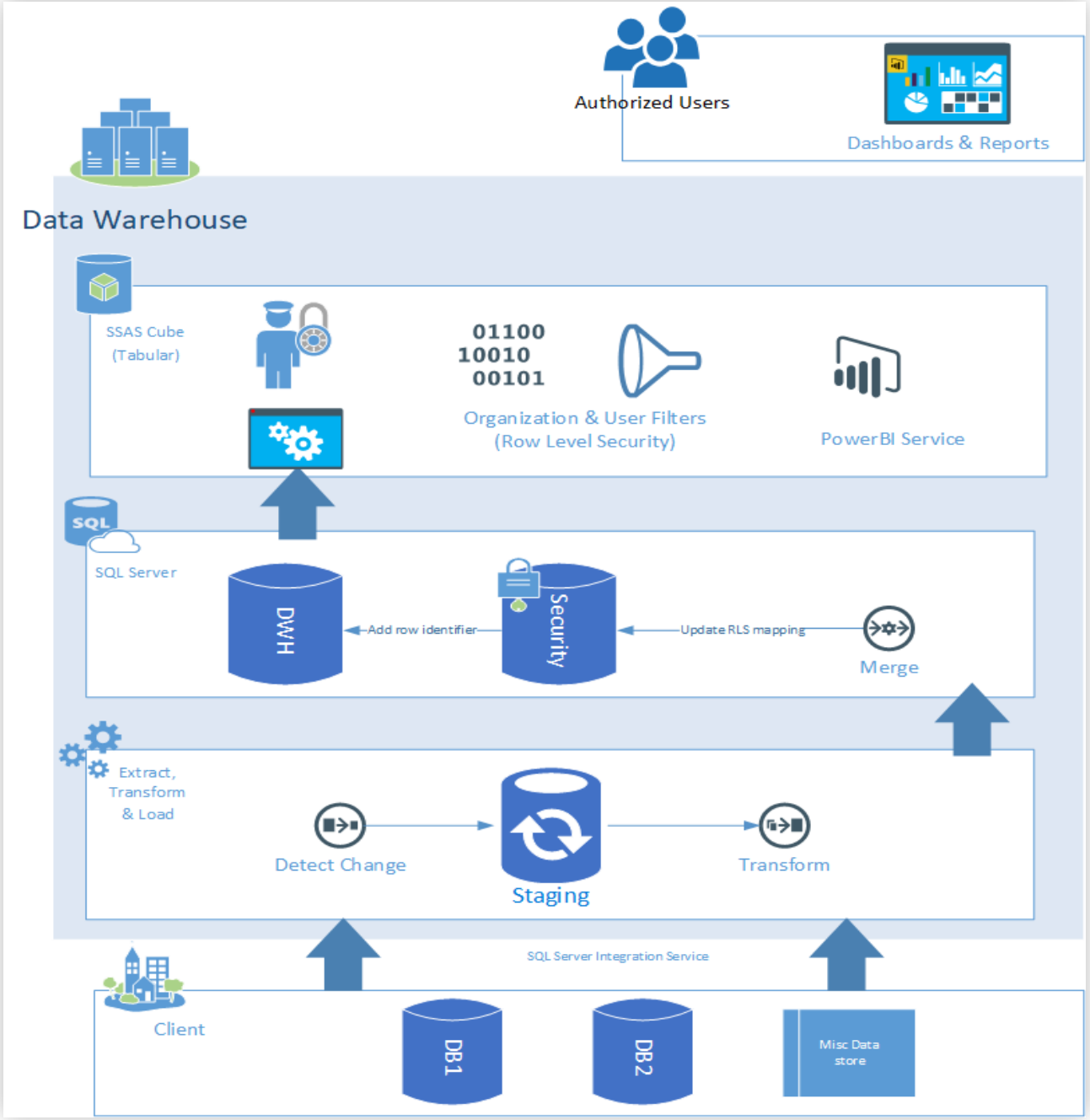
blue margin

Data Warehouse Best-Practices

Best Practice DW System Architecture

Application Layers

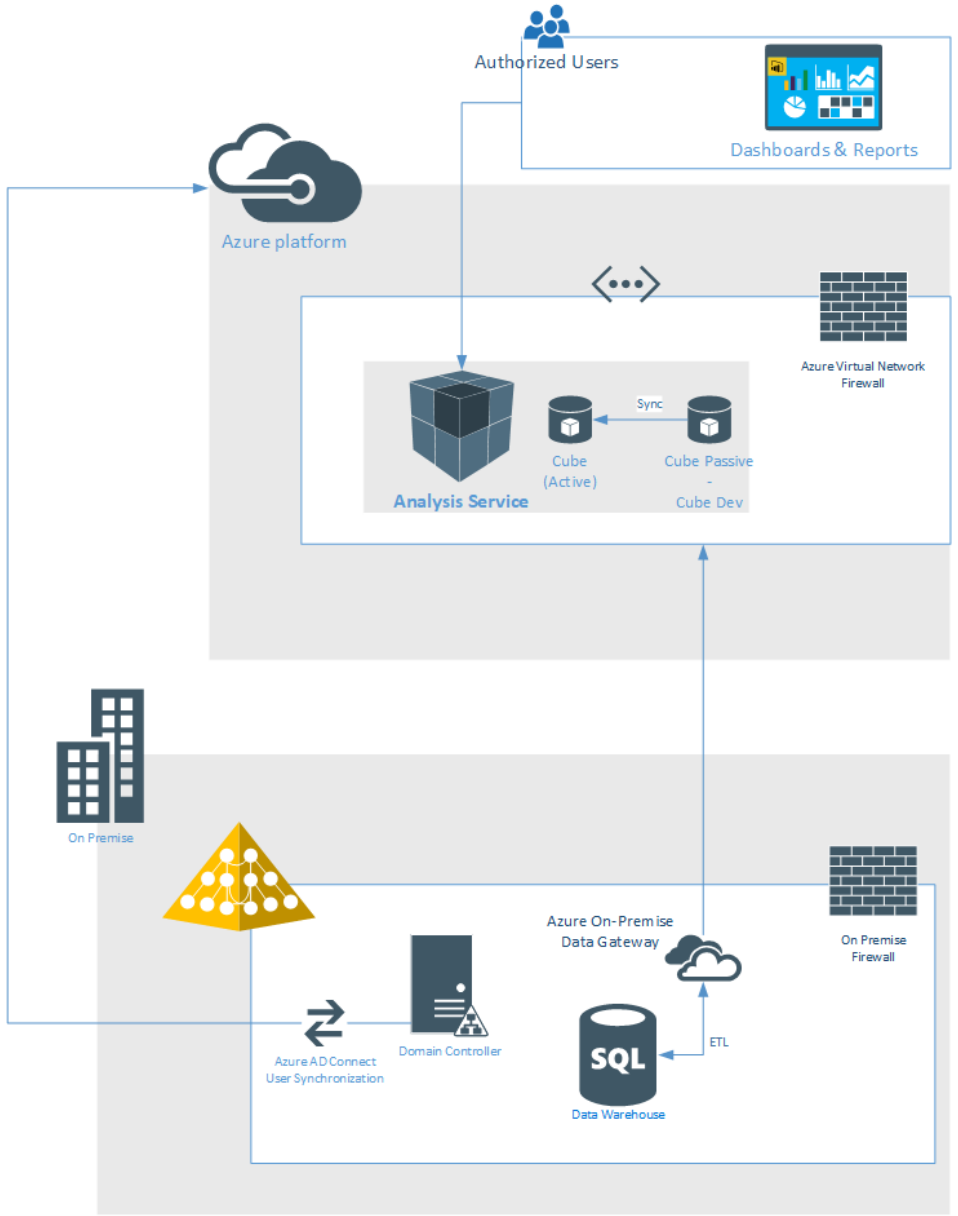
- SQL Server:
  - **Staging** – This data store is for staging data records per source type. The data is cleared after each ETL execution
  - **Security** – This data store contains sensitive data. Only internal Service and Administrator accounts have access. It will store specific data source credentials and client access information
  - **DWH** – Contains the star schema, data source, and Row Level Security schemas, providing the aggregates and calculations that support reporting elements Power BI (or any reporting tool). The Data is also suitable for other client applications
- SQL Server Integration Services (ETL):
  - This application Extracts, Transforms, and Loads company data into the data store
- SQL Server Analysis Services Cube:
  - This application processes and renders data to external client applications
  - The SSAS service will host a Tabular Cube providing the data marts to the Power BI service. This service is accessible to end-users following the pre-define Role Base or Row level Security
- Power BI:
  - The Power BI service will be used to provide reporting and visual analytics to end-users



