert. Zet
€5,47 op je rekening.

Ja, ik wil nu overmaken



Overschrijven



Betaalrekeningen

Hr LWRM WaalTD
NL16 INGB 0748 3119 55

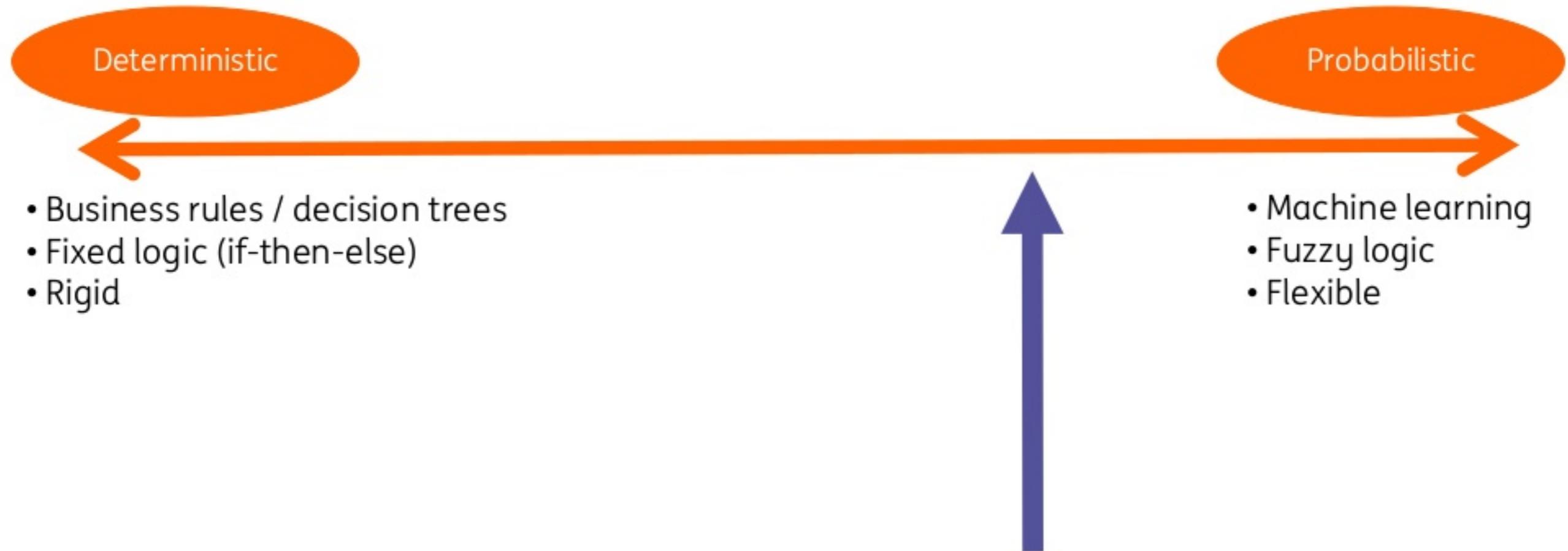
€ -5,47

If only we had realtime capabilities...

