Why Database Matters In Card Fraud Detection

Big Data LDN : 13-14 Nov, 2019



Dheeraj Mudgil

Enterprise Architect [Fraud]

Introduction

Barclays

- Is 329 year old Bank (1690); it is older than the UK Itself
- Moves, lends, invests and protects money for customers and clients worldwide
- Offers Banking, Cards, Corporate Banking, Insurance, Markets, Mortgages, Payments Acceptance, Ecommerce, Wealth & Investment Management ...
- Financial Innovation

Dheeraj Mudgil

- Enterprise / Solutions Architect with software engineering background
- 20+ years in the Industry mainly in Financial sector
- Working with Barclays in Chief Technology Office
- Architecture owner for the Fraud Platform



How much data we generate today?



Trends - General



- Devices
- New Channels



- Automation
- Desire to release products fast



- Social Media & Info Sharing
- Innovative Products



Vulnerabilities



Trends – Customer



- Customer Patience
- Instant / on-demand



- Loyalty towards Product
 - Enabler Options/Choices



- CX Expectations
- Expects protection from Loss



- Complex customer needs
- Complex customer behaviour



Trends – Fraud & Fraudster



Barrier to HPC → Enabler - Cloud



- Sophistication of Fraud Techniques
- Fraud supply chain more organised



Barriers to collect Info

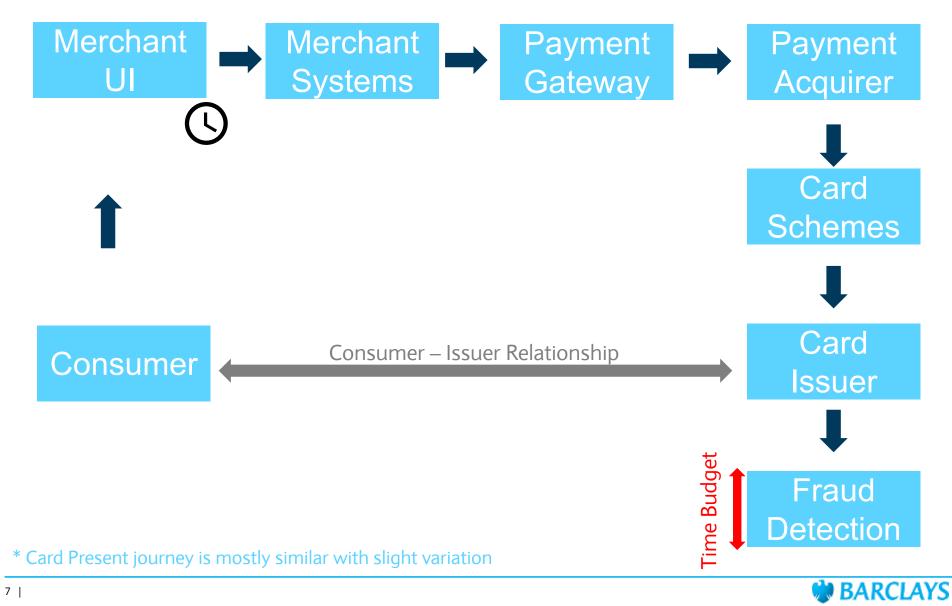
Enabler – Social Media Platforms



• Complex fraudster behaviour



Typical Card Payment Journey (Card Not Present)

