



Securing and Governing a Multi-Tenant Data Lake within the Financial Industry

Presenters: Bradley Smith
Ian Pillay

Data Lake

Security, governance, multi-tenancy



Who are we?

Bradley Smith

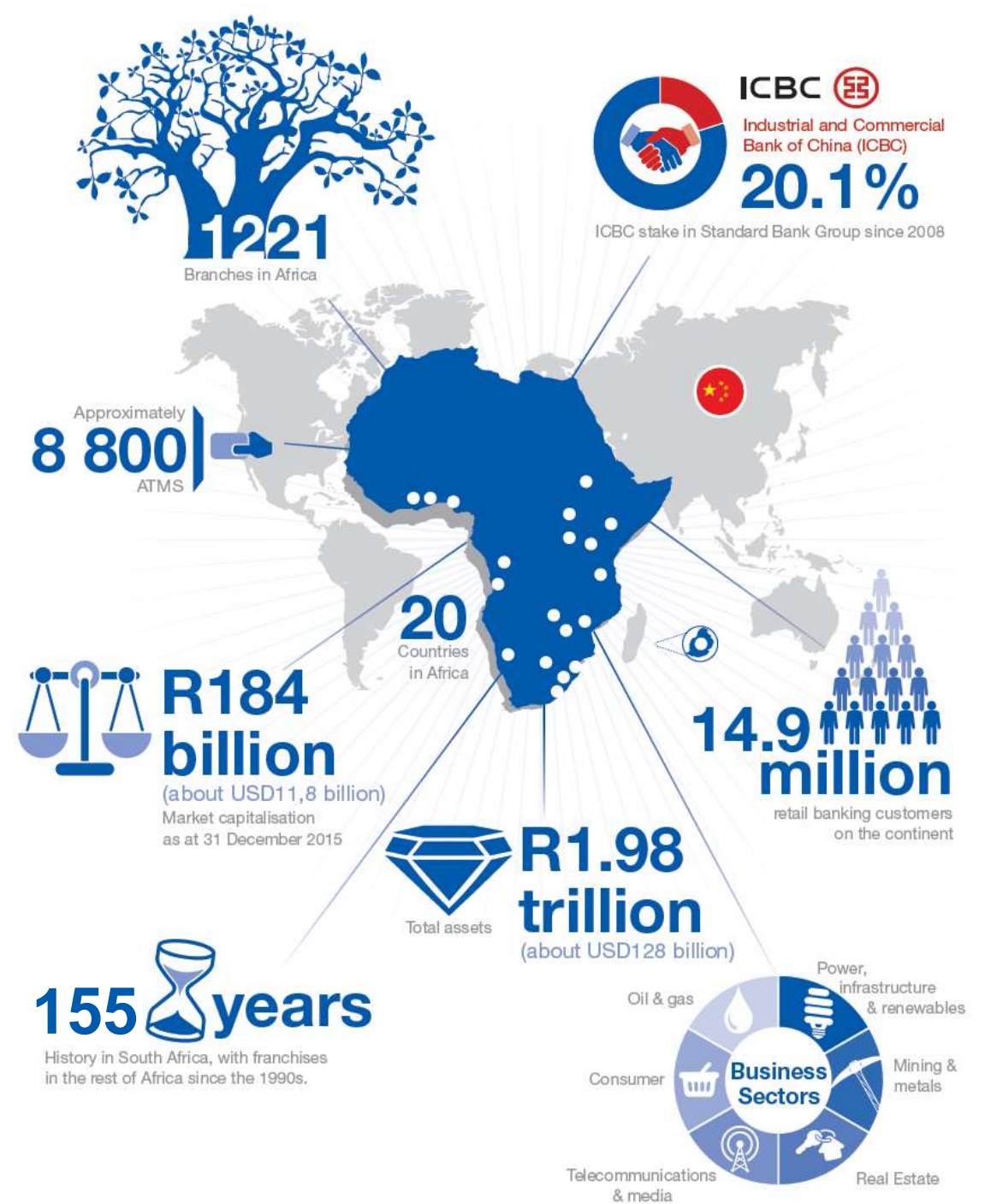
- Hadoop Administrator – Jack of all trades
- Bachelor of Information Science with Honors in Big Data
- Interests: track racing

Ian Pillay

- Hadoop Administrator – Open source connoisseur
- Bachelor of Information & Computer Science
- Interests: game development (Unity 3D, Blender)



Standard Bank



Security

“Transparent security focused on user experience.”