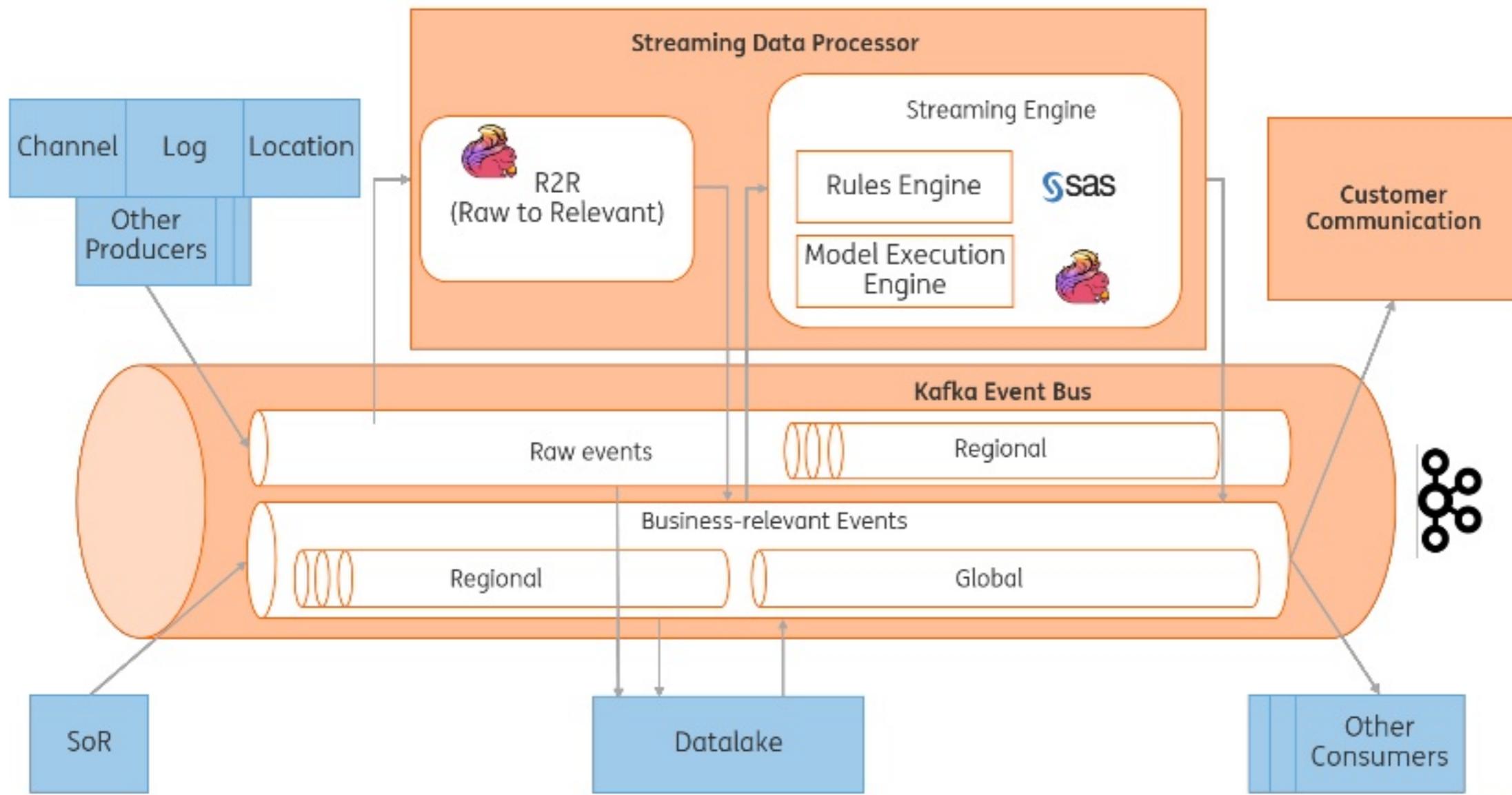


Streaming Data Platform

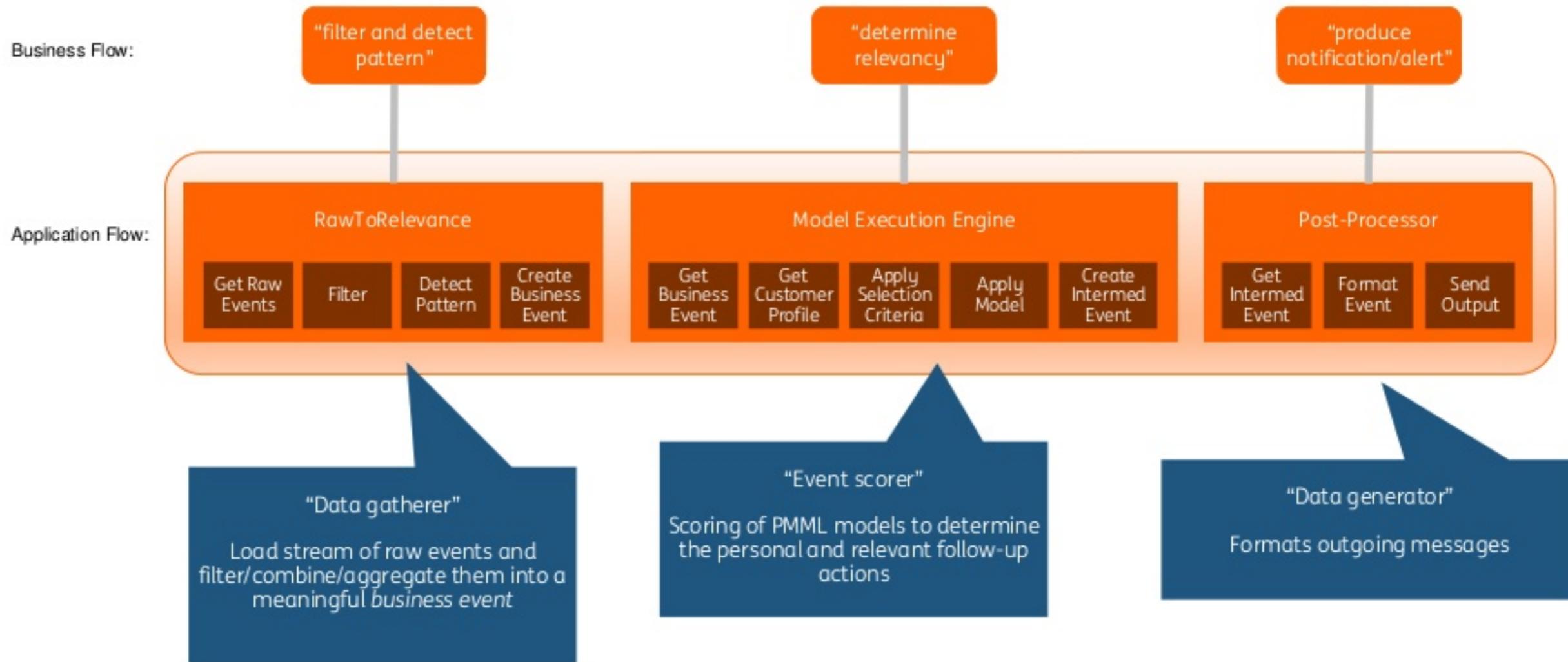
The Streaming Data Platform has to be probabilistic in nature



