When Milliseconds Matter: Architecting Real-Time Analytics Into Operational Systems

Paul DiMarzio
IBM z Systems Big Data and Analytics
Worldwide Portfolio Marketing Manager

#SHAREorg

Session objectives

- Introduce a strategy for “right-time” enterprise decision management
- Discuss the role that predictive analytics plays when right-time means real-time
- Propose an architectural tripod for supporting real-time decisions
- Highlight the importance of infrastructure design in enabling real-time decisions
- Look at performance results for real-time, in-transaction predictive analytics

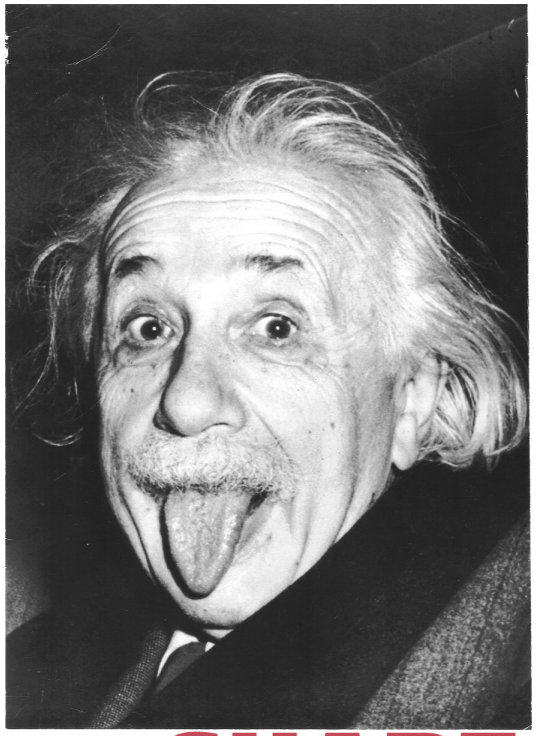
- Review business rationale and architecture for some common real-time decision use cases
  - Countering payment fraud, waste, abuse and financial crimes (banking example)
  - Additional in-process payment analytics (insurance claims example)
  - Predictive customer intelligence / executing the next best action (banking up-sell example)

- Take a quick look into the future, and review

Why we're having this discussion today

"You can't do analytics on the mainframe."

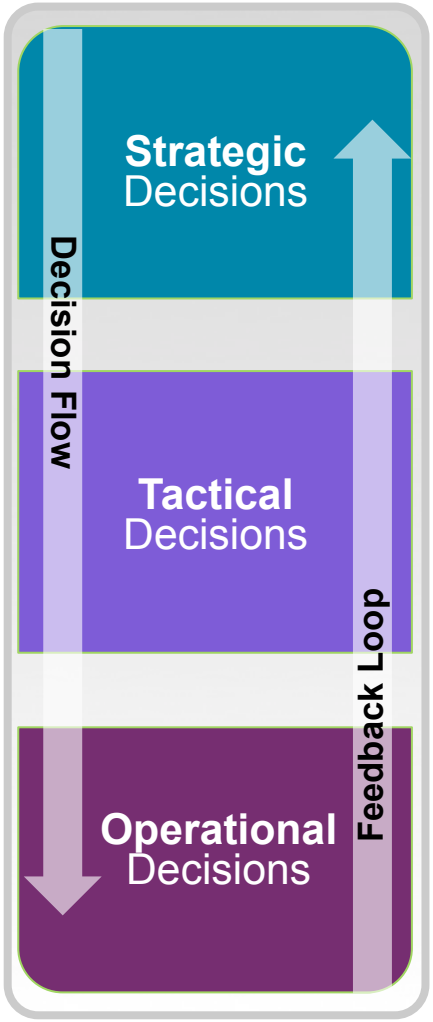
- most any CIO over the past few decades



"Our thinking has created problems which cannot be solved by that same level of thinking."

- a pretty wise dude

An enterprise analytics strategy must be able to support a wide variety of decision types



- Guide overall direction of the enterprise
- ***Should we expand overseas?***

- Manage and control operations
- ***What product should we promote today?***

- Handle every customer interaction
- ***Is this payment request valid?***

How often is this decision made?

What is the value/risk?

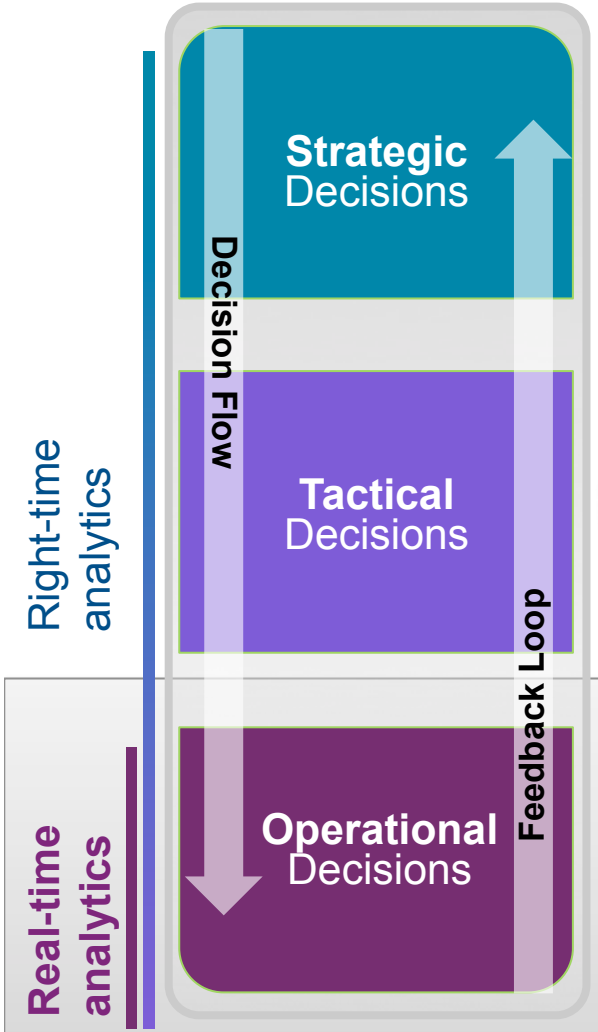
How repeatable is this decision?

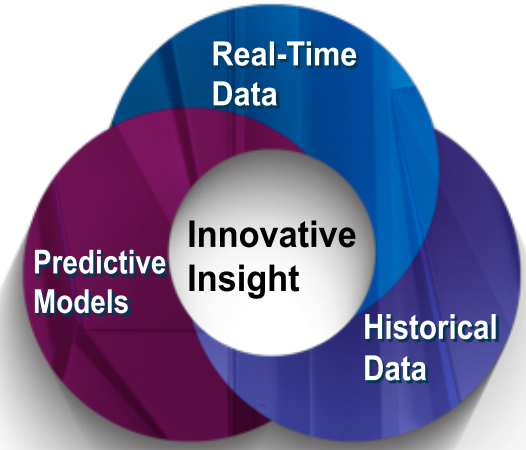
What type of analysis does the decision require?

How quickly must the decision be reached?

Source: James Taylor, Decision Management Solutions: *Transforming Operational Systems From Transactional Systems to Decision Management Systems, 2015*

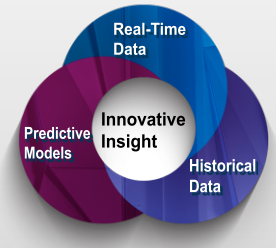
"Real-time" is a very subjective term: understand your use case carefully!





Analyst running reports

"Real-time" objective: reduce time from days/hours to hours/minutes



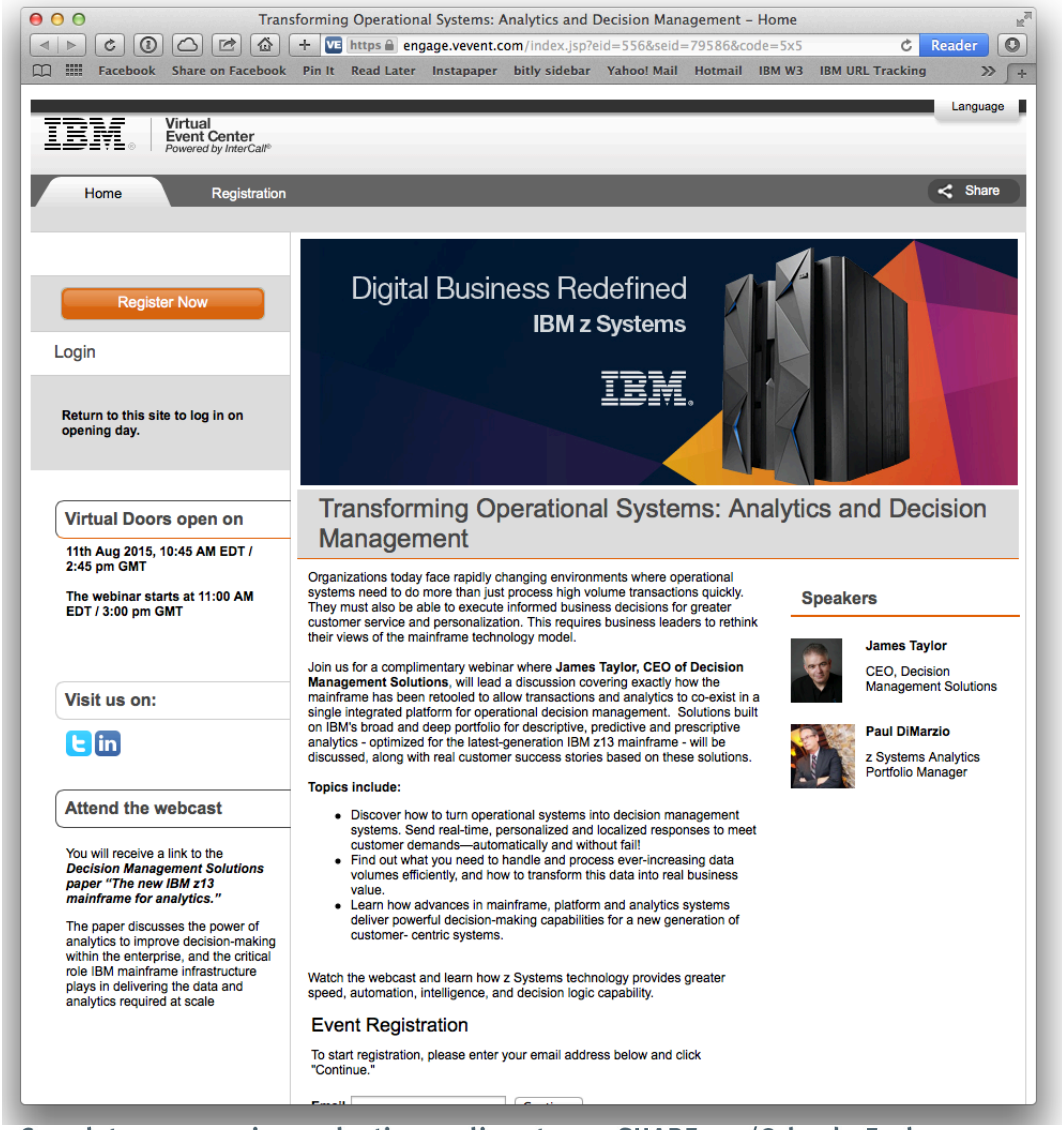
Transactions in flight

Real-time objective: reduce time from seconds to milliseconds – or less!

