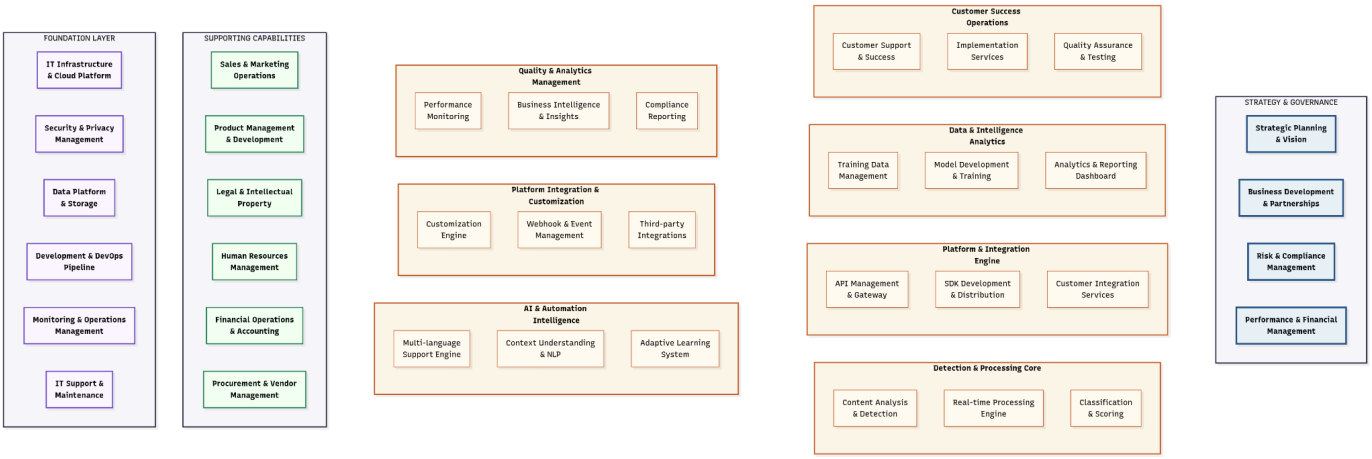
A. Business Capability Model (BCM)



B. Analisis Teknologi

1. CORE AI & MACHINE LEARNING TECHNOLOGIES

1.1. Natural Language Processing (NLP) Stack

Primary NLP Frameworks

- **spaCy 3.5+**: Production-ready NLP library untuk text preprocessing, tokenization, dan analisis linguistik
    - **Business Impact**: Fast processing (1M+ tokens/second), production-stable untuk real-time applications
    - **Use Case**: Text preprocessing, named entity recognition, dependency parsing untuk context analysis
    - **Cost**: Open source, $0 licensing cost
- **Hugging Face Transformers**: State-of-the-art pre-trained language models
    - **Business Impact**: Access ke model terbaik (BERT, RoBERTa, GPT) tanpa perlu train from scratch
    - **Use Case**: Fine-tuning untuk bahasa Indonesia, context understanding, sentiment analysis
    - **Cost**: Open source models, paid inference API $0.0004/1K tokens

Deep Learning Frameworks

- **PyTorch 2.0**: Primary framework untuk model development dan research
    - **Technical Advantage**: Dynamic computational graphs, easier debugging, faster iteration
    - **Business Impact**: Faster time-to-market untuk new models, easier A/B testing
    - **Production Consideration**: TorchServe untuk deployment, TorchScript untuk optimization
- **TensorFlow 2.x + TensorFlow Serving**: Production deployment dan serving
    - **Technical Advantage**: Mature ecosystem, optimized serving, better for large-scale production
    - **Business Impact**: Lower serving costs, better latency untuk high-traffic applications
    - **Integration**: Native support untuk Kubernetes, monitoring, dan auto-scaling

