Mesosphere Guide to

DATA-RICH APPS IN FINANCIAL SERVICES

MESOSPHERE

Summary

Modern data-driven application architectures are getting broad adoption in leading financial services firms. Technologies first developed by webscale companies for personalized experiences are now used by financial services to enable everything from customer insights to cybersecurity and fraud detection. At the heart of this shift are microservices and cloud-native data services that bring new flexibility for organizations to get to actionable insights from enterprise data. Business leaders are recognizing they can capitalize on enterprise data, with Gartner finding that 54% of financial services CEOs now quantify the value of data as a financial asset. For a leading global bank, the key to competitive advantage is a strategic innovation platform enabling faster time to market of new services, quick adoption of new technologies, and complete control of their hybrid cloud infrastructure footprint.

A Data-driven Arms Race

Digital transformation is shaping the financial services industry along five main factors spanning market-driven, operational, and regulatory areas.

1. **Automation and commoditization of services** - From trading to tax preparation, the industry has increasingly moved online. In recent years, digital platforms that provide automated and algorithm-driven financial planning services with little to no human supervision (robo advisors) have become increasingly popular. It has become easier for consumers to self-manage their own banking and finance transactions. The result is that personal connections and customer loyalty are becoming obsolete unless the organization can effectively capture, analyze, and capitalize on digital interactions.

2. **Increased scale of transactions -** With the prevalence of mobile devices, the convenience and low cost of making financial transactions has led to ever-increasing activity. This means more data that banks need to collect and analyze to manage risk and market trends.

3. **Increased cybersecurity risks** - Tactics and tools exclusive to state actors a few years ago are now well within the capabilities of the smallest criminal organizations. Businesses need to develop real-time and data-driven capabilities (e.g., anomaly detection) to prevent money laundering and intrusions by malicious actors.

4. **Customer-centric engagement** - Thanks in part to social media and webscale online retailers, users today expect personalized experiences. Yet for many businesses, data silos between business units, functions, or geographies has meant that the organization does not have a comprehensive view of the customer. To provide competitive services to customers, organizations need to find actionable insights using data across all silos.

5. **New regulations** - Governments recognize the value of data and have compliance requirements for data privacy, data mobility, and market liquidity. To meet these regulatory objectives, organizations need to manage apps and data securely, and run simulations on increasingly complex data sets.

Limitations of Existing Systems

Traditional infrastructure, monolithic big data platforms, and application development practices are not suitable for meeting today's financial services challenges.

Virtual-machine (VM) centric infrastructure suited to an era of monolithic applications cannot meet the modern financial institution's need to manage users and data at scale. Whether the infrastructure is cloud or datacenter-based, the challenge for modern infrastructure is how best to aggregate multiple servers into highly scalable distributed systems. This means cloud-native architectures.

Further up the stack, traditional monolithic big data systems (e.g., those centered around Hadoop) suffer from two main problems. First, they require dedicated infrastructure clusters, which can mean months to provision and configure new hardware before making a data service available. Second, they are not able to easily integrate new big data technologies. In recent years there has been a Cambrian explosion of big data technologies, each optimized for addressing specific challenges (e.g., data ingestion, processing and persistence). As new technologies become mainstream (e.g., Apache Kafka, Apache Cassandra, Apache Spark, Apache Flink and TensorFlow), organizations centered on monolithic big data platforms will increasingly struggle to modernize their infrastructure. This impacts not only the technology but also the talent pool of data engineers they will be able to hire and retain.

On the application level, bringing new end-user services to market quickly is critical to staying competitive. The goal is to empower developers to ship frequently with a DevOps model emphasizing communication and collaboration with product management and other functional areas. Continuous integration tools (e.g., Jenkins & Gitlab) are critical to this approach. These tools are also distributed systems, often running on dedicated infrastructure for each development team. The result can often be high costs for barely utilized infrastructure, or long wait times as developers wait for builds to complete.

Modern Financial Services Architecture Requirements

Successfully navigating these shifts in the financial services industry requires three related capabilities, starting with a solid but highly flexible foundation, and moving away from monolithic stacks. The first is agile infrastructure operations for the broadest set of workloads with hybrid cloud portability. Second is developer agility enabled by an elastic continuous integration and continuous delivery (CI/CD) toolchain to iterate and ship new products and services quickly. Finally is the capability to adopt new data technologies quickly across the entire organization. Together, these capabilities unleash the innovation of various teams and help them to bring new ideas to market quickly.