FOCUS FOR TODAY

Source: James Taylor, Decision Management Solutions: *Transforming Operational Systems From Transactional Systems to Decision Management Systems, 2015*

Shameless webinar plug



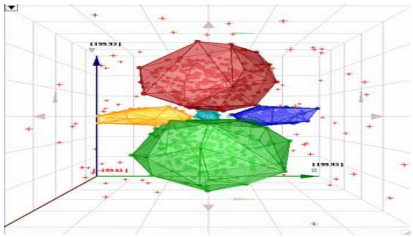
Operational decisions: high volume, low latency, highly repeatable and high value in aggregate

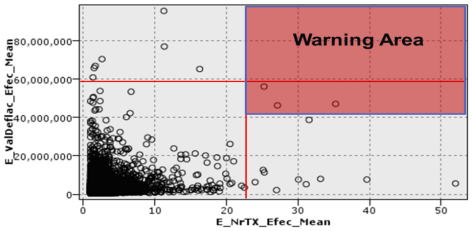
- Require analytics that are closely linked to transactional systems so that decisions can be automated, in **real-time**
- Key considerations: richness of analysis and ability to maintain SLAs

http://bit.ly/1S8rjQQ

Predictive analytics should be a key element of your real-time decisioning strategy…

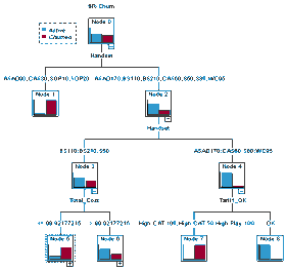
- Predictive models detect patterns
  - Deviation from expected behavior can isolate bad (or good) behavior, trigger additional actions or new targeted marketing and up-sell / cross-sell offers



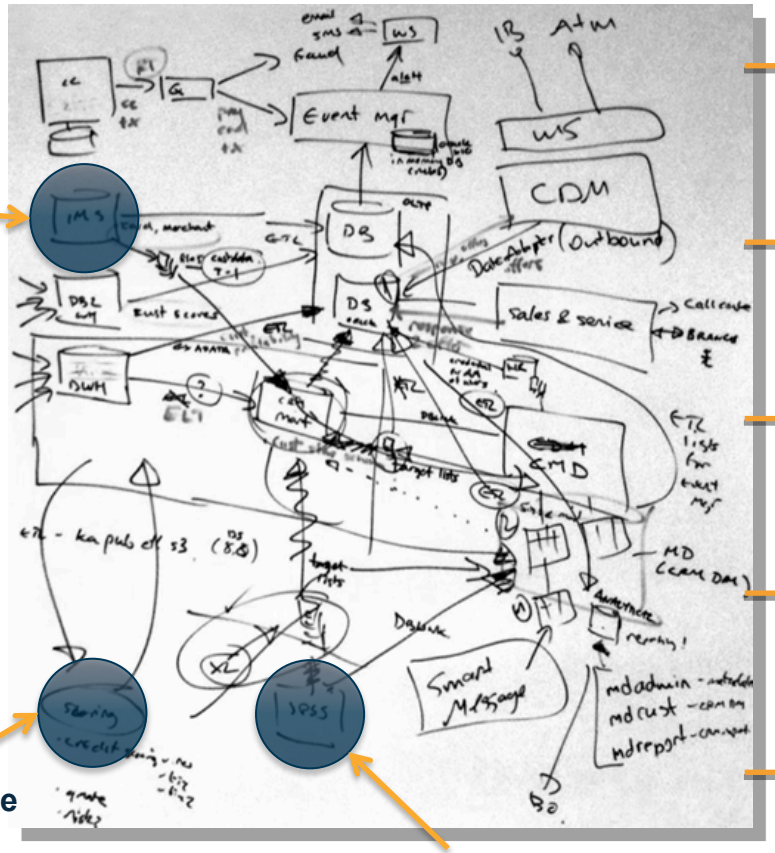


- Like the real world, predictive models are not binary
  - Understanding how closely a pattern of behavior matches a known pattern of bad (or good) behavior can help uncover crimes or non-obvious opportunities

- Predictive models can have many variations
  - Can be built to assess only specific transactions or more generically for all transactions
  - Multiple layers of models can be invoked for increasing sophistication of analysis, triage leading to further inspection of contributing factors and weights



…but, outdated views of infrastructure can impede progress



Significant complexity
Separated data warehouses

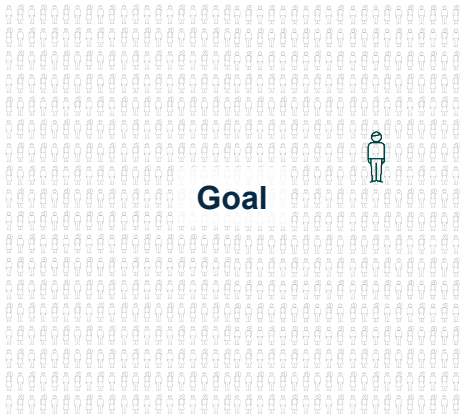
Analytics latency
Transactional data is not readily available

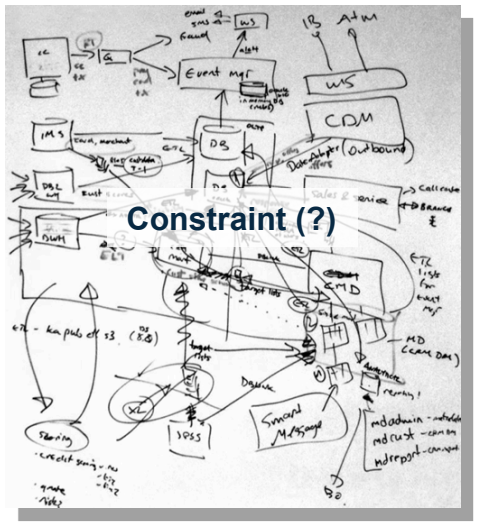
Lack of synchronization
Data is not easily aggregated and fresh

Data duplication
Multiple copies of the same data

Excessive costs
Of moving data around

Unfortunately, you're probably not using predictive analytics in your operations today



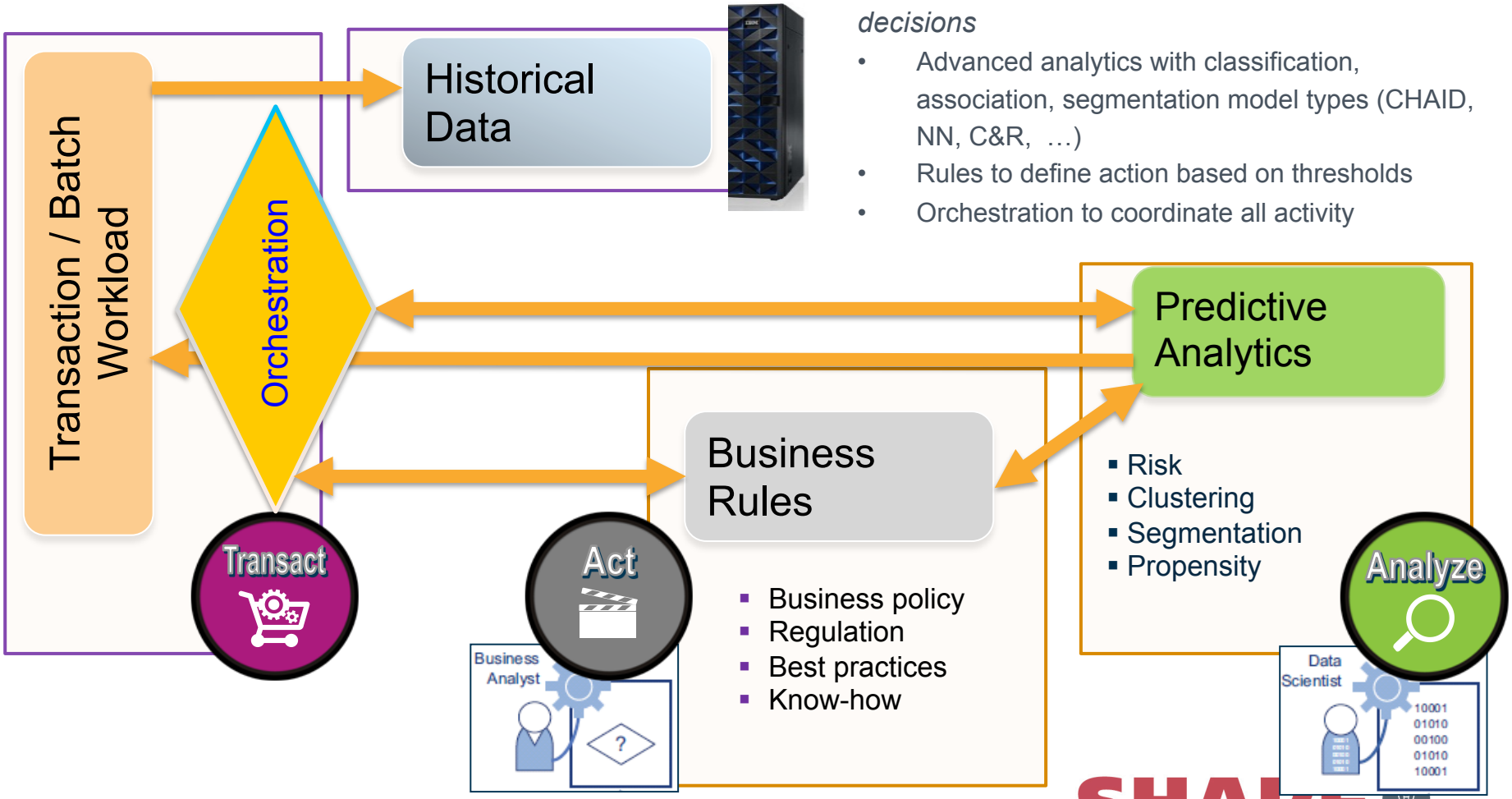


- **What you want**: to service the "demographic of one"
  - Understand opportunities and risks, at the granularity of each individual customer interaction, while interaction is happening
  - More sophisticated analytics, incorporating more data for better decision-making
  - Faster identification and response to new behavior correlations
- **What you do**: sub-optimize based on *perceived* infrastructure constraints
  - To meet SLAs, avoid penalties, and maintain customer expectations for real-time service you:
    - Rely only on simpler techniques, such as rules
    - Analyze a subset of transactions, or analyze post-transaction
    - Make decisions based on aged data
- **What you get**: you miss revenue-generating opportunities, overpay, don't uncover criminal activity in a timely fashion, …

The architecture you want: leverage core capabilities to automate, optimize, govern operational business decisions

Predictive modeling, business rules and orchestration together enable the most effective decisions

- Advanced analytics with classification, association, segmentation model types (CHAID, NN, C&R, …)
- Rules to define action based on thresholds
- Orchestration to coordinate all activity



Starting point: where are the transactions and data that support real-time operations located?

There's a very high probability that your operations and data of record are maintained on a mainframe …

92% of top banks

90% of top airlines

84% of top insurers

71% of Fortune Global 500

92% of top retailers

worldwide rely on the mainframe

Operational Systems of Record:

- Customer
- Accounts
- Payments
- Purchases
- Card
- Claims
- etc.