Specialized Models untuk Content Moderation

- **BERT-based Models**:
    - **IndoBERT**: Pre-trained untuk bahasa Indonesia, understanding context dan nuansa
    - **mBERT**: Multilingual support untuk regional languages
    - **DistilBERT**: Lightweight version, 60% smaller dengan 97% performance retention
    - **Business Case**: IndoBERT untuk accuracy, DistilBERT untuk cost optimization
- **Custom Transformer Architectures**:
    - **RoBERTa**: Optimized BERT dengan better training methodology
    - **ELECTRA**: Efficient pre-training untuk domain-specific fine-tuning
    - **DeBERTa**: Enhanced architecture untuk better context understanding

1.2 Model Training & Optimization Technologies

MLOps Platform

- **MLflow**: Experiment tracking, model versioning, dan lifecycle management
    - **Features**: Model registry, experiment comparison, automated deployment
    - **Business Impact**: 40% faster model development cycle, reproducible experiments
    - **Cost**: Open source core, managed version $1-3 per model/month
- **Weights & Biases (wandb)**: Advanced experiment tracking dan collaboration
    - **Features**: Real-time metrics, hyperparameter optimization, team collaboration
    - **Business Impact**: Improved model performance melalui systematic experimentation

    - **Cost**: $50/month per researcher, enterprise pricing available

Model Optimization & Acceleration

- **ONNX Runtime**: Cross-platform inference optimization
    - **Performance**: 2-10x speedup untuk inference, reduced memory usage
    - **Business Impact**: Lower serving costs, better user experience
    - **Compatibility**: Supports PyTorch, TensorFlow, scikit-learn models
- **TensorRT (NVIDIA)**: GPU-optimized inference engine
    - **Performance**: Up to 40x speedup untuk BERT inference on V100/A100 GPUs
    - **Business Impact**: Handle 10x more requests dengan same hardware cost
    - **Use Case**: High-throughput real-time processing

Data Processing & Pipeline

- **Apache Spark**: Distributed data processing untuk large-scale training data
    - **Capabilities**: ETL processing, feature engineering, distributed training data preparation
    - **Scale**: Process billions of text samples untuk model training
    - **Cost**: $0.10-0.30 per compute hour on cloud
- **Ray**: Distributed computing framework untuk ML workloads
    - **Features**: Distributed training, hyperparameter tuning, serving
    - **Business Impact**: Train larger models faster, better resource utilization
    - **Use Case**: Distributed hyperparameter search, parallel model training

2. REAL-TIME PROCESSING & INFRASTRUCTURE

2.1 Stream Processing Technologies

Message Streaming Platform

- **Apache Kafka**: Core event streaming platform
    - **Throughput**: 10M+ messages/second dengan proper configuration
    - **Features**: Fault tolerance, exactly-once processing, stream processing
    - **Business Impact**: Real-time content processing, zero data loss
    - **Cost**: $50-200/month untuk moderate traffic, scales linearly
- **Apache Kafka Streams**: Stream processing library
    - **Capabilities**: Real-time aggregation, windowing, joins
    - **Use Case**: Real-time analytics, trend detection, adaptive filtering

    - **Performance**: Process 100K+ events/second per instance

Caching & Fast Storage

- **Redis Cluster**: Distributed in-memory database
    - **Use Cases**: API response caching, session storage, real-time counters
    - **Performance**: Sub-millisecond latency, 1M+ operations/second
    - **Business Impact**: <50ms API response time, better user experience
    - **Cost**: $50-500/month depending on memory requirements
- **Apache Cassandra**: Distributed NoSQL untuk time-series data
    - **Use Case**: Store detection history, analytics data, audit logs
    - **Scale**: Petabyte-scale storage, linear scalability
    - **Performance**: 10K-100K writes/second per node

2.2 API & Integration Infrastructure

API Gateway & Management

- **Kong Gateway**: Enterprise API management platform
    - **Features**: Rate limiting, authentication, analytics, plugin ecosystem
    - **Security**: OAuth2, JWT, API key management, IP whitelisting
    - **Business Impact**: Protect dari abuse, monetize API usage, detailed analytics
    - **Cost**: Open source core, enterprise $3K-10K/year
- **AWS API Gateway**: Managed serverless API gateway
    - **Features**: Auto-scaling, DDoS protection, caching, monitoring
    - **Integration**: Native AWS services integration, Lambda triggers
    - **Cost**: $3.50 per million API calls, included caching