System Architecture

A successful Data Warehouse project starts with a solid foundation. Blue Margin provides the technical guidance and resources needed to setup an IT environment on-premises or in the Cloud adequate for a scalable data project, including:

- Cost/benefit analysis for on vs. off-premises
- Azure deployment and integration with the existing client ecosystem (Network and Active Directory integration)
- SQL server configuration (physical or on Virtual Machine)
- Analysis Services configuration
- High availability, scalability
- Storage recommendations and backup plan
- Monitoring and alert configuration



Mixed Architecture

Data Warehouse on-premises for easier Data Governance

Analysis Service and PowerBI in the Cloud for simplified access, administration

Data Modeling

We equip our clients for long term success. Using the Kimball methodology and following industry best practices, the data warehouse accommodates data at the lowest possible granularity to support future business requirements.

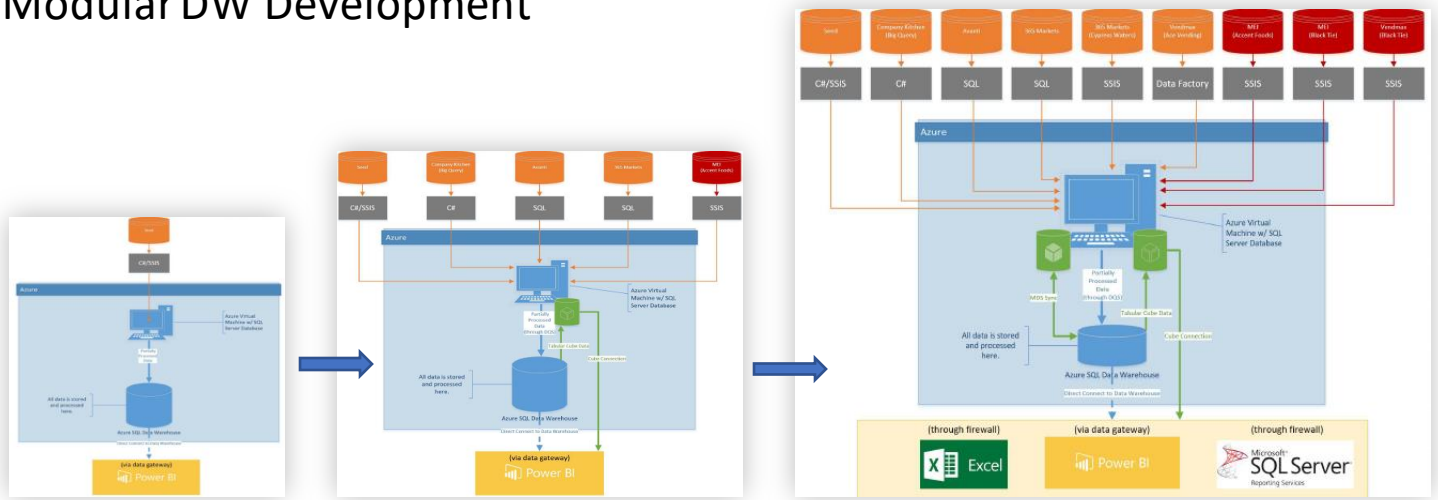
Principles:

- Achieve the goals of the project as simply as possible (outlined in our ETL Guidelines doc). Cut through to elegant solutions that are intuitive, powerful, and robust
- Scrum development methodology (managed on MSFT VSTS)
- Optimal performance (Star Schema, de-normalizing)
- History tracking (Type-2, "Slowly Changing" dimensions)
- Clear and concise code, with in-line documentation
- Standardized naming convention
- Modular data mart design
- Built-in Logging and Alerting
- Automated validation for Source-to-Target. Checksum Validation

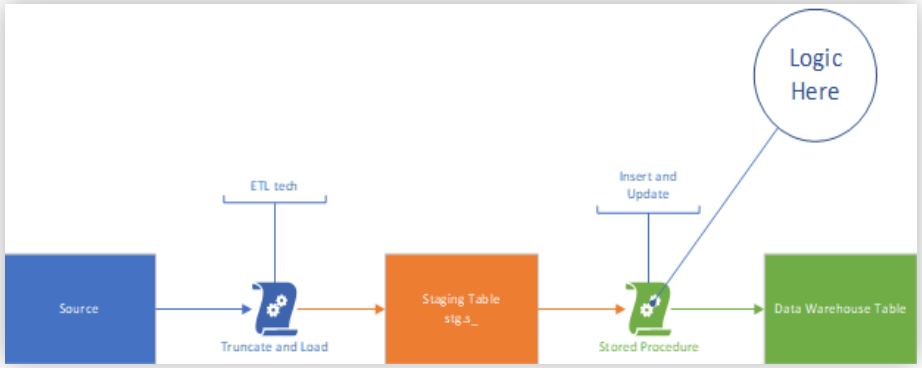
Four-step dimensional model design process – Incorporating business-user requirements and realities of the data:

1. Select the business process (taking orders, shipping, tracking inventory, handling service calls, etc.)
2. Declare the grain (lowest grain possible) – determine what an individual fact row represents
3. Identify dimensions - "How do business users describe the data resulting from measurement events?")
4. Identify the facts - "What is this process measuring?"