The mainframe processes roughly

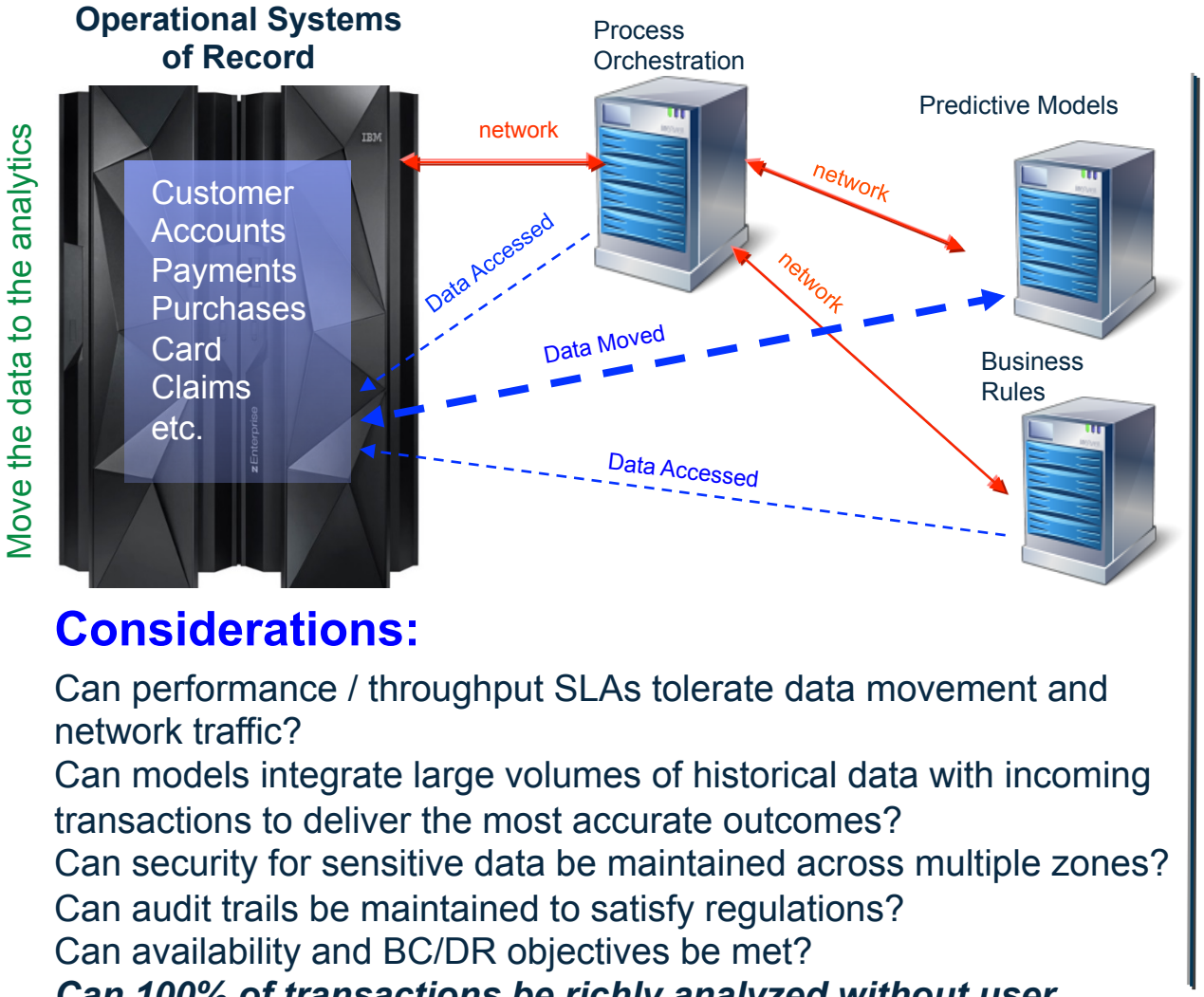
30 billion business transactions

every day

How can we leverage the mainframe for incorporating analytics into operational decisions?

Complete your session evaluations online at www.SHARE.org/Orlando-Eval

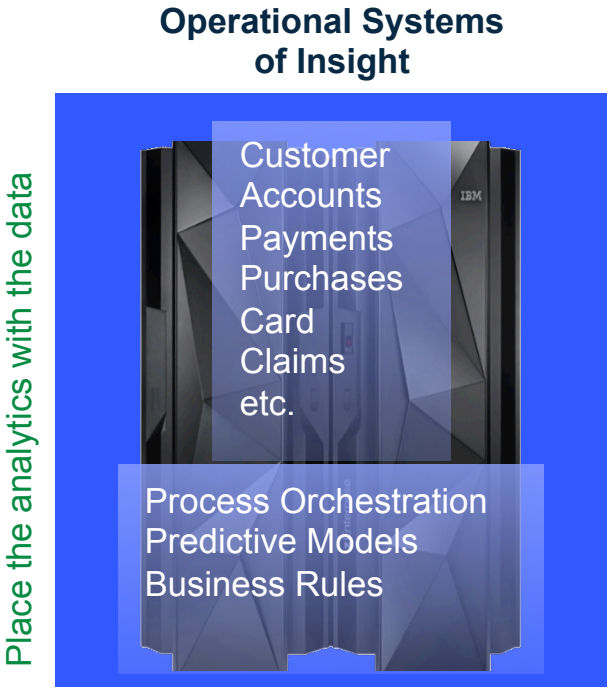
The infrastructure you need: compare approaches



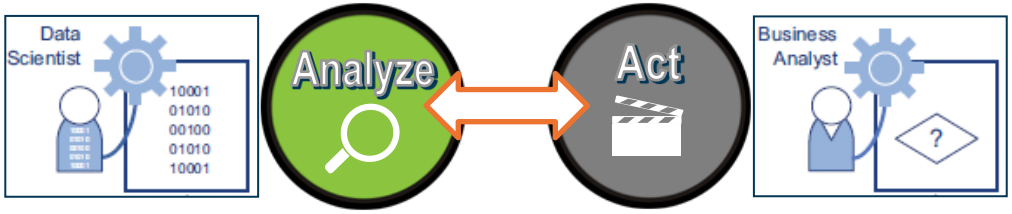
Considerations:

Can performance / throughput SLAs tolerate data movement and network traffic?
Can models integrate large volumes of historical data with incoming transactions to deliver the most accurate outcomes?
Can security for sensitive data be maintained across multiple zones?
Can audit trails be maintained to satisfy regulations?
Can availability and BC/DR objectives be met?
Can 100% of transactions be richly analyzed without user impact?

Better approach?



The core software for integrated z Systems solutions



Predictive Analytics

IBM SPSS Modeler with Scoring Adapter for z Systems

OR

Zementis for z Systems

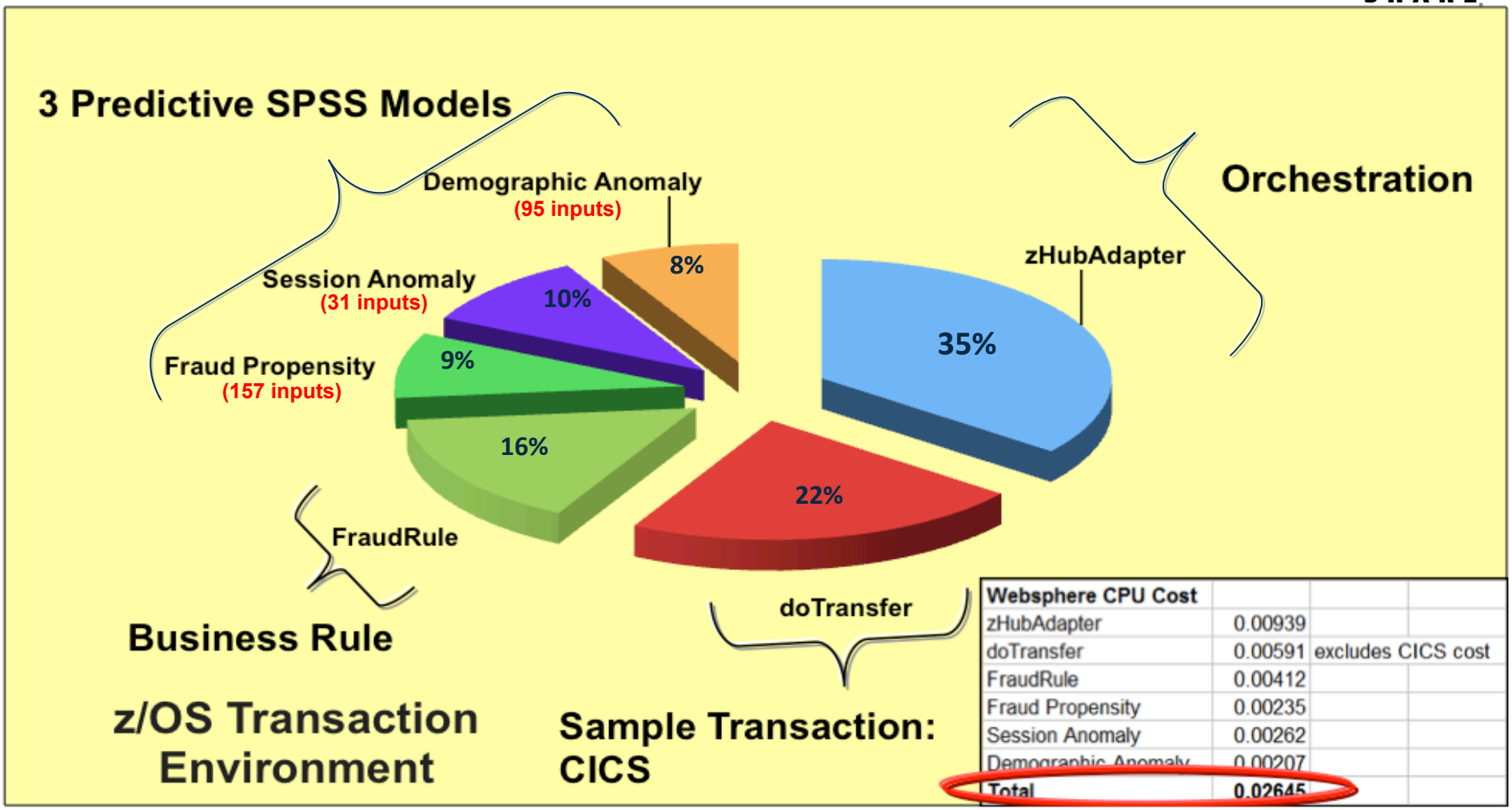
- Delivers better, more profitable decisions, at the point of customer impact
- Improves accuracy by scoring directly within the transactional application against the latest committed data
- Delivers the performance needed to meet operations SLAs
- Avoid data governance and security issues, save network bandwidth, data copying latency, disk storage
- Same high qualities of service as operational systems
- Easier to incorporate scoring into applications

Business Rules

IBM Operational Decision Manager for z/OS

- Automate and manage frequently occurring, repeatable business decisions
- Codifies business policies, practices and regulations
- Enables changes to be easily made by business people
- Automates decision making with the fidelity of an expert
- Centralized, externalized decisions enable consistency and reuse
- Manage business decisions in a natural language
- Decouple development and decision change lifecycle

