Gmail Vasily S. <vasily.souzdenkov@gmail.com>

Resource 2
1 message

Vasily S. <vasily.souzdenkov@gmail.com> Mon, Sep 1, 2025 at 8:35 AM
To: "Vasily S." <vasily.souzdenkov@gmail.com>

TLDR

This framework provides a comprehensive blueprint for building a marketing analytics environment from the ground up. Key components include:

- **Tiered Architecture**: Star schema data warehouses with ETL pipelines using tools like Dataswarm for scalable data processing
- **Multi-Level Reporting**: Tier 0 executive dashboards for leadership, Tier 1 operational dashboards for teams, and Tier 2 exploration dashboards for analysts
- **Performance 5 Framework**: Account simplification, automation, creative diversification, data quality, and results validation as core best practices
- **Advanced Analytics**: MMM (Marketing Mix Modeling), statistical meta-analysis, and predictive modeling for performance optimization
- **Customer Segmentation**: Behavioral, demographic, and value-based segmentation strategies for personalized marketing
- **Quality-First Approach**: Data governance, quality checks, and continuous monitoring throughout the analytics lifecycle

Analytics Architecture and Infrastructure

Foundation Components

Data Warehouse Design: Implement a **star schema** architecture with fact tables (fct_*) capturing events and dimension tables (dim_*) providing descriptive attributes. This pattern enables stable, performant analytical queries and serves as the foundation for downstream reporting.

ETL Pipeline Framework: Utilize **Dataswarm** or similar DAG-based schedulers for batch ETL processing. Key principles:

- **Pipeline as Code**: Define workflows entirely in Python scripts for version control and modularity
- **Operator-Based Execution**: Use specialized operators (DataSync for ingestion, PrestoInsertOperator for transformations)
- **Automatic Dependency Management**: Leverage data annotations to automatically infer task execution order

Infrastructure Layers:

1. **Data Marts**: Create flexible, tool-specific data marts as modules

2. **Dimensional Tables**: Centralized dimension tables for common identifiers
3. **Staging Layer**: Intermediate processing layer for ETL operations
4. **Access Control**: Implement data privacy controls based on content sensitivity

Performance Optimization: Design for scalability with considerations for data volume growth, query performance, and horizontal scaling capabilities.

Architecture Reference

Data Collection and Tracking Framework

Data Quality Management

Technical Rules: Implement automated checks for primary key validation, duplicate detection, and data type consistency. These are **Risk Level 1 (High)** issues that will stop pipelines from running.

Business Rules: Establish validation for business logic compliance, including referential integrity and domain-specific constraints. Categorize as **Risk Level 2 (Mid)** for application-level issues.

Data Governance Framework:

- **Data Lifecycle Management**: Processes for creation, usage, storage, and archival
- **Metadata Management**: Centralized catalog for data discovery and lineage tracking
- **Privacy and Security**: Comply with regulations like HSUD (Highly Sensitive User Data)
- **Data Ownership**: Clear accountability for data assets and quality

Collection Best Practices

Conversions API Implementation:

- Ensure correct technical setup with proper event coverage
- Focus on **data freshness** (real-time processing)
- **Deduplicate events** across Meta Business Tools
- Leverage multiple data sources (CRM, Business Messaging)

Signal Quality Optimization:

- **Maximize event coverage** across the customer journey
- **Increase event effectiveness** through high event match quality
- Implement **custom conversions** for signal segmentation
- **Actions by integration type**: Tailor approaches for Direct, Partner, or Gateway integrations

Data Quality Tools Reference

Metrics and KPI Framework Design

KPI Selection Framework

Hierarchical Approach:

1. **Beacon/Northstar Metric**: Align with overall vision and core values
2. **Service Outcome Metrics**: One or more KPIs for each high-level goal (max 2 primary KPIs recommended)
3. **Operational Metrics**: Tactical metrics that directly map to specific actions

Measurement Standards:

- **90% confidence interval** as default for hypothesis testing
- **Clear numerical goals** with pre-defined p-value thresholds
- **Success criteria**: Use both confidence intervals and point estimates
- **Multiple comparisons correction** when using >3 KPIs

Performance Measurement Framework

Performance 5 Pillars (Industry Best Practice):

1. **Account Simplification**: Reduce complexity to improve efficiency
2. **Automation**: Use AI-driven optimization for campaign management
3. **Creative Diversification**: Differentiate ad creative by concepts/formats
4. **Data Quality**: Optimize data sources like Conversions API
5. **Results Validation**: Verify campaigns with lift experiments or MMM

Metric Types:

- **Descriptive Metrics**: Traffic light systems for best practice adoption
- **Predictive Metrics**: Weighted systems or ML-based models for performance forecasting
- **Composite Metrics**: Aggregated scores (e.g., MAX score over 100 points)

Unified Measurement: Combine MMM (Marketing Mix Modeling), MTA (Multi-Touch Attribution), and experiments using Bayesian statistics to create a single source of truth.

