The Cloud Data Lake Journey
Roadblocks and Successes

Ankur Gupta
EVP Global Sales & Marketing
About Bitwise

Experienced Global Player
- 1200+ strong team of elite data specialists
- Global delivery centers in Pune, India with corporate headquarters in Chicago, USA
- 21 years in business and a certified Minority Business Enterprise

Technology Leadership that Challenges the Status Quo

Bitwise Innovation
Our customers are never satisfied, and neither are we. Which is why we constantly challenge conventional wisdom to develop cutting-edge technology solutions that drive key business results.

Bitwise Expertise
Our elite team of data and digital experts partner with our customers to solve their most complex problems by pushing the boundaries of innovation and delivery.

Bitwise Difference
We take a different approach to providing a customer-centric experience built on our passionate people, technology innovations and agile, collaborative processes.
Why Cloud?

**Business Drivers**

**Consumption Based Expenses** > Pay when consumed – and only pay for what consumed

**Speed and Agility** > Rapid environment provisioning will allow speed to market

**Economies of Scale** > Combined customer volumes = lower prices, reduced TCO

**Focus on Customers** > Focus can shift to customer needs, not managing Data Centers

**Capacity** > Scalable ‘as needed’ capacity

**Global Capability** > Faster response times by deploying globally

**Opportunity to Modernize** > Retire, replace or wrap low-value legacy applications
Choosing The Right Platform

Perform Due Diligence for all Evaluation Criteria to Select the Right Platform

<table>
<thead>
<tr>
<th>Architecture</th>
<th>Performance</th>
<th>Operations</th>
</tr>
</thead>
<tbody>
<tr>
<td>UX &amp; UI – Access Point, Mobile Application – Complexity, Size Data – Structured, Unstructured, Complexity Infrastructure – Life Expectancy</td>
<td>Elasticity Scalability Latency Throughput Resource</td>
<td>Business Integrity Tool Integration Deployment</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Financial</th>
<th>Risk</th>
<th>Security &amp; Compliance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operating Cost Business Value</td>
<td>Organizational Business Criticality Technical Resource Contractual Audit</td>
<td>Jurisdiction Regulation Privacy Encryption</td>
</tr>
</tbody>
</table>

Private Cloud

Hybrid

Microsoft Azure

Other Service Providers (Rackspace, Equinix, NetApp, Akamai, etc)
Weighing Solution Options

No ‘One-Size Fits All’ Approach

- Once we select the appropriate platform, what migration approach will we use?
- What are the best options to meet requirements?
- How do different approaches fit together to define the overall migration strategy?

Sustain
- Maintain application with minimum effort and cost

Re-platform
- Migrate high-cost operational applications

Decommission
- Retire application in a controlled manner, preserving essential data

Remediate
- Refactor high-cost applications which are not easily migrated

Consolidate
- Reduce duplication, disparate applications and drive efficiency

Extend / Enhance
- Improve integration and flexibility through service enablement

Migrate
- Translate to a new language or re-use the code in modern environment

Replace
- Replace legacy applications with modern technology stack

Diagram Concept: The Legacy Lifecycle Options (UKInnovate2010)
## Strategy and Assessment

### Assessment Activities
- Investigation and Discovery
- Strategy Workshop
- Gap Analysis
- Solution Sets
- Identify risks
- Recommendations

### Assessment Areas
- Business Needs
- Technical Architecture
- System, Application / Data Architecture
- Tool Evaluation
- Governance

### Workshop Sessions
- Sources
- Data Model and Retention
- Data Processing, Reconciling and Mode
- ETL Tool
- Metadata
- Reporting
- Storage

### Roadmap
#### Infrastructure & Capacity Planning
1. Conduct Assessment & Identify Cloud Infrastructure (PaaS)
2. Finalize the Cloud to be moved based on cost benefits, Scaling, Ease of portability & Performance
3. Capacity Planning and Strategy
4. Discover Sizing, Application, Data Discovery

#### Solution
1. Discover the “Current State” & “Future State” High Level Architecture (Analytics Systems)
2. Build Future State System (Cloud)
   - Hadoop Data Lake Cluster
   - Hadoop Analytics Cluster
   - Hadoop Metadata Cluster
3. Define Apps and Data Migration Strategy

#### Governance & Security
1. Establish Governance to HDL & Hadoop Analytics Cluster
2. Security Enablement
3. Program Role Definitions
4. Resource Allocation
5. Deployment
6. Decommission Strategy
Practical Scenario:
World Leader in Financial Technology Solutions
Business Case for Cloud

Develop an approach to transform to an end state which is more agile, secure and cost effective resulting in an environment that can rapidly respond to changing business demands.