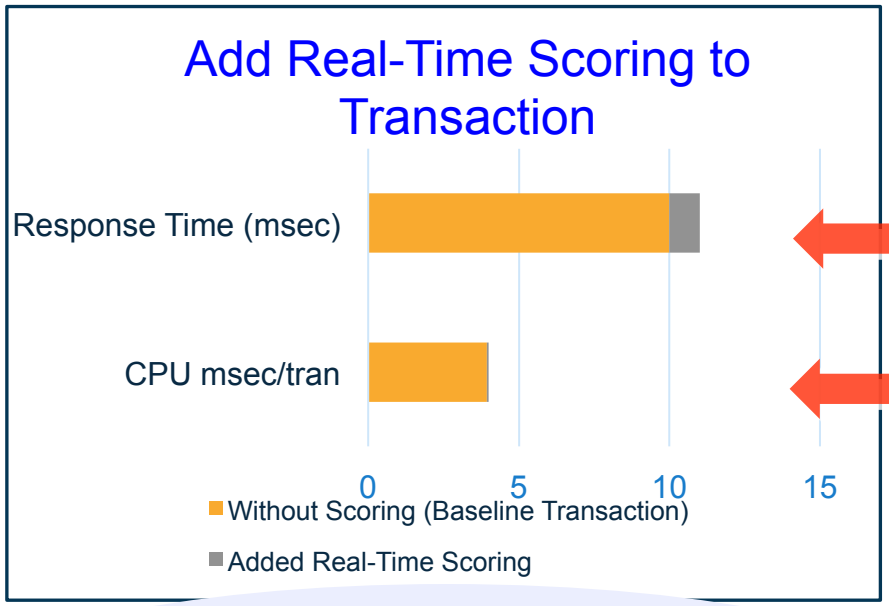
Sample performance findings: full architecture



Websphere CPU Cost		
zHubAdapter	0.00939	
doTransfer	0.00591	excludes CICS cost
FraudRule	0.00412	
Fraud Propensity	0.00235	
Session Anomaly	0.00262	
Demographic Anomaly	0.00207	
Total	0.02645	

- Three advanced SPSS models executed during each transaction with many inputs
- Performance measures show favorable results (26.5 msec CPU time, end to end)
- More optimizations possible: z196 used, interface optimizations, etc.

Sample performance findings: real-time scoring alone



1 msec difference

0.06 msec difference

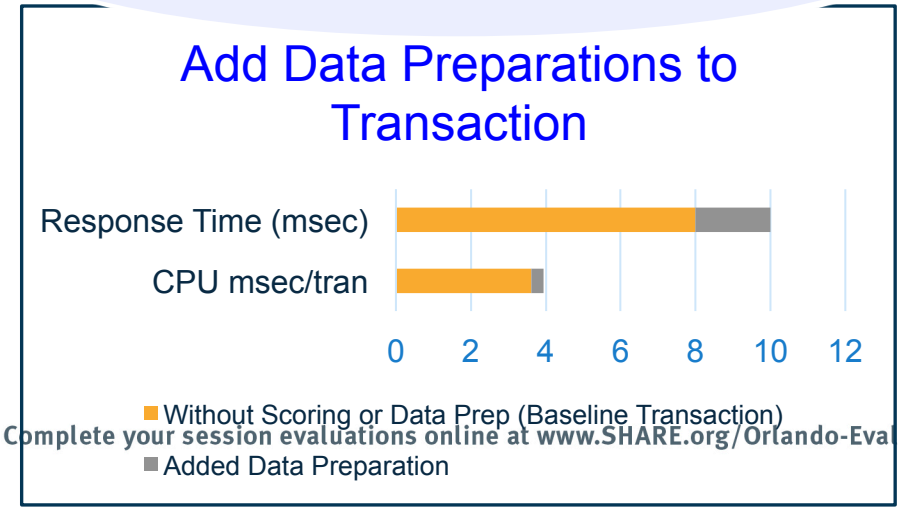
Minimal SLA and CPU impact!

Sample Workload Specifics (Lab Measurements):

- IBM z13
- Transactions > 320,000 in 5 min.
- Predictive Model: Logistic Regression
- Inputs to Model : 12

Some models may require data preparations to generate model inputs, these will vary, one example follows

*** Internal testing shows scoring to be a fairly fixed consistent cost, so relative impact will improve with transactions heavier than 10ms baseline*



Data Preparations, 12 fields total:

- 1 required no data preparation
- 5 required database access
- 6 required calculations

Significant opportunity to optimize data preparation using pre-aggregated data and IBM DB2 Analytics Accelerator

*** Results will vary based on: modeling algorithm, number of inputs, amount, scope and type of data preparations*

Use Case 1:
Countering payment fraud, waste, abuse and financial crimes
(banking example)

The worldwide cost of occupational fraud has risen 9% over the past two years, to $3.7T USD annually

The cost of fraud is considerable – and rising, not falling

- The typical organization **loses 5% of its revenues** to occupational fraud each year *(no change)*
- The **median loss** caused by the occupational fraud cases in our study was **$140K** *(up 4%)*
- **20%** of the cases were **greater than $1M** *(up 2%)*

Fraudsters are good at hiding their behavior

- The frauds reported to us lasted a median of **18 months** before being detected *(no change)*

It is nearly impossible to recover fraudulent losses

- **58%** of the victim organizations had **not recovered any of their losses** due to fraud, and only 14% had made a full recovery

Yet, most organizations still rely on "after the fact" methods

- Over 40% of all cases were detected by a **tip** – more than twice the rate of any other detection method

Sources: Association of Certified Fraud Examiners (ACFE) "Report to Nations", 2012 and 2014

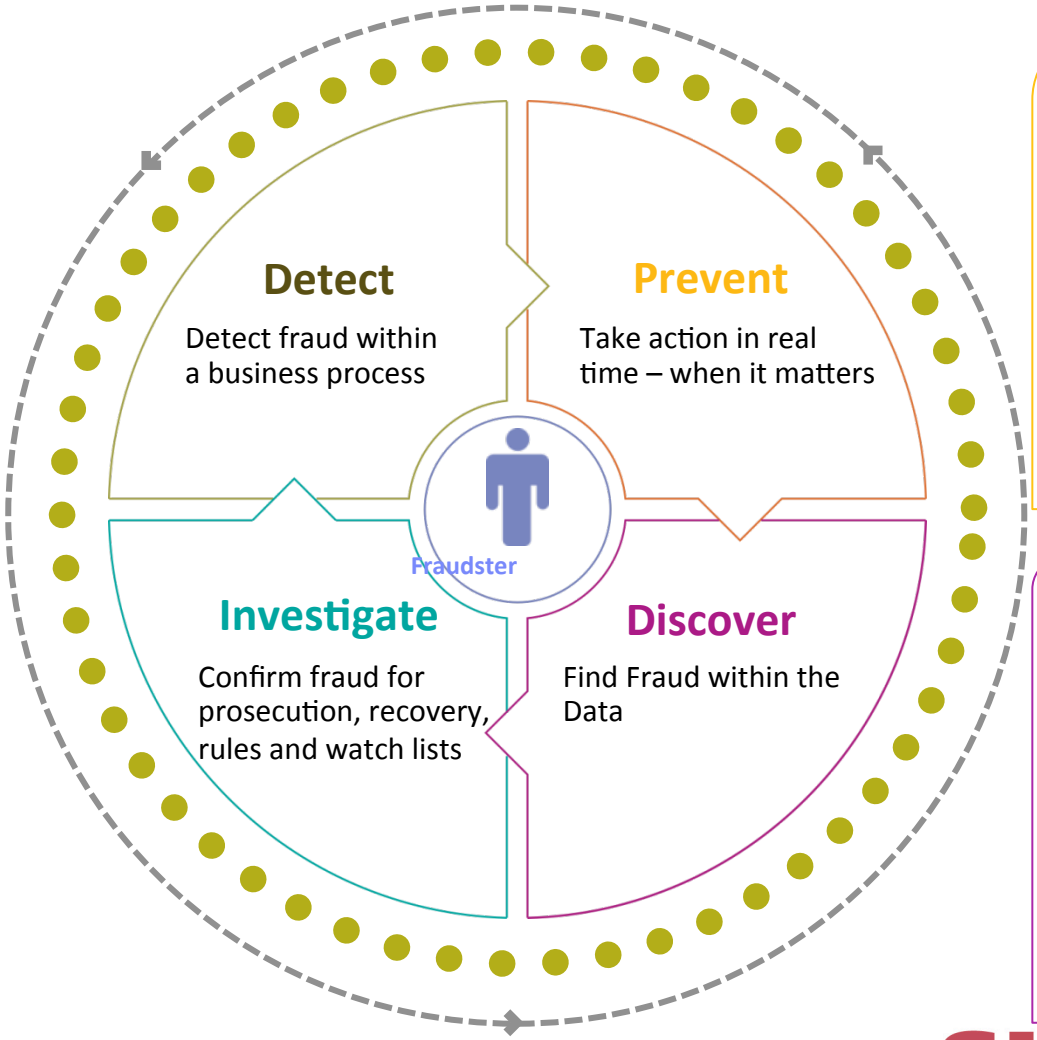
IBM's fraud management point of view

Detect

Detect in real time if a transaction, request, application, document, etc. is potentially fraudulent by applying models and rules in real time

Investigate

Gather data about fraudsters and/or schemes DETECTED or DISCOVERED; build cases for prosecution, recoveries, or denial of payments. Build watch lists and rules



Prevent

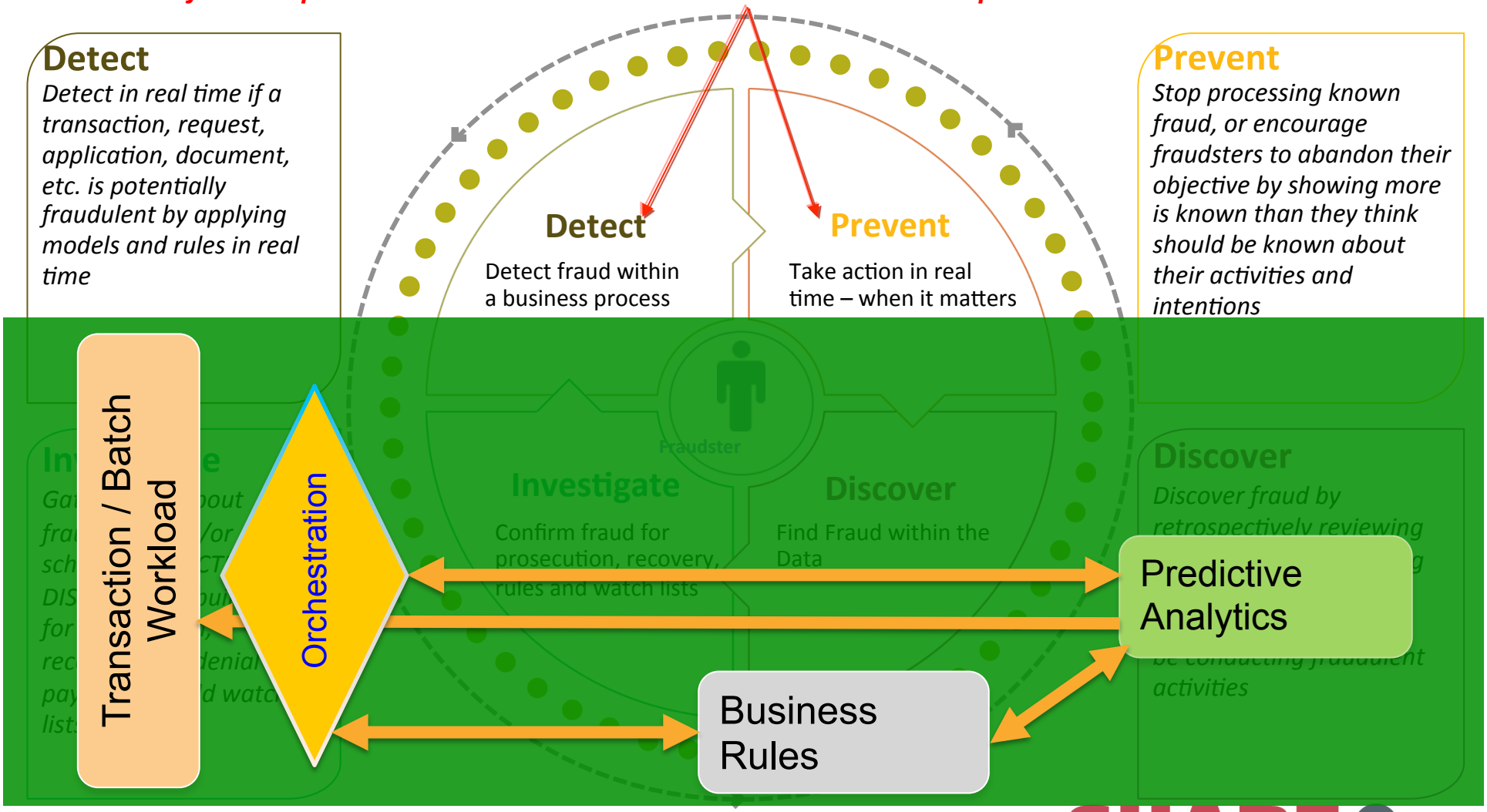
Stop processing known fraud, or encourage fraudsters to abandon their objective by showing more is known than they think should be known about their activities and intentions