Operational agility requirements for today's financial services organizations fall into four main areas. First, hybrid cloud portability, and the ability to provision services in minutes and later destroy those services when no longer needed. Second is scale and performance — for example, data processing for risk management requires the ability to store data for long periods of time and at scale (e.g., petabytes of data spanning multiple months), data encryption can require significant processing power, and real-time analysis requires streaming data processing. Third is broad workload coverage so that the infrastructure can run traditional monolithic apps as well as modern microservices, plus data services and machine-learning algorithms that benefit from hardware acceleration (e.g., GPU, FPGA). Lastly, the architecture needs to support an open ecosystem, so that organizations can make investments in various software technologies based on how they're being used by the business, without having to project future growth or commit years ahead.

Developers are charged with creating new services quickly and bringing them to market. In doing so, they demand a cloud-like experience, lest they head straight to cloud services like AWS and build applications locked into the cloud provider's infrastructure. The goal is to give developers what they need and never make them wait. This requires providing developers with self-service access to a broad set of platform services from container orchestration to data services supported by an elastic CI/CD build pipeline.

The ability for financial services organizations to rapidly transmit, capture and analyze data at scale is critical to staying competitive and managing risk. Real-time data processing enables firms to position new products and services to customers just in time. It also means greater responsiveness in identifying new threats with streaming analytics and machine learning. Additionally, financial services need to facilitate the development of enterprise-wide data reservoirs, with multi-tenancy, so they can isolate data as needed or enable cross-BU or cross-functional data analysis where appropriate. Company-wide data visibility provides opportunities to create new services and better helps the organization understand its posture in complying with regulations.

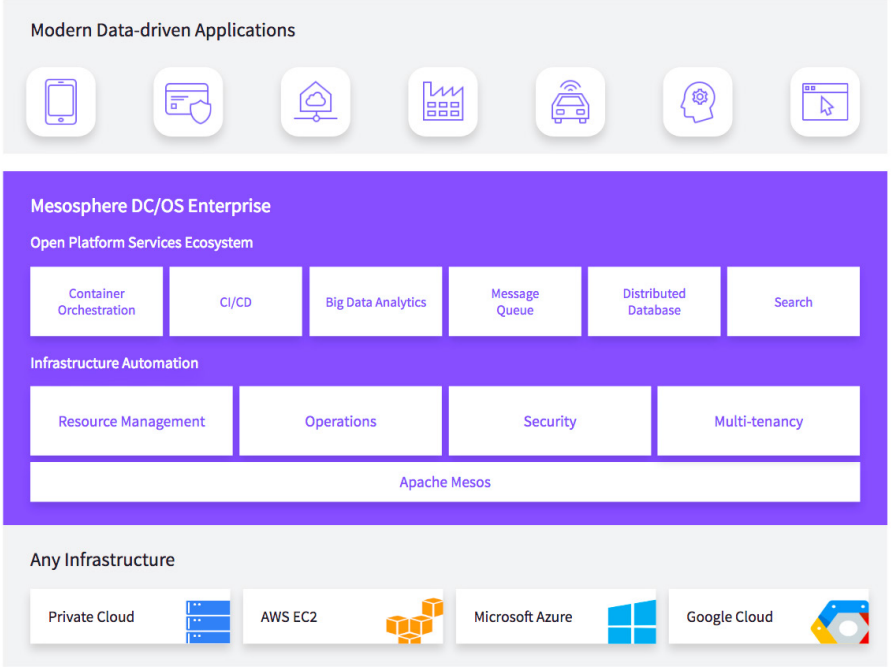
Mesosphere DC/OS Approach

Five of the top ten banks by assets in North America use Mesosphere technologies to bring agility to operations, application development, and data infrastructure. Mesosphere DC/OS enables an open partner ecosystem on a resilient and secure platform that runs on any infrastructure or cloud. This means faster time to value of new services, automated operations, and savings on cloud or infrastructure costs.

Mesosphere DC/OS serves as a platform for container orchestration and data services orchestration, providing a universal toolkit enabling financial services firms to power scalable, data-rich applications. Running containerized microservices with container orchestration on DC/OS, organizations can quickly and efficiently prototype and build business applications that both meet the needs of internal teams and better serve their customers.

In the data services area, frequently used building blocks include Apache Spark, Apache Kafka, and Apache Cassandra. All these cloud-native data services run elastically on DC/OS. With Apache Spark, a firm can analyze market trends and customer data to determine growth opportunities between different business units, while concurrently predicting and protecting against a large variety of risks and threats. Kafka is commonly used as a message queue for highly scalable and lossless data capture and replication between services. Lastly, Cassandra is a cloud-native distributed database that automatically replicates data across nodes and across datacenters.

Mesosphere DC/OS's application-aware scheduling architecture means that even more data services can be integrated in a straightforward fashion. As of 2017, there are over 100 services available in the DC/OS service catalog.



Financial Services Use Cases

Financial services institutions are more connected to their customers than ever before, generating tremendous datasets requiring new data analysis tools. At the same time, there are an increasing number of security, audit, and regulatory requirements that create additional administrative and management overhead. With a proper plan to handle both the scale of data and accompanying complexity, financial services organizations are turning challenges into valuable business opportunities.

