

A Thought Leader Whitepaper

The relationship between Data Transformation and Banking Innovation



Financial Markets & Banking Thought Leaders: Authors & Contributors



Gregory Robinson
LEAD Business & Data Analyst
VP of Financial Services Sector
LEADing Practice Owner & Advisory Board Member



Prof. Mark von Rosing
LEAD Data Engineer
VP of Enterprise Engineering Standards
LEADing Practice Owner & Founding Member



Ken Teske
LEAD Security Architect
LEAD Data Architect
VP of Americas
LEADing Practice Owner & Co-Chairman

Who we are and what we try to do with this Thought Leader Paper

Computacenter and LEADing Practice offer in combination a unique approach to the development and implementation of a Data Lake. In this thought leader whitepaper, we discuss banking's need to rethink their data and information concepts. We discuss the value of learning from banking leaders/outperformers and experiences brought in from other industries. We provide actionable suggestions that can be considered in the digital transformation journey of banks. This whitepaper inspires and supports you in your journey. It conveys how you can build on other outperforming organizations' successes while developing your organization's Data concept/solutions.

The Banking Industry is changing

Even less than a decade ago, a bank needed to have physical offices across the country to service clients. Clients were required to fill out forms and provide identity proof to open accounts. After ATMs arrived, at least a part of banking transactions could be conducted without going into a branch. Online banking further eased banking, but there was still a physical brick and mortar bank in the background that processed and kept track of its business.

Key Take Away

- Explore your data potential
- Learn from the outperformers
- Discover how to use Data as a key component in your Digital Banking Transformation
- Find a way forward of how to store, use and open your data
- Accelerated digital adaptation

Then Fintech companies came along, changing the way money was handled. From online, cashless transactions, they are now venturing into social media, mobile banking and analytics. These forces highlight that traditional banks must seek out new revenue sources driven by customer data monetization. This digitization of banking may seem like a daunting proposition in an era of increased data privacy regulations, but it is essential to sustainable profitability. Data lakes thus become pivotal to banking. Why? Market research reveals that the digital banking

transformation has entered into a second evolutionary state where data usage, regardless of its structure, is driven by banking innovation and transformation across all data types. Diverse data enables open banking platforms, artificial intelligence solutions, Robotic Process Automation (RPA) and interlinks to the Internet of Things (IoT), as figure 1 highlights. Customers gravitate towards these banks because customers are central, not peripheral, in such a dispersed business environment. Outperforming banks exploit this trend by revisiting their data lake/data strategy to discover more significant insights into their markets.