Discover

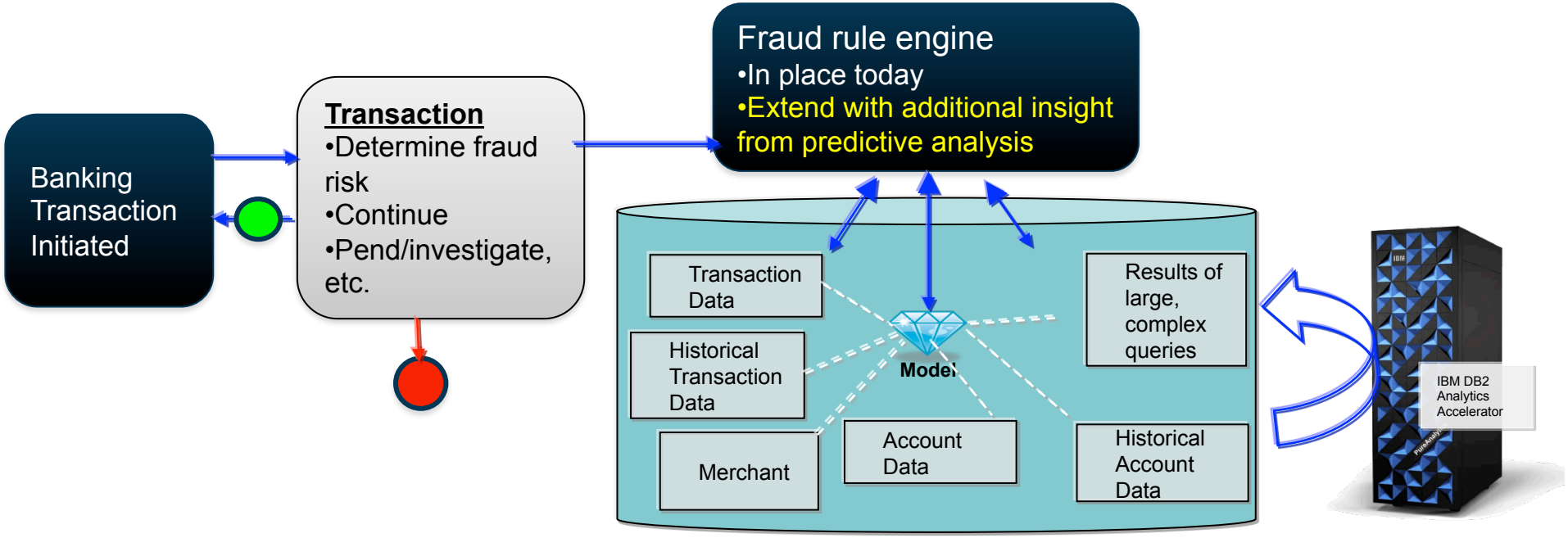
Discover fraud by retrospectively reviewing past data and identifying individuals or organizations that may be conducting fraudulent activities

IBM's fraud management point of view

z Systems optimization focus: best results based on location of operational data



Business view - bank: incrementally enhance existing fraud detect with predictive approach

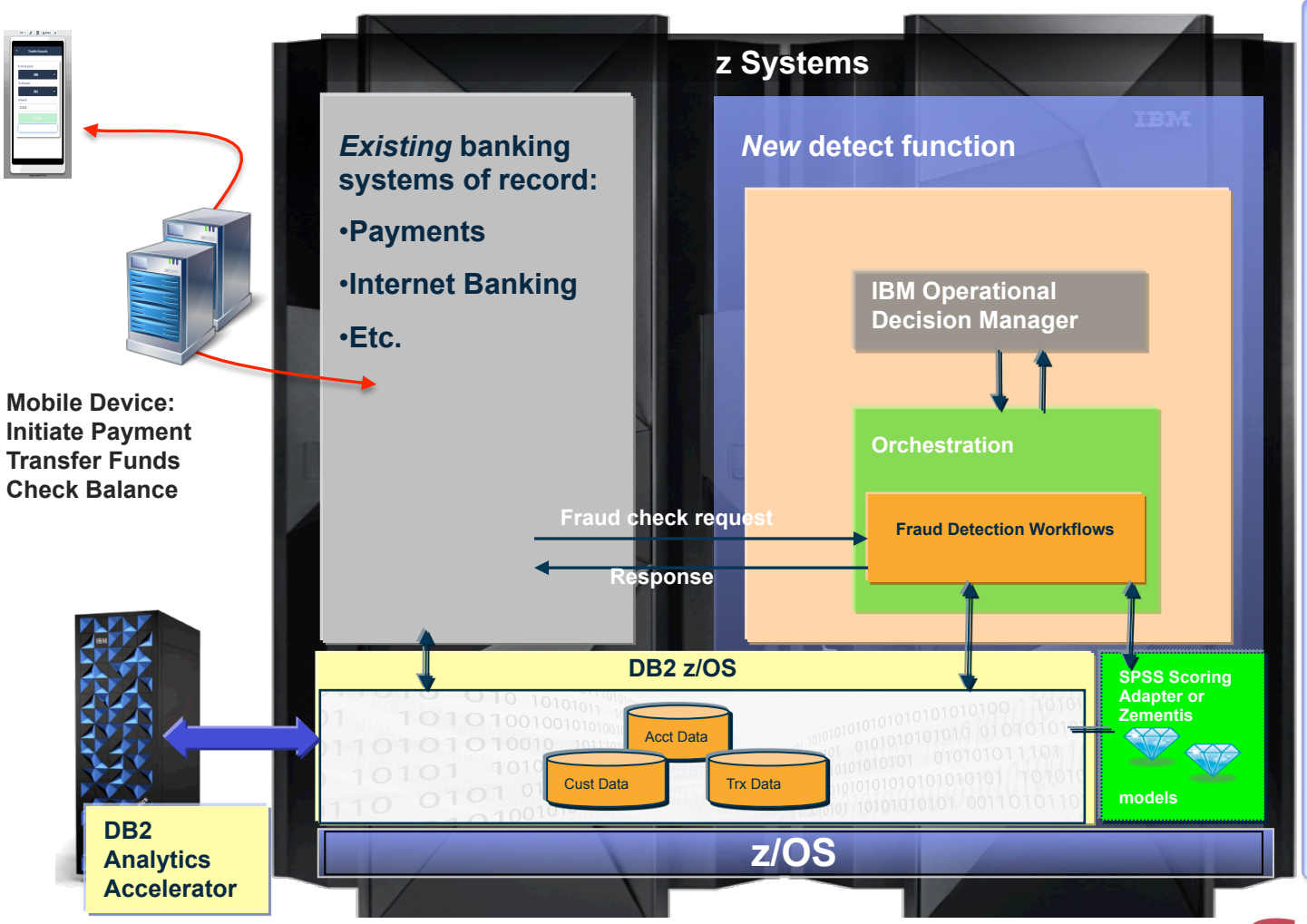


***Business Goal**: reduce loss due to card fraud, reduce card deactivation, grow revenue associated with card purchases, reduce call center costs, improve service yielding preferred card usage*

***Approach**: incorporate aggregate data from geographic location, merchant, issuer and card history into existing card authorization business flow to reduce fraud while preserving transactional SLAs*

- Integrated high performance query optimizations enable client to aggregate data several times a day and use this complex data as part of real-time fraud detection process
- Enhance with predictive scoring integrated with fraud detection transaction for even more preventive capabilities

Architecture view - bank: incrementally enhance existing fraud detect with predictive approach