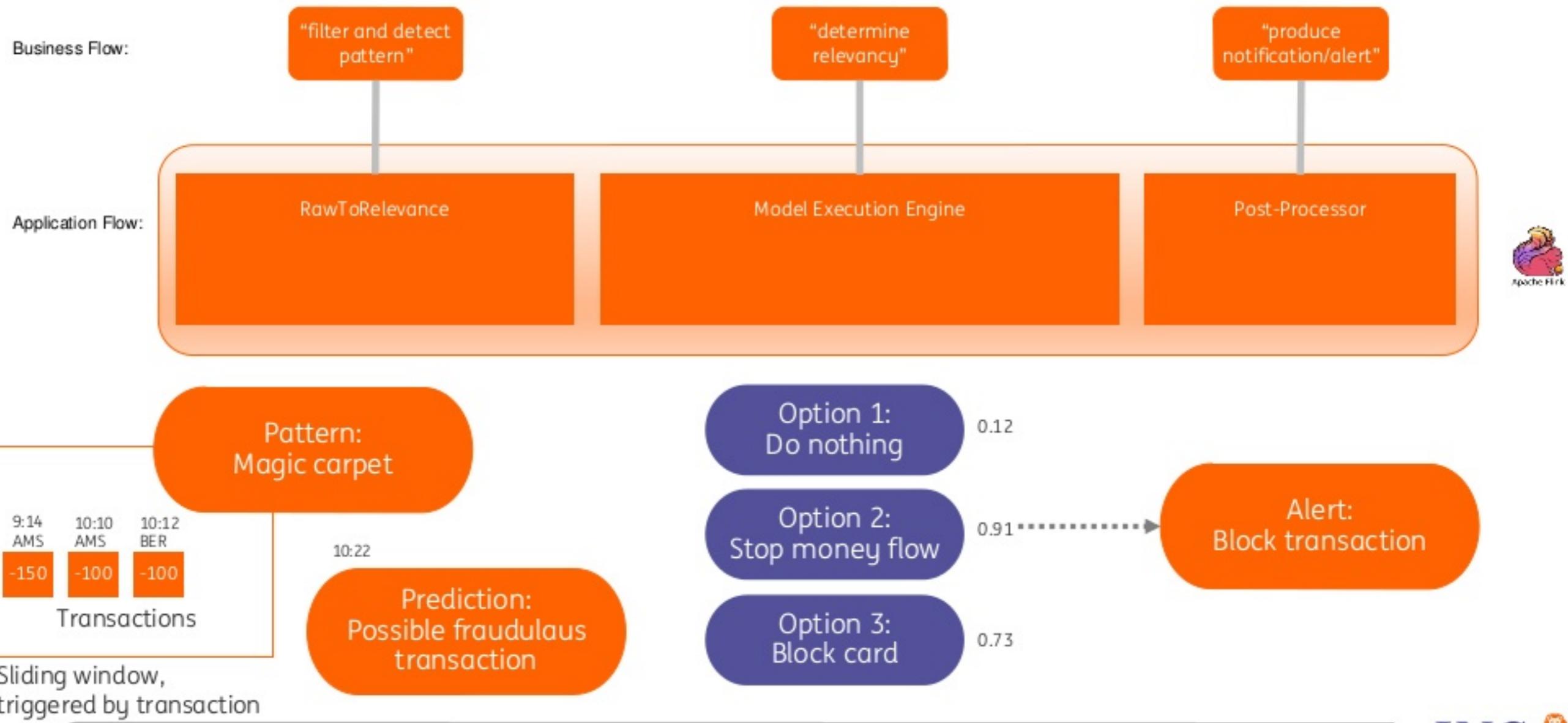
High-level Streaming data platform architecture



