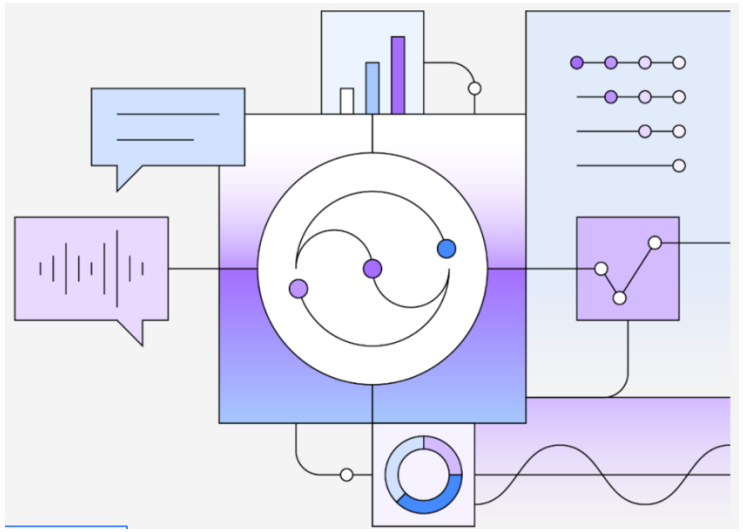
Harnessing Blockchain & DataScience to streamline & transform data management