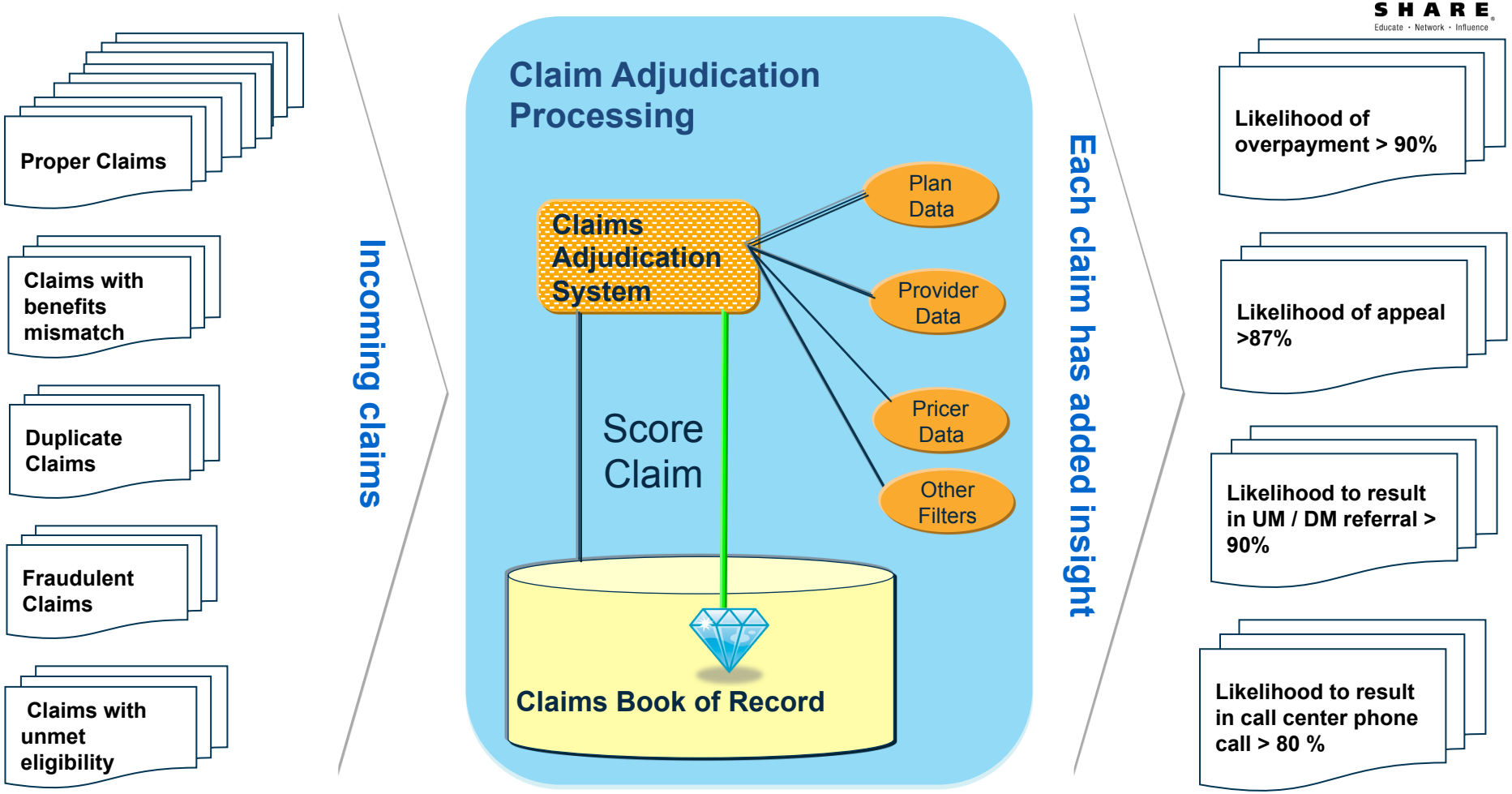
Benefits

Fraud detection rates improved through integrated infrastructure that delivers new capabilities without sacrificing SLAs or compromising security

- Flexibility to invoke processes real-time or scheduled, tight or loose coupled to invoking application
- Prioritize detect operations based on inputs (tran, LOB, amounts, etc)
- Leverage same infrastructure across multiple functions with both standardized processes & variations for specific LOBs as needed
- Can provide as a service across multiple clients, differentiate through context

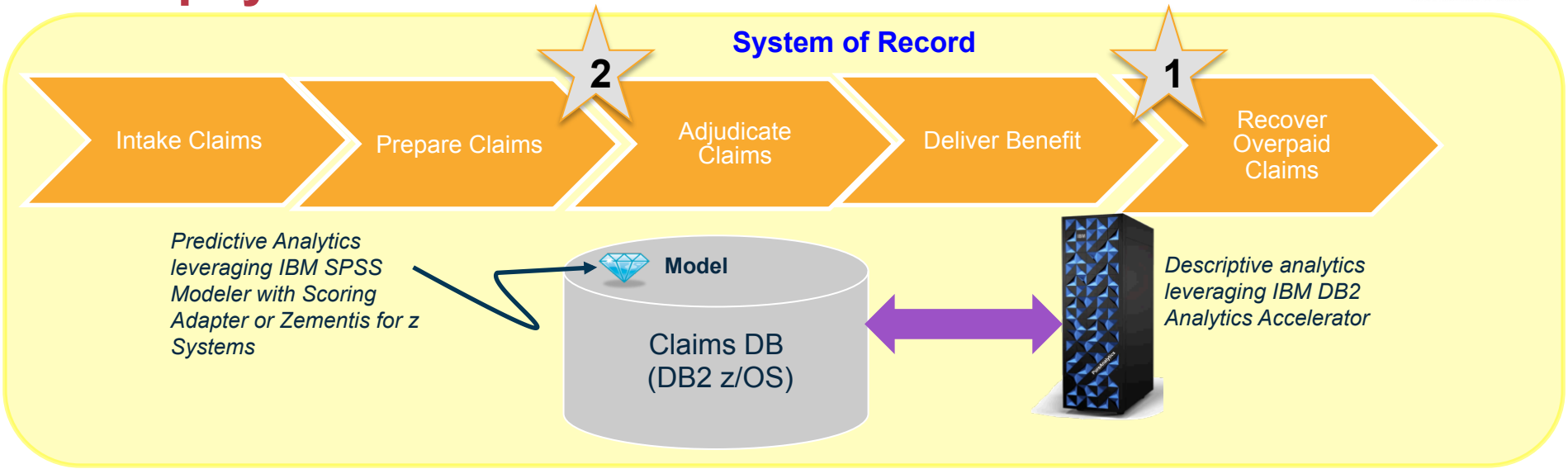
Use Case 2:
Additional in-process payment analytics (insurance claims example)

How can insurers leverage in-process scoring for best advantage?



Quickly, efficiently tag each claim with additional business insight; by extending in-process claims scoring across domains, you transform predictive analytics from a specialized function to one that best leverages transactional systems at a repeatable enterprise scale.

Business view - insurer: minimize loss from claims overpayments



Challenge 1: Complex reports for overpaid claims not completing on time, result is monetary losses

Solution: Integrate optimized analytics of IDAA with overpayment reporting of transactions

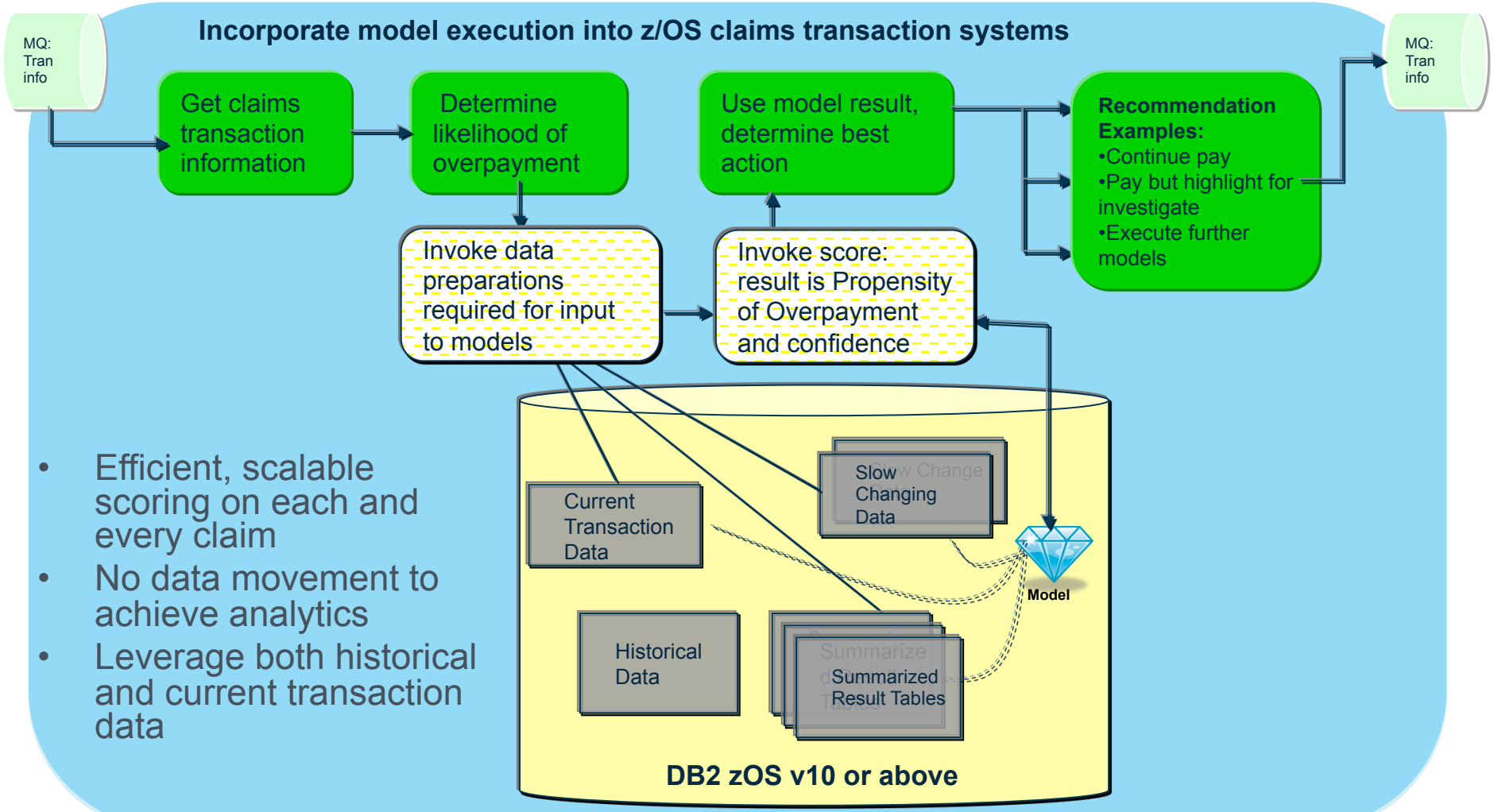
Benefits:
- Up to 2000x improvement in speed of overpayment reports
- LOB users enabled to respond with more agility to overpayment trends
- Informed decisions at the right time

Challenge 2: Stop improper payments *prior* to payment, avoid pay & chase, meet SLAs

Solution: Integrate predictive analytics into claims adjudication for analytics in place

Benefits:
- Very efficient scale for analytics
- Scale requirements only achievable with analytics as part of transaction flow
- Expected results of efficient in-transaction analytics can be multi-million dollars per year

Architecture view - insurer: minimize loss from claims overpayments



- Efficient, scalable scoring on each and every claim
- No data movement to achieve analytics
- Leverage both historical and current transaction data

Use Case 3:
Predictive customer intelligence / executing the next best action
(banking up-sell example)

Real-time predictive customer intelligence leads to smart business decisions at the point of impact

1. Build long-term customer relationships	2. Drive one decision one interaction at a time	3. Maximize customer life-time value

Telco Providers

- When a high-value, long-tenure customer calls with a bill query within six months of the end of his / her contract, offer him / her a complimentary handset upgrade if he / she renews in advance.
- Customer repeatedly called tech support with mobile Internet issues comes to the website to check process to transfer his / her number. Initiate a chat, check problems are now resolved, refund two months data charges by way of apology, and offer discount on a new handset.

Banking Providers

- Customer calls for password reminder on "self-trade" account. Pattern of usage shows limited activity and few returns; offer upgrade to managed investment offering.
- Customer's transactions show multiple trips to Asia; offer international emergency cover, and provide information about commission-free ATM withdrawals from our regional partner.