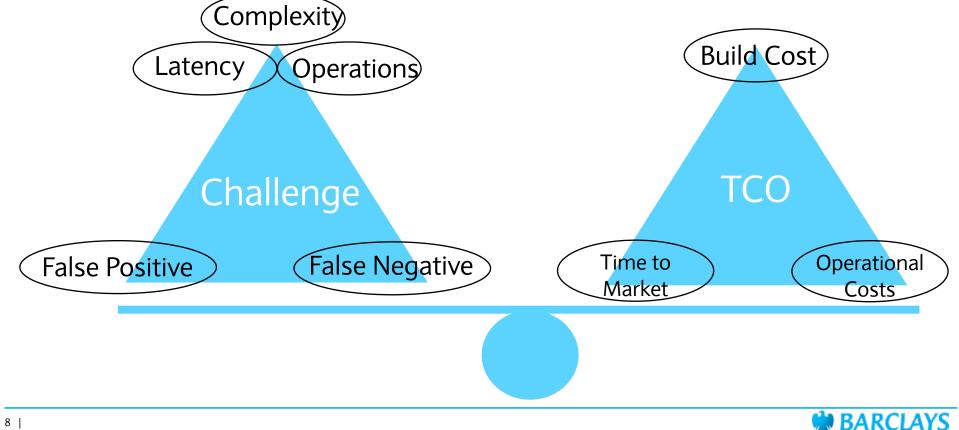


Use-Case Analysis

- **Consumer Swift Checkout & Fraud Protection** ۲
- Merchant Fast Approval & Zero declines ۲
- Issuer Encourage Card usage & Protect Fraud losses •

Key Characteristics:

- Latency Sensitive •
- **Complex Processing Rules** •
- Large Datasets •
- Consistency •



Very Simple !

Set up a Database – To store customer history & patterns

Build an Application – Containing logic & expose via API



Which Database?