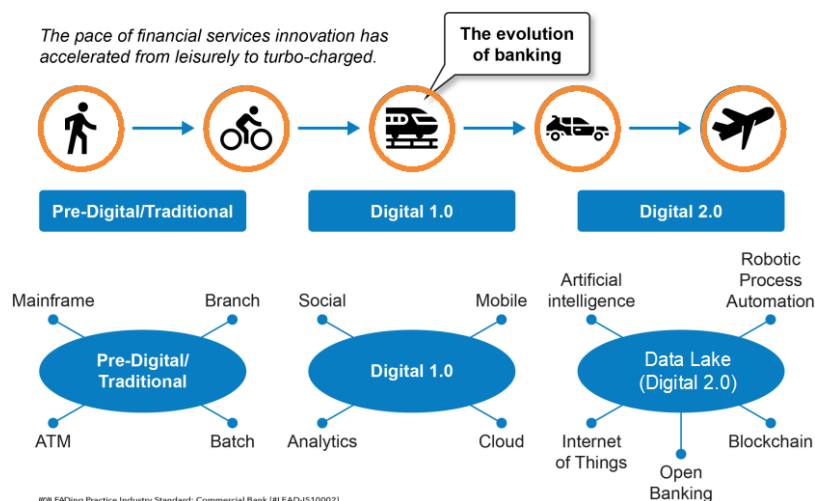


Figure 1: The data-driven digitization of Banking

Banks are swimming in data, and a Data Lake is the answer

Financial institutions generally offer services that cover commercial banking, private banking, personal services, and investment banking. Though data warehouses cater to these service portfolios, the business intelligence was somewhere lacking. The existing solutions take too long to offer insightful results out of the collected data. With data volumes increasing each day, challenges grow and grow. Banks are data-rich but information-poor, which represents the real challenge. The most common challenges include that the data is too scattered. It is impossible to analyze it by customer segment, banking product across different divisions, units, regions and even countries. It takes an enormous number of person-hours to maintain data quality in each of the mentioned areas. It takes considerable effort to prepare data, and thus it is not possible to handle business needs rapidly enough. There lacks real clarity on who is responsible for the different kinds of data, even though that data is the key for banks to generate insights that will enable better-designed products, services, and strategies to win and retain customers. These complications are fuelled by masses of data available in multiple structured, unstructured, semi-structured formats like text, images, videos, and copious other forms. Traditional databases are not equipped to handle such a large variety

of data in the volumes that flow into an organization from access to these valuable external sources. To derive value out of it, banks need to rethink data. Corraling data and amalgamating data into a data lake resolves gaps and inconsistencies by flowing all its content and forms into the data lake. The information can then be extracted from this universal resource and exploited.

Learning from the LEADers

The Outperformer Study from LEADing Practice identified those banks that successfully generate business value from their data do outperform their peers. The outperformers are 2.2 times more likely to transform themselves using data than the underperformers. The study also revealed that banks that implemented a Data Lake outperformed similar banks by about 10% in organic revenue

growth, as shown in figure 2. These leaders achieved new analytics types like machine learning over new sources like log files, data from click-streams, social media, and data from Internet-connected devices stored in the data lake. This success helped them in all data management areas, from capturing, aggregating, analysing, and disseminating to applying data. The leaders were able to identify and act upon opportunities for business growth faster by attracting and retaining customers, boosting productivity, proactively and making informed decisions.

Whether a Fintech company or a financial institution, their common denominator is that they produce data—and lots of it. Due to the Internet’s boom, scalability and massive data infrastructures have allowed companies to prosper beyond greater heights. However, in most cases, data management is unable to keep up with this fast-paced expansion. Traditional banking is becoming customer-centric to create new revenue sources, thus stay relevant. Many questions now must be answered: what customers expect from their bank and which channel do they prefer, online or personal consulting? The digitization of customer data in the banking industry into a data lake solution to determine the answers is critical.

The Data Lake Solution for Banking

The Outperformer Study from LEADing Practice pinpointed a Data Lake as an ideal solution to be deployed in the cloud from capturing, aggregating, analysing, and disseminating to using the data. Why? The cloud provides performance, scalability, reliability, availability, a diverse set of analytic engines, and a massive economy of scale. Meanwhile, 39% of organizations consider cloud as their primary deployment for analytics, 41% for data warehouses, and 43% for Big Data. The top reasons organizations perceived the cloud as an advantage for Data Lakes are better security, faster deployment, better availability, more frequent feature/functionality updates, more elasticity, enhanced geographic coverage, and defined utilization costs. Cloud service providers like AWS, Azure, and SAP Cloud Private program provide a centralized, curated, and secure repository as a service that stores all data, both in its original form and preprepared for analysis. To illustrate, we can refer to AWS



Figure 2: Outperforming with Data

Cloud Data Lake market leadership, which includes the ability to interlink cloud and data lake solutions to enable, as figure 3 illustrates:

- Loading data from diverse sources
- Monitoring those data flows
- Setting up partitions
- Turning on encryption and managing keys
- Defining transformation jobs and monitoring their operation
- Re-organizing data into a columnar format
- Configuring access control settings
- Deduplicating redundant data
- Matching linked records
- Granting access to data sets
- Auditing access over time