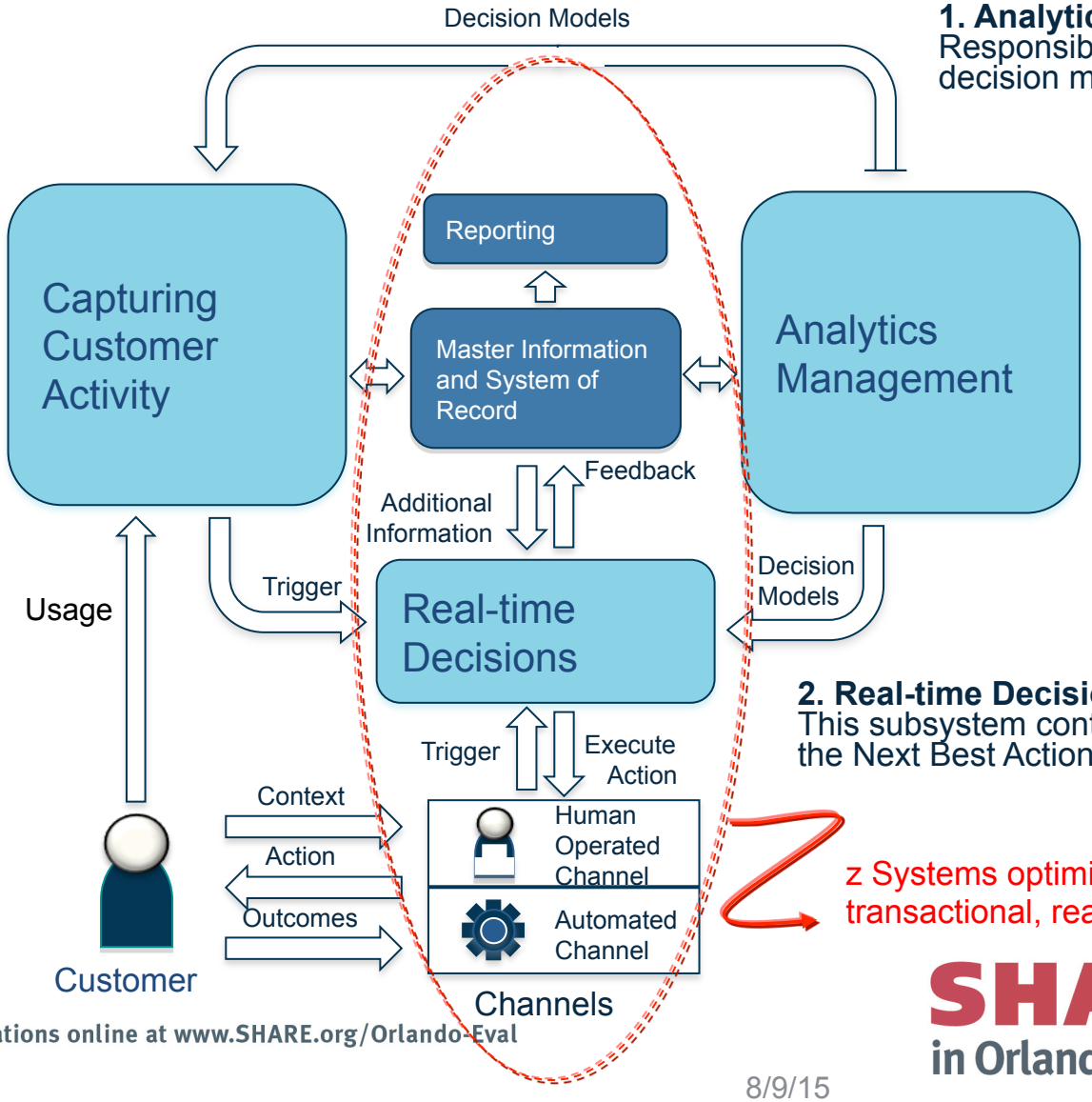
Insurance Providers

- When confirming to a customer that his / her claim will be settled, direct him / her to a repair shop which, although it may take 1 – 2 days longer than other options, has higher quality ratings.
- When a customer calls in with a coverage query, let them know they can save money by combining their two policies into a single multi-car one.

Next best action subsystems: high-level logical view

1. Analytics Management:
Responsible for developing the decision models for the solution.

3. Capturing Customer Activity:
This subsystem gathers information about the customer's activity, including the outcomes of the Next Best Actions offered by this solution.



2. Real-time Decisions:
This subsystem contains the execution of the Next Best Action decision loop.

z Systems optimized functions focus on transactional, real-time execution

Business view - bank: optimized next best action for customer-initiated loan

Short tenure small business LOC requesting car loan

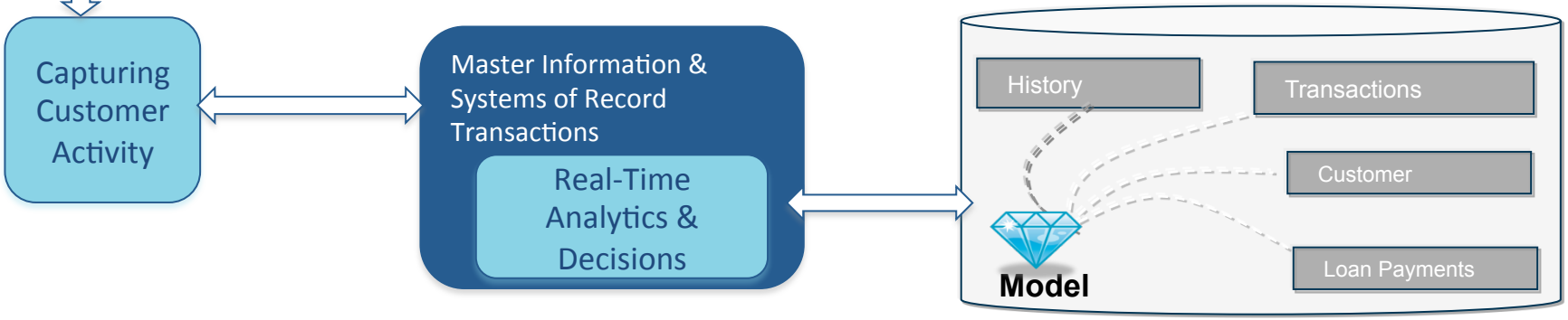
Customer call: need a new business car loan

Advise: eligible for new car loan up to $35,000 with $1,000 down on vehicle

Eligible to leverage current LOC for car loan with improved overall rate and LOC credit increase. If LOC used for loan, then down payment not required

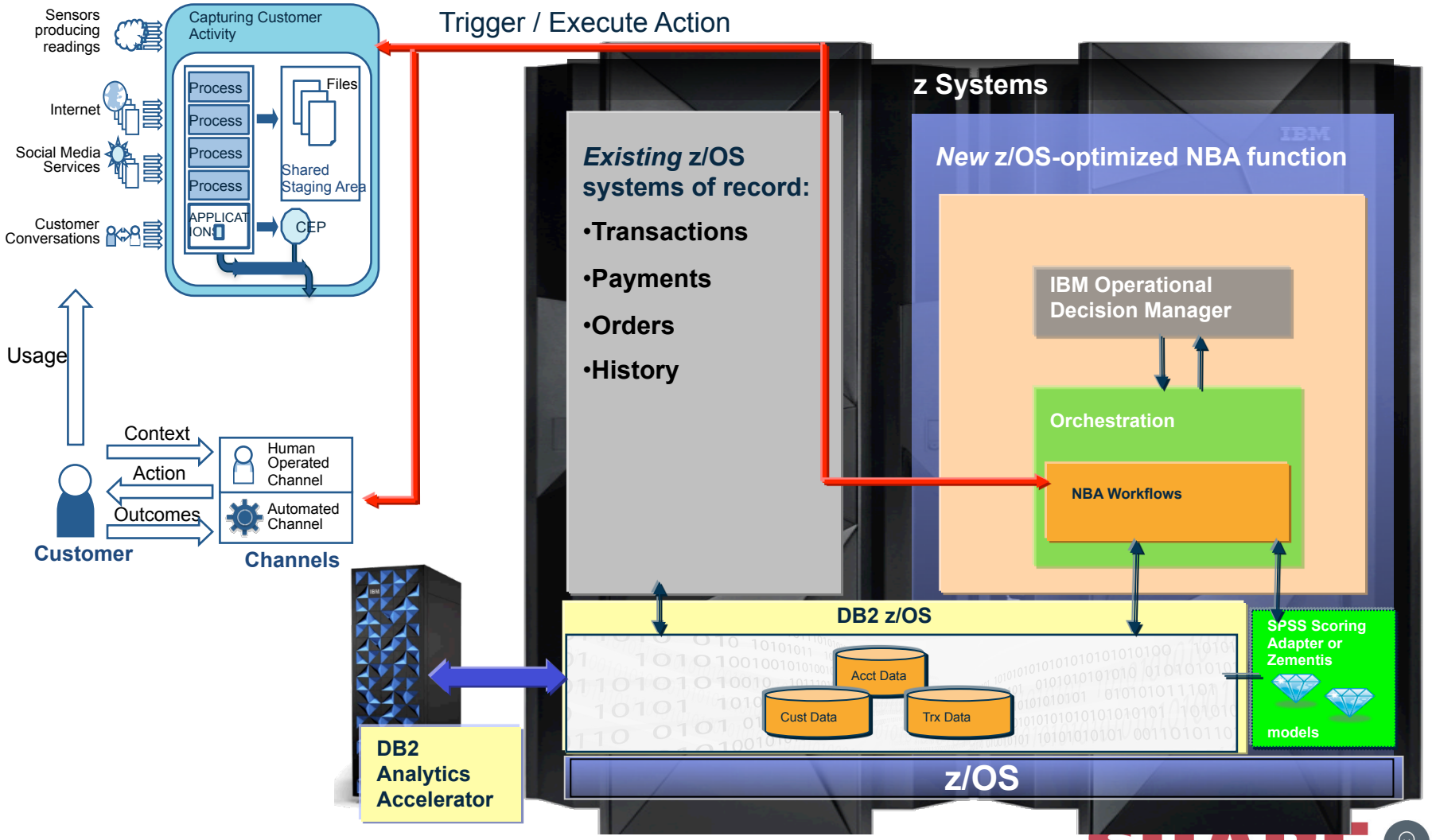
Offer accepted by customer

Offer decision considerations



- Banking clients want to integrate information across all product lines in order to make **real-time, targeted decisions**
- Why real-time? Bank may risk losing customer business or loyalty for other products
- Need to incorporate high value, predictive advanced analytics as part of transactional systems
- Why do we need all the data to score? Reduce bank's risk in approving loan

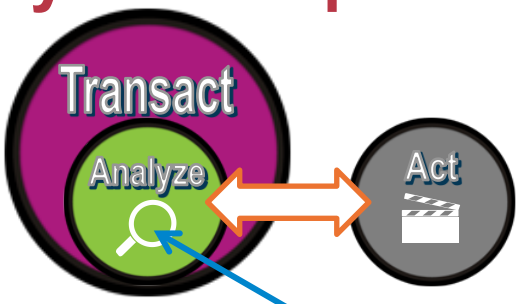
Architecture view: example real-time analytics for next best action