INFRASTRUCTURE **APPLICATIONS – ENTERPRIS** HADOOP ON-PREMISE -HADOOP IN THE CLOUD DATA ANALYST PLATFORMS STREAMING / IN.MEMORY DATA SCIENCE PLATFORMS SALES MARKETING - R2R MARKETING - R20 CLISTOMER SERVICE Zeta SendGrid MEDALLIA zendesk cloudera Hortonworks **AWS** Microsoft Azure aws statabricks Sstriim Microsoft Opentaho alteryx RADIUS App Annie IEM AKNIME 💋 data CLARABRIDGE C INSIDE SALES.COM (/), E BlueYonder [PERSADO] @kahur C Google Cloud BigInsights Lattice MAPR. Pivotal Confluent GridGain COMING MING Repidminer Gainsight NG DATA ATTIV/O Datameer Ouid incorto. 🚫 conversica 🐖 ME ACTIONIC SAUTHRU DigitalGenius afiniti IBM InfoSphere 🔶 tubular 👹 DataFox 🙉 clari A aviso tact.ai **8** ANTIFIND Comparticle Amplero. inter ana ClearStory Origami hazelcast TERRACOT bluedata jethro C A Z E N A 💥 CenturyLink & amperity fuse|machines ENGAGIO msgai 🛄 INTERCOM ENDOR MODE Bottlenose HUMAN CAPITAL LEGAL FINANCE ENTERPRISE BACK OFFICE -SECURITY NoSOL DATABASES -NewSOL DATABASES GRAPH DBs - MPP DBs - CLOUD EDW **BI PLATFORMS** VISUALIZATION MACHINE LEARNING PRODUCTIVITY AUTOMATION TANIUM NETLANEE Szscalar RAVEL How Vie entelo /:napla Scoogle Cloud aws Clustrix 📦 🌔 neo4j 🎄 + a b le a u 🛛 SAP aws Microsoft AWS aws **Ui**Path StackPath 🔀 illumio 🔽 CODE42 🐨 CipherC Azure Machine Learning Q^{Seal} 🗱 slack VERTICA X. Pivotal Amazon Nentur ORACLE Min zuora IBM Data Wasabaura Cu 🙆 Google C 🙆 Google Cloud H,O 🗑 Cockroach LABS 🛛 😣 IBM JUDICATA mongoDB MarkLogic MEMSQL () influxdata A DataRobot ORACLE Oction BREVIA umiata 😳 DIFFBOT Qlik Q Periscope looker ATSCALE aamalo RoWade&Wendy Stella Captricity 1.9 DOTOSTO 8 VOLTOR Pivotal ELEMENT" Rognitic PREMIENT 🔘 clara talla **A**ppZen BlueTalon Custree 🗬 pymetrics 👻 VERSIVE R TRADESHIP ArangoDB Couchbase Citusdata splice VISENZE InfiniteGraph Exasol #snow1 butterai Kasisto 🔓 GoodDuta 🝀 birst mya uncommon de casete Objectivity CHARTIO redislabs SCYLLA. & paradigm4 deeporrosio bonsa **Ø**dremio APPLICATIONS – INDUSTRY COMPUTER VISION SPEECH & NLP HORIZONTAL A DATA TRANSFORMATION - DATA INTEGRATION - DATA GOVERNANCE -MGMT / MONITORING • EDUCATION -GOVERNMENT FINANCE - LENDING FINANCE -REAL ESTATE INSURANCE -ADVERTISING 🙆 Google Cloud 🙂 twillo Microsoft Azure BM Watson Cortana, Face 8 ♦ Informatica MuleSoft 🔥 Informatica AWS O New Relic octufio MediaMeth 🕃 Liulishuo ondeck Affirm INVESTING REDFIN •talend @pentaho IBM @SailPoint AppNexus **O OPENGOV** 🔿 amazon alexa - narrative 🔕 sci O rubrik O APPDYNAMICS Amazon Rekognition snaplogic & TEALIUM sentient 🛞 Voyageries 🗹 devices criteol. 🗣 xAd 🌖 Integral KNEWT 拍拍贷 UIANPU.AI C Datamin Opendoo semanticmaching Lemonade WAVEFRONT * Clover o clarifai 🙆 Mobvoi 💎 🔊 Kreditech AVANT ¢ v⊤s McAfee Skyhigh Security Cloud altervx 📿 TRIFACTA Segment enigma 🔯 👩 dynatrace) Affectiv ø DRACLE OpenX at sceres mark43 ADDEPAR Geclara Connance A Upstarl の商品 SoundHound Inc. CREDIFI NUMERAI C podium 🏥 collibra 🔆 tamr 🛛 🥬 Paxata alooma splunk> SignalFx ¢druuo A Numenta PETUUM theTradeDesk Adaprithms kidaptive INSIKT J upgrade 100Credit A Indiad Ocortical io snips ZAL©NI SEVER AL @ deepomatic dstillery 🛃 Live FN FiscalNote ISENTIUM Moogsoft Qurravel Alation NOCOLOgics -CURIOUSAL OSARC PANORAMA 🕜 WeLab ٥ ۷ \bigcirc StreamSets UNIFI () Gridspace TAPAD mgur TrueAccord S MoneyLion COMPST/ import.io 🔡 Stitch Data OKERA 2 twenty by pagerduty *Numerify knowre > RevenPac 10 OpenDataSoft Ð aire agnifi A CAPE ▲ PAGAYA SEARCH -SOCIAL ANALYTICS LOG ANALYTICS WEB / MOBILE / ---STORAGE CLUSTER SVCS ---- APP DEV --- CROWD ---HARDWARE GPU DBs HEALTHCARE LIFE SCIENCES TRANSPORTATION -AGRICULTURE COMMERCE - INDUSTRIAL -splunk> COMMERCE ANALYTICS ٥ SOURCING Google TPU grm kin≣tica Hootsuite GiloT PREDIX aws flatiron Clover XYRUUS HealthTap UBER TESLA Lightber ndWe COLOT 🕅 KarbonX FARMERS Google Analytics sumologia NETBASE TITCH FIX (Intel)AI GRAPHCORE CLEARPATH lental verily mixpanel A. AMPLIT Di 2Co TACHYUS 🐲 👉 docker Granular PURESTORAGE 🛕 Keen IO coveo LOGGLY 🝠 synthesio 🛛 🚮 MYTHIC WUXINextCODE ZEPHYR Vowork Lucidworks drive.a 📹 RetailNe swiftype BLUERIVER rainforest 🕲 TIMBER sumAll 👝 Airtable Alluvium SCORTE implereach AMOTHE floure MESOSPHERE \î naut nimblestorage 🖸 algolia FarmersEcio scole Movidius iii alphasense kıbāna bitly predata RESCI DISIGOPT Qumulo **RI AZINGOR** OTHER 🙆 Core OS 🚙 CASK HIVE HALD FarmLogs brytlyt PG-Strom omni:us SINFOLIA 🕱 logz.io SimilarWeb eranify custora OPTIMUS •moovit COHESITY DNAnexus ñexar Hoppen celect mavrx CITRINE twoRAR comma.ai duetto **R**SXEVER a Terration N netrodyne HOTH PAIGE DATAVANT M Unbabel 👩 Adardeck 🍠 Second 6 prospera aws Soogle Cloud Hicrosoft Civil Maps **OWKIN** SECURITY FRAMEWORK OUFRY / DATA FLOW DATA ACCESS COORDINATION STREAMING STAT TOOLS AL/ MACHINE LEARNING / DEEP LEARNING SEARCH LOGGING & VISUALIZATION COLLABORATION MONITORING (Pheeloop) Spant SQL 🕺 🐝 Spark 🙆 🍦 python" TensorFlow theano M A Dlib. - learn • jupyter Apache Range \checkmark 🛕 nifi 🌒 mongo DB talend Zeppelin • elasticseard 📕 kibana cassandra KNOX R Caffe Microsoft hadoop Apache Zookeeper DM TK Gerald Stream Publichade Solr® 😚 Rodeo & SciDB YARN TEZ ScalaLab Apache 😋 SINGA A SENTRY 🔒 logstash mxnet 🔆 VELES 🖋 0 Flink 🗞 kafka 😑 druid Sentry FeatureFu Spark 🔋 Seentsco Mriak 00488 🛐 NumPy Aucore ANACONDA MESOS neon' DIMSUM Prometheus Apache Amba 🐼 DL4J Aerosolve SciPy Genoe DSSTNE DATA SOURCES & APIs DATA RESOURCES HEALTH FINANCIAL & ECONOMIC DATA AIR / SPACE / SEA PEOPLE / ENTITIES LOCATION INTELLIGENCE OTHER DATA SERVICES -INCUBATORS & SCHOOLS -RESEARCH Orbitol Issight planet. Bloomberg 🥝 THOMSON REUTERS D | DOW JONES FOURSQUARE qualtrics Q Palantir 👧 facebook research GE Digital acxiem. experian. VALIDIC OpenAI 💜 MIRI DATA.GOV FAirware AIROBÓTICS **OPERA** Mu Sign CAPITAL IQ CBINSCHIS Xignite Quandl EPSILON practice fusio A Inside/view Place 📵 🎯 esri -22 🌔 M.A sense 340 ∆ spire @ mccsown 🛞 data.world " 😰 DataCamp </u> DataElite Crimson Hexagon D 🧭 helium 🛛 🧕 samsara PREMISE @estimize + fitbit GARMIN ۲ Descartes 🗜 factual. 📴 🔨 Mapillary enigma DK Quantcast fractalese kaggle BASIS WINDWARD' Stellusiabs Esgle Alpha StockTwits Stall thinknown SAFE GRAPH n 😢 cuebiq 3 HUMAN API Skinso