Customer Analysis & Customer 360 for Retail Services

When the financial institution is better able to meet its customers' needs, customers are more likely to consume and utilize services, and everyone benefits. Aggregated data from across different financial segments — everything from consumer banking and card services to investment and wealth management — can be combined into large distributed storage environments such as Apache Hadoop Distributed Filesystem (HDFS). These data can then be analyzed with big data analysis tools (such as Apache Spark) to identify behavior patterns and determine which services best fit customer needs and identify potential growth opportunities for the rest of the business.

For example, analyzing captured consumer spending data from a credit card division against retail banking data can yield insights and opportunities to investment or insurance divisions.

Cyber Security

Large financial institutions have tens of thousands of security and networking devices that each generate metrics and logs-based data that should be analyzed in real time. But often the volume of data generated makes gaining insights or identifying threats challenging. Existing log aggregation systems cannot scale and are simply incapable of handling this volume of data effectively, which results in very short retention periods for alerts and messages — data that can contain valuable insights into existing threats to the institution. Big data platforms running on DC/OS provide a highly scalable mechanism to collect, aggregate, and analyze this data to proactively identify non-signature based threats or patterns using data analysis tools like Apache Spark or exotic custom solutions that leverage TensorFlow with GPU acceleration.

Self Service and Automation

Different teams and different team members typically each have their own workflows and desire to control their own environment. Using the Mesosphere DC/OS platform, an organization can quickly and simply build out a self-service infrastructure so that developers can deploy their own applications and associated instances of a variety of data services.

This approach removes developer roadblocks by allowing users to provision production-ready big data services in a secure and compliant way. Services are deployed according to best practices and have many of the failure recovery functions automated, reducing the administrative overhead typically involved in maintaining such an infrastructure. DC/OS co-locates many of these services on existing capacity on the cluster. Cassandra, Kafka, HDFS, or Elastic clusters can be carved from an existing Mesosphere DC/OS cluster quickly and painlessly, without the hassle or administrative overhead of procuring additional resources. DC/OS also includes many security capabilities that are critical for financial institutions, such as certificate authority, authentication, authorization, auditing with LDAP, and SAML2, in addition to secrets management. DC/OS can then provide true multi-tenancy without compromising performance, security, or compliance. Different business units share resources safely and securely, and once a service is no longer needed release the resources back to the pool for use by other teams or business units.

Application and Service Development

One of the larger costs for any large financial services organization is time and money spent developing new applications and services. The Mesosphere DC/OS platform can serve as the foundation for comprehensive continuous integration/continuous deployment (CI/CD) pipeline via the use of a simple integration between the Jenkins CI automation tool and DC/OS's native container orchestration engine. Running tools like Jenkins on DC/OS provides elastic build bursting capabilities, so developers aren't waiting for builds before apps can be tested. Additionally, rolling out CI to different teams becomes straightforward because development teams can share infrastructure capacity.

Large-Scale Analytics

Mesosphere DC/OS, leveraging Apache Mesos capabilities, is one of the pioneering leaders in multi-resource scheduling. DC/OS is resource-aware; a single cluster with pools of nodes optimized for different purposes — CPUs, GPUs, memory, and disk — can schedule many different types of tasks and jobs using the Dominant Resource Fairness (DRF) algorithm, which efficiently allocates resources across a distributed cluster in a fair way. Utilizing GPU acceleration, users can perform large-scale computations to perform a variety of analytic and machine learning tasks, from real-time analysis to gain business insights from large datasets to training and developing predictive models to better guide business decisions. Adding GPUs to the DC/OS cluster allows users to transparently run GPU-efficient jobs in the same environment and through the same workflows as their more traditional CPU-based analysis jobs.

Major Global Bank Accelerates Time to Value of New Services

A top-ten global bank needed a strategic innovation platform to break existing silos holding back innovation. Staying ahead of the competition meant it needed to accelerate time to value of new initiatives and facilitate the adoption of new big data technologies.

The firm chose Mesosphere DC/OS Enterprise to build this internal platform because it met each of the firm's design criteria:

- **Decompose opinionated monolithic stacks** to improve maintainability, security, scalability, efficiency, and enable the bank to bring together the best and latest technologies (e.g., Apache Spark, machine learning, Functions-as-a-Service)
- **Provide self-services to developers and data engineers** to accelerate time to value, and extend the platform with the firm's internal marketplace of services (which all run on DC/OS)
- **Complete control of hybrid infrastructure footprint** with datacenter-to-cloud portability, and position the firm to leverage multiple public cloud infrastructure providers based on cost and performance
- **Pool infrastructure operations** both from a compute capacity and staffing perspective
- **Platform must be mature** enough to be resilient and production-proven at scale

Adoption of Mesosphere DC/OS was driven from multiple organizations. The internal real-time data analytics team was interested in DC/OS's data services orchestration capabilities. Application development teams were interested in DC/OS for its ability to provide an elastic build farm to enable CI/CD. Enterprise architecture leads recognized DC/OS's potential for breaking silos between functions and business units with an open, consistent, and elastic compute platform. Below are details of how Mesosphere DC/OS is being used by this leading financial institution.