Performance 5 Framework

Dashboard and Reporting Structure

Dashboard Tiering Strategy

Tier 0 - Executive/Topline Dashboards:

- **Purpose**: Business-critical metrics for senior leadership (VP+)
- **Audience**: Senior executives, directors, team leadership
- **Design**: Report-style with limited/no filtering options
- **Content**: Northstar metrics, goal tracking, high-level health indicators
- **Performance**: <10 second load time, <80 widgets per tab

Tier 1 - Product/Operations Dashboards:

- **Purpose**: Team and product performance monitoring

- **Audience**: Directors, product teams, engineering
- **Design**: Semi-flexible with critical filtering options
- **Content**: Operational metrics, system health, product KPIs

Tier 2 - Deep Dive/Exploration Dashboards:

- **Purpose**: Detailed analysis and investigation
- **Audience**: Analytics teams, data scientists, engineers
- **Design**: Fully flexible with extensive filtering capabilities

Design Principles

Four Core Principles:

1. **Simplify**: Focus on data, minimize interface distractions
2. **Prioritize**: Make data visualizations the primary experience
3. **Intentional**: Every element should serve the data story
4. **Systemize**: Standardize patterns for consistency

Layout Guidelines:

- **Information Hierarchy**: Most important metrics top-left
- **Gestalt Principles**: Use proximity and similarity for logical grouping
- **5 Section Maximum**: Avoid cognitive overload
- **BAN (Big Aggregate Numbers)**: Include KPIs at the top as visual anchors

Performance Standards:

- Avoid Class D (large) queries on Tier-0 pages
- Implement data quality widgets and bug filing capabilities
- Use standardized color palettes and semantic color meanings

Dashboard Design Reference

Statistical Analysis and Modeling

Advanced Analytics Methods

Marketing Mix Modeling (MMM):

- **MEOW Package**: Media Effect Optimization Workflow for budget allocation
- **Training Time**: 1-5 minutes for model development
- **Optimization**: Channel-level lift calculation and budget reallocation
- **Constraint Handling**: Account for fixed spending (e.g., on-platform spend)

Time Series Analysis:

- **ARIMA Models**: For trend and seasonality analysis
- **BSTS (Bayesian Structural Time Series)**: For intervention impact assessment
- **VAR (Vector Autoregression)**: For multivariate time series relationships

Experimental Design Framework:

- **Power Analysis**: Determine sample size and study length a priori
- **Stopping Conditions**: Pre-defined rules for data collection completion
- **Multiple Comparisons**: Bonferroni correction for >3 KPIs
- **Test/Control Splits**: 50/50 default, 90/10 for calendar-driven campaigns

Machine Learning Applications

Predictive Modeling:

- **Class Imbalance Handling**: SMOTE, undersampling for minority classes
- **Feature Engineering**: Behavioral, demographic, and engagement features
- **Model Validation**: Cross-validation, precision/recall optimization
- **Lifetime Value Prediction**: Advanced techniques for customer value forecasting

Meta-Analysis Techniques:

- **Bayesian Approaches**: Random effects modeling for study aggregation
- **Statistical Standards**: Minimum 10 studies for lift analysis, 15 for A/B tests
- **Methodology Types**: Statistical, descriptive, and multi-cell analysis

MMM Platform Reference

Segmentation and Audience Analytics

Customer Segmentation Strategies

B2B Segmentation Methods:

1. **Firmographics**: Organizational placement and company characteristics
2. **Customer Value Tiering**: Revenue-based prioritization
3. **Needs-Based**: Speed vs. information preferences
4. **Customer Sophistication**: Technical capability and support requirements
5. **Psychographic/Behavioral**: Values, goals, and behavioral patterns

Advertiser Intent Framework:

- **X-Axis**: Access to business outcome data
- **Y-Axis**: Facebook optimization capabilities
- **8 Segments**: From direct optimization to strategic support categories
- **MECE Principle**: Mutually exclusive, collectively exhaustive segments