Enterprise security model:

Proactive

Prevent

Predict

Reactive

Detect

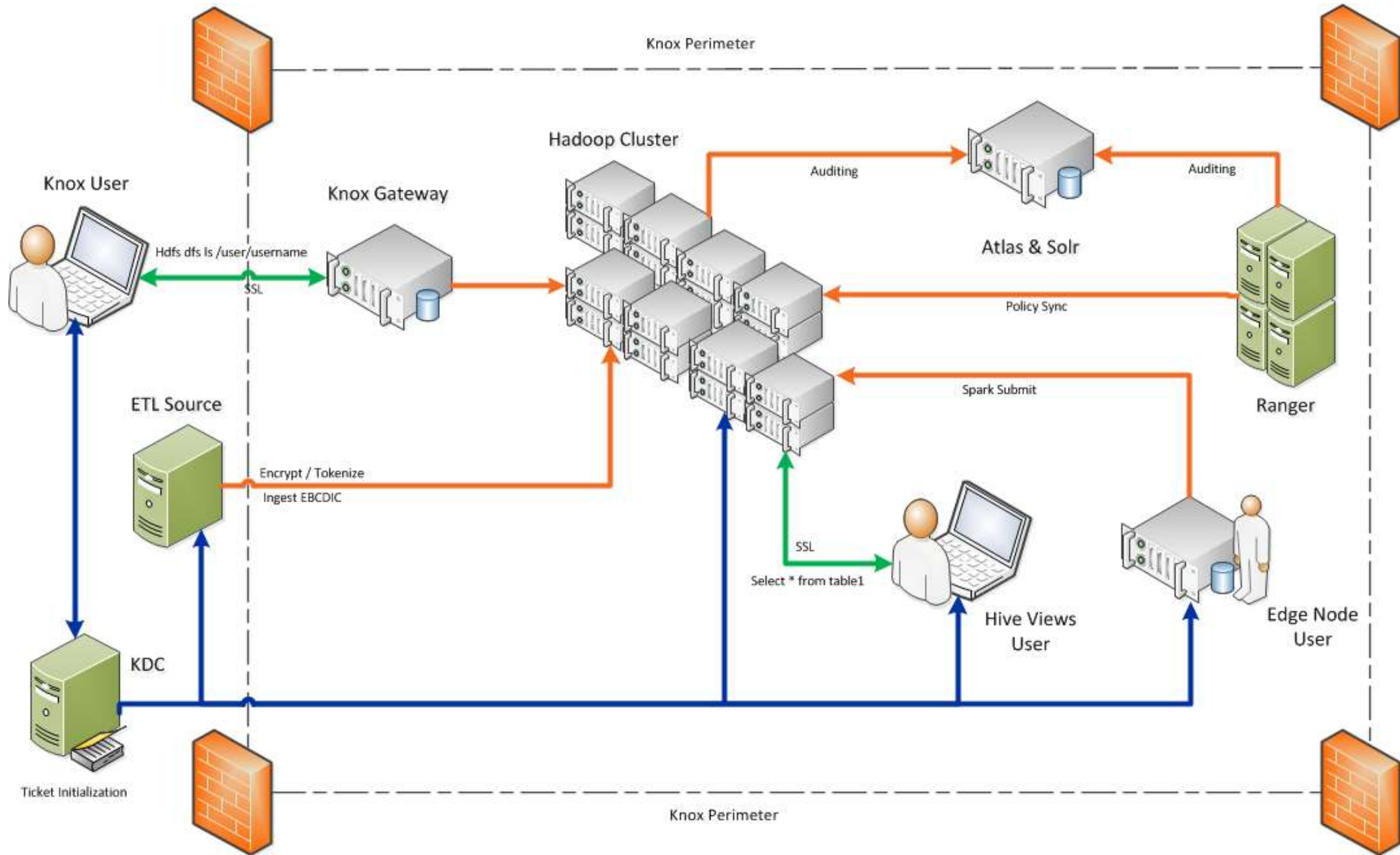
Respond and contain

Defining requirements - compliance