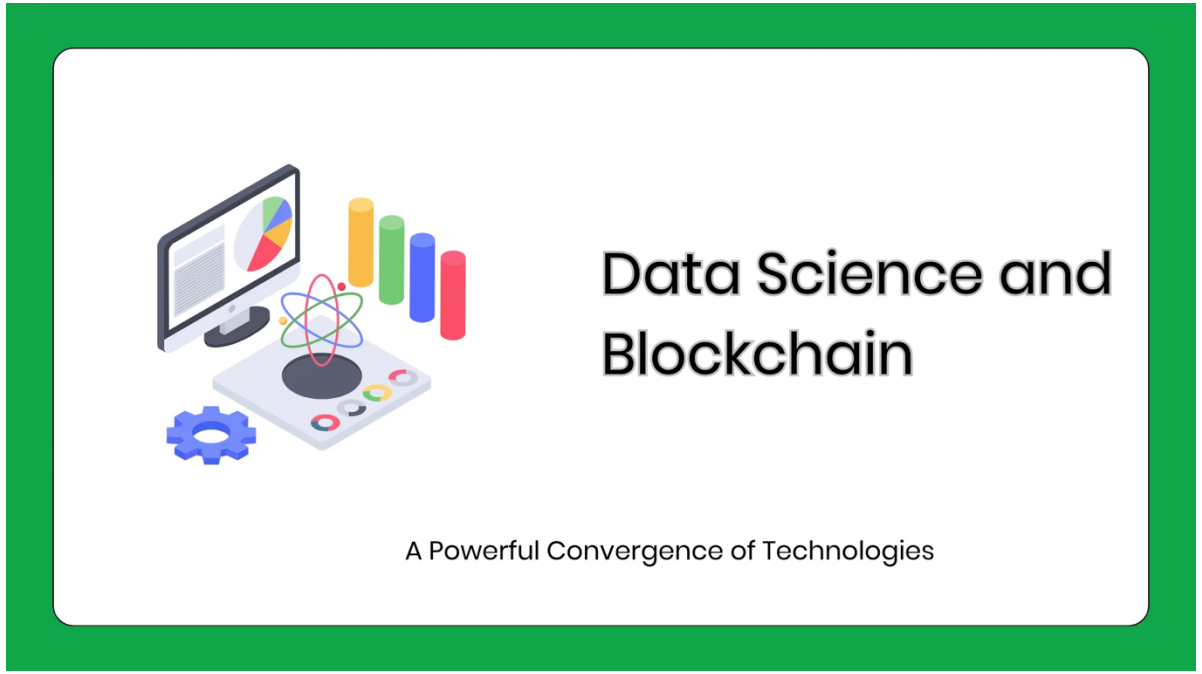
By. Muhammad Adeel Sultan Khan

Agenda

- ❖ Intersection of Blockchain & Data Science
- ❖ What is Data Science
- ❖ What is Blockchain
- ❖ Features of Blockchain
- ❖ Blockchain Process Flow
- ❖ Blockchain as a Service (BaaS)
- ❖ Relationship between Blockchain & Data Science
- ❖ Benefits of Blockchain & Data Science
- ❖ How Blockchain is used by Industry leaders
- ❖ What are Smart Contracts in Blockchain
- ❖ Advantages & Disadvantages of Blockchain's role in Data Science
- ❖ Future trends & opportunities
- ❖ Blockchain & Data Science Case Studies
- ❖ Agtech industry issues & Blockchain
- ❖ Conclusion

Intersection of Data Science & Blockchain

- Data science involves extracting insights and knowledge from structured and unstructured data through various methodologies and algorithms. Blockchain technology is a decentralized, secure, and transparent system of recording transactions across a network of computers.
- By combining the analytical power of data science with the security and immutability of blockchain, businesses can revolutionize how they manage and utilize data
- By leveraging the analytical capabilities of data science within the secure framework of blockchain, organizations can enhance data integrity, transparency, and security.
- This fusion enables improved decision-making processes, streamlined operations, and enhanced customer trust through verifiable data transactions. Moreover, the incorporation of smart contracts in blockchain technology can automate and optimize data management processes, reducing costs and improving efficiency.



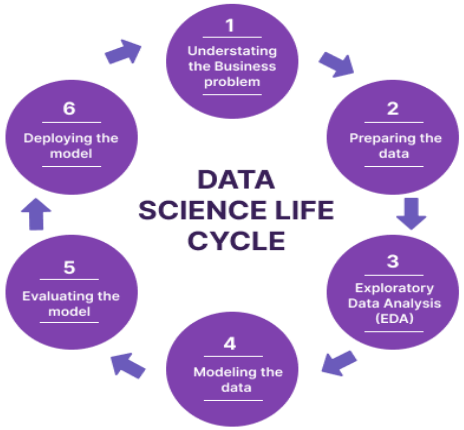
What is Data Science

- Data science combines math and statistics, specialized programming, advanced analytics, artificial intelligence (AI) and machine learning to uncover actionable insights hidden in an organization's data. These insights can be used to guide decision making and strategic planning

- The accelerating volume of data collection has made data science one of the fastest growing field across every industry

- The data science lifecycle involves various roles, tools, and processes, which enables analysts to glean actionable insights. Typically, a data science project undergoes the following stages:

  - **Data ingestion**: data collection—both raw structured and unstructured data from all relevant sources
  - **Data storage and data processing**: This stage includes cleaning data, deduplicating, transforming and combining data using ETL
  - **Data analysis**: exploratory data analysis to examine biases, patterns, ranges, and distributions of values in data
  - **Communicate**: provide reports, data visualization for business to take strategic decisions



What is Blockchain

- Blockchain is the world's most popular and fastest-growing technology. It has been used in many industries, such as finance, supply chain, healthcare, and more. Blockchain is a digital ledger that records transactions on a distributed public database. It is also called "Distributed Ledger Technology" or "DLT". The blockchain consists of blocks that are linked together and secured with cryptography.

- ***Blockchain*** is a distributed, decentralized, public ledger that records and stores transactions across a network of computers

- ***Blockchain*** makes it impossible to change, hack, or cheat the system

- ***Blockchain*** is a digital ledger of transactions that is duplicated and distributed across the entire network of computer systems

