

Cricket Stat Pro

IPL data analysis pipeline using Apache Spark

[Click here](#)



Made with GAMMA



Efficient ETL Operations



Data Extraction

70% enhancement from Kaggle and CSV sources



Transformation

Complex operations with Spark Data Frames and SQL



Apache Spark

Comprehensive pipeline for IPL data analysis

Performance Metrics

70%

Data Extraction

Improvement in extraction
efficiency

55%

Storage Efficiency

Boost in AWS S3 storage
optimization



Technology Stack



Apache Spark

Core processing engine for data transformations and SQL operations



AWS S3

Scalable cloud storage for optimized data access



Data Sources

Kaggle datasets and CSV files integration



Optimized Data Storage



Transform

Spark processing



Deploy

AWS S3 storage



Analyze

Seamless access

Ensuring seamless access for future analysis and visualization