Platform Cloud for Developers and Data Engineers

Highly strategic to this financial firm is enabling developers and data engineers to adopt the latest platform services to build their applications. Running Jenkins on Mesosphere DC/OS, developers are able to build, test, and integrate microservices and quickly deploy them to production with an elastic CI/CD pipeline. DC/OS also provides production-proven container orchestration to help these applications scale in production.

Enterprise-wide Data Reservoir

Breaking down silos was critical for this leading bank to identify and create new business opportunities. With Mesosphere DC/OS, data services like Apache Spark, Apache Kafka, Apache Cassandra, and Apache HDFS run elastically on shared infrastructure in a multitenant model. The bank can maintain data access controls but also facilitate cross-pollination of insights across BU's where needed.

- **Customer 360** - Internal marketing analytics platform captures multi-channel data, paid search, and website traffic information. The service ingests marketing data from a variety of sources and builds an integrated customer profile that is accessible firm-wide. Analytics in this area may include graph databases, and relationship mapping.

- **Advanced prospecting analytics & planning** - Fast reporting of potential customers based on a wide range of criteria. Internal services identify high net worth potential customers in a market. Data taken from a corporate database is transferred via Kafka towards Spark and/or Elastic for analytics.

- **Visibility and control of prospecting execution** - Web-based portal used by the private banking business unit focused on high net-worth individuals. The system expands the prospecting pool by aggregating data from multiple sources across the firm, including client, prospect, and lead information.

Security

Banking infrastructure security is more critical than ever as digital financial transactions continue to grow. The bank is developing advanced security and anomaly detection capabilities running on Mesosphere DC/OS to identify threats in real-time. All cyber device-captured data in this bank are streamed into logs and then to the internal platform. This platform processes these logs with Apache Spark, Apache Kafka, and other tools using HDFS storage.

Firmwide Risk Analytics

An internally developed platform is used to perform market risk analytics and compliance analytics. Key aspects of this platform include batch processing accelerated by GPU (Mesosphere DC/OS is able to prioritize these workloads to run on GPUs in the resource pool). Additional analytics include Value at Risk (VaR) simulations and anomaly detection for anti-money laundering operations. The bank also regularly runs stress test to ensure compliance with regulatory standards (e.g., Basel III).

Working with Mesosphere

Founded by engineering leads at early webscale companies and the co-inventor of the Apache Mesos distributed systems kernel, Mesosphere aims to deliver a “datacenter or cloud as a computer” operating model to businesses. This approach provides automation critical to running and scaling cloud native applications and data services, without having to significantly scale the required infrastructure and operations personnel.

Recognizing the shift towards distributed systems can be new for many organizations, Mesosphere provides software, support, training and professional services working closely with strategic customers to get to desired outcomes. Level and duration of professional services support can vary based on the customer’s experience with Linux, Apache Mesos, distributed systems, and cloud native architectures in general. In some cases, organizations already experts in Apache Mesos or the open source Mesosphere DC/OS distribution look to Mesosphere for mission critical support. Alternatively, customers also choose to have Mesosphere solution architects embedded with internal enterprise architecture and engineering teams to ensure the implementation of distributed systems best practices, in addition to mission-critical support and training.

Conclusion

There is a rapid evolution of the business technology used in financial services towards data-rich applications. The most successful banks and financial services firms will be able to develop new value-added services leveraging real-time data insights, machine learning, and other technologies. They will also be able to bring services to market faster using best practices for application development, shifting towards microservices and continuous integration and continuous delivery pipelines.

Firms will need to decide how to adopt technologies like advanced analytics and machine learning, fast becoming key areas for competitive advantage. A key question is what roles cloud providers play for this highly strategic capability. Relying completely on cloud providers to integrate new technologies on their stack can simplify adoption, but can also lead to lock-in of critical applications and customer data. The increasingly blurry line between technology and financial companies is also a factor — a bank may find itself competing with its cloud provider if they choose to enter its business.

Many firms will choose to separate decisions on which data/analytics technologies to adopt and where the technology will run (datacenter or cloud). The former are strategic decisions, and the latter are cost/performance decisions. The opportunities presented by data-rich applications and services can be huge. Financial services firms must position themselves to adopt these rapidly changing technologies as part of a larger business technology strategy built on a flexible architecture.