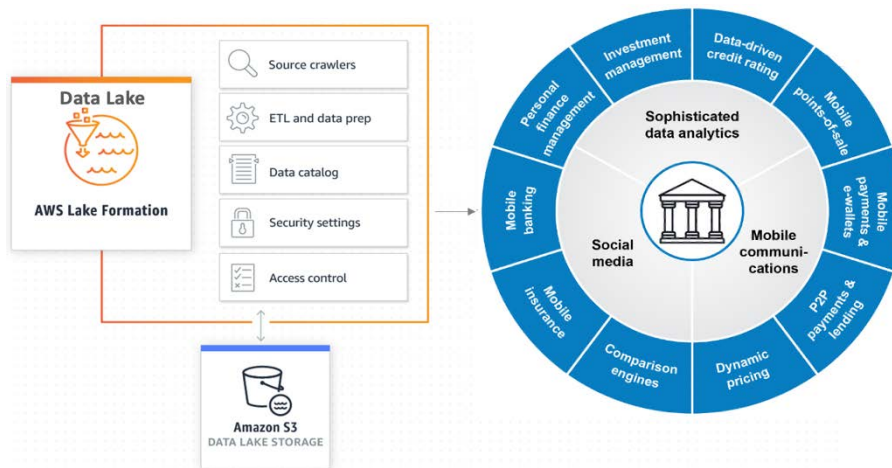


Figure 3: AWS Cloud Data Lake market leadership

The data lake

collects and catalogues these structured and unstructured data from various database and objects storage from defining data sources, data access and security policies.

Key Findings and Benefits

A data lake enables an outperforming organization to capture raw data, determine specific data types, sources, and analytics types to derive greater profitability. Data lakes provide easy, flexible ways of generating answers to the most relevant strategic questions. Our findings are based on the lessons learned from the leaders achieving data transformation in Banking. Therefore, we present the following detailed recommendations that, like with a cooking recipe, can be followed step by step to ensure a valuable output in the shortest time possible.

Where to Start - Opportunity Analysis

We have considered how a bank's success and lifespan is primarily based upon how they manage their data to predict what's next. Cloud Computing, Data Lakes, Serverless Applications and Databases have allowed organizations to breach new frontiers in the tech and finance space. Nevertheless, the main questions many banks ask themselves are how to exploit a data lake concept. We recommend approaches such as the Data Lake Opportunity Analysis in figure 4. The suggested areas of examination would provide instruments that enable rapid exploration, evaluation, and insight into any aspect of your organization's potential around a Data Lake. These instruments would include identifying the maturity of your data strategy, roadmaps, related data portfolio, costs associated, the IT systems involved, operations, capabilities, functions, processes, services. It can also connect to the most relevant innovation and transformation subjects. It analyses identifying Data Lake business, information and technology needs, wants, issues, problems, and pain points to determine the existing maturity, and any potential issues at hand.

The suggested way to identify the Data Lake Opportunity Analysis' needed information is through various stakeholders involved in a workshop. The workshop is best facilitated by external experts that are not engaged in your daily challenges. These experts can help explore how your bank can use a data lake from an independent perspective as a part of your transformation journey. It is crucial that such an opportunity assessment workshop also helps you with many other valuable and needed information. Among them would be developing suggested enhancement areas, based on comparing the Data Lake Opportunity Analysis results to Data Lake best practices, industry practices or leading practices from the outperformers. You should not re-invent the wheel. Such a move would be too time-consuming, expensive with a failure rate of over 70%. The best way forward is to ensure that the partner you choose should have reference content with documented best, industry and outperforming practices. What you need are repeatable patterns that can be reused, adapted to your needs and create specific solutions for your organization.