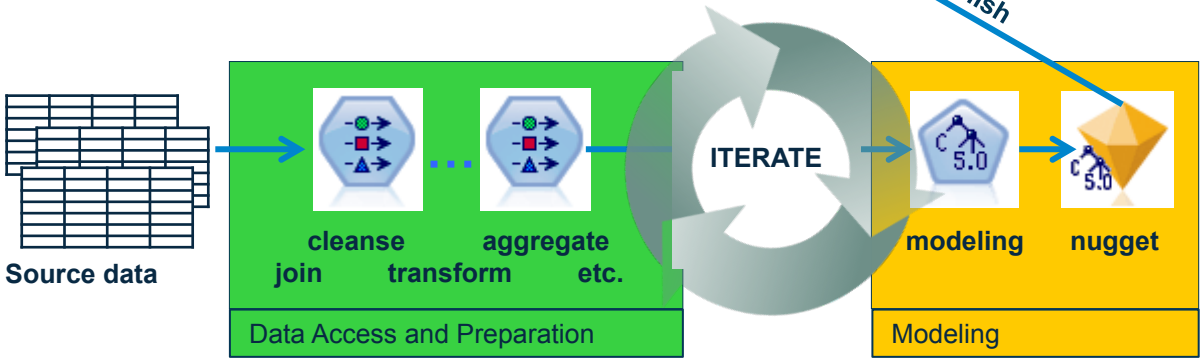
Wrap-up

What's next: streamline data preparation, modeling to improve accuracy and responsiveness

z Systems: unique ability to score transactions in real-time without breaking SLAs



z Systems: ability to build richer, more accurate models orders of magnitude faster



Business as Usual: Input data from DB2 z/OS copied to SPSS Modeler, which then performs all data preparation and modeling algorithms

Significant data movement, processing consumption limits modeling options

Transformation Acceleration (new)

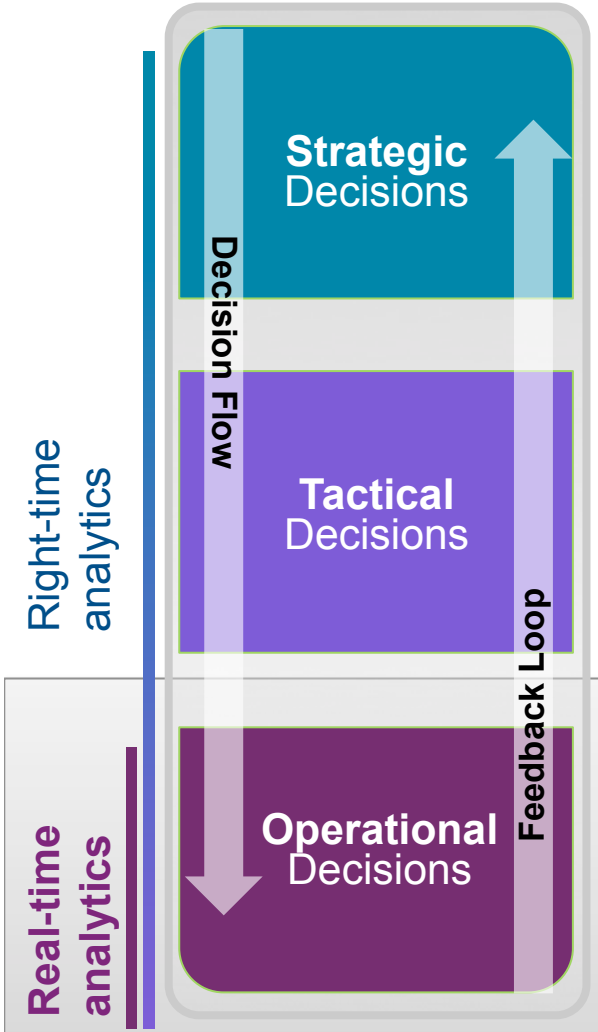
- Input data sourced from accelerated tables in DB2 Analytics Accelerator
- All data preparation performed within accelerator using accelerator-only tables

Modeling Acceleration (future)

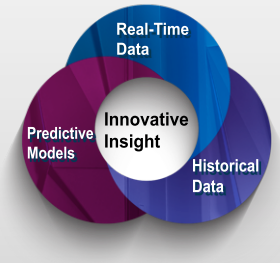
- Modeling algorithm called and executed in Accelerator on an accelerator-only table

Adapt to change faster, increase effectiveness of campaigns, generating more up-sell revenue

Recap: the key elements behind real-time operational decisions



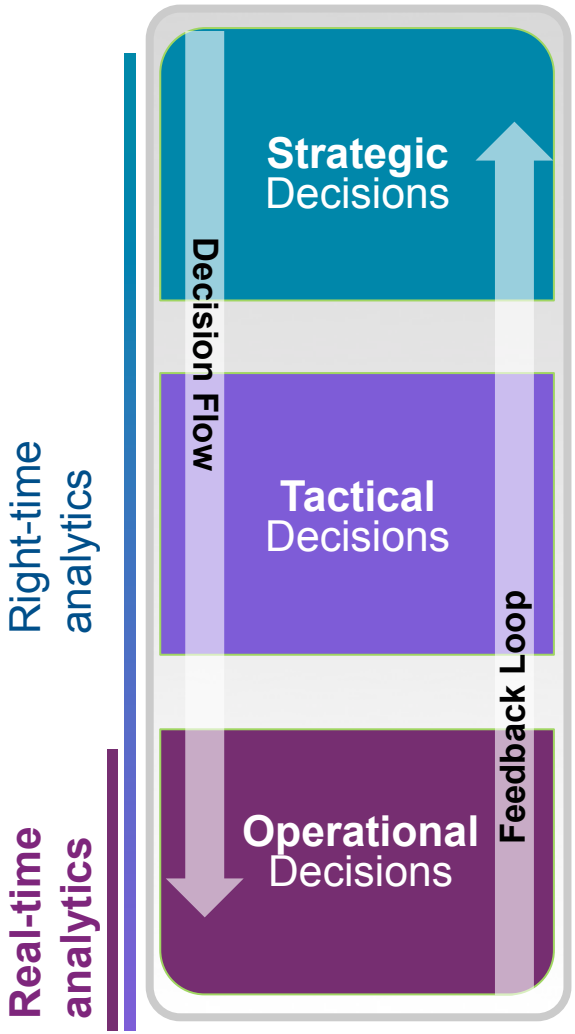
- Predictive modeling, business rules and process flows together enable the most effective decisions
- Integrating analytics with transactions transforms operational systems into Decision Management Systems

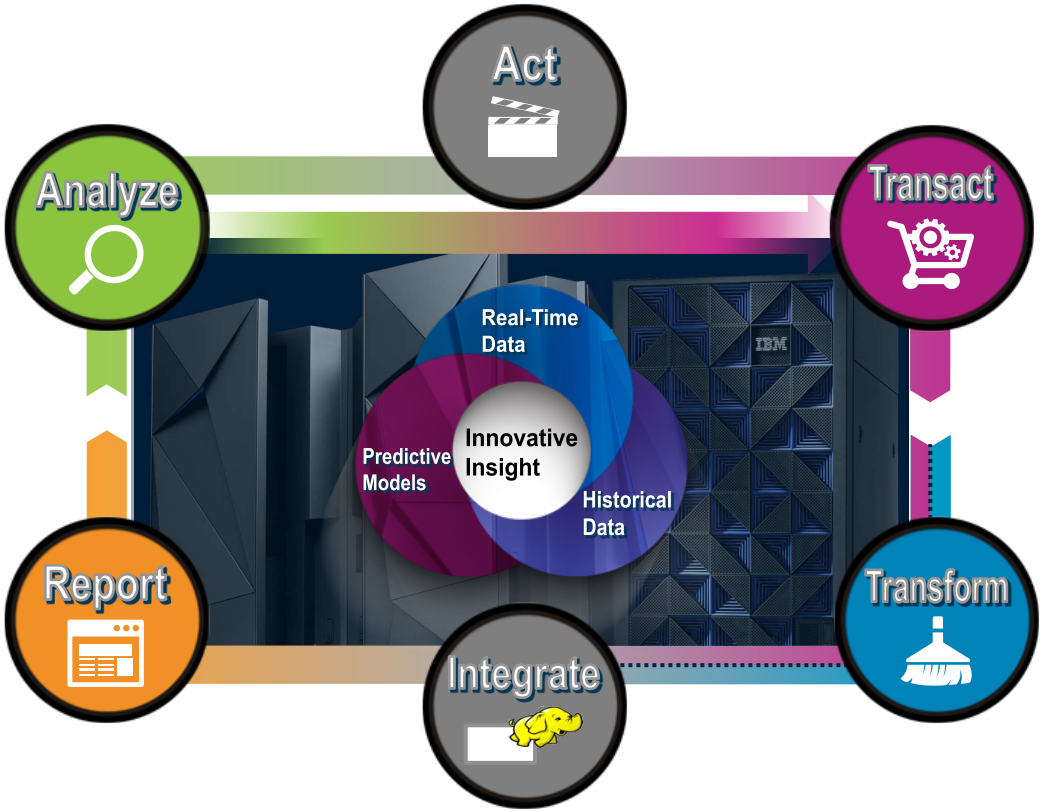


Transactions in flight

Real-time objective:
reduce time from seconds to milliseconds – or less!

z Systems: vision, strategy and technology to fuse transactions and analytics to support enterprise decisions





Best of class integrated data life cycle management for: Generating customer insights, fighting fraud and financial crimes, …

Integrating analytics with transactions on z Systems

Business Advantages

- Integrate advanced analytics as part of each and every transaction with negligible impact to transaction SLAs
- Access to most current data for best analytic outcome, reduced false positives
- Actionable insight on every transaction, real-time or batch
- Analytics in the flow of business to stop fraud, increase customer loyalty, increase revenue and reduce risk

IT Advantages

- Avoid costly ETLs for analytics with fewer copies of data to manage, secure, and make highly available
- Reduced network costs through avoiding off-platform calls during transaction interaction
- z Systems governance for integrating new analytic models
- Leverage investments in z Systems data infrastructure, particularly data sharing and DB2 Analytics Accelerator
- Extremely efficient scoring within DB2 or Java minimizes IT consumption

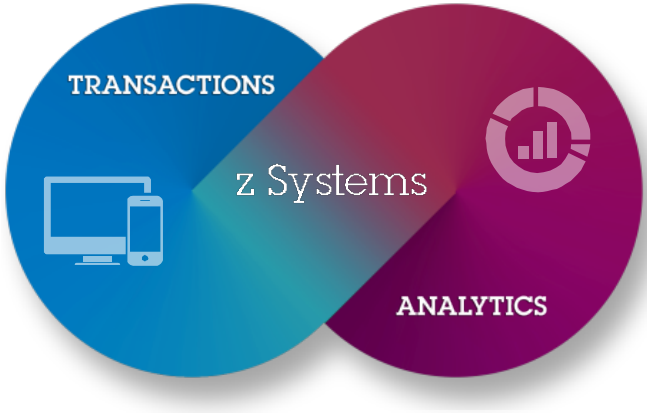
Session summary

- Predictive analytics should be integral to any real-time decision architecture
  - Don't let outdated views of infrastructure cause you to sub-optimize operational business solutions!
  - An architecture that combines rules and predictive analytics can deliver the most effective real-time decisions

- When operational systems are on the mainframe, IBM and Zementis products deliver optimized support for real-time decision management while maintaining operational SLAs
  - Common use cases are countering payment fraud, waste, abuse and financial crimes; payment optimizations; and predictive customer intelligence
  - IBM is now optimizing the model creation phase for the mainframe as well

- The z Systems portfolio enables "right-time" insights throughout the hierarchy of enterprise decision types

Insights on every transaction



Analytics as part of the flow of business

Complete your session evaluations online at www.SHARE.org/Orlando-Eval

SHARE
Educate · Network · Influence

THANK YOU

Complete your session evaluations online at www.SHARE.org/Orlando-Eval

SHARE in Orlando 2015

IBM Corporation
Systems and Technology Group
Route 100
Somers, NY 10589

Produced in the United States of America
February 2015