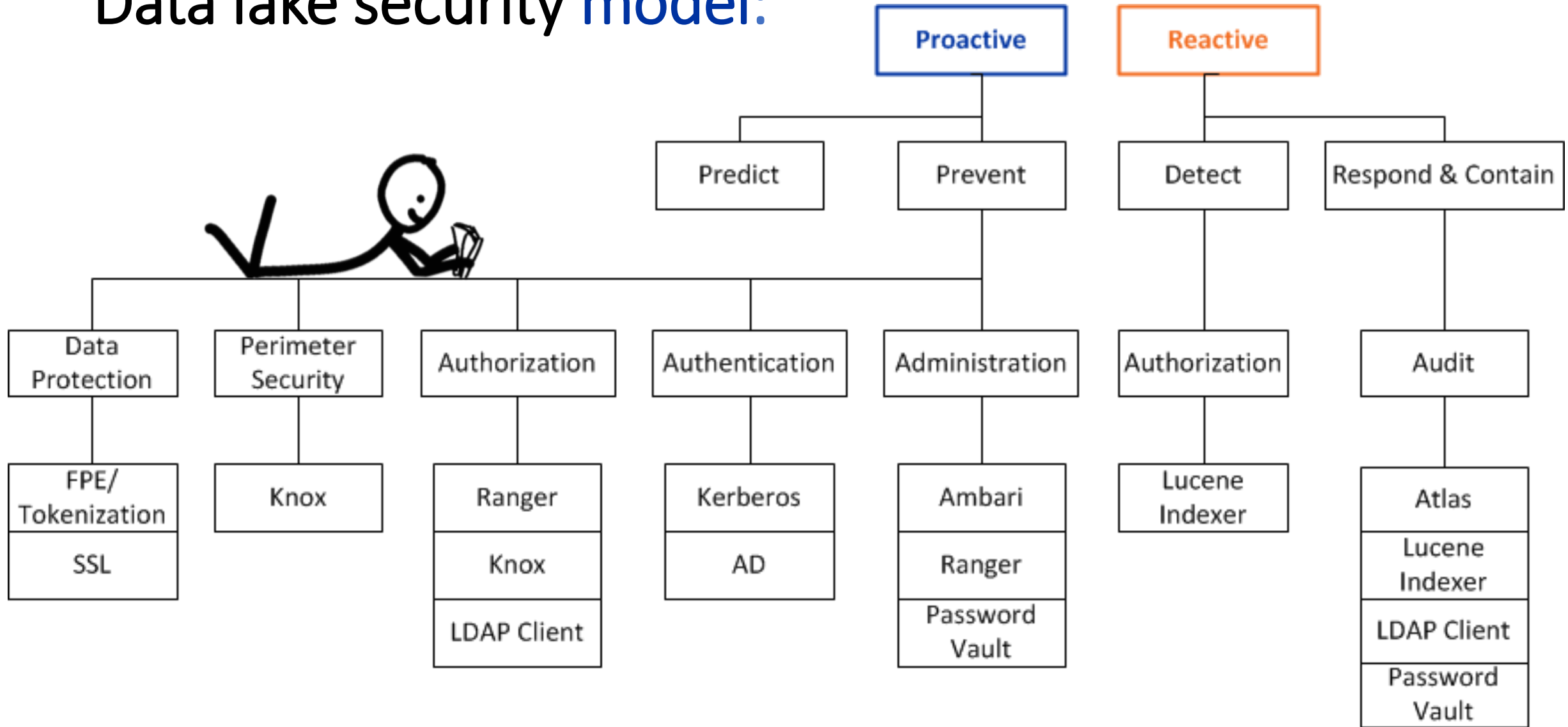
- POPI – Protection of Personal Information Act
- GDPR – General Data Protection Regulation
- PCI DSS – Payment Card Industry Data Security Standard