Vision

Speed of Delivery
- Microservices
- APIs/SDKs
- Cloud Delivery Models
- Invest in CI/CD: Automation
- Agile Delivery

User Experience
- User-centric design methodology adoption
- Transform workforce experience
- Cloud collaboration capabilities
- Additional mobile enablement
- Wireless enterprise

Flagship Products / Platforms
- External Facing Portal
- Customer Service Portal
- Enterprise Integration Platform
- Analytics and Data Services

SEAMLESS INTEGRATION
Cloud Data Lake Infrastructure

Reference Architecture – Conceptual

- Foundation of highly durable data storage and streaming of any type of data
- Metadata Index and workflow which helps us categories and govern data stored in data lake
- Search Index and workflow which enables data discovery
- Set of security controls – Governance through technology
- API and end User interface that expose these features to internal and external users

Tools And Technology – Example for AWS

- Tools change per provider

Deployment & Administration
- App Service
- Analytics

Networking
- Compute
- Storage
- Database

Global Infrastructure

Compute
Analytics
Storage
Database

Tools & Technology Example
- AWS
- Reference Architecture
- Conceptual Deployment & Administration
- Compute
- Storage
- Database

Confirmation of highly durable data storage and streaming of any type of data

- Metadata Index and workflow which helps us categories and govern data stored in data lake
- Search Index and workflow which enables data discovery
- Set of security controls – Governance through technology
- API and end User interface that expose these features to internal and external users
Cloud Data Lake System Architecture

### Key Objectives

**Internal/External Users**
Build a cloud based big data strategy that the enterprise as well as external users can leverage to build better analytics and better business decisions.

**Reporting and Analytics**
Service a variety of use cases for reporting and analytics where internal reporting needs, internal analytical needs besides merchant, vendor reporting are taken care of.

**Cost Advantage and Agility**
Be compatible with expanding data and business intelligence needs while leveraging the cost advantages and agility provided by the Cloud Platform.

### “Make or Break” System Components

**Tokenization**
Securing PCI/PII Data

**EDH**
Raw/Processing Layer (for internal enterprise users and public customers)

**Actionable Insights**
Analytics Layer (Querying)
Key Moments / Lessons Learned

Countless roadblocks along the way as well as successes and ‘aha’ moments. Here are some key moments...

### Roadblocks

**Security** – How to secure PCI and PII data. Current functionality used data encryption, but how do we encrypt in Cloud?

**Data Types** – Current data formatting uses decimals, but Cloud provider not compatible with decimal data types, so the initial system kept breaking.

**ETL in Cloud** – ETL in Cloud provider is still maturing. How do we build robust ETL pipelines?

**Reporting Layer** – All business rules written in Microstrategy. Cloud provider offers limited connectivity to Microstrategy. How do we consume the data in Microstrategy?

### Successes

**Tokenization Solution** – We developed a solution to secure PCI/PII data that Tokenizes on ground and pushes to Cloud (only Tokenized data is propagated in Cloud).

**Hadoop Migration** – We were able save significant time moving Hadoop system to Cloud by ‘lift and shift’ of Hadoop Hive tables using Cloud PaaS.

**Assessment** – Complete and accurate assessment was essential for collecting details for all application and database objects and developing migration roadmap.

**No ‘Big Bang’** – Step by step migration approach based on small POCs, then using successful POCs as frameworks to implement.
Strategy & Assessment Report

Cloud Assessment Reports
• System Architecture
• Technical Architecture
• Space Forecasting
• Tool Evaluation

System Architecture Report Outline

1. EXECUTIVE SUMMARY
   A. ENTERPRISE LEVEL OBJECTIVES
   B. ASSESSMENT SCOPE
   C. ASSESSMENT OBSERVATIONS SUMMARY

2. ASSESSMENT APPROACH
   A. PHASE 1: INVESTIGATION & WORKSHOPS
   B. PHASE 2: DISCOVERY & GAPS
   C. PHASE 3: SOLUTION SETS
   D. PHASE 4: RECOMMENDATION & REVIEW

3. ASSESSMENT FINDINGS
   A. SUMMARY OF CURRENT SYSTEMS
   B. CURRENT STATE OVERVIEW

4. DISCOVERY STATISTICS
   A. APPLICATION AND DATABASE INVENTORY

5. CHALLENGES / GAPS / REQUIREMENTS
   A. CHALLENGES
   B. VISION

6. PROPOSED ARCHITECTURE
   A. SOLUTION SETS
   B. DATA ACQUISITION
   C. DATA INTEGRATION
   D. DATA STORAGE AND PROCESSING
   E. DATA ANALYTICS / EXPLORATION LAYER
   F. DATA CONSUMPTION
   G. DATA ARCHITECTURE
   H. STREAMING (REAL-TIME) ANALYTICS
   I. SECURITY & COMPLIANCE
   J. INFORMATION GOVERNANCE

K. BACKUP AND ARCHIVAL
L. MONITORING, LOGGING AND DIAGNOSTICS FOR APPLICATIONS, DATA AND JOBS
M. JOB SCHEDULER

7. MIGRATION
   A. MIGRATION STRATEGY / PLAN
   B. MIGRATION APPROACH / METHODOLGY
   C. FUTURE STATE ORGANIZATION
   D. MIGRATION - DEVELOPMENT STRATEGY
   E. MIGRATION – TESTING STRATEGY
   F. MIGRATION – DEPLOYMENT STRATEGY
   G. MIGRATION PLAN ROADMAP & MILESTONE
   H. APPLICATION & DATA LANDSCAPE IN THE CLOUD- END-TO-END
And the Journey continues....

Thank You

Inquiries?
Call us at 847.969.1500
Email sales@bitwiseglobal.com

www.bitwiseglobal.com