BIG DATA & AI LANDSCAPE 2018

V1 - Last updated 6/19/2018

10

© Matt Turck (@mattturck), Demilade Obayomi (@demi_obayomi), & FirstMark (@firstmarkcap) mattturck.com/bigdata2018

(EXL DataKind

How to Select:

•

FIRSTMARK 📂

. . .

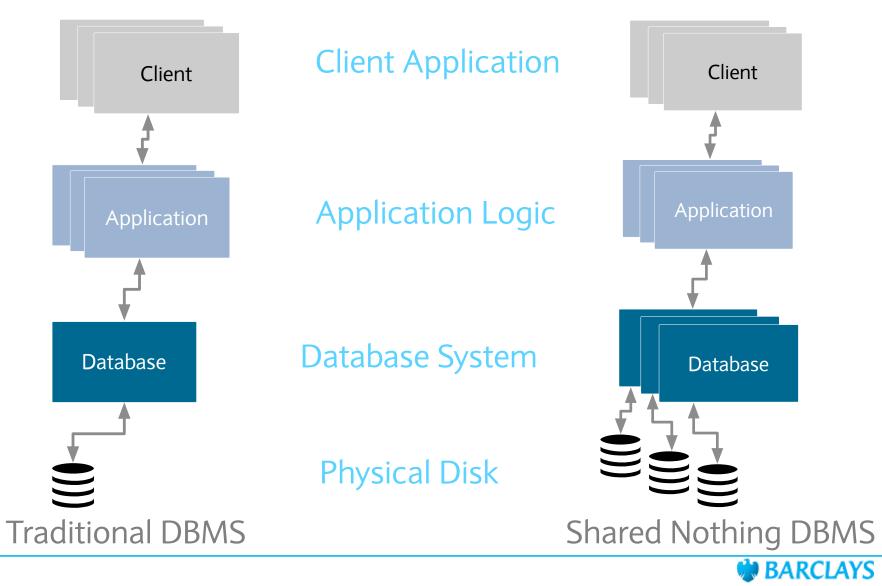
- Known Lets use what I am familiar with
- Generic Lets pick a ۲ general purpose database
 - Safety Look at Industry reports (Gartner etc.)
 - Enterprise Constraints – No more than one/two DB allowed
- **Retrofit -** We already have something lets use it
- Luck Toss a coin / dice!