- ***Blockchain*** builds trust between supply chain actors

- It improves efficiency by coordinating people, organizations, and institutions

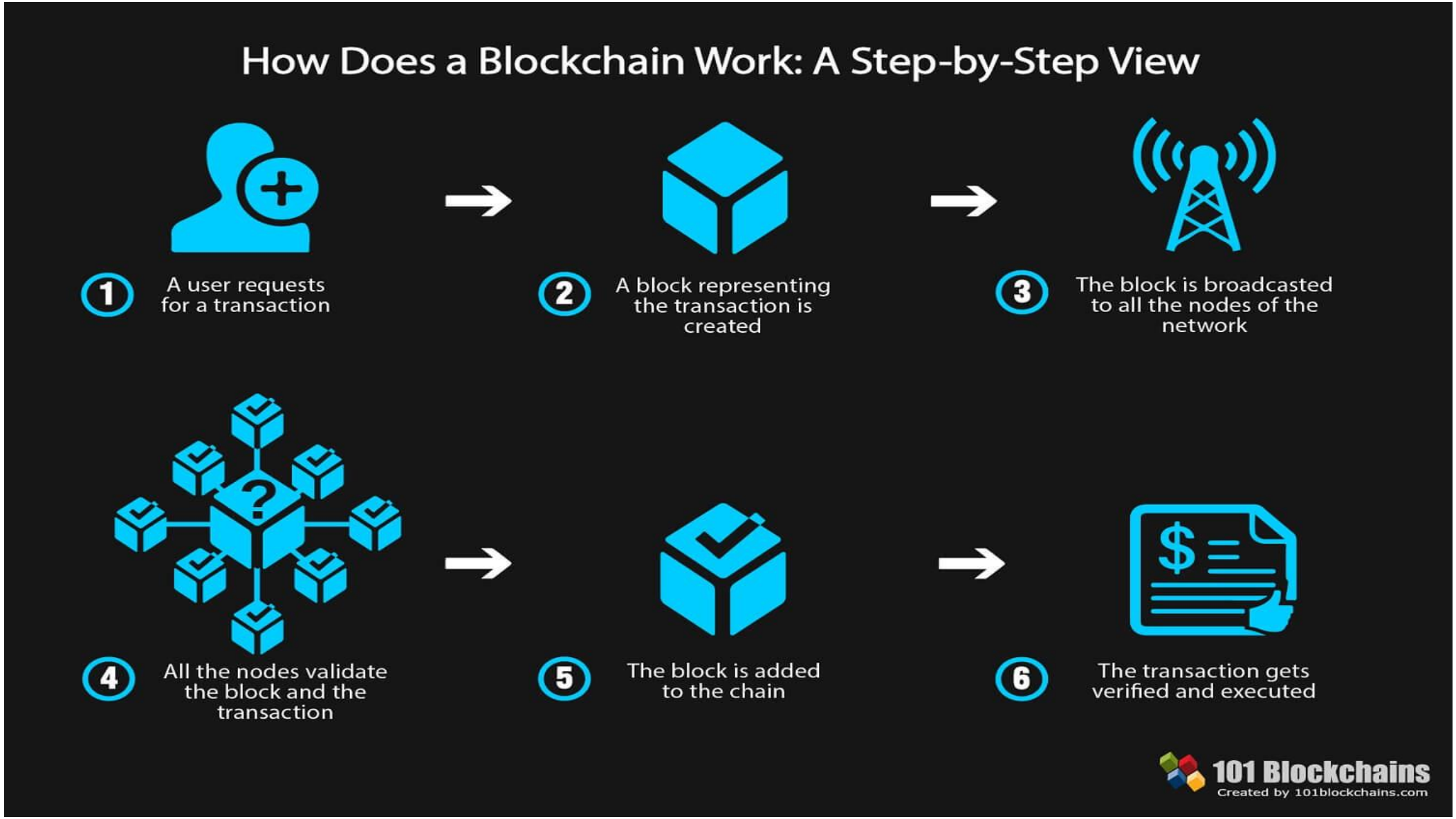
- The size of blockchain innovations is expected to grow from an estimated $17.57 billion in 2023 to nearly

$826 billion by 2032, 52.8% (CAGR)