5 pillars of enterprise security:

Pillar	Intent	Tool
Administration	How do I set policy?	Apache Ambari / Apache Ranger Password vault
Authentication	Who am I?	Kerberos / LDAP
Authorization	What can I do?	Apache Ranger / LDAP client Apache Knox
Audit	What did I do?	Apache Ranger / Lucene indexer LDAP Client
Data Protection	How can I encrypt data?	FPE / Tokenization / SSL



Data lake security model:

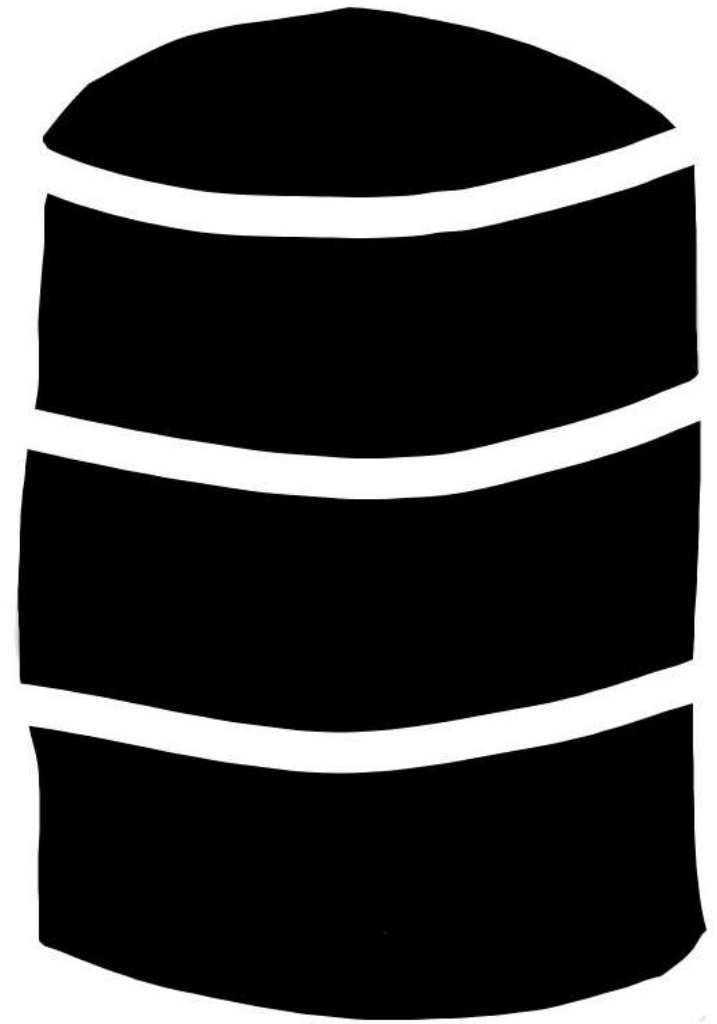
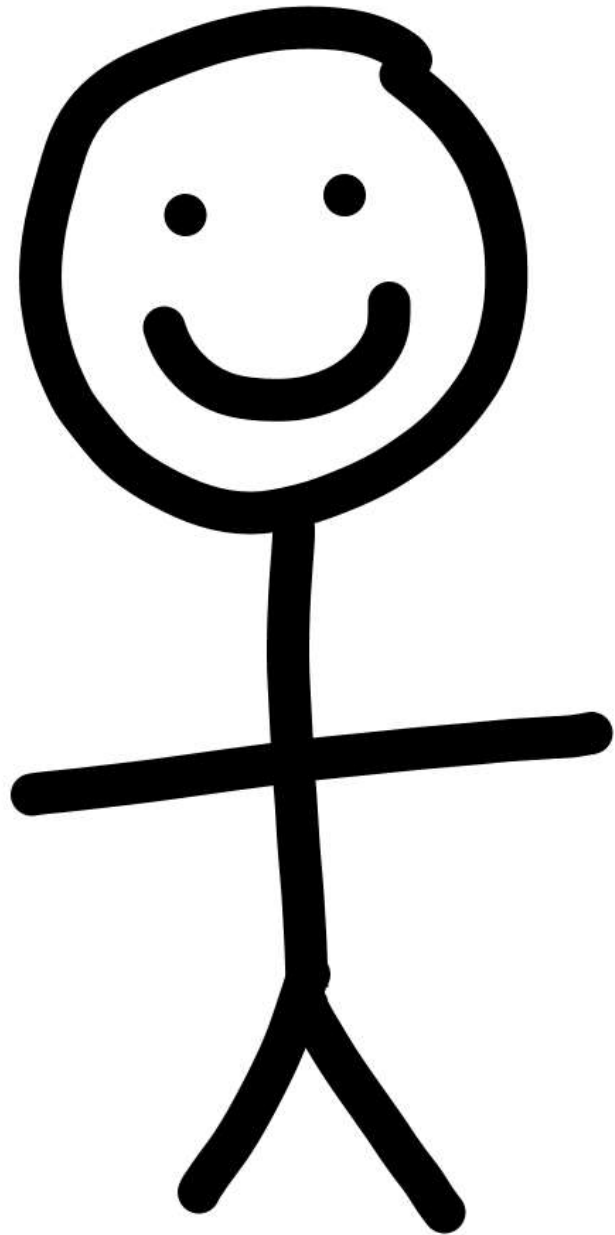