Credit for Infographic – https://mattturck.com/bigdata2018/

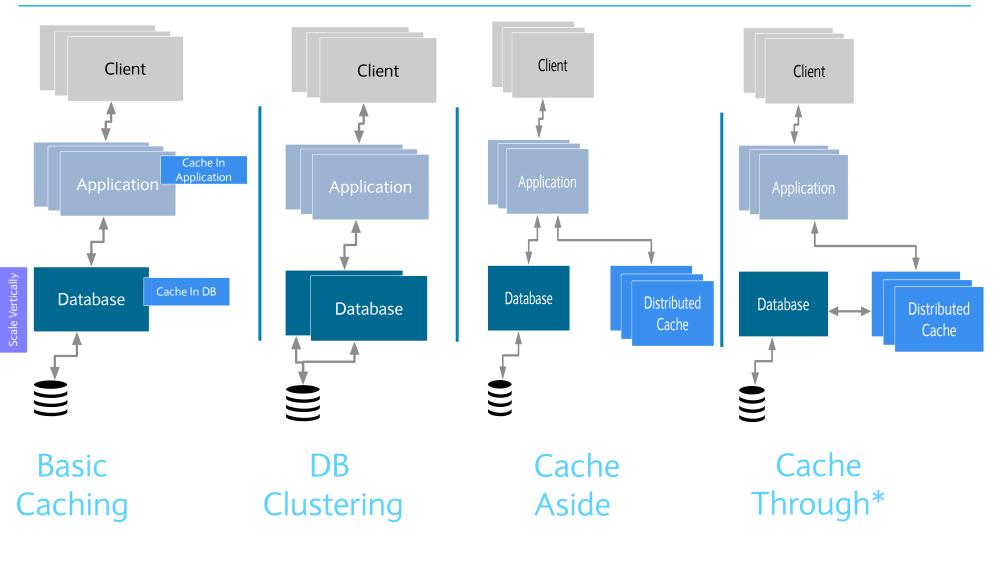
DroneDeploy



Simple Solution



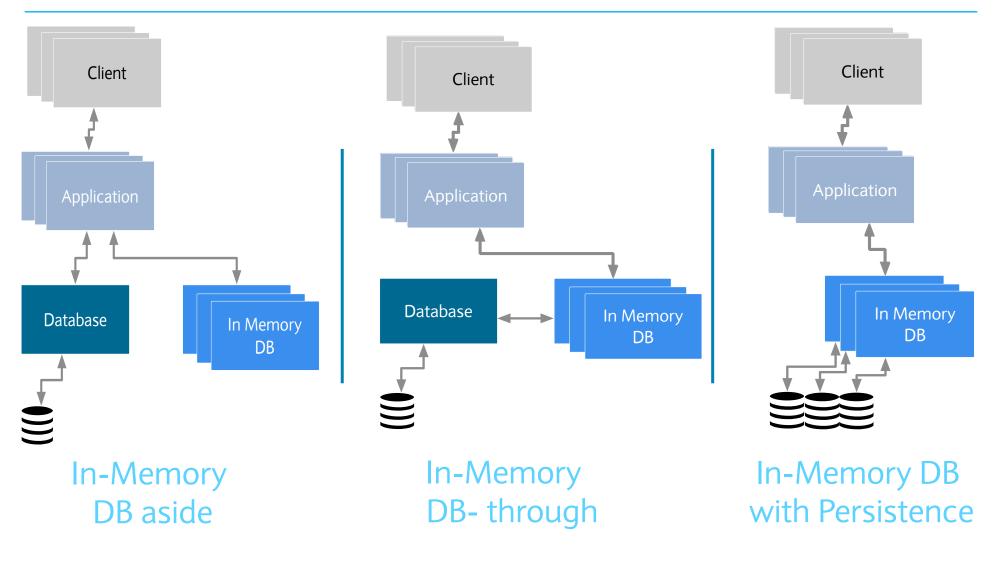
Common Patterns [Simplistic View]