Streaming Data Platform Architecture



Streaming Data Platform Architecture – Example 1



Streaming Data Platform Architecture – Example 1

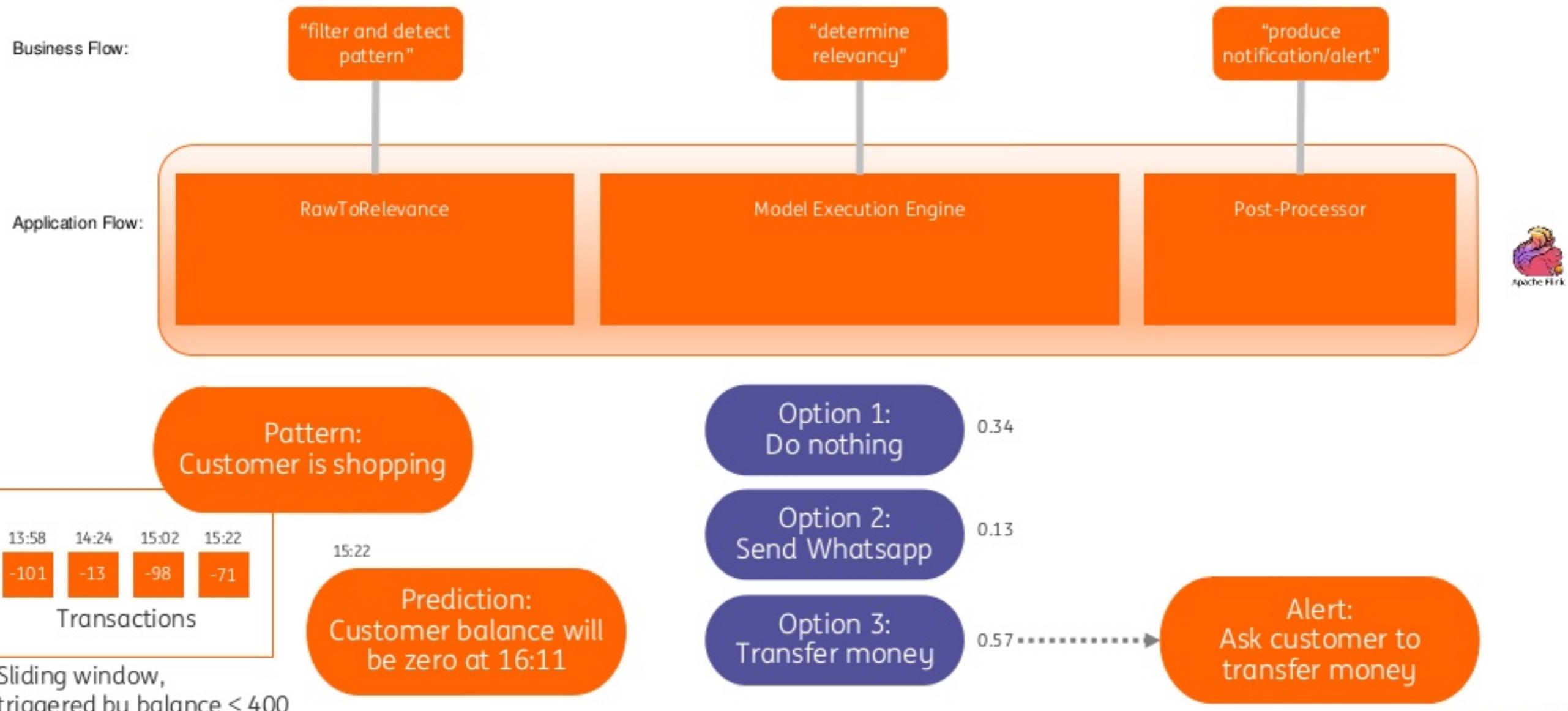
```
model SimpleModel version "1.0" triggered by event Withdrawal version "1.0.0"
  using features A.D {
    predict pmml("ccc.pmml", A.D); }

model modelMagicCarpet version "1.0.0" {
  triggered by event Withdrawal version "1.0.0" {
    double fBalance := A.A;
    map prediction := pmmlPredict("place_time.pmml", fBalance);
    string @customer.previousScore := get(prediction, "result");
  }
  publish fraudscoringtopic version "1.0.0" {
    string scoreResult := get(prediction, "result");
    string scoreConfidence := get(prediction, "confidence");
    string scoreExplanation := "scorePublishCount: " + (string)coalesce(feature.scorePublishCount,
      0);
    string scoreAffinity := "previousScore: " + feature.previousScore;

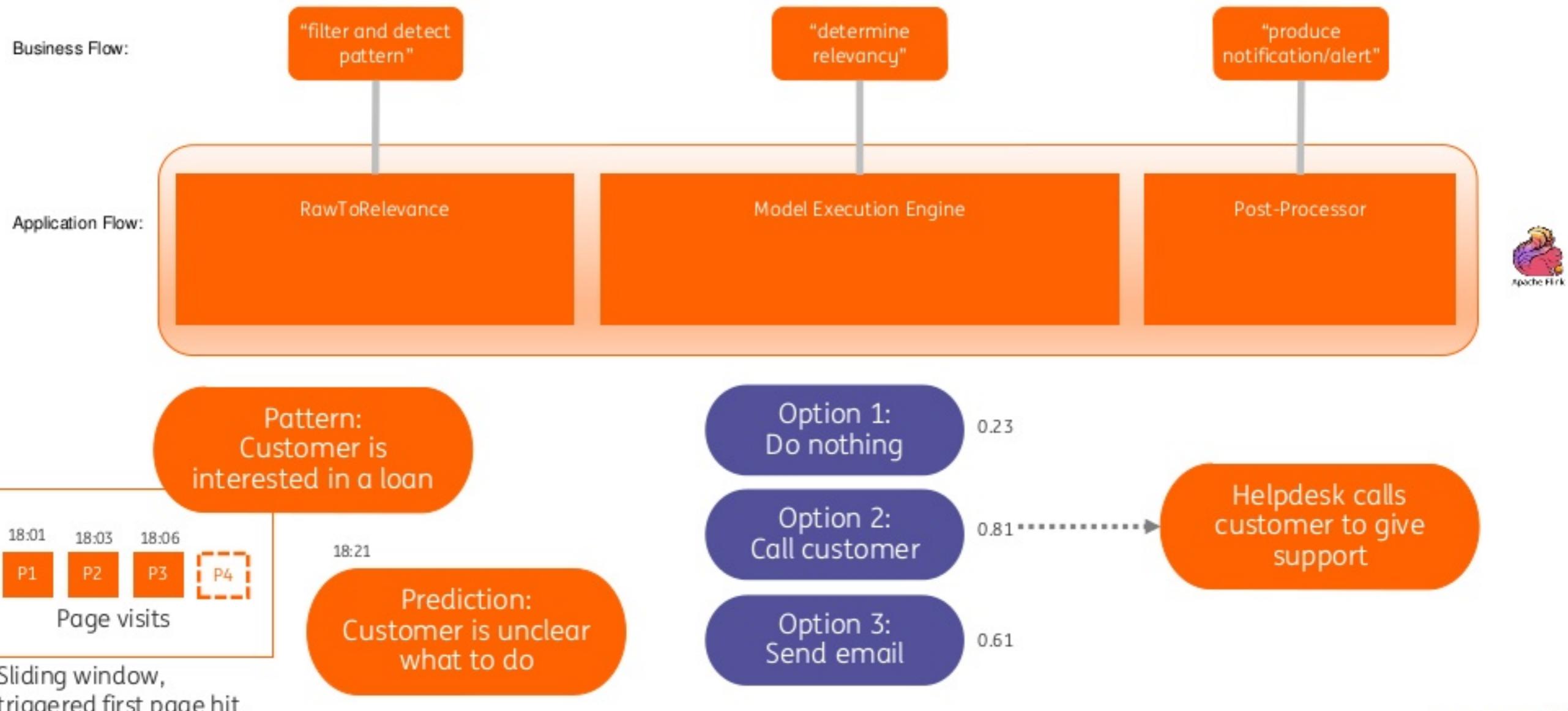
    long @customer.scorePublishCount := coalesce(feature.scorePublishCount, 0) + 1;
  }
}
```