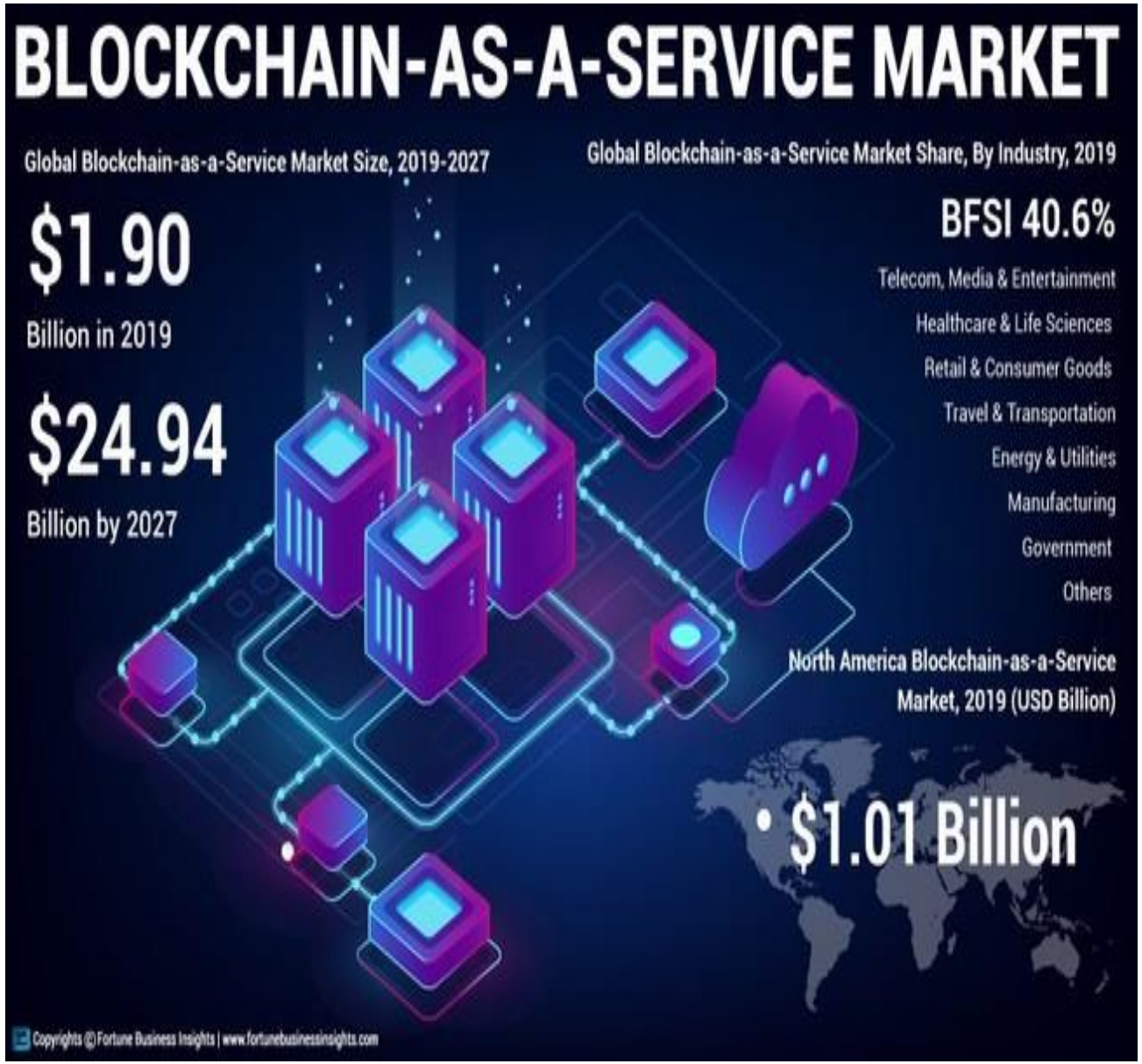
Features of Blockchain

1. **Decentralized:** The blockchain network has no single person looking after the network, instead a group of nodes is responsible for maintaining the network.
2. **Enhanced Security:** As the blockchain network is decentralized one can change the characteristics of the network for their benefit and using encryption adds another layer to the security of the blockchain.
3. **Transparency:** Every node in the blockchain network has a copy of the digital ledger and to add a transaction every node needs to check the validity of the transaction. If the majority of the node thinks that the transaction is valid only then the transaction can be added to the block. Thus, making the entire network transparent and corruption-proof.
4. **Distributed Ledgers:** The ledger on the blockchain network is maintained by all the nodes. So there is no single point of failure. Even if one node fails other nodes have the same copy of data so the network continues to function.
5. **Faster Settlement:** Blockchain offers faster settlement in comparison to traditional banking services. This helps to transfer money relatively faster using blockchain