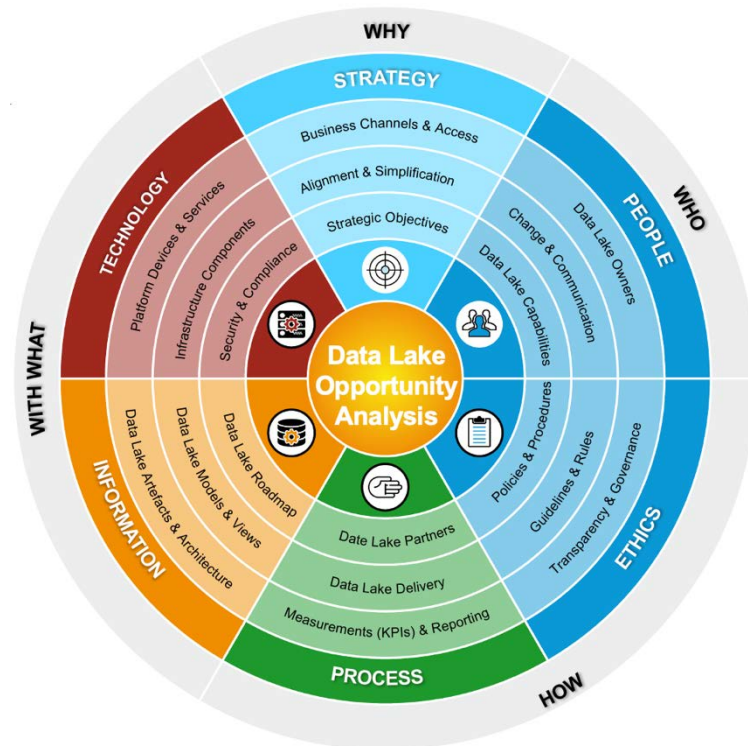


Figure 4: Data Lake Opportunity Analysis

Because you need to ensure that you can learn from other organizations and leaders across various industries to create maximum value in the shortest time possible, check that the reference content can be used across the entire Data Lifecycle. That lifecycle encompasses the Understanding of your unique situation, the critical aspects of Data Innovation and Data Transformation, and how you can use it to enable your Continuous Data Improvement. Figure 5 gives an overview of such reference content. We have identified and captured how most organisations would need to undertake such a journey through the reference content.

	Degree of Change: High	Degree of Change: Medium/High	Degree of Change: Low/Medium	Fit Gap & Root Cause Analysis (Needs & Wants) & Performance & Value Expectations	
UNDERSTAND	DATA LAKE INNOVATION		DATA LAKE TRANSFORMATION	CONTINUOUS Data Lake IMPROVEMENT	
Analyse & Research	Data Lake Design	Develop Data Lake	Execute Data Lake	Data Lake Management	
<ul style="list-style-type: none"> Understand the Data Lake potential, set ambition levels, and prioritize solution Data Lake Opportunity Analysis (S) Data Strategy (RC) Data Transformation Roadmap (S) Data Portfolio (S) Banking Data Capability Map (RC) Banking Data Requirement Management (RC) 	<ul style="list-style-type: none"> Design and validate your Data initiatives to confirm costs and benefits Data Lake Idea/Design Thinking Data Lake Concept (RC) Banking Data Process & Flows (RC) Data Modelling (S) Data Lake Use Cases (RC) Platform Design (RC) Data Lake Architecture (S) 	<ul style="list-style-type: none"> Develop and build Data Lake solutions into your banking business Data Lake Orchestration (S) Relationship between Data Lake and Banking Use Case Data Lake & Service Model (RC) Data Security Concept (RC) Banking Data Rules (RC) Identity & Access Mgt (RC) Multi/Hybrid Cloud Integration (RC) 	<ul style="list-style-type: none"> Undertake rapid prototyping and testing of the Data Lake solution; evaluate results Data Lake Composition concept (S) Banking Data Tagging (RC) Data Lake & Change Management (RC) Data Lake Meta Model (RC) Workplace Integration (RC) Data Lake Capability Development (RC) Data Visualization (S) 	<ul style="list-style-type: none"> Scale and roll out proven Data Lake solutions, addressing technical and org. barriers Data Lake Governance Framework (S) Data Lake & Change Management (RC) Data Lake Management Roadmap (S) Data Lake Management Portfolio (S) Data Maturity Dev. (RC) Data Lake Management Meta Model (RC) 	<ul style="list-style-type: none"> Optimize and manage Data Lake-enabled solutions, updates; monitor results, and adapt as needed Data Lake Opportunity Assessment (S) Data Lake & Change Management (RC) Data Lake Req. Management (RC) Data Lake Transformation Benchmarking (S) Data Lake Meta Model (RC)