Not actual ING code

Streaming Data Platform Architecture – Example 2



Streaming Data Platform Architecture – Example 3

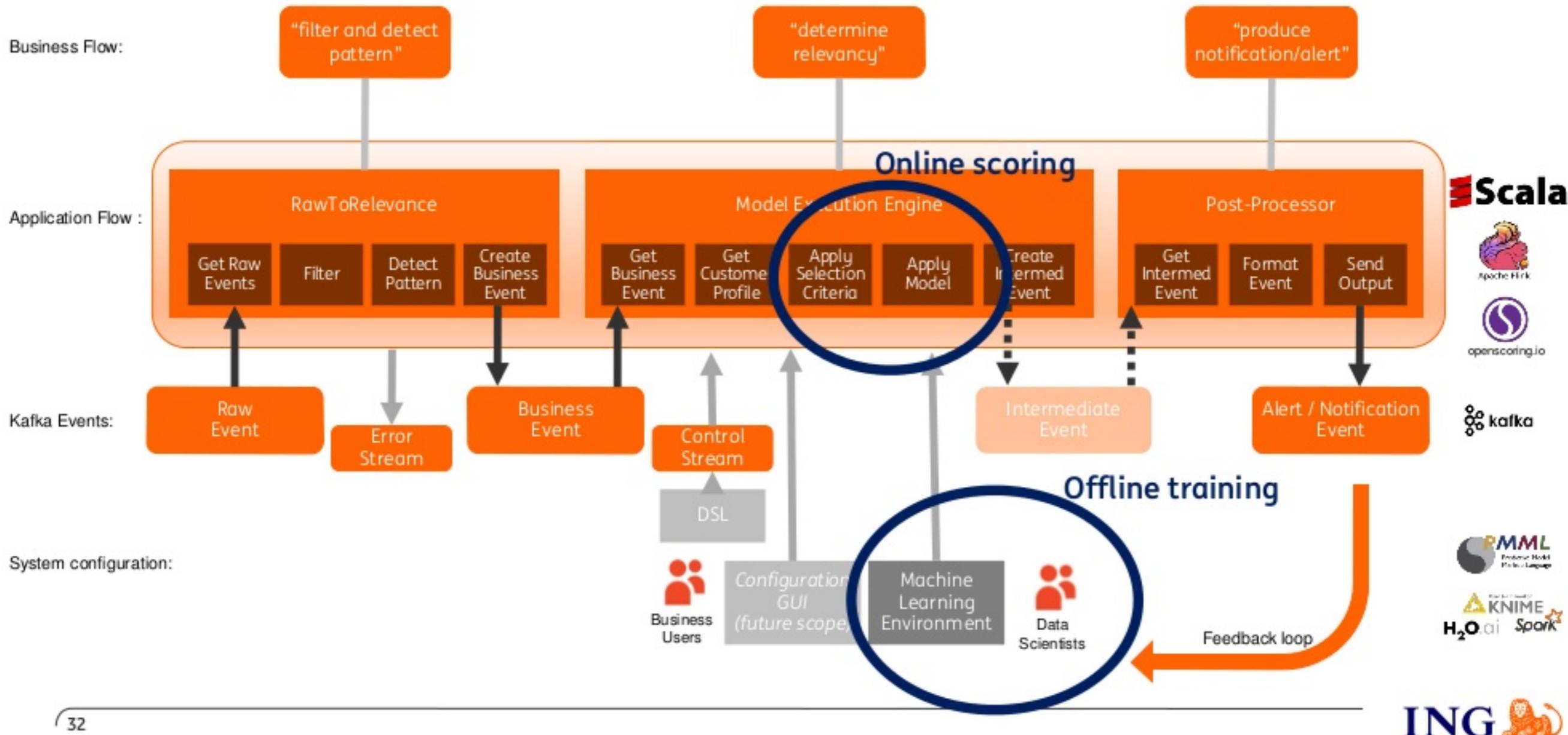


Streaming Data Platform Architecture – Example 3

```
val clickDataStream: DataStream[ClickEvent] = env.fromCollection(dummyData).assignTimestampsAndWatermarks(  
    new AscendingTimestampExtractor[ClickEvent] {  
        def extractAscendingTimestamp(element: ClickEvent): Long = element.eventTime.getMillis  
    })  
val input: KeyedStream[ClickEvent, String] = clickDataStream.keyBy((event: ClickEvent) -> event.accNum)  
  
val steps = List[String]("page1", "page2", "page3", "page4")  
val start: Pattern[ClickEvent, _] = Pattern.begin[ClickEvent]("start").within(Time.seconds(1))  
.where(event => event match {  
    case c: ClickEvent => c.processStep.equals(steps.head)  
    case _ => false  
})  
.next("Step2").where(event => event match {  
    case c: ClickEvent => c.processStep.equals(steps(1))  
    case _ => false  
})  
.next("Step3").where(event => event match {  
    case c: ClickEvent => c.processStep.equals(steps(2))  