Blockchain Process Flow



Blockchain as a Service (BaaS)



Relationship between Blockchain and Data Science

	Blockchain	Data science
Features	Decentralized, transparent, and immutable	Extracts insights from structured and unstructured data
Tools	Distributed ledgers, cryptographic algorithms, smart contracts	Programming languages, statistical libraries, visualization tools, machine learning frameworks
Benefits	Secures data, prevents unauthorized access, and maintains audit trails	Extracts insights from data to make decisions

Benefits of Blockchain & Data Science

1. **Data Integrity and Quality:** Blockchain ensures that data remains unaltered and authentic by maintaining an immutable ledger. Data Science relies on high-quality, accurate data to generate insights and build predictive models. Blockchain can enhance the quality of data science outcomes by providing verified and tamper-proof data.

2. **Enhanced Data Security:** Blockchain uses cryptographic techniques to secure data, making it difficult for unauthorized parties to access or manipulate the data. Data Science benefits from increased data security as sensitive information used in data science models and analytics can be protected from breaches and unauthorized changes.

3. **Data Provenance and Transparency:** Blockchain provides a transparent and traceable record of data origins and changes. This is crucial for validating the source of data and understanding its history. Data Science utilizes provenance data to validate the reliability of data sources and track changes over time, improving the credibility of data-driven conclusions.

4. **Decentralized Data Storage:** Blockchain distributes data across a network of nodes rather than relying on a central repository. This can reduce the risks associated with single points of failure. Data Science can leverage decentralized storage solutions to access and analyze distributed data sources, potentially leading to more comprehensive insights.

5. **Data Sharing and Collaboration:** Blockchain enables secure and controlled sharing of data among parties, with clear permissions and access controls. Data Science benefits from improved collaboration opportunities, where data scientists and researchers can share and access data across organizations while maintaining data integrity and security.

Benefits of Blockchain & Data Science

1. Enables Data Traceability It can be used to store data in an immutable way and make it easier to trace, ensuring the integrity of the data.

2. Allows for Real-Time Analysis Transactions are stored in blocks that are linked together in chronological order inside of chains. This makes it possible for people to make sense of data without a central database or third-party verification. Blockchain technology allows for real-time analysis by allowing users to input data into the system and then instantly see what happens with it as it moves through the system.

3. Makes Data Sharing Easier This type of information is stored in blocks and each block has a timestamp, which makes the data tamper-proof.

4. Ensures High-Quality Data Blockchain is an incorruptible digital ledger that stays updated on the fly and contains a record of every transaction that has ever happened, giving it a vast amount of trust.

5. Enhanced Data Integrity ensures data cannot be changed without leaving an immutable record of the change.

6. Encrypted transaction immutable record of transactions between two parties without a central authority to verify the transaction.

7. Builds Trust Blockchain can help to create a more transparent system that relies on the community more than any single person within it.

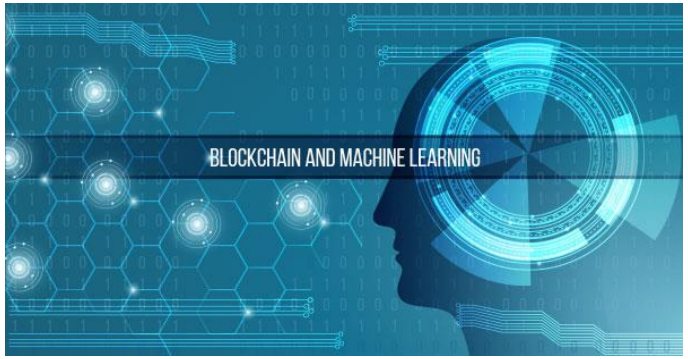
8. Data Lakes Blockchain uses the source of the data to record it in a specific block with a specific cryptographic key.