Container Orchestration

- **Kubernetes**: Container orchestration platform
    - **Features**: Auto-scaling, rolling deployments, service discovery, health checks
    - **Business Impact**: Zero-downtime deployments, automatic scaling, cost optimization
    - **Cost**: Managed Kubernetes $70-200/month per cluster
- **Istio Service Mesh**: Advanced traffic management
    - **Features**: Load balancing, A/B testing, canary deployments, observability
    - **Use Case**: Gradual model rollouts, traffic splitting untuk testing
    - **Business Impact**: Safer deployments, better observability

3. DATA MANAGEMENT & ANALYTICS

3.1 Database Technologies

Primary Data Storage

- **PostgreSQL 14+**: Main relational database
    - **Extensions**: pg_vector untuk vector similarity search, TimescaleDB untuk time-series
    - **Features**: ACID compliance, advanced indexing, JSON support
    - **Use Case**: User data, configuration, structured analytics
    - **Performance**: 10K-50K transactions/second dengan proper tuning
- **MongoDB**: Document database untuk unstructured data
    - **Use Case**: Content metadata, user-generated rules, configuration documents
    - **Features**: Flexible schema, horizontal scaling, aggregation pipeline
    - **Scale**: Multi-terabyte collections, sharded clusters

Search & Analytics

- **Elasticsearch 8.x**: Full-text search dan analytics engine
    - **Use Case**: Content search, real-time analytics, log analysis
    - **Features**: Full-text search, aggregations, machine learning features
    - **Performance**: Sub-second search pada billions of documents
    - **Cost**: $45-200/month per node depending on instance size
- **ClickHouse**: Columnar database untuk analytics
    - **Use Case**: Real-time analytics, business intelligence, reporting
    - **Performance**: 100x faster than traditional databases untuk analytical queries
    - **Features**: SQL compatibility, compression, distributed queries

3.2 Data Pipeline & ETL

Workflow Orchestration

- **Apache Airflow**: Workflow management platform
    - **Use Case**: Data pipeline orchestration, model training schedules, ETL processes
    - **Features**: DAG-based workflows, retry logic, monitoring, scheduling
    - **Business Impact**: Reliable data processing, automated model retraining
- **dbt (data build tool)**: Data transformation framework
    - **Use Case**: SQL-based data transformations, data quality testing
    - **Features**: Version control, documentation, lineage tracking
    - **Business Impact**: Reliable analytics data, faster insights

Data Processing Engines

- **Apache Spark**: Distributed computing untuk big data
    - **Use Case**: Large-scale data processing, feature engineering, batch inference
    - **Performance**: Process terabytes of data dalam minutes to hours
    - **Cost**: $0.10-0.50 per compute hour, auto-scaling available

4. CLOUD INFRASTRUCTURE & SECURITY

4.1 Cloud Platform Strategy

Multi-Cloud Approach

- **Primary: AWS**
    - **Services**: EC2, EKS, SageMaker, S3, RDS, ElastiCache, Lambda
    - **Advantages**: Mature ML services, global presence, enterprise support
    - **Cost**: Typically 20-30% premium but better service reliability
- **Secondary: Google Cloud Platform**
    - **Services**: GKE, Vertex AI, BigQuery, Cloud Storage, Cloud SQL
    - **Advantages**: Advanced AI/ML tools, competitive pricing untuk compute
    - **Use Case**: Development/testing environment, specific AI workloads

Compute & Scaling

- **Kubernetes Engine**: Managed Kubernetes untuk application deployment
    - **Features**: Auto-scaling, node auto-provisioning, cluster upgrades
    - **Cost Optimization**: Spot instances (70% cost reduction), preemptible VMs
    - **Business Impact**: Elastic scaling based on demand, cost control
- **Serverless Computing**: AWS Lambda, Google Cloud Functions
    - **Use Case**: Event-driven processing, API backends, scheduled tasks
    - **Benefits**: Pay-per-execution, automatic scaling, zero infrastructure management
    - **Cost**: $0.20 per 1M requests, very cost-effective untuk sporadic workloads