    case _ => false  
})  
// step4  
.next("Step4").where(event => event match {  
    case c: ClickEvent => c.processStep.equals(steps(3))  
    case _ => false  
}).within(Time.seconds(30))  
  
val pStream: PatternStream[ClickEvent] = CEP.pattern(input, start)  
val result = pStream  
.flatSelect(  
    (pattern: mutable.Map[String, ClickEvent], timestamp: Long, out: Collector[String]) -> {  
        out.collect("TimeoutEvent")  
    })  
    (pattern: mutable.Map[String, ClickEvent], out: Collector[ClickEvent]) -> {  
        println("RESULTS:::")  
        pattern.get("Step4").foreach(event => out.collect(event))  
    }  
)  
  
env.execute("CEP cash loan timeout pattern")
```

Not actual ING code

Streaming Data Platform Architecture



Technology



ING chose Apache Flink for its Streaming Data Platform

Benefits for ING:

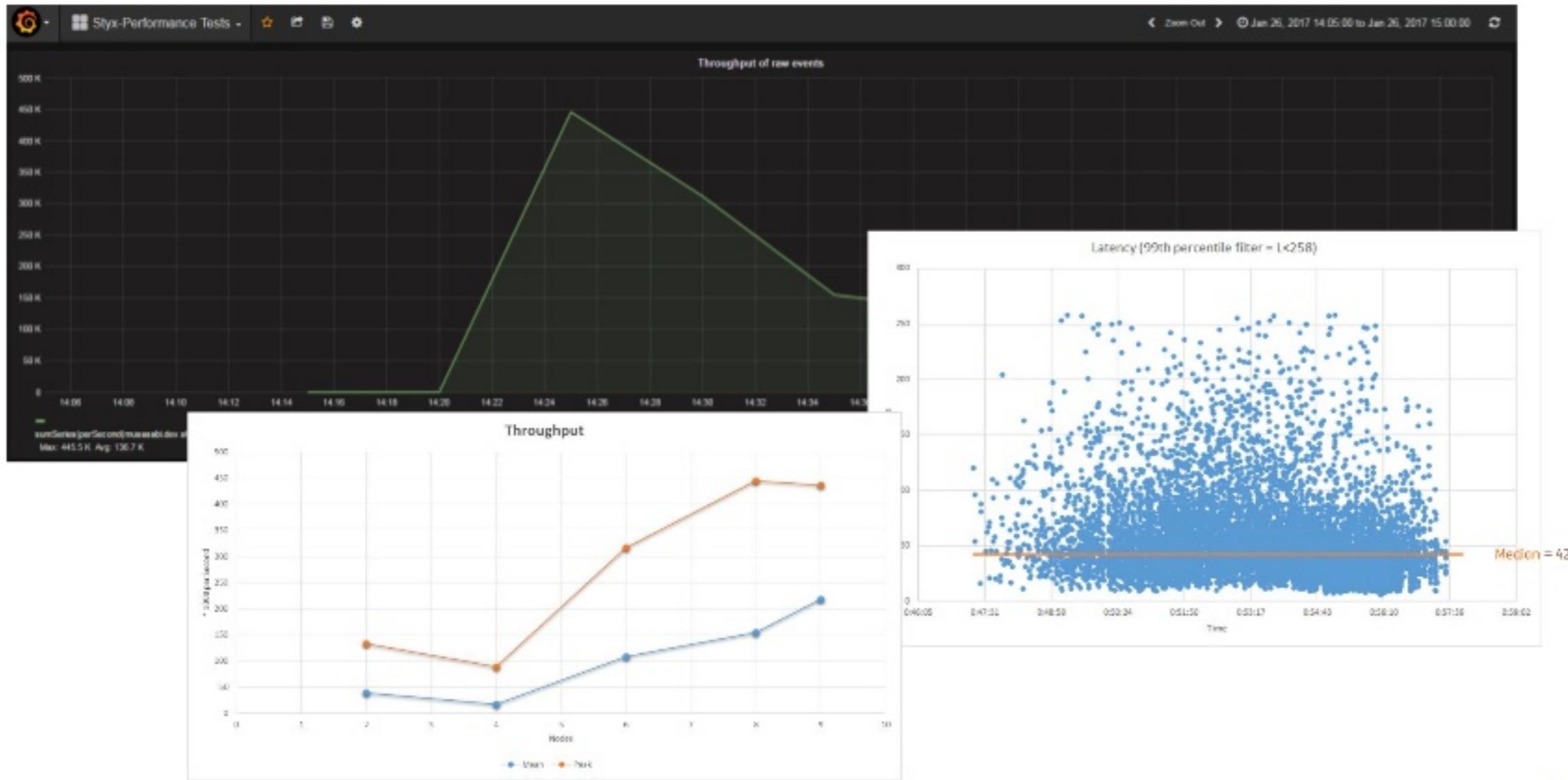
- true streaming: per-event processing, no micro-batching
- open source, active community
- high volumes, low latency
- connectors to Kafka
- enterprise ready
- Scala api