Leading Practice Data Lake Reference Content | RELEASE 10031411

Figure 5: Data Lake Reference Content

What to do – a successful pattern emerges

We have addressed the immense failure-rate in data transformation and digital banking innovation initiatives. Our latest Outperformer study survey of Banking C-Level executives and directors found that digital transformation (DT) issues and challenges are their #1 concern in 2021. Over 70% of all banking data transformation initiatives do not reach their goals or create anything close to the expected value. In our quest to identify the reason for the high failure rate, we determined that while many Banks are right in what they do (being a bank), they lack expertise in innovation and transformation projects. We have captured and implemented the successful patterns for this undertaking through the entire lifecycle. The Data Lake Framework is a non-linear, iterative agile process involving four phases; **Understand, Data Lake Innovation, Data Lake Transformation and Continuous Improvement**. It can help guide the various stakeholders and practitioners to work with the Data Lake Framework phases and sprints found in the blocks illustrated in figure 6, the Data Lake Design Cycle. The relating Artefacts enable the needed Innovation and Transformation at the same time.

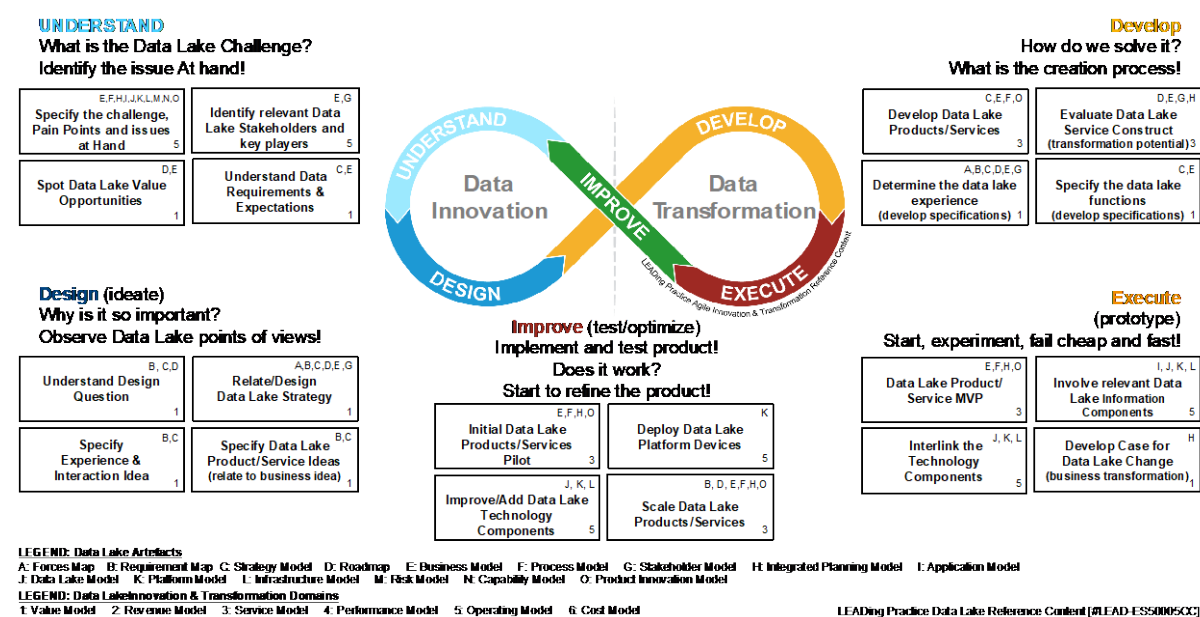


Figure 6: The Data Lake Design Cycle

Conclusion - The benefits of Learning from others

There are many benefits of learning from others, both in the banking industry's best practices and from standard, relevant industry practices. We also need to add the *leading* practices of the outperformers. Though your organization has the knowledge and hard work that has delivered your desired