* Other variations also exists

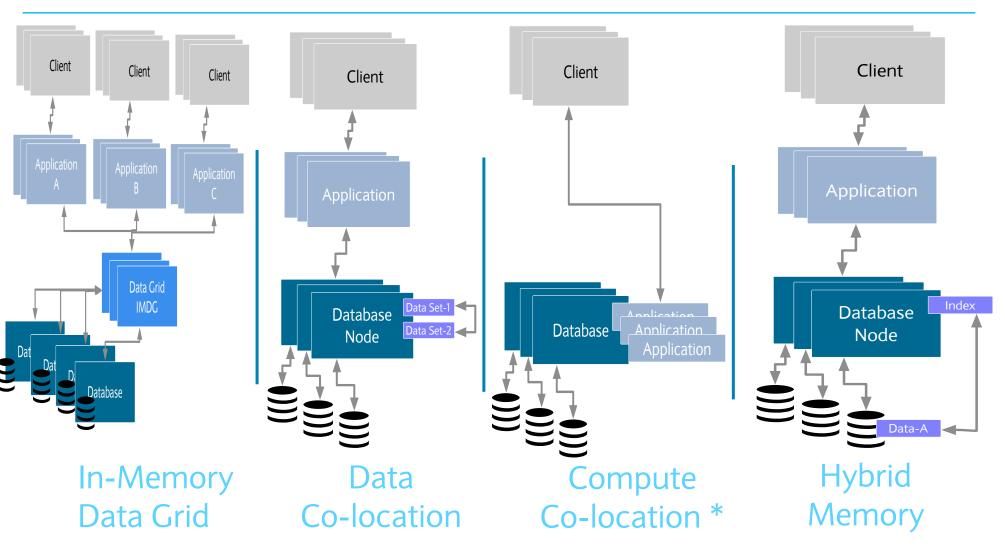
unrestricted

Common Patterns [Simplistic View]





Common Patterns [Simplistic View]



* Deploying application in database is roughly similar



Complexity of Financial Fraud

1 st Party Fraud	3rd Party Fraud	2nd Party Fraud	Mule	s Scams .	
Financial Crime					
				—— Key Concep	ts
Card Fraud	٠	Invoice Fraud	٠	Identity Fraud	
Account Take	overs •	Bank Account Frau	d•	Plastic Fraud	
Cheque Fraud	•	Cyber Fraud	٠	Mandate Fraud	
Payment Fraud		Cash Point Fraud	٠		

Extending use-case: Implications?

- Complexity & Criticality
- Additional data needs
- Scale, throughput and other NFRs...



Fraud Types

Regulatory requirements



- Impact on the use-case ?
- Impact on Architecture ?

- Strong Customer Authentication (SCA)
- Transaction Risk Assessment (TRA)
- Confirmation of Payee
- Payment Validations
- Contingency Reimbursement Model (CRM)