9. Make Predictions (Predictive Analysis) Data can be analyzed to reveal valuable insights and trends and can be used to predict future trends

10. Cost reduction It has helped in cost reduction by reducing costs associated with third parties, intermediaries, and brokers. It also helps in increasing the speed and transparency of transactions, which reduces costs associated with compliance.

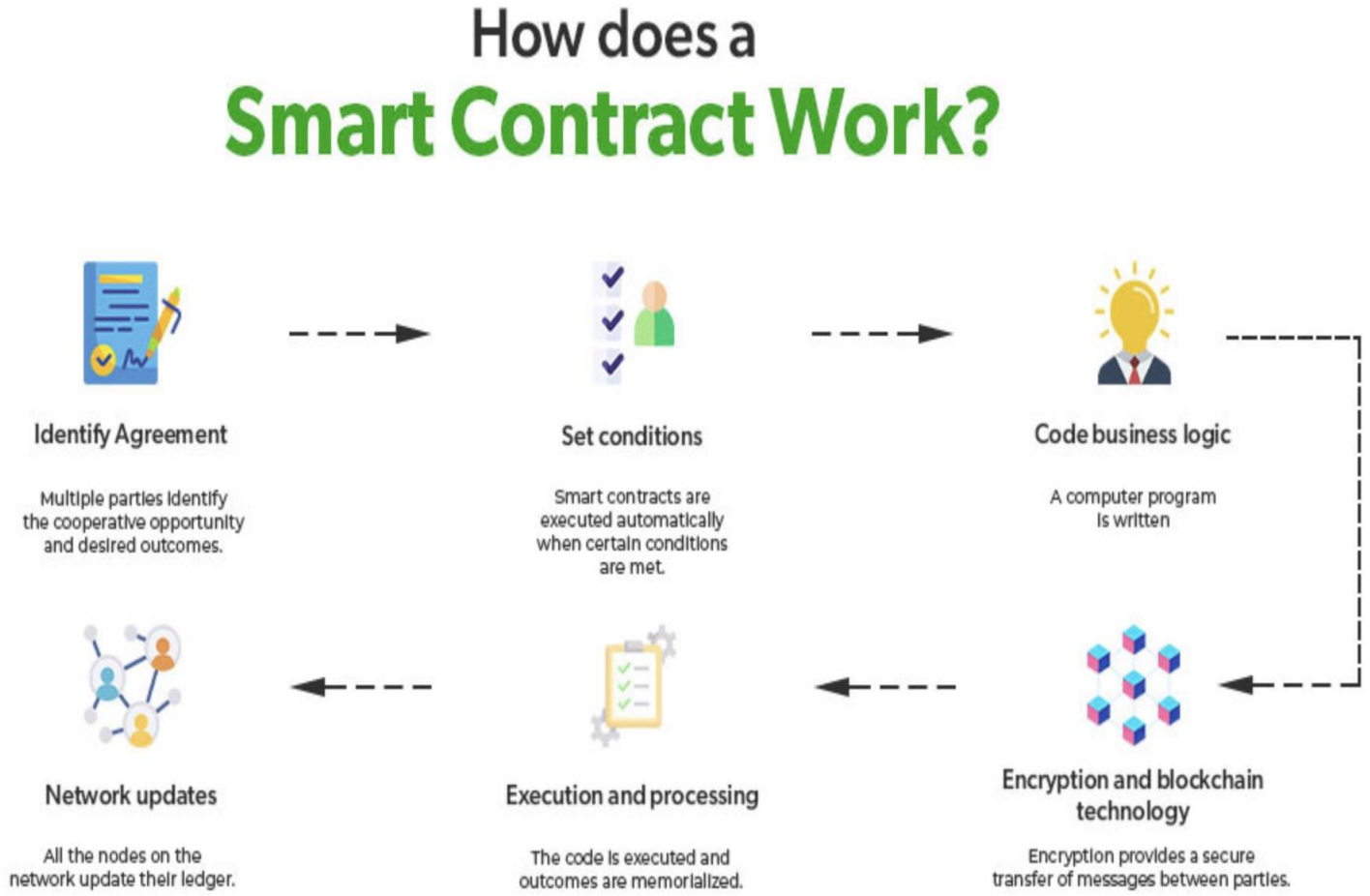
How Blockchain is used by Industry leaders