Security challenges:

- Kerberos
- Siloed security teams
- Data encryption & tokenization
- SSL
- Integration conflicts
- Elevated accounts

Governance

“Responsible approach to an autonomous experience.”





How we govern:

- Policies
 - Data
 - Resources
- Compliance
 - POPI – Protection of Personal Information Act
 - GDPR – General Data Protection Regulation
 - PCI DSS – Payment Card Industry Data Security Standard

Chinese Wall



Governance challenges:

- Internal enterprise policies
 - Master Data Management
 - Homegrown metadata solution (OPMD – Operational Metadata)
- Global data ownership
 - Auditing
 - Alerting
 - Reporting

Multi-tenancy

“All for **one**, and one for **all**”

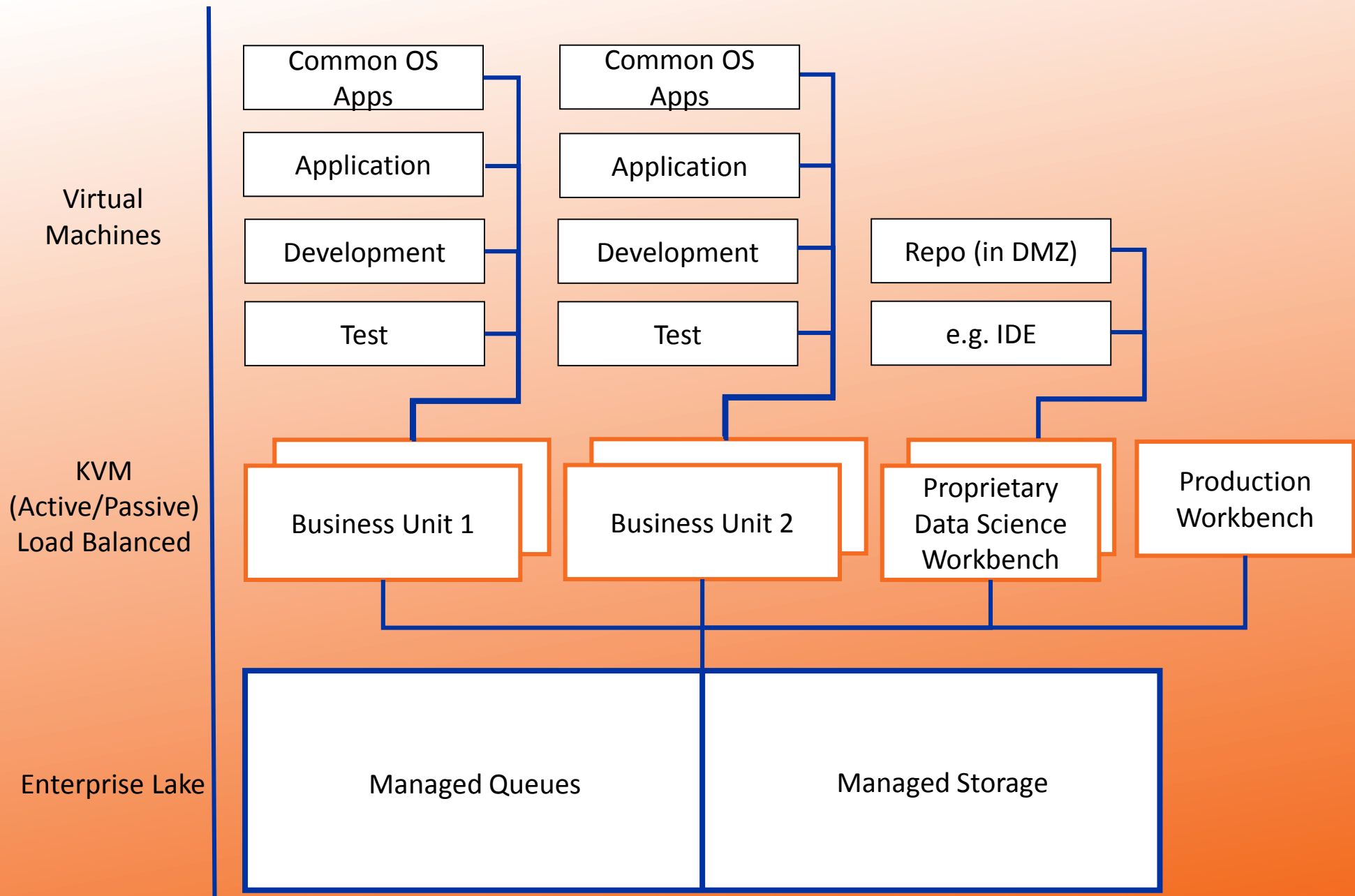
Data
Lake

Analytics
Platform



QUEUES

RESOURCES



Multi-tenancy challenges:

- Cloud computing
- Queue performance
 - Transparent real time performance dashboards
- Data science workbenches
 - Distributed application management
- Legacy users
- API framework
- Web services

Disruptive technologies

“We don’t know what we don’t know.”

Disruptive technologies:

- Quantum computing
- Adversarial machine learning
- Cryptocurrencies
- Blockchain
- Mixed reality
- Cloud computing
- Emerging API threats
- IOT
- Cybersecurity wars
- Autonomous cars

Q & A

You now have 100% elasticity on your Q



Speaker Information

- Bradley Smith
 - Twitter: @Tryxster
 - LinkedIn: <https://www.linkedin.com/in/bradgsmith/>
 - Email Address: bradley.g.smith@outlook.com
- Ian Pillay
 - Twitter: @IanJamie26
 - LinkedIn: <https://www.linkedin.com/in/ian-pillay-development/>
 - Email Address: miraj26@gmail.com