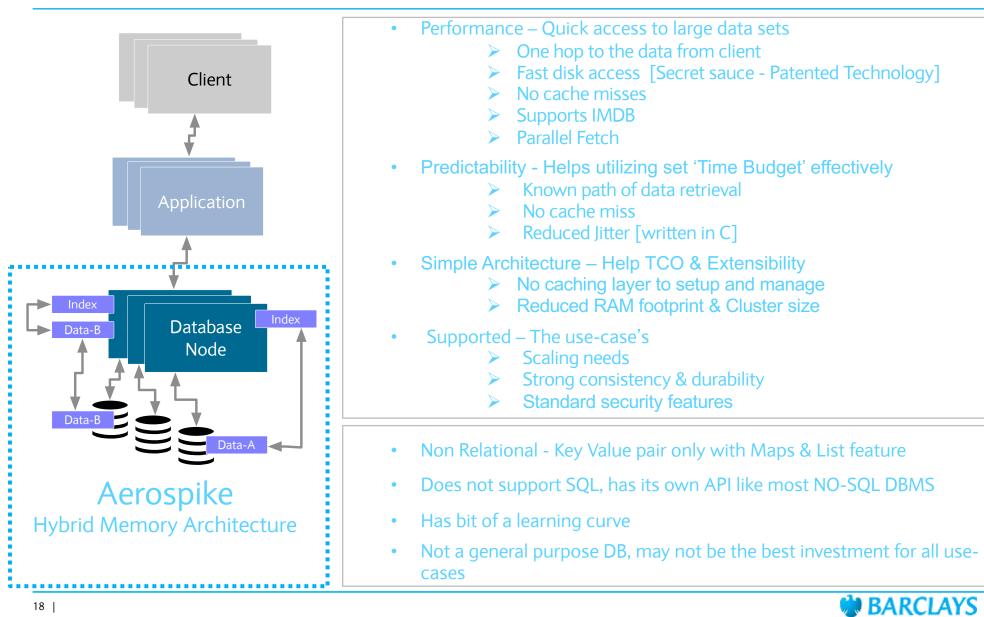
- Impact on the Data handling ?
- Impact on DB technology choice ?

Summary

- Database Matters in Fraud use-case
 - It's extremely important component in the Fraud detection process
- Each DB Technology
 - Has its pros and cons
 - Scale differently to others
 - Each technology may offer unique feature as its core strength but may have weakness in other areas
- No Silver Bullet in selecting a DB
 - 80% use-cases may fall into general purpose where selection may be relatively easy
 - Remaining 10 20% use-case are more complex and difficult

- Functionally similar use-cases
 - May only be functionally alike but
 - Additional use-cases may increase complexity and scale exponentially
 - NFRs may completely warrant changing the solution
 - Operational Environment can change / dictate solution
- What may work
 - Deep understanding of use-case
 - Understanding Data itself
 - Meticulously matching detailed use-case requirements with product features

Barclays Fraud Use-Case



DB Selection – Lessons / Suggestions

- Choosing right DB technology is more difficult (for extreme 10-20% use-cases)
- Aspects frequently misunderstood or often taken lightly
 - Scale & <u>Throughput</u> requirements
 - Sensitivity towards <u>Latency</u>
 - > <u>Jitter</u> in DB operations and its Implication on overall application behaviour
 - Matching <u>CAP theorem</u> aspects with use-case
 - > Data characteristics and usage [Hot /Cold, All Hot, volume, variety etc.]
 - TCO calculation considering all factors and environments
- Key aspects to consider to narrow down choices
 - Pedigree One specific problem the product was originally designed to solve i.e. product's core competence and matching with the use-case
 - Jepsen Report If consistency is important; wonderful resource for DB community
 - References Drawing comparable from industry use-cases

- Operational aspects
 - Knowing your operational environment and constraints imposed by overall landscape, wider architecture, practices, pipelines etc.
 - Security & Monitoring
 - > Patching strategy of DB and its alignment with the UNIX node patching
 - Vendor Support & SLAs
 - Other standard aspect such as Resilience, Cloud/On-Premise, DR, Active Active, Rolling Upgrades, Backup, Accessibility to data/API etc.
- Investment in following pays off
 - PoC/s for right DB selection
 - > Training the Build & Operational teams on product
 - Right Data-Modelling for the application/s
- Things change
 - Better products to support your use-case may be available tomorrow, so keep eyes open and be aware of Vendor-Lock-In

2 of 2





Dheeraj Mudgil Contact Details : dheeraj.mudgil@barclaycard.co.uk LinkedIn : <u>https://uk.linkedin.com/in/dheeraj-mudgil-36ba71a</u>