Features that we use:

- state management and fault tolerance: savepointing, checkpointing → exactly-once semantics
- time windowing: event time, flexible (e.g. sum all transactions in the past minute)
- Complex Event Processing

Performance and scalability tests



PMML (Predictive Model Markup Language) bridges the gap between data scientists and data engineers

PMML is the de facto standard for model representation

- XML-based interchange format for predictive models
- Support of hot-replacing an algorithm (configuration) at runtime by just providing an updated PMML model and interpreting (scoring) it with openscoring.io
- PMML models can be created by a large number of tools, including: Spark, SAS, R, Knime, Python, H2O.ai
- Large number of supported machine learning models
- Working together with the open source community on FlinkML
- Next steps: PFA, TensorFlow



| Type | Model support |
|------------------------|--|
| Simple business rules | <ul style="list-style-type: none">• Rulesets, score cards |
| Complex business rules | <ul style="list-style-type: none">• Sequences, association rules |
| Predictive models | <ul style="list-style-type: none">• Naive Bayes• k-Nearest Neighbors• Regression: linear, logistic, regularized• Trees: decision trees, random forest• Support Vector Machines |
| Advanced models | <ul style="list-style-type: none">• Bayesian Network• Neural Network (Deep Learning)• Gaussian Process |
| Other | <ul style="list-style-type: none">• Time Series• Cluster Models |