- AI, Machine Learning, Predictive Analytics & Blockchain --- step towards ***"Disruptive Innovation"***
- Since Blockchain allows all participants to see all data across the network, it enables BI/AI/ML at every network
- Blockchain is used in banking and enables faster cross-border payments, without the need for third-party authorization
- In the healthcare sector, blockchain technology is already being used to protect medical records and bring more transparency to the healthcare industry. Once a transaction has been added to the blockchain, it is then verified by the network and cannot be erased
- Blockchain technology is one of the most effective solutions for improving data security because blockchain is encrypted in such a way that it automatically provides proper validation and protects data from being modified or altered by hackers.



What are Smart Contracts in Blockchain

- Smart contracts are digital contracts stored on a blockchain that are automatically executed when predetermined terms and conditions are met.
- Smart contracts are used to automate the execution of an agreement so that all participants can be immediately certain of the outcome, without any intermediary's involvement or time loss. They can also automate a workflow, triggering the next action when predetermined conditions are met.



https://www.geeksforgeeks.org/smart-contracts-in-blockchain/

Advantages and Disadvantages of Blockchain's role in Data Science

Advantages	Disadvantages
Facilitates easier compliance with regulatory requirements through an immutable and verifiable record of transactions.	Implementing and maintaining blockchain solutions can result in increased expenses, with regard to infrastructure, development, and energy consumption.
Permits transparent data transactions and audit trails, resulting in easier tracking and verification of data origins and changes.	Blockchain-oriented scalability issues can result in slower transaction processing times.
Removes centralised control through data storage across multiple nodes.	Erroneous data entries can remain present, indefinitely, impacting data analysis outcomes.
Ensures accuracy and consistency of data, over time, through an immutable ledger.	Compatibility issues between blockchain and data science can arise, complicating integration efforts.
Allows new models for data monetisation, which can securely sell or trade data.	Uncertainties concerning data ownership, compliance, and jurisdiction can pose risks and hinder adoption.

https://zebpay.com/blog/blockchain-data-science-enhancing-decentralised-tech

Future Trends and Opportunities

Prediction	Trend	Opportunity
Enhanced Data Privacy and Security	Rise in utilization of privacy-preserving technologies.	Better mechanisms for improved sensitive data protection.
Decentralised Data Marketplaces	Growth of platforms to ensure data transactions using blockchain.	Development of new business models for data monetisation and democratisation.
AI and Machine Learning Integration	Combining blockchain with AI and machine learning to increase data integrity.	Development of decentralised AI models that leverage secure data sources.
Automated Smart Contracts	Increased usage for automation of complex data-driven processes.	Streamlined operations, reduced manual intervention, and less errors regarding contract execution.
Interoperability and Cross-Chain solutions	To enable seamless data exchange between different blockchain networks.	Increased flexibility and data integration across different blockchain networks.

https://zebpay.com/blog/blockchain-data-science-enhancing-decentralised-tech

Blockchain & Data Science Case Studies

❖ **IBM Food Trust:**

- Overview – IBM Food Trust uses blockchain to track the journey of food products, from farm to table.
- Integration – Blockchain and data science are combined to analyze and visualize supply-chain data for creating immutable records.
- Impact – Increases transparency, reduces food fraud, and improves the ability to rapidly address contamination issues.

❖ **Chainalysis:**

- Overview – Chainalysis provides blockchain analysis and related tools to detect and prevent financial crimes.
- Integration – Uses blockchain data for forensic analysis, along with data science techniques to identify suspicious activity and patterns.
- Impact – Improves the detection of money laundering and frauds and increases the security and reliability of crypto transactions.