outputs to date, over 70% of digital transformations fail. Most organizations we work with or have analysed had believed that their solution will succeed and are so different that they don't need good reference content. Regrettably, the outcomes are failure arising from the worst practices or even anti-patterns. Most organizations don't consider the reference content as a steppingstone for learning from others. Tailoring and applying such reference content delivers the bulk of what your organization needs, allowing you to focus on your unique situation, reducing risks, costs and time. Our research has revealed up to a 40% time reduction. A project can save about 40-50% of the overall project budget by exploiting the reference content. Such factors alone evidence that most organizations should be reconsidering their 're-invent everything yourself strategy'. Our evidence shows that such a 'do it yourself' and 'we are the best' mindset needlessly and dangerously brings in complexity and destroys more value than most organizations could imagine. Reference content is a starting point and therefore adds to increased innovation and transformation loops from the outset. It works. Figure 7 from our latest outperformer study reveals the difference in using reference content. It

not only saved time and reduced cost but also increased the overall value output by about 120%.

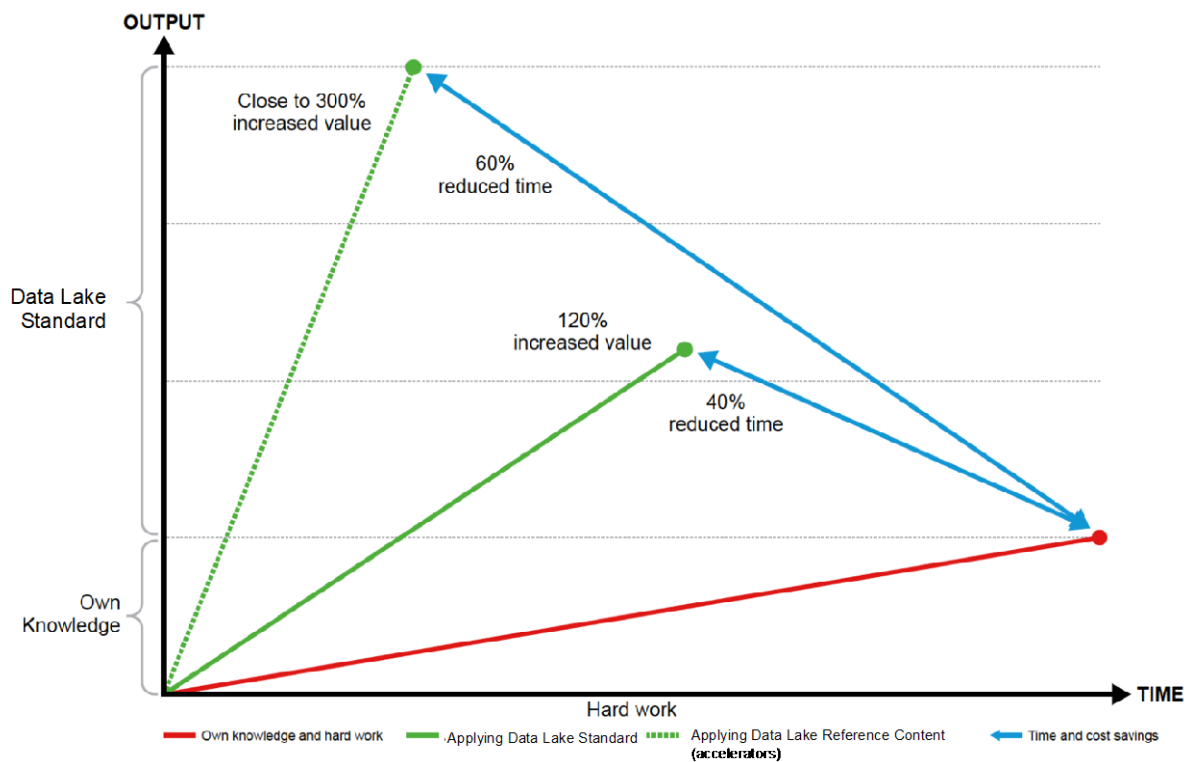


Figure 7: How Reference Content works

Questions & Comments

If you have any questions, comments or wish to benchmark yourself against the leaders and outperformers who have taken this journey. Or just wish to discuss various use cases, please feel free to contact Christian & Ken: Banking_Outperformers@LEADingPractice.com