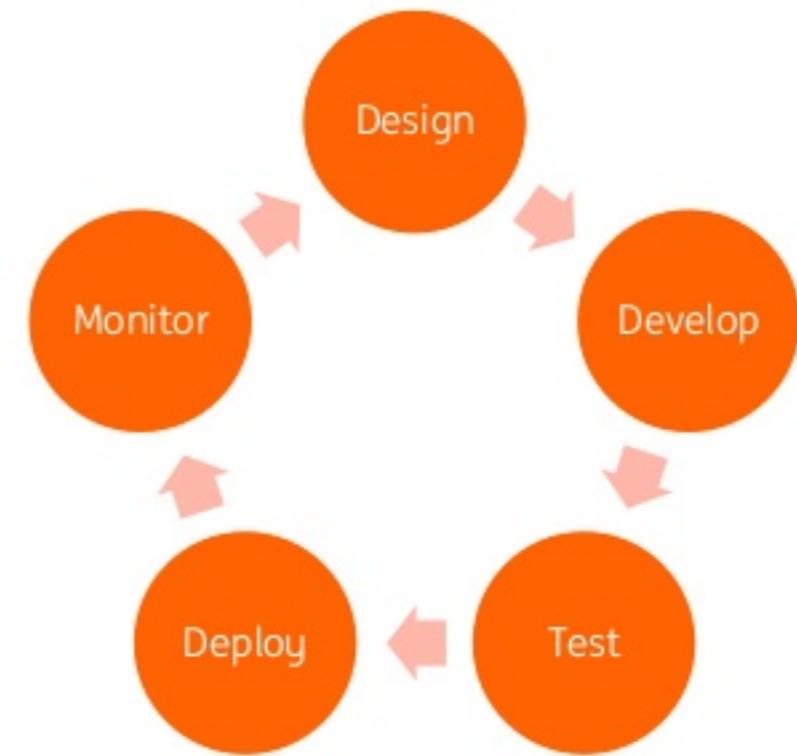
PMML Example

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<PMML xmlns="http://www.dmg.org/PMML-4_3" version="4.3">
  <Header>
    <Application name="JPML-SkLearn" version="1.1.4"/>
    <Timestamp>2017-01-11T13:52:59Z</Timestamp>
  </Header>
  <MiningBuildTask>
    <Extension name="estimator" value="repr(estimator)">LogisticRegression(C=1.0, class_weight=None, dual=False, fit_intercept=True,
      intercept_scaling=1, max_iter=100, multi_class='ovr', n_jobs=1,
      penalty='l2', random_state=None, solver='liblinear', tol=0.0001,
      verbose=0, warm_start=False)</Extension>
  </MiningBuildTask>
  <DataDictionary>
    <DataField name="CUST_AGE" optype="continuous" dataType="double"/>
    <DataField name="AVG_MONTH_ACCOUNT_BALANCE" optype="continuous" dataType="double"/>
    <DataField name="AVG_MONTH_DEPOSIT_BALANCE" optype="continuous" dataType="double"/>
    <DataField name="CUSTSEG" optype="categorical" dataType="string">
      <Value value="HI"/>
      <Value value="LO"/>
      <Value value="ME"/>
    </DataField>
    <DataField name="Y" optype="categorical" dataType="double">
      <Value value="0"/>
      <Value value="1"/>
    </DataField>
  </DataDictionary>
  <TransformationDictionary>
    <DefineFunction name="logit" optype="continuous" dataType="double">
      <ParameterField name="value" optype="continuous" dataType="double"/>
      <Apply function="/">
        <Constant dataType="double">1</Constant>
        <Apply function="+">
          <Constant dataType="double">1</Constant>
          <Apply function="exp">
            <Apply function="*">
              <Constant dataType="double">.45</Constant>
            </Apply>
          </Apply>
        </Apply>
      </Apply>
    </DefineFunction>
  </TransformationDictionary>
</PMML>
```

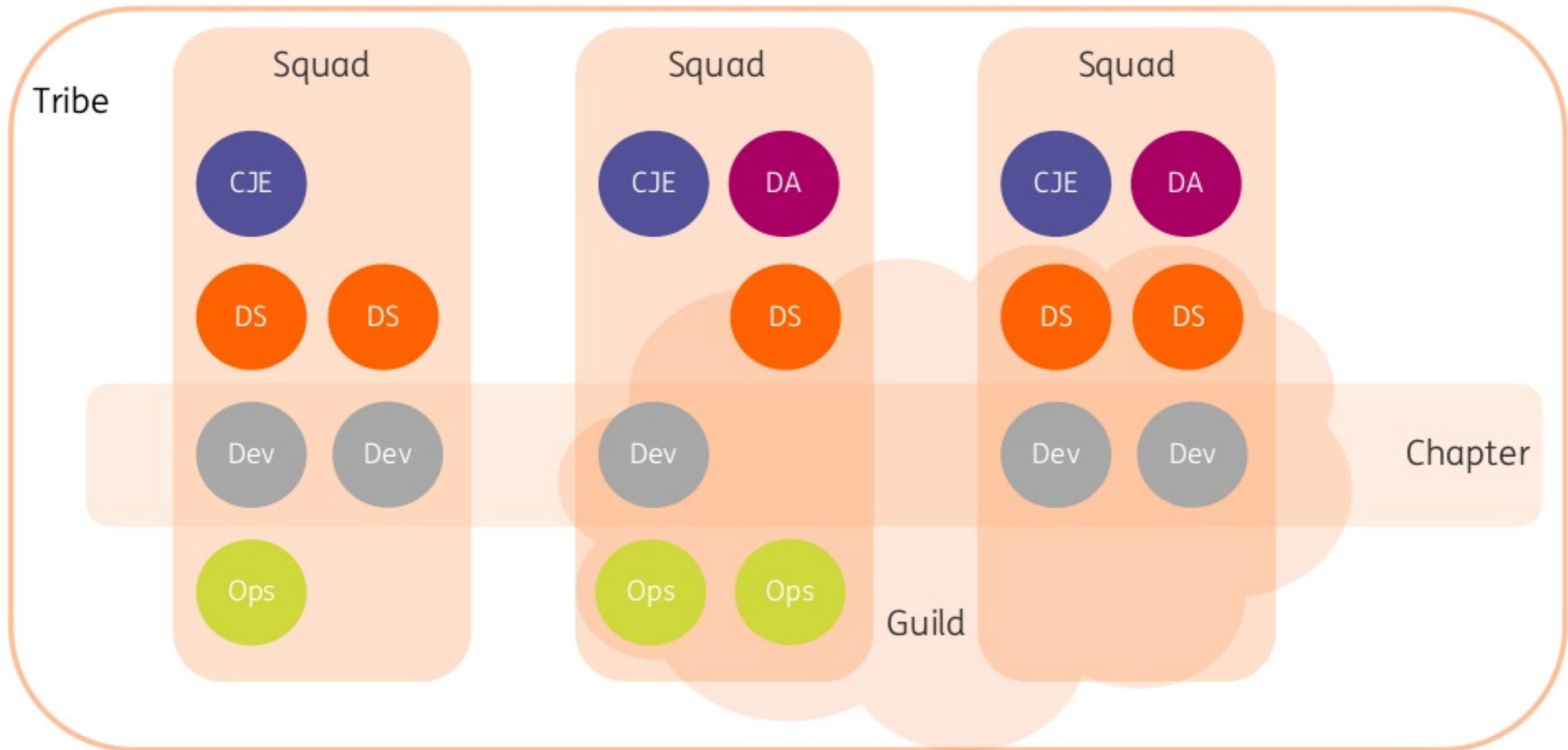
Not actual ING code

Flink fits the development & operations model with the ING way of working

- Cross functional squads
 - Business, dev, ops, data scientist
 - Reusing ING standard building blocks and processes
 - Infrastructure, monitoring
- Continuous integration & delivery in DTAP environment
 - Develop: Gitlab → Jenkins → Artifactory
 - Using Flink's savepointing mechanism
 - No data loss, guaranteed
 - Limited downtime
- Scalability, resilience and robustness
 - Flink gives us automatic restart on failure, using checkpointing mechanism
 - Rolling updates of infrastructure and middleware leads to no downtime



In ING's one way of working, 'business' and 'IT' go hand-in-hand



Conclusions

Why is Streaming Analytics with Flink the right choice for ING, moving forward?

- ING is an “**IT company** with a banking licence”, whose competitive advantage is in its use of data and analytics. To excel, we’ll handle bigger volumes of streaming data with higher performance demands.
- ING build a streaming analytics **platform** as a generic, multi-purpose solution with *fraud detection* and *actionable insights* as its first use cases.
- Enterprise-readiness and **security** of Flink is continuously improving; although major improvement points are on the topics of data retention and lineage (GDPR requirements).
- With **PMML** we have a standard format that is very well suited for **multidisciplinary** teams, where data scientists work on the same user stories as engineers.
- We set up a streaming analytics **community** within ING, and work together with industry leaders (Lightbend, DataArtisans).

Questions



We're hiring!

Follow us to stay a step ahead

ING.com



@ING_News



LinkedIn.com/company/ING



SlideShare.net/ING



Facebook.com/ING



YouTube.com/ING



Flickr.com/ING Group