❖ **Ocean Protocol:**

- Overview – Ocean Protocol is a decentralized data marketplace that uses blockchain to facilitate data sharing and monetization.
- Integration – Uses blockchain for secure data transactions and smart contracts, including data science to explore and analyze insights from shared data.
- Impact – Enables safe and fair data exchange, increases data privacy, and provides incentives for data providers.

❖ **MediLedger Project:**

- Overview – The MediLedger Project emphasizes on improving the security and efficiency of the medical supply chain via blockchain.
- Integration – Integrates blockchain and data science to provide secure, tamper-proof records for scrutinizing supply chain information.
- Impact – Increases transparency, diminishes fraud, and betters compliance, resulting in increased data accuracy and integrity.

❖ **SelfKey:**

- Overview – SelfKey uses blockchain for decentralized identity management.
- Integration – Leverages blockchain for secure identity storage and data science to analyze identity data and prevent fraud.
- Impact – Increases privacy and control over personal data, simplifies the verification process, and reduces risks of identity theft.

https://zebpay.com/blog/blockchain-data-science-enhancing-decentralised-tech

Agtech Industry Current Issues

- **Food Fraud**
  - Global Food Traceability Center estimates food fraud costs the global food industry $10 - $15 billion annually
- **Lack of Trust**
  - “It’s about trust -- building trust with math and true consensus,” Andy Kennedy, interim director, Global Food Traceability Center
- **Manual Process**
  - To much paper work in keying entries, bills documents, storage and quality information
- **Insufficient Quality & Storage historical data**
  - Data about grain storage conditions & quality info is not available

Some Key Agtech Players using Blockchain

- https://www.ripe.io/

  By designing a **transparent digital food supply chain**, the startup harnesses quality food data to create the Blockchain of Food – mapping the food journey

- https://www.agridigital.io/

  A blockchain-based and integrated commodity management solution for the global grains industry. The platform helps to **process complex agricultural transactions through smart contracts**.

- http://www.agriledger.io/

  A UK social enterprise project **supporting farmers in tracing food origins**, getting easier access to financing, and storing transactions data.

- https://www.worldcovr.com/

  Provide crop insurance to protect against loss of yield using satellites to monitor the rainfall and **trigger payouts automatically**.

- https://demeter.life/

  A central hub to **rent and farm micro fields anywhere in the world** – with no middlemen, complexity or the overhead of a big organization

- https://etherisc.com/

  A blockchain startup offering crop insurance to farmers via its **decentralized insurance applications**.

Conclusion: Harnessing Data Science and Blockchain for business success

The integration of blockchain and data science provides a great milestone in technological innovation. Blockchain's ability to deliver secure, transparent and immutable data is a powerful feature, while data science ability to transform the data into actionable insights and predictive models makes the data more useful. This combination address challenges such as data integrity, privacy, and productivity.

The advancements in artificial intelligence, machine learning, and the Internet of Things (IoT) further enhance the synergy between data science and blockchain .*The synergy between these technologies promises to drive future progress, leading us towards disruptive innovation starting from decentralized data marketplaces to enhanced AI paradigms.*

***Harnessing data science with blockchain* can create new opportunities for businesses to innovate and stay competitive**

References

- https://www.prnewswire.com/news-releases/blockchain-as-a-service-baas-market-to-hit-usd-24-94-bn-by-2027-rising-demand-for-decentralized-software-services-to-boost-market-growth-fortune-business-insights-301053885.html
- https://101blockchains.com/blockchain-as-a-service/
- https://www.startus-insights.com/innovators-guide/8-blockchain-startups-disrupting-the-agricultural-industry/
- https://www.ncbiotech.org/news/blockchain-bringing-transparency-traceability-food-and-more
- https://medium.com/@Intersog/what-happens-when-you-combine-blockchain-and-machine-learning-2afafc9654d2
- https://medium.com/womenintechnology/exploring-the-intersection-of-data-science-and-blockchain-an-expert-analysis-21a11bf408dc