4.2 Security & Compliance Infrastructure

Identity & Access Management

- **OAuth 2.0 + OpenID Connect**: Industry standard authentication
    - **Implementation**: Auth0, AWS Cognito, atau custom dengan Keycloak
    - **Features**: SSO, MFA, role-based access control

    - **Business Impact**: Enterprise-grade security, compliance ready
- **HashiCorp Vault**: Secrets management platform
    - **Use Case**: API keys, database credentials, encryption keys
    - **Features**: Dynamic secrets, encryption as a service, audit logging
    - **Security**: Hardware security module (HSM) integration

Data Protection & Privacy

- **Encryption Technologies**:
    - **At Rest**: AES-256 encryption untuk all stored data
    - **In Transit**: TLS 1.3 untuk all network communication
    - **Application Level**: Field-level encryption untuk sensitive data
- **Privacy Engineering**:
    - **Differential Privacy**: Add noise to analytics untuk privacy protection
    - **Data Anonymization**: Remove atau mask PII dari training datasets
    - **GDPR Compliance**: Automated data subject rights, consent management

5. MONITORING & OBSERVABILITY

5.1 Application Performance Monitoring

Metrics & Monitoring

- **Prometheus + Grafana**: Metrics collection dan visualization
    - **Metrics**: Custom business metrics, system metrics, model performance
    - **Alerting**: PagerDuty integration untuk critical issues
    - **Cost**: Open source, $50-200/month untuk managed services
- **Jaeger**: Distributed tracing untuk microservices
    - **Use Case**: Track requests across services, identify bottlenecks
    - **Business Impact**: Faster issue resolution, performance optimization

Logging & Analysis

- **ELK Stack**: Elasticsearch, Logstash, Kibana untuk log management
    - **Use Case**: Centralized logging, error tracking, security monitoring
    - **Scale**: Handle millions of log events per day
    - **Features**: Real-time search, alerting, dashboard
- **Structured Logging**: JSON-based logging dengan correlation IDs
    - **Benefits**: Better searchability, correlation across services

    - **Tools**: Winston (Node.js), structlog (Python), logrus (Go)

5.2 Model Performance Monitoring

ML Model Monitoring

- **Evidently AI**: Model drift detection dan data quality monitoring
    - **Features**: Data drift, concept drift, model performance degradation
    - **Business Impact**: Proactive model maintenance, prevent accuracy degradation
    - **Integration**: Real-time monitoring pipeline, automated retraining triggers
- **WhyLabs**: Data dan model observability platform
    - **Features**: Data profiling, anomaly detection, model explainability
    - **Use Case**: Monitor training data quality, detect bias, explain predictions

6. DEVELOPMENT & DEPLOYMENT

6.1 Development Workflow

Version Control & Collaboration

- **Git dengan GitHub/GitLab**: Source code management
    - **Features**: Branch protection, code review, automated testing
    - **Workflows**: GitFlow untuk feature development, semantic versioning
- **Docker**: Containerization untuk consistent deployments
    - **Benefits**: Environment consistency, easier scaling, dependency management
    - **Strategy**: Multi-stage builds, security scanning, size optimization

CI/CD Pipeline

- **GitHub Actions / GitLab CI**: Automated testing dan deployment
    - **Pipeline**: Code → Test → Build → Deploy → Monitor
    - **Features**: Parallel testing, conditional deployments, rollback capabilities
    - **Security**: Secrets management, vulnerability scanning

6.2 Infrastructure as Code

Infrastructure Automation

- **Terraform**: Infrastructure provisioning dan management
    - **Benefits**: Reproducible infrastructure, version control, planning

    - **Use Case**: Multi-cloud infrastructure, disaster recovery setup
- **Helm Charts**: Kubernetes application packaging

    - **Benefits**: Templated deployments, configuration management, upgrades
    - **Use Case**: Application deployment, environment management