Behavioral Analytics Implementation

Behavior-Based User Segmentation (BUS):

- **Tree Structure**: Hierarchical representation using user behaviors and features

- **Reward Function**: Measures behavior representability of active users for marginal users
- **Optimization Goal**: Maximize marginal user representation through segment collection

Clustering Algorithms:

- **K-Means**: For spherical clusters with numerical data
- **Hierarchical Clustering**: For varying densities and relationship visualization
- **DBSCAN**: For non-spherical clusters and noise handling

Value Optimization Approaches:

- **Customer Lifetime Value (CLV)**: Predict long-term customer worth
- **Signal Segmentation**: Different conversion values for same user actions
- **Audience Value Optimization**: High/low value user classification

Segmentation Framework

Executive vs Operational Reporting

Executive Reporting Framework

Tier 0 Executive Characteristics:

- **Audience**: Senior executives, VP+ level, pillar leadership
- **Content**: Business-critical metrics, goal monitoring, health indicators
- **Format**: One-pager style with minimal filtering
- **Update Frequency**: Real-time or daily for critical metrics
- **Performance**: <10 second load time requirement

Executive Dashboard Components:

- **Goal Tracking**: Longitudinal and holdout percentage goals
- **Health Monitoring**: Core product/org health metrics
- **Efficiency Metrics**: OpEx improvements, automation rates
- **Business Impact**: Revenue, conversion, and engagement outcomes

Operational Reporting Framework

Tier 1 Team/Product Dashboards:

- **Audience**: Directors, product teams, engineering teams
- **Content**: Product performance, operational metrics, system health
- **Format**: Semi-flexible with critical filtering options
- **Features**: Drill-down capabilities, dimensional breakdowns

Tier 2 Deep Dive Dashboards:

- **Audience**: Analytics teams, data scientists, engineers
- **Content**: Granular analysis, exploration capabilities
- **Format**: Fully flexible with extensive filtering

- **Use Case**: Root cause analysis, detailed investigations

Differentiation Strategies

Information Hierarchy:

- **Executive**: High-level summaries, trend indicators, goal status
- **Operational**: Detailed breakdowns, actionable insights, diagnostic tools
- **Exploratory**: Raw data access, flexible segmentation, ad-hoc analysis

Communication Approach:

- **Executive**: "What happened?" and "Are we on track?"
- **Operational**: "Why did it happen?" and "What should we do?"
- **Analytical**: "How can we dig deeper?" and "What patterns exist?"

Quality Standards:

- **Tier 0**: Highest data quality priority, rigorous audit processes
- **Tier 1**: Lower priority, less rigorous audit reminders
- **Tier 2**: Standard quality checks, content lifecycle management

Dashboard Tiering Reference

Implementation Roadmap and Best Practices

Implementation Phases

Phase 1: Foundation (Months 1-3)

- Establish data warehouse architecture with star schema design
- Implement core ETL pipelines using Dataswarm framework
- Set up basic data quality monitoring and governance
- Create Tier 0 executive dashboard with core KPIs

Phase 2: Core Analytics (Months 4-6)

- Deploy Performance 5 measurement framework
- Implement customer segmentation strategies
- Build Tier 1 operational dashboards
- Establish statistical analysis capabilities (MMM, experimentation)

Phase 3: Advanced Capabilities (Months 7-12)

- Add predictive modeling and machine learning
- Implement advanced segmentation (behavioral, value-based)
- Create Tier 2 exploration dashboards
- Deploy automated alerting and anomaly detection

Success Metrics

Technical KPIs:

- Dashboard load time <10 seconds (Tier 0)
- Data quality score >95% for critical datasets
- Pipeline success rate >99.5%
- User adoption rate >80% for target audiences

Business KPIs:

- Marketing ROI improvement from baseline
- Campaign optimization cycle time reduction
- Decision-making speed improvement
- User satisfaction scores for dashboard consumers

Critical Success Factors

1. **Executive Sponsorship**: Ensure leadership buy-in and resource allocation
2. **Cross-Functional Alignment**: Coordinate between engineering, analytics, and business teams
3. **Iterative Development**: Start simple, gather feedback, and continuously improve
4. **Quality First**: Prioritize data accuracy and reliability over feature quantity
5. **User-Centric Design**: Design dashboards for specific audiences and use cases
6. **Scalable Architecture**: Build for growth in data volume and user base

This framework provides a comprehensive foundation for building a world-class marketing analytics environment that scales from executive decision-making to detailed operational analysis.