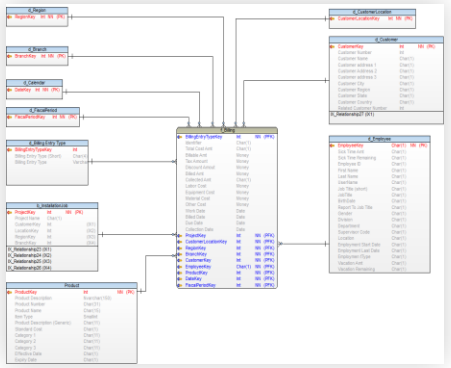
Modular DW Development



Simplified Generic ETL Logic placement



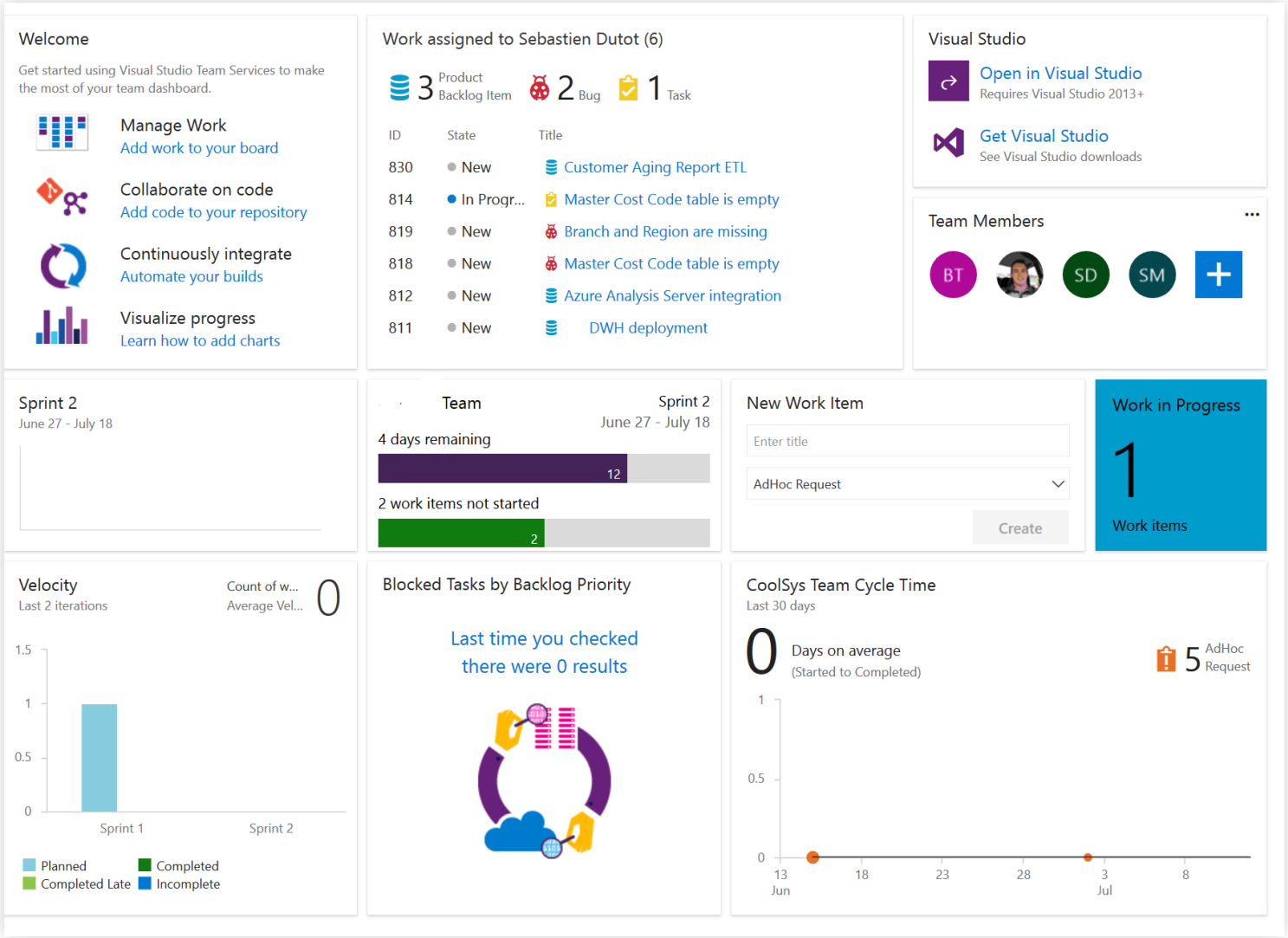
Star Schema



Development & Client Deliverables

To ensure up to standard and accurate solution, Blue Margin uses Visual Studio Team Services (VSTS) to store project workflow and related code. This strategy ensures:

- Measurable and realistic sprint iteration
- Simplified communication channel
- Shared location for business requirements and quality testing
- Support for weekly project checkpoints internally, and weekly status updates with clients
- Visibility into progress by project managers
- Code versioning and history



VSTS Project dashboard overview

Documentation and Knowledge Transfer

In order to give clients ownership and autonomy with their data, we provide comprehensive documentation and training:

- Executive summary overview of the high-level components and functionality of the data warehouse
- Complete systems documentation with technical architecture diagrams, ERD's, and source to target mappings
- Naming convention and in-line code documentation
- A strict "No Black Boxes" policy

Naming Convention sample

Schemas:

- 3 characters
- One schema for each function of the data warehouse
  - stg = staging, staging tables (truncate and load)
  - rpt = reporting, usually only views
  - etl = transforms, usually only stored procedures
  - mon = Monitoring of ETL
  - ndw = data warehouse schema (n = first letter of client name)

Detailed Documentation